Naval Security

ACOUSTIC DETECTION OF UNDERWATER MINES

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

UC#07 — 2026/04 (v6.0)　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 specialized virtual "**Golden Ear**" (expert in underwater acoustics) operating 24/7/365 with the same quality of decision, or to design undetectable mines by simulation.
- 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

Hardware
SUPERMICRO server 41.47 Tflops FP64 960 threads, provided by our partner BECHTLE

Variable to Predict: The model identifies detected object: **ROCK** | **MINE**

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

Reference cases: **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 | 85% (80% for Training, 20% for Validation)	
ROCK	MINE
82 | 46.59%	94 | 53.41 %

External Test Dataset: 32 signals | 15%	
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 **9 gradual rules** without chaining.
- Each rule uses from 2 to 5 predictors among the **12 predictors** that XTRACTIS automatically identified as significant (out of the 60 Potential Predictors).
- Only a few rules are triggered at a time to compute the decision.

☑ **High Predictive Capacity**: It has good Real Performance (on unknown data).

☑ **Ready to Deploy**: It computes real-time predictions up to 7.6 million 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**			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 v14.1.59775

1. We launch 2,000 inductive reasoning strategies. Due to the small number of reference cases, each strategy is applied to 40 different 5-fold-partitions of the Learning Dataset to get a reliable assessment of the descriptive and predictive performance, 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: 2,402 rules share 15 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, from a large synthetic reference dataset:

4. We build a synthetic dataset composed of 17,600 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 this new dataset (100% Training): XTRACTIS induces 2,000 IVEs.
6. The top-IVE selected is the one that has the best performance (on Training and original cases), and the best intelligibility, i.e., the fewer predictors and rules.

Total number of induced unitary models
402,000 IVEs

Criterion for the induction optimization
Matthews Correlation Coefficient (MCC)

Validation criterion for the top-model selection
MCC

Duration of the process @ Induction Speed FP64
8 hours @41.47 Tflops

Environmental footprint
33.6 kWh | **1.88** kg of CO_2

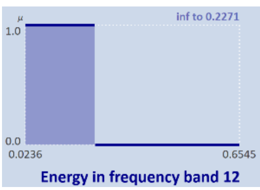
TOP-MODEL STRUCTURE

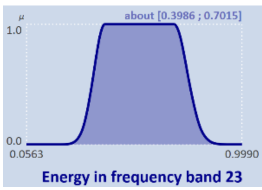
AI Act Compliant Decision System

The top-IVE model has very good intelligibility given the complexity of the phenomenon (it scores 4.64 out of 5): its 9 rules combine 12 predictors with 3.4 predictors per rule on average. Its Structure Report reveals the entire internal logic of the decision system and ensures that the model is understandable. It is a transparent model that can be audited by a professional expert and certified before deployment to end-users.

PREDICTORS

- 12 energy measures in frequency bands (out of 60)
- Ranked by impact significance
  (2 strong & 10 medium signals):
  #1 Energy in frequency band 45, #2 Energy in frequency band 36…
- Labeled by fuzzy and binary classes
  Examples: **binary interval** "inf to 0.2271";
  fuzzy interval "about [0.3986 ; 0.7015]"





RULES

- 9 connective fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules)
- 2 to 5 predictors per rule (on average, 3.4 predictors per rule)
- Example: **fuzzy rule R2** uses 3 predictors and concludes "ROCK". 8 other fuzzy rules complete this model.

IF	Energy in frequency band 17	IS	sup to ~0.5622
AND	Energy in frequency band 23	IS	~[0.3986 ; 0.7015]
AND	Energy in frequency band 27	IS	~[0.4032 ; 0.7008]
THEN	Detected Object	IS	ROCK

*Literally, the detected object is Rock **if** the energy in frequency band #11 is over around 0.5622, **and** in band #23 is between approximately 0.3986 and 0.7015 **and** in band #27 is between approximately 0.4032 and 0.7008.*

TOP-MODEL PERFORMANCE

The top-IVE performance, measured in Training on synthetic data and on original cases, then in External Test on reference data, guarantees the model's predictive and real performance.

