



Fraud Detection

PHYSICOCHEMICAL IDENTIFICATION OF WINES' ORIGIN

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

UC#25 – 2026/09 (v2.0)

xtractis.ai

PROBLEM DEFINITION

GOALS Design an AI-based decision-making system that accurately identifies a wine from its physico-chemical analysis, in order to detect fraudulent bottles of wine, in a rational and explainable way, and help preserve the wine's protected designation of origin.

PROS & BENEFITS

- Identify the specific physicochemical parameters significantly characterizing each wine and enhance expert knowledge by helping wine-specializing chemists understand the cause-and-effect relationships between these parameters and the wine origin.
- Help anti-fraud services detect counterfeit wines and understand the underlying strategy of fraudsters in the way they counterfeit wine.
- Avoid false alarms thanks to transparent and accurate automatized wine composition analysis.

REFERENCE DATA

Source:
Stefan Aeberhard,
M. Forina, James Cook
University, Australia

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 predicts the Wine's Origin among 3 modalities:

WINE A | WINE B | WINE C

Potential Predictors 13 physicochemical compounds describe each wine (all 13 are numeric variables):

Alcohol, Malic Acid, Ash, Alkalinity of Ash, Magnesium, Total Phenols, Flavonoids, Nonflavonoid Phenols, Proanthocyanidins, Color Intensity, Hue, OD280/OD315 of Diluted Wines, Proline.

Observations 178 reference cases.

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: 151 Wine cases 84.83%			External Test Dataset: 27 Wine cases 15.17%		
Training (121 80%), Validation (30 20%)			WINE A	WINE B	WINE C
WINE A	WINE B	WINE C	7 25.93%	9 33.33%	11 40.74%
41 27.15%	50 33.11%	60 39.74%			

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 8 gradual rules without chaining.
 - Each rule uses from 2 to 4 predictors among the 7 variables that XTRACTIS automatically identified as significant (out of the 13 Potential Predictors).
 - Only a few rules are triggered at a time to compute the decision.
- ☑ **High Predictive Capacity** It has a perfect 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  WINE A  WINE B  WINE C Automated Decision (identify Wine Origin)
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:



- 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 performances, respectively from Training and Validation Datasets.
- 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)**.
- Among the 6,000 induced CVEs, the top-CVE with the best predictive performance remains complex: 1,054 rules share 13 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:

- We build a synthetic dataset composed of 10,117 new cases simulated by deduction from the top-CVE, around the 151 original learning cases but distinct from them.
- We apply 2,000 induction strategies to the same single partition of this new dataset (100% Training | Validation on original points): XTRACTIS induces 2,000 IVEs.
- The top-IVE selected is the one that has the best performance (on Validation and Training) 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

Average F₁-Score

Validation criterion for the top-model selection

Average F₁-Score

Duration of the process @ Induction Speed FP64

8 hours @ 1.13 Tflops

Environmental footprint

4.80 kWh

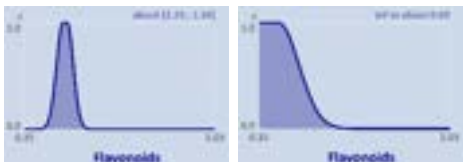
0.27 kg of CO₂

TOP-MODEL
STRUCTURE

The top-IVE has a very good intelligibility score (4.96 out of 5): its **8 rules** combine **7 predictors**, with 2.9 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 certified by the regulator before deploying to end-users, thus AI Act-compliant!

PREDICTORS

- 5 chemical compounds and 2 physical features (out of 13)
- Ranked by individual contribution (2 strong signals, 4 medium signals & 1 weak signal):
#1 **Alcohol** /.../ #7 **OD280/OD315 of Diluted Wines**
- Labeled by fuzzy classes
Examples: **fuzzy interval** "about [1.29 ; 1.39]";
fuzzy interval "inf to about 0.83"



RULES

- 8 fuzzy connective rules without chaining (aggregated into 3 disjunctive fuzzy rules)
- 2 to 4 predictors per rule (on average, 2.9 predictors per rule)
- Example: **fuzzy rule R1** uses 3 predictors and concludes "WINE A". 7 other rules complete this model.

IF	Alcohol (*)	IS	sup to about 13.08
AND	Flavonoids (mg/L)	IS	inf to about 0.83
AND	Color intensity	IS	sup to about 7.15
THEN	Wine Origin	IS	WINE A

Literally, the analyzed wine is of origin "A" if the alcohol by volume is above around 13.08° and the flavonoids concentration is under around 0.83 mg/L and the color intensity is above around 7.15.

