Security & Smart Cities

# PREDICTION OF ROBBERY CRIMES IN AMERICAN CITIES

Benchmark vs. Random Forest, Boosted Tree & Neural Network

UC#23 – 2024/06 (v1.2)

xtractis.ai

## PROBLEM DEFINITION

**GOAL**

**Design an AI-based decision system that accurately predicts the number of robberies committed in a year in a city, given the hyper-complexity of the phenomenon (highly nonlinear behavior) in order to identify sources of crime and warn about the criminality level in a rational and explainable way.**

**PROS & BENEFITS**

- Allow city administrators and police departments to understand the causal relationships between socio-economic factors or security policy and future criminal activities.
- Find the truly influential parameters for assessing criminality to define an effective city policy.
- Carry out dedicated police actions to lower criminality, in accordance with population or macro-economic changes.

**REFERENCE DATA**

Source:
Michael Redmond, Computer Science, La Salle University, Philadelphia, USA

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 **Robberies per 100k population** which is a continuous variable ∈ [0 ; 2,264]. The Per Capita attribute allows to manage little communities and big cities as well.

**Potential Predictors:** **102 variables from 1990 US Census and crime data** from the 1995 FBI UCR: population characteristics such as origins, racial type, incomes, wages, graduation, employment, marital status, family type and size, …

LEMAS predictors from the 1990 US LEMAS survey were not used due to lack of information: 84.16% of missing values - police departments with less than 100 officers were omitted in data.

**Observations:** **2,214 cities, each is associated with a value of robberies per 100k population.** 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: 1,881 cases \| 85% | External Test Dataset: 333 cases \| 15% |
|---|---|
| 70% for Training, 15% for Validation, 15% for Test | |

**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 18 gradual rules without chaining.
- Each rule uses from 2 to 4 predictors among the 14 variables that XTRACTIS automatically identified as significant (out of the 102 features characterizing robbery crime records).
- 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).

☑ **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 | Automated Prediction (warn about crime level) |
|---|---|---|---|---|---|---|
| | | | | | | Robberies per 100k inhabitants: 0 – 2,264 |
| SOFTWARE ROBOTS | XTRACTIS® REVEAL<br>Delivers the decision system + its Structure & Performance Reports | | | XTRACTIS® PREDICT<br>Delivers the decision + the Prediction Report explaining its reasoning | | |

## TOP-MODEL INDUCTION

### INDUCTION PARAMETERS

Powered by:

XTRACTIS® REVEAL v13.0.47764

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
**RMSE**

Validation criterion for the top-model selection
**RMSE**

Duration of the process @ Induction Speed FP64
**~2 days** @ 1Tflops

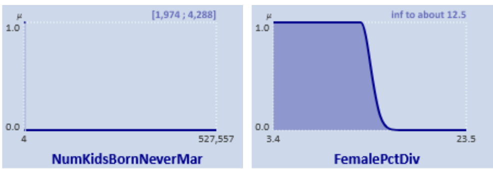
### TOP-MODEL STRUCTURE

The top-IVE has an excellent intelligibility as it has **18 rules** combining **14 predictors**, with 2.3 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

- 14 parameters out of 102
- Ranked by impact significance
  (7 strong, 6 medium and 1 weak signal):
  #1 PctHousLess3BR
  #2 PctKidsBornNeverMar…
- Labeled by fuzzy and binary classes
  Examples: **binary interval** "[1,974 ; 4,288]"
  **fuzzy interval** "inf to about 12.5"




#### RULES

- 18 connective fuzzy rules without chaining (aggregated into 16 disjunctive fuzzy rules)
- 2 to 4 predictors per rule (on average, 2.3 predictors per rule)
- Example: fuzzy rule R2 uses 3 predictors and concludes {2}. 17 other rules complete this model.

| | | | |
|---|---|---|---|
| IF | FemalePctDiv | IS | inf to ~12.5 |
| AND | PctKidsBornNeverMar | IS | inf to ~4.1 |
| AND | PctForeignBorn | IS | inf to ~18.7 |
| THEN | Robberies per 100k Population | IS | 2 |

*Literally, the city has a very low robbery level (2 per 100k inhabitants) if the percentage of divorced women is less than around 12.5%, and the percentage of kids born to never-married people is less than around 4.1%, and the percentage of foreign-born people is less than about 18.7%.*

### 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 | DESCRIPTIVE | PREDICTIVE | REAL | REAL |
|---|---|---|---|---|
| Dataset | 70% Training | 15% Validation | 15% Test | External Test |
| RMSE | 100 (4.44%) | 100 (4.40%) | 138 (6.08%) | 132 (5.85%) |
| Correlation | 0.880 | 0.923 | 0.866 | 0.893 |

