



# Scalable Neurosymbolic Knowledge Discovery from Tabular Data

Erkan Karabulut  
e.karabulut@uva.nl

8 December 2025



# Table of Contents

## Part I: Scaling Knowledge Discovery

- Rule mining for knowledge discovery and interpretable machine learning
- Using Neurosymbolic methods to scale rule mining

## Part II: Using Prior/Background Knowledge for Knowledge Discovery

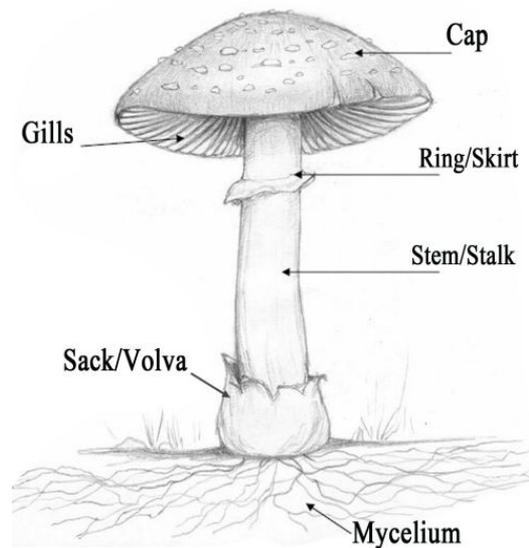
- Semantic association rules from tables with auxiliary knowledge graphs
- Transfer learning from tabular foundation models to boost knowledge discovery

# **Part I: Scaling Knowledge Discovery (co-occurrence patterns)**

# Association Rule Mining

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

[archive.ics.uci.edu/dataset/73/mushroom](https://archive.ics.uci.edu/dataset/73/mushroom)



<https://www.yellowelanor.com/portfolio/mushroom-identification-basics/>

# Association Rule Mining

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

Association rule **examples:**

*cap-shape(bell) → poisonous(yes)*

*cap-color(brown) ∧ poisonous(no) → habitat(woods)*

Patterns/rules can also include: negations, disjunctions, existential quantifiers ...

# Association Rule Mining

[nature](#) > [nature machine intelligence](#) > [perspectives](#) > [article](#)

Perspective | Published: 13 May 2019

## Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

[Cynthia Rudin](#)

[Nature Machine Intelligence](#) 1, 206–215 (2019) | [Cite this article](#)

93k Accesses | 7299 Citations | 538 Altmetric | [Metrics](#)

A [preprint version](#) of the article is available at arXiv.

**Relevance** of rule learning in current AI

## Knowledge Discovery

Find non-trivial, implicit, previously unknown and potentially useful (interesting) patterns in the data. Think of medical uses ...

## Interpretable Inference

Interpretable by design ML models rather than probabilistic explanation of black box models.

**Bonus:** inference with rules is the fastest!

# Association Rule Mining

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

How to measure how “good” the rules are?

$R1 = \text{cap-shape}(\text{convex}) \rightarrow \text{cap-color}(\text{brown})$

$\text{support}(R1) = P(\text{cap-shape}=\text{convex} \wedge \text{cap-color}=\text{brown})$   
 $= 5/8 = 62.5\%$ .

$\text{confidence}(R1) = P(\text{cap-color}=\text{brown} \mid \text{cap-shape}=\text{convex})$   
 $= 5/5 = 100\%$ .

# Association Rule Mining

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

**Do we optimize for high values?**

**High support** → Trends in the data

$$\text{support}(\text{cap-color}=\text{brown} \rightarrow \text{habitat}=\text{woods}) \\ = 6/8 = 75\%.$$

**Low support** → anomalies, minority classes ...

$$\text{support}(\text{cap-shape}(\text{bell}) \rightarrow \text{poisonous}(\text{yes})) \\ = 2/8 = 25\%.$$

# Association Rule Mining

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

**Do we optimize for high values?**

**High confidence x strong association**

$R3 = \text{cap-shape}(\text{convex}) \rightarrow \text{habitat}(\text{woods})$   
 $\text{confidence}(R3) = 100\%$

**But ...**

$\text{support}(\text{habitat}(\text{woods})) = \sim 85\%$ .

