xtractis® BY INTELLITECH

Defense

PASSIVE MAGNETIC IDENTIFICATION OF LAND MINES

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

2024/02 (v3.0) xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI-based decision system that accurately detects land mines and identifies the type of mine only from a few variables, to instantly deliver the appropriate rational decision.

PROS & BENEFITS

- Find the cause-and-effect relationships between the relevant predictors among the 3 parameters of this study, and the actual presence of a mine and its type.
- Enhance demining technicians and military experts' knowledge by understanding the strategies to identify the type of mine.
- Help engineers to design enhanced mine detectors, manual or autonomous, embedding a classifier making explainable and accurate automated decisions.
- Assist the military profession in making a more reliable decision, thanks to rapid, systematic, explainable, and safer detection process with a passive magnetic sensor.

REFERENCE DATA

Source:
Cemal YILMAZ, Department of Electrical and Electronics Engineering, Gazi University, Ankara, Turkey.

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 the soil content:
NO MINE | ANTI-PERSONNEL MINE | ANTI-TANK MINE

Predictive Variables

3 potential predictors characterizing each situation: VOLTAGE value due to magnetic distortion, SOIL TYPE, SENSOR HEIGHT from the ground.

Observations

338 object detection experiments; each is associated with a no-mine case or a mine case specifying the type of mine. 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 unknown data and for benchmarking.

Learning Dataset: 287 cases | 84.91% 80% for Training, 20% for Validation		
NO MINE	ANTI-PERSONNEL MINE	ANTI-TANK MINE
60 | 20.91%	167 | 58.19%	60 | 20.91%

External Test Dataset: 51 cases | 15.09%		
NO MINE	ANTI-PERSONNEL MINE	ANTI-TANK MINE
11 | 21.57%	30 | 58.82%	10 | 19.61%

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 **29 gradual rules without chaining**, **each rule uses 1 to 3 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	Automated Decision (detect & identify mine): NO MINE / ANTI-PERSONNEL MINE / ANTI-TANK 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
v13.0.44978

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: 2,503 rules sharing 3 predictors.

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

4. We build a synthetic dataset composed of 14,350 new cases simulated by deduction from the top-CVE, around the 287 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 robust as the top-CVE, but more intelligible: 29 rules sharing 3 predictors.

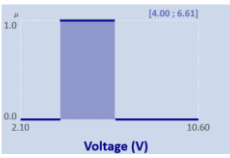
Total number of induced unitary models	Criterion for the induction optimization	Validation criterion for the top-model selection	Duration of the process (Induction Power FP64)
402,000 IVEs	Average F_1-Score	Average F_1-Score	9,5 Hours (1 Tflops)

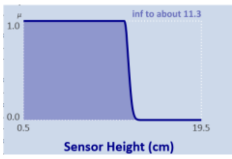
TOP-MODEL STRUCTURE

The top-IVE model has an excellent intelligibility as it combines the 3 predictors preserved by XTRACTIS into 29 rules, aggregated into 3 disjunctive rules. 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

- 3 out of 3 parameters (2 continuous + 1 nominal)
- Ranked by impact significance
  (1 strong predictor, 2 medium predictors):
  #1 Voltage /#2 …
- Labeled by binary and fuzzy classes.
  Examples: **binary interval** "[4.00 ; 6.61]"
  fuzzy interval "inf to about 11.3"





RULES

- 29 connective fuzzy rules without chaining
  (aggregated into 3 disjunctive fuzzy rules)
- 1 to 3 predictors per rule (on average, 2.5 predictors per rule)
- Example: fuzzy rule **R14** uses 3 predictors and concludes **ANTI-PERSONNEL MINE**. 28 other rules complete this model, including 3 binary rules.

IF	Voltage	IS in	[4.00 ; 6.61] V
AND	Sensor Height	IS	inferior to ~11.3 cm
AND	Soil Type	IS in	{Dry and Limy, Dry and Sandy, Humid and Humus, Humid and Limy, Humid and Sandy}
THEN	Object	IS	ANTI-PERSONNEL MINE

Literally, the detected Object is Anti-Personnel Mine if the Voltage value due to magnetic distortion is between 4.00V and 6.61V, and the Sensor is under approximately 11.3cm to the ground, and the soil is of any type except Dry and Humus.