Perf. Type	Quality of CVE Copy		REAL
Dataset	100% Training (Synthetic Data)	176 original cases (remaining unknown to the top-IVE)	External Test
Matthews Corr. Coeff. (MCC)	99.7%	98.9%	88.1%
Classification Error	0.13%	0.57%	6.25%

EXPLAINED PREDICTIONS FOR 2 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v14.1.59775

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

DEDUCTIVE INFERENCE OF RULES

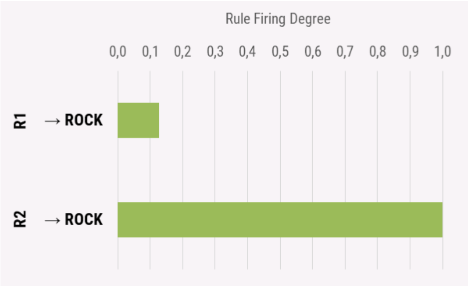
REAL-TIME DECISION

ECHO #3	
Energy in frequency band 9	0.5598
Energy in frequency band 12*	0.7060
Energy in frequency band 17	0.6759
Energy in frequency band 20	0.8619
Energy in frequency band 23	0.4293
Energy in frequency band 27	0.5070
Energy in frequency band 28	0.8533
Energy in frequency band 36	0.3043
Energy in frequency band 37	0.6116
Energy in frequency band 42	0.2587
Energy in frequency band 45	0.2111
Energy in frequency band 49	0.0130
Actual Value	ROCK

For this signal, 2 rules are triggered:

R2 at 1.000, and **R1** at 0.128,
to agree on {ROCK}.

The other 7 rules are not activated.



NUMBER OF TRIGGERED RULES
2 / 9

FUZZY PREDICTION
{ ROCK | 1.000 }

FINAL PREDICTION
{ ROCK }

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

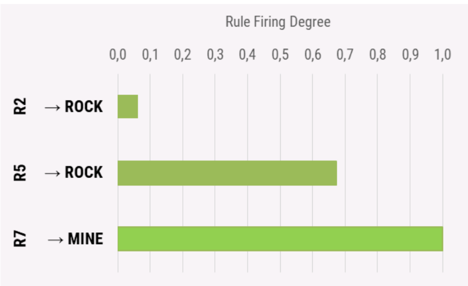
ROCK

ECHO #108	
Energy in frequency band 9	0.0610
Energy in frequency band 12	0.2473
Energy in frequency band 17	0.6588
Energy in frequency band 20	0.8672
Energy in frequency band 23	0.8385
Energy in frequency band 27	0.4139
Energy in frequency band 28	0.3269
Energy in frequency band 36	0.4619
Energy in frequency band 37	0.4234
Energy in frequency band 42	0.3324
Energy in frequency band 45	0.2137
Energy in frequency band 49	0.0453
Actual Value	MINE

For this signal, 3 rules are triggered:

R7 at 1.000 to conclude on {MINE},
R5 at 0.673 and **R2** at 0.062 to conclude {ROCK}.

The other 6 rules are not activated.