Many other feature will imply *habitat(woods)* anyways.

$\text{cap-color}(\text{brown}) \rightarrow \text{habitat}(\text{woods})$  100% confidence  
 $\text{odor}(\text{almond}) \rightarrow \text{habitat}(\text{woods})$  100% confidence

...

# Association Rule Mining

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

**Association strength:**

$$\frac{\text{confidence}(A \rightarrow C) - \text{confidence}(A' \rightarrow C)}{\max(\text{confidence}(A \rightarrow C), \text{confidence}(A' \rightarrow C))}$$

$$\max(\text{confidence}(A \rightarrow C), \text{confidence}(A' \rightarrow C))$$

range [-1, 1].

-1: dissociation, 0: independence, 1: association.

$$\begin{aligned} \text{Assoc\_strength}(\text{habitat}(\text{woods}) \rightarrow \text{cap-color}(\text{brown})) \\ = (1 - 0) / 1 = 100\%. \end{aligned}$$

Again, one very common pattern implies another common pattern. Not that interesting!

# Association Rule Mining

## Key takeaway

→ We define what is a “good” pattern and often require a hypothesis over the data.

Developing new “interestingness” functions is a prominent research direction in knowledge discovery.

# Association Rule Mining (Apriori)

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

## How to find such patterns?

→ Given minimum support and confidence thresholds, find all such patterns that satisfy the thresholds.

Apriori (1994) is still the most commonly used tabular rule miner today!!

# Association Rule Mining (Apriori)

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

1 item(set)	occurrence
cap-shape(bell)	2
cap-shape(convex)	5
cap-shape(flat)	1
...	...

2 item(set)	occurrence
cap-shape(convex), cap-color(brown)	5
...	...

Minimum support threshold = 50%

Minimum confidence threshold = 80%

Step 1: Gradually find item(sets) that satisfies the min. support threshold

Agrawal, R. Srikant, and Ramakrishnan Srikant. "R. Fast algorithms for mining association rules." Proceedings of the 20th international conference on very large data bases, vldb. 1994.

# Association Rule Mining (Apriori)

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

2 item(set)	occurrence
cap-shape(convex), cap-color(brown)	5
...	...