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
Average F_1-Score	99.05%	98.84%	98.07%	90.19%
Classification Error	0.93%	1.13%	1.87%	9.80%

EXPLAINED PREDICTIONS FOR 3 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v13.0.44978

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

DEDUCTIVE INFERENCE OF RULES

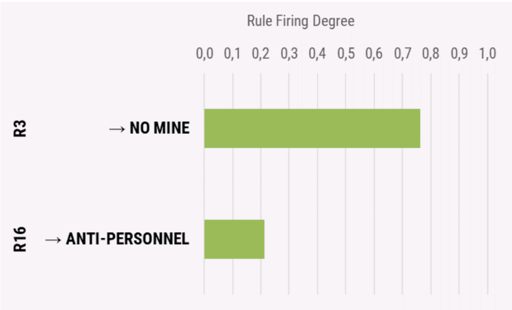
AUTOMATED DECISION

Situation #244

actual value = NO MINE	
Voltage (V)	3.52
Sensor Height (cm)	12.7
Soil Type	Dry and Limy

Real Time

For this situation, 2 rules are triggered:
R3 is fired at 0.764 to conclude NO MINE, and
R16 is fired at 0.212 to conclude ANTI-PERSONNEL



NUMBER OF TRIGGERED RULES
2 / 29

FUZZY PREDICTION
{ NO MINE | 0.764,
ANTI-PERSONNEL | 0.212 }

FINAL PREDICTION
{ NO MINE }

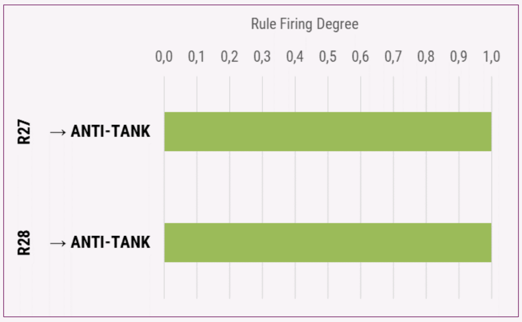
The system delivers a correct detection compared to the experiment:
NO MINE

Situation #253

actual value = ANTI-TANK MINE	
Voltage (V)	4.66
Sensor Height (cm)	20.0*
Soil Type	Dry and Sandy

Real Time

For this situation, 2 rules are triggered:
both **R27** and **R28** are fired at 1.000
to conclude ANTI-TANK



NUMBER OF TRIGGERED RULES
2 / 29

FUZZY PREDICTION
{ ANTI-TANK | 1.000 }

FINAL PREDICTION
{ ANTI-TANK }

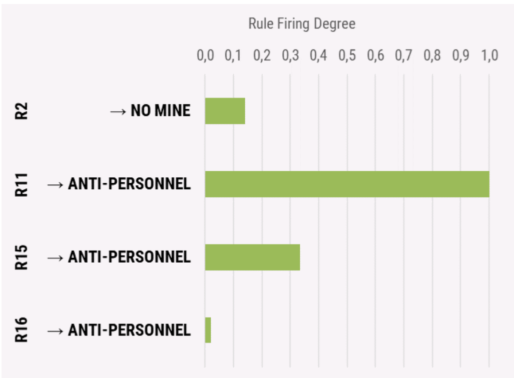
The system delivers a correct identification compared to the experiment:
ANTI-TANK MINE

Situation #300

actual value = ANTI-PERSONNEL MINE	
Voltage (V)	3.81
Sensor Height (cm)	10.9
Soil Type	Humid and Sandy

Real Time

For this situation, 4 rules are triggered:
R2 is fired at 0.141 to conclude NO MINE,
R11 is fired at 1.000, **R15** at 0.333
and **R16** at 0.020 to conclude ANTI-PERSONNEL



NUMBER OF TRIGGERED RULES
4 / 29

FUZZY PREDICTION
{ ANTI-PERSONNEL | 1.000,
NO MINE | 0.141 }

FINAL PREDICTION
{ ANTI-PERSONNEL }

The system delivers a correct identification compared to the experiment:
ANTI-PERSONNEL MINE

**Predictor value is out of the variation Range of the model (<2.65 % OOR for case #253) but inside the allowed extrapolation range. XTRACTIS will refuse to give a result for an extrapolation outside the allowed extrapolation range. It is one situation of the "Refusal" prediction.*