# EXPLAINED PREDICTIONS FOR 2 UNKNOWN CASES

Powered by: XTRACTIS® PREDICT v13.0.47764

**CASE**
(from the External Test Dataset,
i.e., not used in the Learning Dataset)

**DEDUCTIVE INFERENCE OF RULES**

**AUTOMATED DECISION**

## Beavercreek - OHIO

| actual value = 53 | |
|---|---|
| racepctblack | 0.9 |
| racepctwhite | 96.4 |
| racepctasian | 2.3 |
| pctUrban | 100.0 |
| NumUnderPov | 1,170 |
| PctEmplManu | 18.2 |
| FemalePctDiv | 6.4 |
| TotalPctDiv | 5.8 |
| NumKidsBornNeverMar | 77 |
| PctKidsBornNeverMar | 0.3 |
| PctImmigRecent | 15.1 |
| PctHouseLess3BR | 16.8 |
| PctForeignBorn | 3.4 |
| PopDens | 1,261 |

Real Time

For this city, 5 rules are triggered
to conclude to 41 Robberies per 100k population,
a very low robbery:

| # Rule | Conclusion | Firing Degree |
|---|---|---|
| R2 → | 2 | 1.000 |
| R1 → | 2 | 0.407 |
| R3 → | 15 | 1.000 |
| R4 → | 43 | 1.000 |
| R5 → | 105 | 1.000 |
| PREDICTION = | 41 | |

NUMBER OF TRIGGERED RULES
5/18

FUZZY PREDICTION
{ 2 | 1.000,
15 | 1.000,
43 | 1.000,
105 | 1.000}

FINAL PREDICTION
41

The decision system delivers an accurate prediction compared to the actual value.

VERY LOW robbery level


## Atlanta - GEORGIA

| actual value = 1,495 | |
|---|---|
| racepctblack | 67.1 |
| racepctwhite | 31.1 |
| racepctasian | 0.9 |
| pctUrban | 100.0 |
| NumUnderPov | 102,000 |
| PctEmplManu | 9.4 |
| FemalePctDiv | 18.8 |
| TotalPctDiv | 17.1 |
| NumKidsBornNeverMar | 43,400 |
| PctKidsBornNeverMar | 15.6 |
| PctImmigRecent | 19.9 |
| PctHouseLess3BR | 63.7 |
| PctForeignBorn | 3.4 |
| PopDens | 2,886 |

Real Time

For this city, 3 rules are triggered
to conclude to 1,288 Robberies per 100k population,
a rather high robbery:

| # Rule | Conclusion | Firing Degree |
|---|---|---|
| R7 → | 233 | 0.018 |
| R16 → | 1,317 | 0.520 |
| R17 → | 1,542 | 0.017 |
| PREDICTION = | 1,288 | |

NUMBER OF TRIGGERED RULES
3/18

FUZZY PREDICTION
{ 1,317 | 0.520,
233 | 0.018,
1,542 | 0.017}

FINAL PREDICTION
1,288

The decision system delivers an accurate prediction compared to the actual value.

RATHER HIGH robbery level

# TOP-MODELS BENCHMARK

**MODELING PARAMETERS**

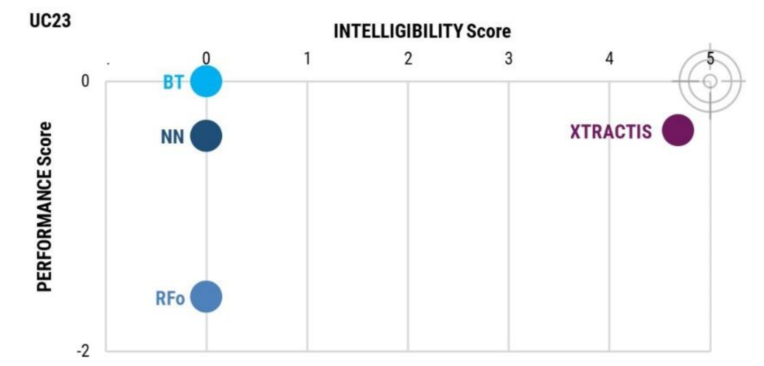
| | XTRACTIS | RANDOM FOREST | BOOSTED TREE | NEURAL NETWORK |
|---|---|---|---|---|
| **MODELS RELEASE** | 2023/11 | 2023/11 | 2023/11 | 2023/11 |
| **ALGORITHM VERSION** | XTRACTIS **REVEAL** 13.0.47764 | Python 3.9 \| LightGBM 3.3.2 | Python 3.9 \| LightGBM 3.3.2 | Python 3.9 \| TensorFlow 2.10.0 \| 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 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-MODEL STRUCTURE**