*cap-shape(convex) → cap-color(brown), 100% conf.*

*cap-color(brown) → cap-shape(convex), ~71% conf.*

Minimum support threshold = 50%

Minimum confidence threshold = 80%

Step 2: Create combinations of implications

# Association Rule Mining - Research Problems

## Research problems

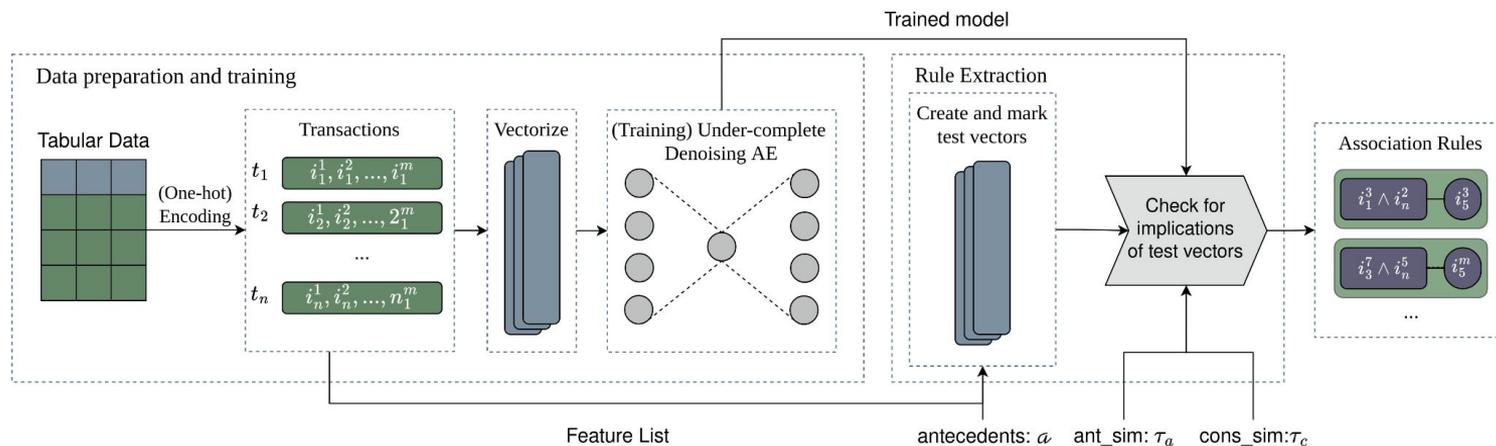
1. **Rule explosion:** Combinatorial explosion in the number of rules based on columns
2. **Interpretability:** A table of 20 columns with max antecedent length of 3 can produce  $C(20,1) \times 3^1 + C(20,2) \times 3^2 + C(20,3) \times 3^3 \sim 1.85$  million rules!

**How to find a small number of interesting patterns fast?**

# Our Approach - Aerial

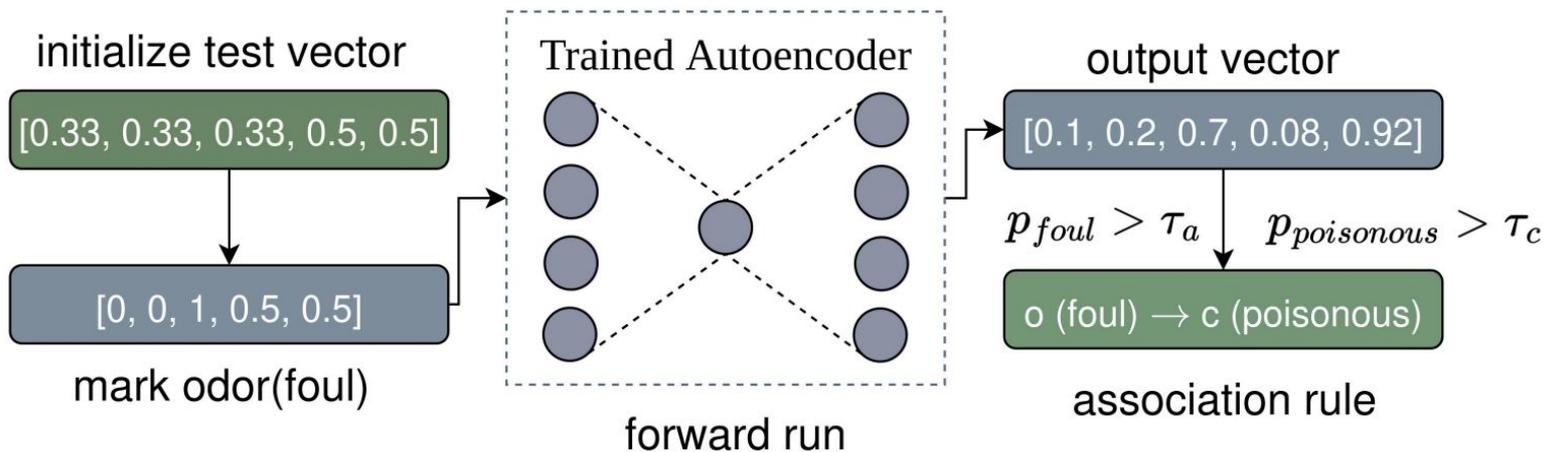
Steps:

1. Use an under-complete Autoencoder to create a neural representation of the data.
2. Extract association rules from the neural representation.



