XTRACTIS®

CRM

PREDICTION OF TELECOM CUSTOMER CHURNING

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

UC#22 – 2025/06 (v2.0) xtractis.ai

PROBLEM DEFINITION

GOAL

Design an AI-based decision-making system that accurately evaluates customer attrition, to assign a churning score to a customer in a rational way and adapt anti-churning strategies.

PROS & BENEFITS

- Identify parameters actually involved in customers' decision to unsubscribe and help analysts understand cause-and-effect relationships between specific characteristics, their combination, and a churning risk.
- Help the CRM team focus only on meaningful cases and take earlier and more personalized anti-churn actions thanks to rapid, systematic, and explainable alerts.
- Reduce the turnover of Telecom company's customers.

REFERENCE DATA

Source:
IBM Sample Data Sets, Telco Customer churn.

Dataset
www.kaggle.com, 2017

Variable to Predict: **CHURNING SCORE** ∈ [0 ; 1]

Potential Predictors: **18 variables describe customers and their using habits** (17 nominal + 1 numeric): gender, senior citizen, partner, dependents, tenure, multiple lines, internet service, online security, online backup, device protection, technical support, streaming TV, streaming movies, contract, paperless billing, payment method, monthly charges, total charges.

Observations: **7,003 reference customer files** (40 duplicates removed from the original dataset). Data are divided into a Learning Dataset for model induction using Training, Validation, and Test Datasets, and an External Test Dataset to check the top- model's performance on real data and for benchmarking.

Learning Dataset*: 5,952 cases | 85.0% 70% for Training, 15% for Validation, 15% for Test	
Retained Customer	Leaving Customer
4,382 | 73.6%	1,570 | 26.4%

External Test Dataset**: 1,051 cases | 15.0%	
Retained Customer	Leaving Customer
774 | 73.6%	277 | 26.4%

**6 | 0.01% missing values; **5 | 0.03% 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 33 gradual rules without chaining aggregated into 16 disjunctive rules.
- Each rule uses from 1 to 7 predictors among the 10 variables that XTRACTIS automatically identified as significant (out of the 18 features characterizing consumers).
- Only a few rules are triggered at a time to compute the decision.

☑ **High Predictive Capacity** It has a quite 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	Churning Score 0 1 Automated Warning
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
v13.0.45610

1. We launch 2,000 inductive reasoning strategies; each strategy is applied to the same single partition of the learning dataset (70% Training / 15% Validation / 15% Test) to get a reliable assessment of the descriptive and predictive performances, respectively from Training and Validation Datasets.
2. Each strategy thus generates one unitary model called **Individual Virtual Expert** (IVE).
3. Among the 2,000 induced models, the top-IVE selected is the one that has the best predictive performance, close to its descriptive performance, and with the best intelligibility, i.e. with the fewer predictors and rules.

Total number of induced unitary models
2,000 IVEs

Criterion for the induction optimization
Gini

Validation criterion for the top-model selection
Gini

Duration of the process
@ Induction Speed FP64
17 hours & 20 minutes
@ 1.13Tflops

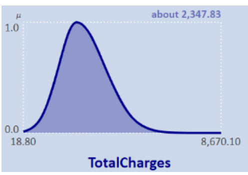
TOP-MODEL STRUCTURE

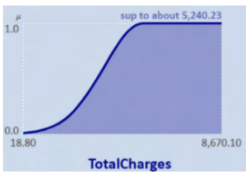
The top-IVE model has an excellent intelligibility as it has **33 rules** combining **10 predictors**, with 2.9 predictors per rule on average.

Its Structure Report reveals all the internal logic of the decision system and ensures that the model is understandable by the human expert. It is a transparent model that can be audited and certified before deployment to end-users.

PREDICTORS

- 10 features out of 18 (9 nominal + 1 numeric)
- Ranked by individual contribution
  (2 strong, 6 medium & 2 weak signals):
  #1 TotalCharges
  #2 InternetService…
- Labeled by modalities or fuzzy classes
  Examples: **fuzzy number** "about 2,347.83";
  fuzzy interval "sup to about 5,240.23"





RULES

- 33 connective fuzzy rules without chaining
  (aggregated into 16 disjunctive fuzzy rules)
- 1 to 7 predictors per rule
  (on average, 2.9 predictors per rule)
- Example: **fuzzy rule R26** uses 5 predictors and concludes {0.941}. 32 other rules complete this model.