TOP-MODELS BENCHMARK

MODELING PARAMETERS

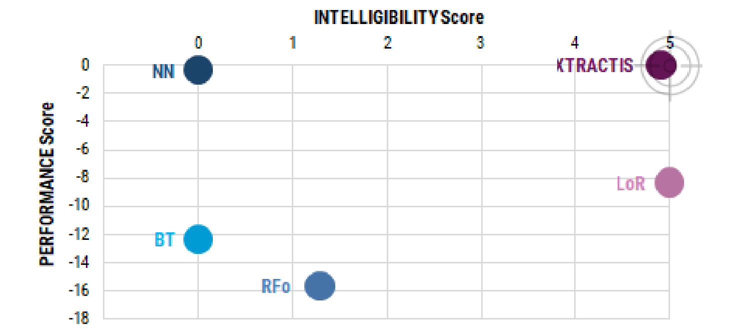
	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREES	NEURAL NETWORK
MODELS RELEASE	2023/03	2023/03	2023/03	2023/03	2023/03
ALGORITHM VERSION	XTRACTIS **REVEAL** 13.0.44978	Python 3.9.10 | Scikit-Learn 1.1.2	Python 3.9.10 | LightGBM 3.3.2	Python 3.9.10 | LightGBM 3.3.2	Python 3.9.10 | TensorFlow 2.10.0 | Keras 2.10.0
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	2,000 ML strategies on Training / Validation data	2,000 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-MODEL STRUCTURE

	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREES	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 3 Potential Predictors)	**3**	**6** 1 nominal predictor with 6 modalities split into 6 predictors	**2**	**3**	**8** 1 nominal predictor with 6 modalities split into 6 predictors
AVERAGE NUMBER OF PREDICTORS PER RULE / EQUATION	**2.5** per rule	**2.7** per equation	**1.8** per rule	**2.1** per rule	**22.1** per equation
STRUCTURE OF THE DECISION SYSTEM	**29** fuzzy rules without chaining aggregated into 3 disjunctive rules. Only a few rules are triggered at a time to compute a decision	**3** linear equations	**72** trees **1,117** binary rules	**3** chains of **66** trees each **1,355** binary rules. Tree #N corrects the error of the N-1 previous trees	**4** hidden layers | **91** hidden nodes **94** equations. 91 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)		**4.91**	**5.00**	**1.29**	**0.00**	**0.00**
CVE Real Performance (Average F_1-Score) in External Test	55.66%	90.19	81.04	75.26	76.68	89.87
Gap to CVE Leader in External Test		**0.00**	**-9.15**	**-14.93**	**-13.51**	**-0.32**
IVE Real Performance (Average F_1-Score) in External Test	55.66%	90.19	82.67	73.86	78.99	89.87
Gap to IVE Leader in Test		**0.00**	**-7.52**	**-16.33**	**-11.20**	**-0.32**
Top-IVE Average Real Performance	55.66%	90.19	81.86	74.56	77.84	89.87
PERFORMANCE Score(4)		**0.00**	**-8.33**	**-15.63**	**-12.36**	**-0.32**



(1) For all algos: on the same Learning Dataset. All Models are optimized according to their validation Average F_1-Score.
(2) All top-models are selected according to their validation Average F_1-Score while checking that it remains close to their training Average 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	$i \in [1 ; n]$ 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 & 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