# Aerial - Rule Extraction (Step 2)

odor = {creosote, fishy, foul}, class = {edible, poisonous}  $\tau_a = 0.5$ ,  $\tau_c = 0.8$

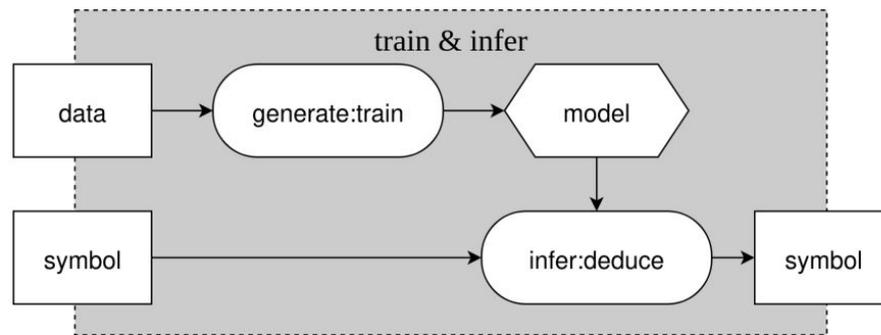


# Scalable Rule ‘Learning’ with Aerial

## Why this works?

1. No counting co-occurrences!  
→ 1000s of times faster execution time on large tables.
2. Successful reconstruction necessitates/implies strong associations!
3. Small model: 2-3 layers of encoder and decoder for mid-sized tables (~100 columns, <100.000 rows).

# Scalable Rule ‘Learning’ with Aerial



**Figure 2:** Boxology [van Bekkum et al., 2021] diagram of neurosymbolic ARM approaches such as Aerial+: i) a neural *model* of *data* (i.e., tabular data) is learned, ii) an algorithm (symbolic) *infers* rules (symbols) from the model using hypotheses (symbols, as in test vectors of Aerial+).

# Scalable Rule ‘Learning’ with Aerial

Successful reconstruction necessitates/implies strong associations!

cap-shape	cap-color	odor	habitat	poisonous
bell	brown	foul	woods	yes
convex	brown	none	woods	no
convex	brown	none	woods	no
convex	brown	none	woods	no
bell	red	foul	urban	yes
convex	brown	none	woods	no
flat	brown	almond	woods	no
convex	brown	none	woods	no

Encoding ‘*habitat(woods)*  $\rightarrow X$ ’ won’t help reconstructing the data!

Because *habitat(woods)* is there in the data almost always anyways. Knowing that won’t help reconstructing any other data point!

Redundant rules are eliminated by design!

# Aerial - Quantitative Insights

Linear  $O(n)$  time in training, polynomial  $O(k^{a+1})$  time in rule extraction, for  $k$  features and  $a$  antecedents.

---

**Algorithm 1:** Aerial+'s rule extraction algorithm from a trained autoencoder

---

**Input:** Trained autoencoder:  $AE$ , max antecedents:  $a$ , similarity thresholds  $\tau_a, \tau_c$

**Output:** Extracted rules  $\mathcal{R}$

```

1  $\mathcal{R} \leftarrow \emptyset, \mathcal{F} \leftarrow AE.input\_feature\_categories;$ 
2 foreach  $i \leftarrow 1$  to  $a$  do
3    $\mathcal{C} \leftarrow \binom{\mathcal{F}}{i};$ 
4   foreach  $S \in \mathcal{C}$  do
5      $\mathbf{v}_0 \leftarrow \text{UniformProbabilityVectorPerFeature}(\mathcal{F});$ 
6      $\mathcal{V} \leftarrow \text{MarkFeatures}(S, \mathbf{v}_0)$ 
7     foreach  $\mathbf{v} \in \mathcal{V}$  do
8        $\mathbf{p} \leftarrow AE(\mathbf{v});$ 
9       if  $\min_{f \in S} p_f < \tau_a$  then
10         $S.low\_support \leftarrow \text{True};$ 
11        continue with the next  $\mathbf{v};$ 
12        foreach  $f \in \mathcal{F} \setminus S$  do
13          if  $p_f > \tau_c$  then  $\mathcal{R} \leftarrow \mathcal{R} \cup \{(S \rightarrow f)\};$ 
14   $\mathcal{F} \leftarrow \{f \in \mathcal{F} \mid f.low\_support = \text{False}\};$ 
15 Return  $\mathcal{R};$ 

```

---

# Aerial - Quantitative Insights

A small number of high quality rules with full data coverage!