TOP-MODEL
PERFORMANCE

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

Perf. Type	Quality of CVE Copy	
Dataset	100% Training (Synthetic Data)	151 original points
Average F ₁ -Score	100.00%	100.00%
Classification Error	0.00%	0.00%



EXPLAINED PREDICTIONS FOR 3 UNKNOWN CASES

Powered by:  XTRACTIS®
PREDICT v14.1.61285

CASE

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

SAMPLE #178

Alcohol	14.13
Ash	2.74
Magnesium	96
Flavonoids	0.76
Color Intensity	9.20
OD280/OD315 of Diluted Wines	1.60
Proline	560

Actual Value

WINE A

DEDUCTIVE INFERENCE OF RULES

For this sample, 2 rules are triggered:
R1 is fired at 1.000 to conclude WINE A, and
R6 is fired at 0.019 to conclude WINE C

The 6 other rules are not activated.



REAL-TIME DECISION

NUMBER OF TRIGGERED RULES

2 / 8

FUZZY PREDICTION

{ WINE A | 1.000,
WINE C | 0.019 }

FINAL PREDICTION

{ WINE A }

The system delivers the correct decision
compared to the actual case



SAMPLE #52

Alcohol	13.83
Ash	2.60
Magnesium	94
Flavonoids	2.99
Color Intensity	5.60
OD280/OD315 of Diluted Wines	3.37
Proline	1,265

Actual Value

WINE B

For this sample, only 1 rule is triggered:
R3 is fired at 1.000 to conclude WINE B

The 7 other rules are not activated.



NUMBER OF TRIGGERED RULES

1 / 8

FUZZY PREDICTION

{ WINE B | 1.000 }

FINAL PREDICTION

{ WINE B }

The system delivers the correct decision
compared to the actual case



SAMPLE #125

Alcohol	11.87
Ash	2.39
Magnesium	82
Flavonoids	3.03
Color Intensity	2.80
OD280/OD315 of Diluted Wines	3.64
Proline	380

Actual Value

WINE C

For this sample, 2 rules are triggered:
R5 and **R6** are fired at 1.000 to conclude WINE C

The 6 other rules are not activated.



NUMBER OF TRIGGERED RULES

2 / 8

FUZZY PREDICTION

{ WINE C | 1.000 }

FINAL PREDICTION

{ WINE C }

The system delivers the correct decision
compared to the actual case



TOP-MODELS BENCHMARK

MODELING PARAMETERS		XTRACTIS 	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
	MODELS RELEASE	2026/09	2025/08	2025/09	2025/08	2025/09
	ALGORITHM VERSION	XTRACTIS REVEAL v.14.1.61285	XTRACTIS BENCHMARK module embedding Python 3.9.10 Scikit-Learn 1.5.2 LightGBM 4.5.0 PyTorch 2.5.1 +cpu Keras 3.7.0			
	CROSS-VALIDATION TECHNIQUE	40 × 5 folds for each CVE model. Then for each IVE model: 100% Training Validation on original points	40 × 5 folds for each CVE model			
	NUMBER OF EXPLORED STRATEGIES ⁽¹⁾	2,000 induction strategies for the CVE on Training / Validation data. 3 aggregation operators tested. 2,000 induction strategies for the IVE on synthetic data	2,000 ML strategies on Training / Validation data. Aggregation operator: Relative Majority			
	TOP-MODEL SELECTION ⁽²⁾	Top-CVE with Average of Normalized Possibilities if IVE expressed ≥ 50% aggregator selected among the 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	NUMBER OF PREDICTORS (out of 13 Potential Predictors)	7	13	11	13	13
	AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION	2.9 per rule	13.0 per equation	2.3 per rule	3.0 per rule	13.0 per equation
	STRUCTURE OF THE DECISION SYSTEM	8 fuzzy rules without chaining (aggregated into 3 disjunctive fuzzy rules) Only a few rules are triggered at a time to compute a decision	3 linear equations	12 trees without chaining 59 binary rules	3 chains of 11 trees each 224 binary rules Tree #N corrects the error of the N-1 previous trees	2 hidden layers 16 hidden nodes 19 equations 16 unintelligible synthetic variables