IF	TechSupport	IS	No
AND	StreamingMovies	IS	Yes
AND	Contract	IS	Month-to-month
AND	PaymentMethod	IS	{Bank transfer (automatic), Electronic check}
AND	TotalCharges	IS	sup to ~$5,240
THEN	Churning Score	IS	{0.941}

Literally, churning risk is at 0.941 (very high), if no technical support is provided, and the customer uses movies streaming, and the contract is renewed month to month, and payment is made through automatic bank transfer or electronic check and the total charges are higher than around 5,240 dollars

TOP-MODEL PERFORMANCE

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

Performance Type Dataset	DESCRIPTIVE 70% Training	PREDICTIVE 15% Validation	REAL 15% Test	REAL External Test
Gini	**70.37%**	**70.52%**	**69.44%**	**68.74%**
% churn at 26th centile	63.01%	65.67%	60.94%	59.85%

EXPLAINED PREDICTIONS FOR 2 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v13.0.45610

CASE
(from the External Dataset,
i.e., not used in Training/Validation/Test)

DEDUCTIVE INFERENCE OF RULES

AUTOMATED DECISION

Customer #5318-IXUZF

actual value = 1.000

TotalCharges	183.75
SeniorCitizen	No
Partner	No
InternetService	Fiber Optic
DeviceProtection	No
TechSupport	No
StreamingTV	Yes
StreamingMovies	No
Contract	Month-to-month
PaymentMethod	Bank transfer (automatic)

Real Time

For this customer case, 5 rules are triggered with different firing degrees as follows:

# Rule	Conclusion	Firing Degree
R10 →	0.120	0.441
R13 →	0.329	0.024
R20 →	0.754	1.000
R21 →	0.782	1.000
R23 →	0.867	1.000
Prediction	0.711	

NUMBER OF TRIGGERED RULES
5 / 33

FUZZY PREDICTION
{ 0.754 | 1.000,
0.782 | 1.000,
0.867 | 1.000,
0.120 | 0.441,
0.329 | 0.024 }

FINAL PREDICTION
0.711

The decision system delivers a rather high churning score approaching the actual value.

Customer #8067-NIOYM

actual value = 0.000

TotalCharges	490.65
SeniorCitizen	No
Partner	Yes
InternetService	No
DeviceProtection	*No internet service*
TechSupport	*No internet service*
StreamingTV	*No internet service*
StreamingMovies	*No internet service*
Contract	Two years
PaymentMethod	Credit card (automatic)

Real Time

For this customer case, only **R5** is triggered with a firing degree of 1.000:

# Rule	Conclusion	Firing Degree
R5 →	0.000	1.000
Prediction	0.000	

NUMBER OF TRIGGERED RULES
1 / 33

FUZZY PREDICTION
{ 0.000 | 1.000 }

FINAL PREDICTION
0.000

The decision system delivers the same "no-churning" diagnosis as in the observed case.

TOP-MODELS BENCHMARK

MODELING PARAMETERS

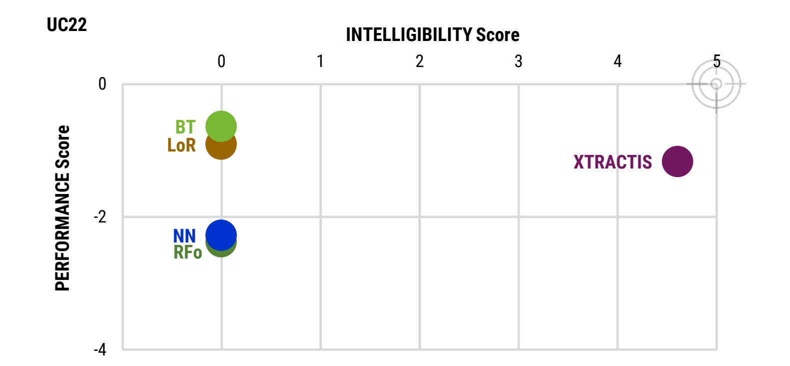
	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
MODELS RELEASE	2023/07	2023/07	2023/07	2023/07	2023/07
ALGORITHM VERSION	XTRACTIS **REVEAL** 13.0.45610	Python 3.9.10 | Scikit-learn 1.3.0	Python 3.9.10 | LightGBM 3.3.2	Python 3.9.10 | LightGBM 3.3.2	Python 3.9.10 | TensorFlow 2.10.0 | Keras 2.10.0
CROSS-VALIDATION TECHNIQUE	All explored strategies for all algorithms use the same single-split of the Learning Dataset: 70% Training | 15% Validation | 15% Test				
NUMBER OF EXPLORED STRATEGIES(1)	2,000 induction strategies	2,000 data analysis strategies	2,000 ML strategies	2,000 ML strategies	2,000 ML strategies
TOP-MODEL SELECTION(2)	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs	Top-IVE among 2,000 IVEs