Algorithm	#Rules	Time (s)	Cov.	Support	Conf.	Algorithm	#Rules	Time (s)	Cov.	Support	Conf.
<b>Congressional Voting Records</b>						<b>Breast Cancer</b>					
BAT	1913	208	1	0.06	0.45	BAT	787.1	162.18	1	0.07	0.41
GW	2542	186	1	0.05	0.48	GW	1584	129.18	1	0.08	0.42
SC	7	186	0.46	0.01	0.43	SC	33.6	137.66	1	0.03	0.27
FSS	10087	272	1	0.01	0.71	FSS	6451.6	225.71	1	0.02	0.36
FP-G   HMine	1764	0.09   0.04	1	0.29	0.88	FP-G   HMine	94	0.01   0.01	1	0.34	<b>0.87</b>
ARM-AE	347	0.21	0.03	0.23	0.45	ARM-AE	131	0.09	0.01	0.19	0.27
<b>Aerial+</b>	149	0.25	1	0.32	<b>0.95</b>	<b>Aerial+</b>	50	0.19	1	0.39	0.86
<b>Mushroom</b>						<b>Chess</b>					
BAT	1377.2	225.57	1	0.1	0.62	BAT	2905.9	235.34	1	0.17	0.64
GW	1924.1	184.56	1	0.11	0.63	GW	5605.25	255.56	1	0.31	0.65
SC	1.33	281.84	0.07	0.02	0.48	SC	1	545.71	0	0	0.7
FSS	794.9	352.99	1	0.04	0.38	FSS	32.75	380.73	0.4	0	0.36
FP-G   HMine	1180	0.1   0.07	1	0.43	0.95	FP-G   HMine	30087	12.43   0.7	1	0.46	0.93
ARM-AE	390	0.33	0	0.22	0.23	ARM-AE	22052	26.98	0.02	0.39	0.54
<b>Aerial+</b>	321	0.38	1	0.44	<b>0.96</b>	<b>Aerial+</b>	16522	0.22	1	0.45	<b>0.95</b>
<b>Spambase</b>											
BAT	0	424	No rules found								
GW	0	508	No rules found								
SC	0	643	No rules found								
FSS	0	677	No rules found								
FP-G   HMine	125223	21.4   2.14	1	0.64	0.92						
ARM-AE	85327	254	0.03	0.31	0.38						
<b>Aerial+</b>	43996	1.92	1	0.62	<b>0.97</b>						

# Aerial - Quantitative Insights

Small number of rules carry equal or higher predictive power as exhaustive mining!