TOP-MODEL SCORES		Random ⁽³⁾	XTRACTIS	LoR	RFo	BT	NN
	INTELLIGIBILITY Score ⁽⁴⁾		4.96	0.60	4.49	0.00	0.00
	CVE Real Perf. (Average F ₁ -Score) in External Test	63.78	100.00	100.00	96.59	96.59	96.59
	Gap to CVE Leader in External Test	-36.22	0.00	0.00	-3.41	-3.41	-3.41
	IVE Real Perf. (Average F ₁ -Score) in External Test	63.78	100.00	100.00	96.59	100.00	96.59
	Gap to IVE Leader in External Test	-36.22	0.00	0.00	-3.41	0.00	-3.41
	Average Real Performance in External Test	-36.84	100.00	100.00	96.59	98.30	96.59
			0.00	0.00	-3.41	-1.71	-3.41

UC25

(1) For all algos: on exactly the same splits of the Learning Dataset. All Models are optimized according to their Validation Average F₁-Score.
(2) All top-models are selected according to their Validation Average F₁-Score while checking that it remains close to their Training Average F₁-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. Perfect results of XTRACTIS and LoR on External Test could be explained by a low number of unknown External Test points.

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 & 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).

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; F₁-Score for a Binomial Classification; Average F₁-Score or Average F₂-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

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

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):

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

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):

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

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):

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

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):

$$\text{Pen4}(T_i) = \min(0, 1 - \text{number of trees per chain})$$

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

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

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 T_i 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	Average Sensitivity	Min. PPV	Average PPV	Min. F ₁ -Score	Average F ₁ -Score	Weighted Av. F ₁ -Score	Refusal
RANDOM MODEL									
Number of Random Permutations (P-value) = 100,000 (0.001)									
Performance against chance (External Test)	33.33%	63.64%	63.78%	63.64%	63.78%	63.64%	63.78%	62.96%	
XTRACTIS TOP-MODEL									
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
CVE - Predictive Performance (Validation)	0.66%	98.33%	99.44%	98.04%	99.35%	99.01%	99.39%	99.34%	0 (0.00%)
CVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
IVE - Predictive Performance (151 orig. pts)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
IVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	0 (0.00%)
LOGISTIC REGRESSION TOP-MODEL									
CVE - Descriptive Performance (Training)	1.99%	95.00%	98.33%	95.35%	97.80%	97.44%	98.02%	98.01%	
CVE - Predictive Performance (Validation)	2.65%	93.33%	97.78%	93.18%	97.07%	96.47%	97.34%	97.34%	
CVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Descriptive Performance (Training)	1.99%	95.00%	98.33%	95.35%	97.80%	97.44%	98.02%	98.01%	
IVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
RANDOM FOREST TOP-MODEL									
CVE - Descriptive Performance (Training)	0.66%	98.33%	99.44%	98.04%	99.35%	99.01%	99.39%	99.34%	
CVE - Predictive Performance (Validation)	2.65%	96.00%	97.56%	96.67%	97.41%	96.67%	97.48%	97.34%	
CVE - Real Performance (External Test)	3.70%	88.89%	96.30%	91.67%	97.22%	94.12%	96.59%	96.27%	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	3.70%	88.89%	96.30%	91.67%	97.22%	94.12%	96.59%	96.27%	
BOOSTED TREE TOP-MODEL									
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Predictive Performance (Validation)	1.99%	97.56%	97.96%	96.72%	98.09%	97.52%	98.02%	98.02%	
CVE - Real Performance (External Test)	3.70%	88.89%	96.30%	91.67%	97.22%	94.12%	96.59%	96.27%	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
NEURAL NETWORK TOP-MODEL									
CVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Predictive Performance (Validation)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
CVE - Real Performance (External Test)	3.70%	88.89%	96.30%	91.67%	97.22%	94.12%	96.59%	96.27%	
IVE - Descriptive Performance (Training)	0.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	
IVE - Real Performance (External Test)	3.70%	88.89%	96.30%	91.67%	97.22%	94.12%	96.59%	96.27%	

The entirety of this document is protected by copyright. All rights are reserved, particularly the rights of reproduction and distribution. Quotations from any part of the document must necessarily include the following reference:
Zalila, Z., Idagrai Labs & Xtractis (2019-2026). XTRACTIS® the General Reasoning AI for Trusted Decisions. Use Case #25 | Fraud Detection: Physicochemical Identification of Wines' Origin – Benchmark vs. Logistic Regression, Random Forest, Boosted Tree & Neural Network. IDAGRAI LABS, September 2026, v2.0, France, 6p.