NUMBER OF TRIGGERED RULES
3 / 9

FUZZY PREDICTION
{ MINE | 1.000,
ROCK | 0.673 }

FINAL PREDICTION
{ MINE }

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

MINE

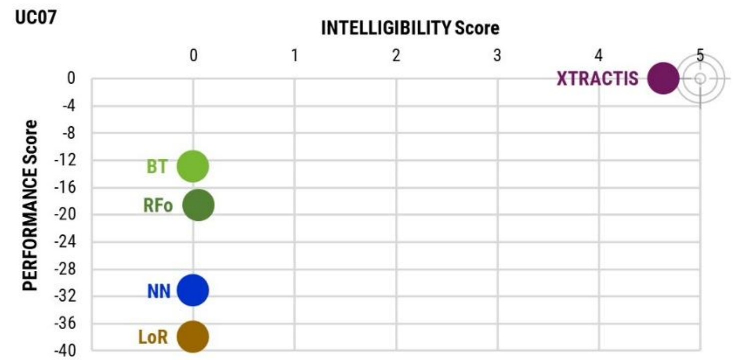
**Predictor value is out of the variation range (OOR) of the model but inside the allowed extrapolation range (8.15% OOR for Case ECHO #3). XTRACTIS will refuse to give a result for an extrapolation far from the allowed extrapolation range. It is one situation of the" Refusal" prediction.*

TOP-MODELS BENCHMARK

		XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELING PARAMETERS	MODELS RELEASE	2026/04	2026/04	2026/04	2026/04	2026/04
	ALGORITHM VERSION	XTRACTIS **REVEAL** 14.1.59775	Python 3.9.10 | Scikit-Learn 1.5.2		Python 3.9.10 | LightGBM 4.5.0	Python 3.9.10 | PyTorch 2.5.1 +cpu | Keras 3.7.0
	NUMBER OF EXPLORED STRATEGIES[1]	**CVEs**: 2,000 induction strategies on Training / Validation data; 3 aggregation operators tested. **IVEs**: 2,000 induction strategies on 100% of the synthetic data	2,000 ML strategies on Training / Validation data Aggregation operator: *Relative Majority*			1,089 ML strategies on Training / Validation data
	VALIDATION TECHNIQUE	**CVEs**: 40 × 5 folds. **IVEs**: validation on all original cases kept aside.	40 × 5 folds for each CVE model			
	TOP-MODEL SELECTION[2]	**Top-CVE**: selected among 6,000 CVEs, with *Simple Majority* aggregator. **Top-IVE**: selected 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 1,089 CVEs

		XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
TOP-MODEL STRUCTURE	**NUMBER OF PREDICTORS** (out of 60 Potential Predictors)	**12**	**60**	**50**	**50**	**60**
	AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION	**3.4** per rule	**60.0** per equation	**4.3** per rule	**3.4** per rule	**60.0** per equation
	STRUCTURE OF THE DECISION SYSTEM	**9** fuzzy rules without chaining (aggregated into 2 disjunctive fuzzy rules) Only few rules are triggered at a time to compute a decision	**1** linear equation	**19** trees without chaining **299** binary rules	**1** chain of **31** trees **280** binary rules Tree #N corrects the error of the N-1 previous trees	**4** hidden layers | **116** hidden nodes **117** equations 116 unintelligible synthetic variables

TOP-MODEL SCORES	Random[3]	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score[4]		4.64	0.00	0.05	0.00	0.00
CVE Real Performance (MCC_%) in External Test	*62.4*	87.5	49.8	75.7	74.9	56.6
Gap to CVE Leader in External Test	*-25.1*	**0.0**	**-37.7**	**-11.8**	**-12.6**	**-30.9**
IVE Real Performance (MCC_%) in External Test	*62.4*	88.1	49.8	62.6	74.9	56.6
Gap to IVE Leader in External Test	*-25.7*	**0.0**	**-38.3**	**-25.5**	**-13.2**	**-31.5**
Average Real Performance (MCC_%) in External Test	*62.4*	87.8	49.8	69.2	74.9	56.6
PERFORMANCE Score[4]		**0.00**	**-38.00**	**-18.65**	**-12.90**	**-31.20**



(1) For all algos: on the same Learning Dataset. All Models are optimized according to their Validation Matthews Correlation Coefficient (MCC).
(2) All top-models are selected according to their Validation MCC while checking that it remains close to their Training MCC.
(3) Baseline performance 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∈[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 & Boosted Tree: Python, LightGBM; Neural Network: Python, PyTorch, 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).
- For binomial classification, the Matthews Correlation Coefficient (MCC) is preferred to the F_β-Score because it is independent of prevalence in the dataset (ratio of class labeled "1") and symmetrical (insensitive to which class is labeled "0" and to which class is labeled "1").