Dataset	Algorithm	# Rules or Items		Accuracy		Exec. Time (s)	
		Exhaustive	Aerial+	Exhaustive	Aerial+	Exhaustive	Aerial+
Congressional Voting Records	CBA	3437	<b>1495</b>	91.91	<b>92.66</b>	0.34	<b>0.14</b>
	BRL	2547	<b>57</b>	<b>96.97</b>	<b>96.97</b>	15.37	<b>9.69</b>
	CORELS	4553	<b>61</b>	<b>96.97</b>	<b>96.97</b>	3.04	<b>0.17</b>
Mushroom	CBA	27800	<b>2785</b>	<b>99.82</b>	<b>99.82</b>	1.75	<b>1.30</b>
	BRL	5093	<b>493</b>	<b>99.87</b>	99.82	244	<b>167</b>
	CORELS	23271	<b>335</b>	90.14	<b>99.04</b>	61	<b>2</b>
Breast Cancer	CBA	695	<b>601</b>	66.42	<b>71.13</b>	<b>0.08</b>	0.28
	BRL	2047	<b>290</b>	71.13	<b>71.46</b>	16.82	<b>14.5</b>
	CORELS	2047	<b>369</b>	73.69	<b>75.82</b>	1.42	<b>0.40</b>
Chess	CBA	49775	<b>34490</b>	<b>94.02</b>	93.86	24.31	<b>6.24</b>
	BRL	19312	<b>1518</b>	<b>96.21</b>	95.93	321	<b>119</b>
	CORELS	37104	<b>837</b>	81.1	<b>93.71</b>	106	<b>3.87</b>
Spambase	CBA	125223	<b>33418</b>	84.5	<b>85.42</b>	23.87	<b>7.56</b>
	BRL	37626	<b>5190</b>	72.78	<b>84.93</b>	1169	<b>431</b>
	CORELS	275003	<b>1409</b>	85.37	<b>87.28</b>	1258	<b>5.23</b>

# PyAerial

The screenshot shows the GitHub README for the PyAerial project. At the top, there are links for 'README' and 'MIT license'. The main heading is 'pyaerial: scalable association rule mining'. Below this, there are several status badges: 'python 3.9,3.10,3.11,3.12', 'pypi v1.0.4', 'CI passing', 'license MIT', 'last commit last sunday', 'Ubuntu 24.04 LTS', 'macOS Monterey 12.6.7', and 'Stars 13'. A navigation bar contains links for 'Install', 'Usage', 'Functions', 'Debug', 'Cite', 'License', and 'Contribute'. The main text describes PyAerial as a Python implementation of the Aerial scalable neurosymbolic association rule miner for tabular data, highlighting its use of an under-complete denoising Autoencoder and its ability to extract high-quality association rules with full data coverage.

# PyAerial

## Installation

You can easily install **pyaerial** using pip:

```
pip install pyaerial
```

```
from aerial import model, rule_extraction, rule_quality
from ucimlrepo import fetch_ucirepo

# load a categorical tabular dataset from the UCI ML repository
breast_cancer = fetch_ucirepo(id=14).data.features

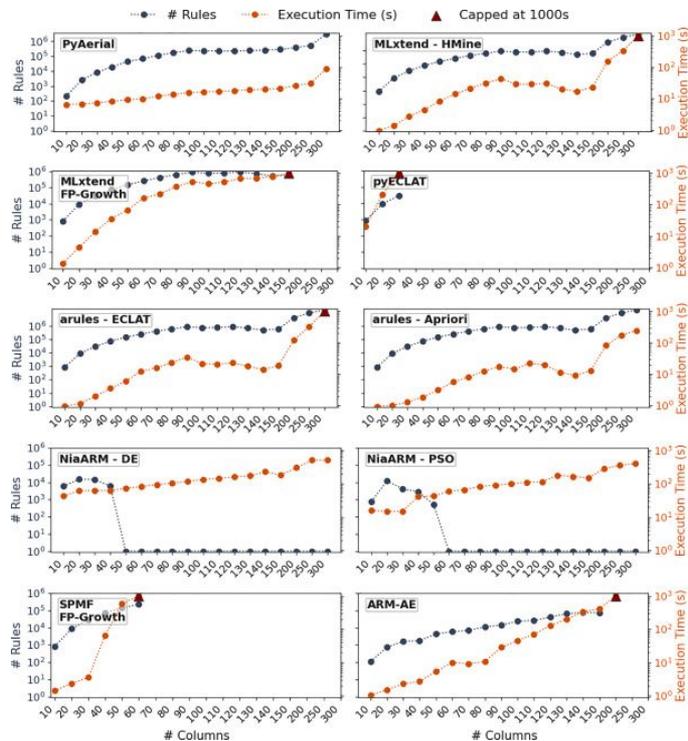
# train an autoencoder on the loaded table
trained_autoencoder = model.train(breast_cancer, device="cuda")

# extract association rules from the autoencoder
association_rules = rule_extraction.generate_rules(trained_autoencoder)

# calculate rule quality statistics (support, confidence, zhangs metric) for each rule
if len(association_rules) > 0:
    stats, association_rules = rule_quality.calculate_rule_stats(association_rules, trained_aut
    print(stats, association_rules[:1])
```