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

<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):
  $$Pen1(T_i) = \min(0\,,1 - \log_{10} \textit{number of predictors})$$
  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):
  $$Pen2(T_i) = \min\left(0\,, 0.01 - \frac{\textit{average number of rules or equations per modality to predict}}{100}\right)$$
  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):
  $$Pen3(T_i) = \min\left(0\,, \frac{9 - 3 \times \textit{average number of predictors per rule or equation}}{7}\right)$$
  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):
  $$Pen4(T_i) = \min(0\,,1 - \textit{number of chained trees})$$
  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	Average Sensitivity	Min. PPV	Average PPV	Min. F_1-Score	**Average F_1-Score**	Weighted Av. F_1-Score	Refusal
RANDOM MODEL									
Nb of Random Permutations (P-value) = 100,000 (0.001%)									
Performance against chance	*35.29%*	*50.00%*	*55.66%*	*50.00%*	*55.66%*	*50.00%*	***55.66%***	*64.71%*	
XTRACTIS TOP-MODEL									
CVE - Descriptive Performance (Training)	4.88%	90.00%	94.71%	89.39%	95.07%	93.65%	94.73%	95.13%	0 (0.00%)
CVE - Predictive Performance (Validation)	5.23%	92.81%	95.94%	84.51%	93.63%	91.60%	94.53%	94.85%	0 (0.00%)
CVE - Real Performance (External Test)	9.80%	86.67%	92.22%	73.33%	89.88%	84.62%	**90.19%**	90.49%	0 (0.00%)
IVE - Descriptive Performance (Training)	0.93%	98.80%	99.23%	97.87%	98.86%	98.54%	99.05%	99.07%	30 (0.62%)
IVE - Predictive Performance (Validation)	1.13%	98.72%	98.96%	97.55%	98.72%	98.24%	98.84%	98.87%	34 (0.72%)
IVE - Real Performance (Test)	1.87%	97.87%	98.23%	96.42%	97.90%	97.14%	98.07%	98.13%	38 (0.80%)
IVE - Real Performance (287 original points)	6.67%	92.17%	94.02%	82.86%	92.43%	89.23%	93.01%	93.43%	2 (0.70%)
IVE - Real Performance (External Test)	9.80%	86.67%	92.22%	73.33%	89.88%	84.62%	**90.19%**	90.49%	0 (0.00%)
LOGISTIC REGRESSION TOP-MODEL									
CVE - Descriptive Performance (Training)	14.63%	77.25%	90.19%	64.77%	83.72%	77.03%	85.56%	85.72%	
CVE - Predictive Performance (Validation)	17.07%	75.45%	87.37%	62.07%	81.33%	73.47%	83.08%	83.32%	
CVE - Real Performance (External Test)	19.61%	70.00%	86.97%	58.82%	79.20%	71.43%	**81.04%**	80.74%	
IVE - Descriptive Performance (Training)	15.33%	77.25%	89.08%	63.95%	82.97%	75.34%	84.81%	85.04%	
IVE - Real Performance (External Test)	17.65%	73.33%	88.08%	62.50%	80.50%	74.07%	**82.67%**	82.64%	
RANDOM FOREST TOP-MODEL									
CVE - Descriptive Performance (Training)	9.41%	86.67%	89.27%	82.81%	90.70%	85.48%	89.84%	90.64%	
CVE - Predictive Performance (Validation)	14.29%	78.33%	84.34%	73.44%	85.16%	75.81%	84.67%	85.82%	
CVE - Real Performance (External Test)	25.49%	63.64%	75.66%	46.67%	76.05%	53.85%	**75.26%**	75.60%	
IVE - Descriptive Performance (Training)	11.85%	85.00%	87.52%	76.12%	87.60%	80.31%	87.42%	88.29%	
IVE - Real Performance (External Test)	27.45%	63.64%	74.55%	43.75%	74.84%	51.85%	**73.86%**	73.88%	
BOOSTED TREES TOP-MODEL									
CVE - Descriptive Performance (Training)	6.62%	83.33%	90.87%	91.57%	94.72%	87.72%	92.62%	93.32%	
CVE - Predictive Performance (Validation)	12.20%	76.67%	85.54%	80.70%	87.76%	78.63%	86.58%	87.76%	
CVE - Real Performance (External Test)	23.53%	64.64%	76.77%	50.00%	77.38%	56.00%	**76.68%**	77.31%	
IVE - Descriptive Performance (Training)	2.09%	93.33%	97.02%	97.08%	98.44%	95.73%	97.70%	97.90%	
IVE - Real Performance (External Test)	19.61%	63.64%	78.99%	63.64%	78.99%	63.64%	**78.99%**	80.39%	
NEURAL NETWORK TOP-MODEL									
CVE - Descriptive Performance (Training)	4.88%	93.33%	95.07%	88.89%	94.52%	91.06%	94.77%	95.15%	
CVE - Predictive Performance (Validation)	3.14%	96.67%	96.78%	92.06%	96.55%	94.31%	96.64%	96.88%	
CVE - Real Performance (External Test)	9.80%	90.00%	90.30%	76.92%	90.01%	83.33%	**89.87%**	90.39%	
IVE - Descriptive Performance (Training)	3.48%	95.81%	96.94%	89.23%	95.80%	92.80%	96.31%	96.56%	
IVE - Real Performance (External Test)	9.80%	90.00%	90.30%	76.92%	90.01%	83.33%	**89.87%**	90.39%	