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 or Matthews Correlation Coefficient (MCC) 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

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

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):

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

  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):

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

  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):

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

  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

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

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	Sensitivity	Specificity	PPV	NPV	F_1-Score	**Matthews Correlation Coefficient**	Refusal
RANDOM MODEL								
Nb of Random Permutations (P-value) = 100,000 (0.001)								
Performance against chance (External Test)	*18.75%*					*82.35%*	***62.4%***	
XTRACTIS TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.0%	0 (0.00%)
CVE - Predictive Performance (Validation)	8.57%	93.62%	88.89%	90.72%	92.31%	92.15%	82.8%	1 (0.57%)
CVE - Real Performance (External Test)	6.25%	94.12%	93.33%	94.12%	93.33%	94.12%	**87.5%**	0 (0.00%)
IVE - Descriptive Performance (Synthetic Training)	0.13%	100.00%	99.72%	99.75%	100.00%	99.88%	99.7%	0 (0.00%)
IVE - Predictive Performance (176 original cases)	0.57%	98.94%	100.00%	100.00%	98.80%	99.47%	98.9%	0 (0.00%)
IVE - Real Performance (External Test)	6.25%	100.00%	86.67%	89.47%	100.00%	94.44%	**88.1%**	0 (0.00%)
LOGISTIC REGRESSION TOP-MODEL								
CVE - Descriptive Performance (Training)	11.93%	87.23%	89.02%	90.11%	85.88%	88.65%	76.1%	
CVE - Predictive Performance (Validation)	19.89%	77.66%	82.93%	83.91%	76.40%	80.66%	60.4%	
CVE - Real Performance (External Test)	25.00%	76.47%	73.33%	76.47%	73.33%	76.47%	**49.8%**	
IVE - Descriptive Performance (Training)	15.34%	80.85%	89.02%	89.41%	80.22%	84.92%	69.8%	
IVE - Real Performance (External Test)	25.00%	76.47%	73.33%	76.47%	73.33%	76.47%	**49.8%**	
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)	11.36%	91.49%	85.37%	87.76%	89.74%	89.58%	77.2%	
CVE - Real Performance (External Test)	12.50%	82.35%	93.33%	93.33%	82.35%	87.50%	**75.7%**	
IVE - Descriptive Performance (Training)	0.57%	98.94%	100.00%	100.00%	98.80%	99.47%	98.9%	
IVE - Real Performance (External Test)	18.75%	88.24%	73.33%	78.95%	84.62%	83.33%	**62.6%**	
BOOSTED TREE TOP-MODEL								
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.0%	
CVE - Predictive Performance (Validation)	7.95%	92.55%	91.46%	92.55%	91.46%	92.55%	84.0%	
CVE - Real Performance (External Test)	12.50%	88.24%	86.67%	88.24%	86.67%	88.24%	**74.9%**	
IVE - Descriptive Performance (Training)	0.57%	98.94%	100.00%	100.00%	98.80%	99.47%	98.9%	
IVE - Real Performance (External Test)	12.50%	88.24%	86.67%	88.24%	86.67%	88.24%	**74.9%**	
NEURAL NETWORK TOP-MODEL								
CVE - Descriptive Performance (Training)	0.57%	100.00%	98.78%	98.95%	100.00%	99.47%	98.9%	
CVE - Predictive Performance (Validation)	6.25%	96.81%	90.24%	91.92%	96.10%	94.30%	87.5%	
CVE - Real Performance (External Test)	21.88%	88.24%	66.67%	75.00%	83.33%	81.08%	**56.6%**	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	21.88%	88.24%	66.67%	75.00%	83.33%	81.08%	**56.6%**	