**Features:** ARM with item constraints, classification rules, visualizations ...

# PyAerial



E. Karabulut, P. Groth, and V. Degeler. "PyAerial: Scalable association rule mining from tabular data". SoftwareX, 31:102341, 2025. ISSN 2352-7110.

GitHub: <https://github.com/DiTEC-project/pyaerial>



**Reconstruction success implies associations!**

# Research Questions

1. How to extend “associations by reconstruction success” to numerical data?
  - a. During rule extraction, start with larger intervals (implies many other features), gradually narrow it down, until no rules found.
  - b. Or make use of optimization methods to find “optimal” intervals for given rule quality criteria
  
2. How to extract negation rules (or any other logical relation)?
  - a. If high reconstruction probability means co-occurrence, low reconstruction probability means negation?
  - b. Patterns over absence of features can be as valuable as their presence.

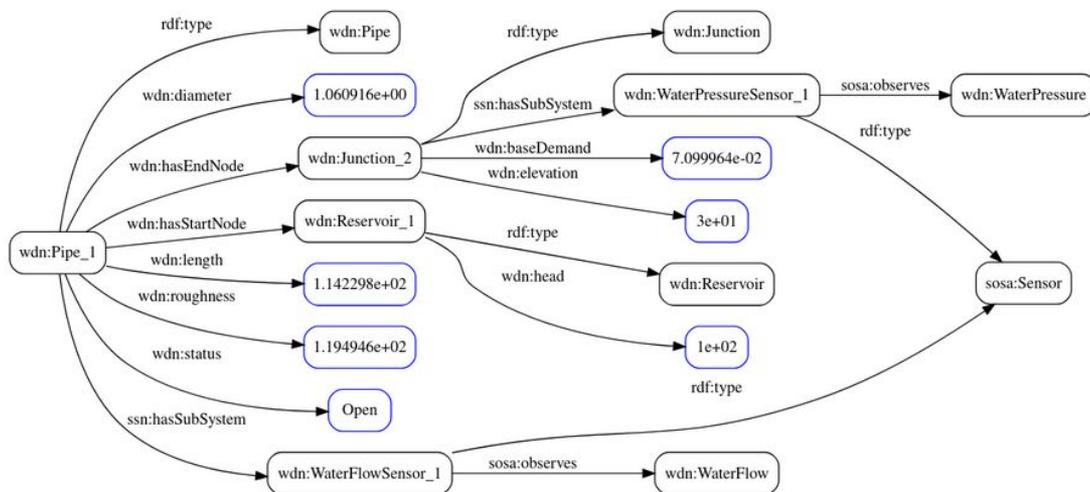
# Part II: Using Prior/Background Knowledge for Knowledge Discovery

# Incorporating Back. Knowledge/Context

Tables do not exist in isolation.