TOP-MODEL STRUCTURE

	XTRACTIS	LOGISTIC REGRESSION	RANDOM FOREST	BOOSTED TREE	NEURAL NETWORK
NUMBER OF PREDICTORS (out of 18 Potential Predictors)	**10**	**39** 10 nominal variables are decomposed into 31 binary variables	**17**	**18**	**39** 10 nominal variables are decomposed into 31 binary variables
AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION	**2.9** per rule	**39.0** per equation	**3.6** per rule	**4.1** per rule	**31.1** per equation
STRUCTURE OF THE DECISION SYSTEM	**33** fuzzy rules without chaining (aggregated into 16 disjunctive fuzzy rules) Only a few rules are triggered at a time to compute a decision	**1** linear equation	**78** trees without chaining **1,421** binary rules	**1** chain of **33** trees **584** binary rules Tree #N corrects the error of the N-1 previous trees	**5** hidden layers | **135** hidden nodes **136** equations 135 unintelligible synthetic variables

TOP-MODEL SCORES

	Random(3)	XTRACTIS	LoR	RFo	BT	NN
INTELLIGIBILITY Score(4)		4.61	0.00	0.00	0.00	0.00
IVE Real Perf. (Gini) in Test		69.44%	70.83%	69.63%	71.77%	69.49%
Gap to IVE Leader in Test		**-2.33**	**-0.94**	**-2.14**	**0.00**	**-2.28**
IVE Real Perf. (Gini) in External Test	*10.29%*	68.74%	67.86%	66.12%	67.46%	66.46%
Gap to IVE Leader in External Test		**0.00**	**-0.88**	**-2.62**	**-1.28**	**-2.28**
IVE Average Real Performance		69.09%	69.35%	67.88%	69.62%	67.98%
PERFORMANCE Score(4)		**-1.17**	**-0.91**	**-2.38**	**-0.64**	**-2.28**



(1) For all algos: on the same Learning Dataset. All models are optimized according to their Validation Gini.
(2) All top-models are selected according to their Validation Gini while checking that it remains close to their Training Gini.
(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∈[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, 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

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

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):

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

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):

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

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):

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

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):

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

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):

$$\mathrm{Pen5}(T_i) = -5$$

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	Lift / Lift max at 26th centile	% of churners at 26th centile	Max Separability	ROC AUC	**Gini**	Refusal
RANDOM MODEL						
Number of Random Permutations (P-value) = 100,000 (0.001%)						
Performance against chance (External Test)	*1.29 / 3.79*	*33.94%*	*33.94%*	*55.15%*	*10.29%*	
XTRACTIS TOP-MODEL						
Descriptive Performance (Training)	2.39 / 3.79	63.01%	76.68%	85.18%	70.37%	0 (0.00%)
Predictive Performance (Validation)	2.49 / 3.78	65.67%	79.15%	85.26%	70.52%	0 (0.00%)
Real Performance (Test)	2.31 / 3.78	60.94%	75.00%	84.72%	**69.44%**	0 (0.00%)
Real Performance (External Test)	2.27 / 3.79	59.85%	76.90%	84.37%	**68.74%**	0 (0.00%)
LOGISTIC REGRESSION TOP-MODEL						
Descriptive Performance (Training)	2.33 / 3.79	61.44%	76.41%	84.49%	68.99%	
Predictive Performance (Validation)	2.34 / 3.78	61.80%	76.69%	84.31%	68.61%	
Real Performance (Test)	2.36 / 3.78	62.23%	76.41%	85.42%	**70.83%**	
Real Performance (External Test)	2.24 / 3.79	59.12%	77.91%	83.93%	**67.86%**	
RANDOM FOREST TOP-MODEL						
Descriptive Performance (Training)	2.37 / 3.79	62.55%	76.30%	85.39%	70.78%	
Predictive Performance (Validation)	2.29 / 3.78	60.52%	77.32%	84.67%	69.35%	
Real Performance (Test)	2.25 / 3.78	59.40%	76.26%	84.81%	**69.63%**	
Real Performance (External Test)	2.22 / 3.79	58.39%	75.45%	83.06%	**66.12%**	
BOOSTED TREE TOP-MODEL						
Descriptive Performance (Training)	2.48 / 3.79	65.41%	79.23%	87.50%	74.99%	
Predictive Performance (Validation)	2.45 / 3.78	64.81%	77.54%	85.05%	70.10%	
Real Performance (Test)	2.37 / 3.78	62.66%	76.27%	85.89%	**71.77%**	
Real Performance (External Test)	2.27 / 3.79	59.85%	76.36%	83.73%	**67.46%**	
NEURAL NETWORK TOP-MODEL						
Descriptive Performance (Training)	2.38 / 3.79	62.64%	76.60%	84.69%	69.38%	
Predictive Performance (Validation)	2.42 / 3.78	63.95%	77.02%	84.99%	69.98%	
Real Performance (Test)	2.31 / 3.78	60.94%	76.69%	84.74%	**69.49%**	
Real Performance (External Test)	2.24 / 3.79	59.12%	76.49%	83.23%	**66.46%**	