| | XTRACTIS | RANDOM FOREST | BOOSTED TREE | NEURAL NETWORK |
|---|---|---|---|---|
| **NUMBER OF PREDICTORS** (out of 102 Potential Predictors) | **14** | **80** | **101** | **102** |
| **AVERAGE NUMBER OF PREDICTORS PER RULE OR EQUATION** | **2.3** per rule | **6.7** per rule | **19.0** per rule | **66.7** per equation |
| **STRUCTURE OF THE DECISION SYSTEM** | **18** fuzzy rules without chaining (aggregated into 16 disjunctive fuzzy rules)<br>Only a few rules are triggered at a time to compute a decision | **5** trees without chaining<br>**349** binary rules | **1** chain of **159** trees<br>**16,747** binary rules<br>Tree #N corrects the error of the N-1 previous trees | **3** hidden layers \| **56** hidden nodes<br>**57** equations<br>56 unintelligible synthetic variables |

**TOP-MODEL SCORES**

| | *Random*(3) | XTRACTIS | RFo | BT | NN |
|---|---|---|---|---|---|
| **INTELLIGIBILITY Score**(4) | | 4.68 | 0.00 | 0.00 | 0.00 |
| IVE Real Perf. (RMSE_%) in Test | | 6.08 | 7.24 | 5.53 | 5.86 |
| **Gap to IVE Leader in Test** | | **-0.55** | **-1.71** | **0.00** | **-0.33** |
| IVE Real Perf. (RMSE_%) in External Test | *14.85* | 5.85 | 7.15 | 5.67 | 6.15 |
| **Gap to IVE Leader in External Test** | | **-0.18** | **-1.48** | **0.00** | **-0.48** |
| IVE Average Real Performance | | 5.97 | 7.20 | 5.60 | 6.01 |
| **PERFORMANCE Score**(4) | | **-0.37** | **-1.60** | **0.00** | **-0.41** |




(1) For all algos: on the same Learning Dataset. All Models are optimized according to their Validation RMSE.
(2) All top-models are selected according to their Validation RMSE while checking that it remains close to their Training RMSE.
(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]<br>n = number of AI Techniques benchmarked in terms of data-driven modeling = 5 |
| Benchmark #k | $B_k$ | k∈[1 ; p]<br>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 $T_i$

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

*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):

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

*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 trees per chain, here for BT only):

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

*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):

$$Pen5(T_i) = -5$$

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 | Correlation | MAE | **RMSE** | Refusal |
|---|---|---|---|---|
| **RANDOM MODEL** | | | | |
| *Number of Random Permutations (P-value) = 100,000 (0.001)* | | | | |
| *Performance against chance (External Test)* | *0.223* | *199 (8.79%)* | *336 (14.85%)* | |
| **XTRACTIS TOP-MODEL** | | | | |
| Descriptive Performance (Training) | 0.880 | 64 (2.83%) | 100 (4.44%) | 0 (0.00%) |
| Predictive Performance (Validation) | 0.923 | 63 (2.76%) | 100 (4.40%) | 0 (0.00%) |
| Real Performance (Test) | 0.866 | 82 (3.60%) | **138 (6.08%)** | 0 (0.00%) |
| Real Performance (External Test) | 0.893 | 81 (3.56%) | **132 (5.85%)** | 0 (0.00%) |
| **RANDOM FOREST TOP-MODEL** | | | | |
| Descriptive Performance (Training) | 0.887 | 53 (2.32%) | 97 (4.31%) | |
| Predictive Performance (Validation) | 0.889 | 66 (2.92%) | 115 (5.10%) | |
| Real Performance (Test) | 0.801 | 89 (3.92%) | **164 (7.24%)** | |
| Real Performance (External Test) | 0.807 | 87 (3.83%) | **162 (7.15%)** | |
| **BOOSTED TREE TOP-MODEL** | | | | |
| Descriptive Performance (Training) | 0.999 | 2 (0.08%) | 9 (0.41%) | |
| Predictive Performance (Validation) | 0.913 | 61 (2.69%) | 105 (4.64%) | |
| Real Performance (Test) | 0.892 | 72 (3.20%) | **125 (5.53%)** | |
| Real Performance (External Test) | 0.899 | 72 (3.19%) | **128 (5.67%)** | |
| **NEURAL NETWORK TOP-MODEL** | | | | |
| Descriptive Performance (Training) | 0.896 | 59 (2.62%) | 96 (4.24%) | |
| Predictive Performance (Validation) | 0.906 | 63 (2.79%) | 106 (4.66%) | |
| Real Performance (Test) | 0.875 | 75 (3.31%) | **133 (5.86%)** | |
| Real Performance (External Test) | 0.858 | 75 (3.53%) | **139 (6.15%)** | |