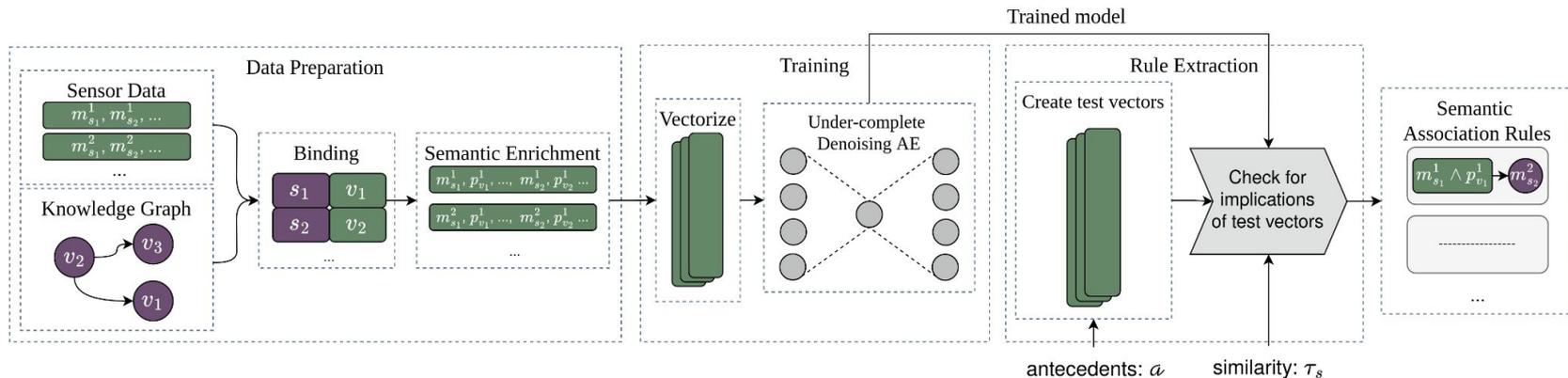
time	value	type	name
2017-12-31 23:30:00.000	187.2	demand	s_Junction_2
2017-12-31 23:30:00.000	3,981.6	flow	s_Pipe_1
2017-12-31 23:00:00.000	4,406.4	flow	s_Pipe_1
2017-12-31 23:00:00.000	205.2	demand	s_Junction_2
2017-12-31 22:30:00.000	4,791.6	flow	s_Pipe_1
2017-12-31 22:30:00.000	205.2	demand	s_Junction_2
2017-12-31 22:00:00.000	212.4	demand	s_Junction_2
2017-12-31 22:00:00.000	5,173.2	flow	s_Pipe_1
2017-12-31 21:30:00.000	5,508	flow	s_Pipe_1
2017-12-31 21:30:00.000	270	demand	s_Junction_2

Sensor measurements table



Knowledge Graph

# Semantic Association Rules



**Numerical association rules:**  $\text{sensor1.value} \geq 18 \wedge \text{sensor1.value} \leq 20 \rightarrow \text{sensor2.value} \geq 20 \wedge \text{sensor2.value} \leq 22$

**Semantic association rules:**  $\text{type}(s1, \text{WP}) \wedge \text{type}(p1, \text{Pipe}) \wedge \text{placed\_in}(s1, p1) \wedge p1.\text{diameter} \geq A1 \wedge s1.\text{value} \geq 18 \wedge s1.\text{value} \leq 20 \rightarrow \text{type}(s2, \text{WP}) \wedge \text{type}(j1, \text{Junction}) \wedge \text{placed\_in}(s2, j1) \wedge \text{connected}(j1, p1) \wedge s2.\text{value} \geq 20 \wedge s2.\text{value} \leq 22$

# Semantic Association Rules - Insights

Same results as before:

- Concise set of high-quality rules (as the # sensors increase) with full data coverage
- Better scalability, etc.

In addition: semantics enable learning more generalizable rules.

	# Rules		Support		Rule Cov.		Confidence	
	w-s	wo-s	w-s	wo-s	w-s	wo-s	w-s	wo-s
<b>LeakDB</b>								
FP-Growth	103K	9K	<b>0.41</b>	0.19	<b>0.43</b>	0.2	0.95	0.97
Aerial	554	2.5K	<b>0.54</b>	0.25	<b>0.59</b>	0.3	0.91	0.87
<b>L-Town</b>								
FP-Growth	25K	5K	<b>0.86</b>	0.36	<b>0.9</b>	0.38	0.96	0.96
Aerial	1K	2.5K	<b>0.59</b>	0.39	<b>0.65</b>	0.45	0.91	0.88
<b>LBNL</b>								
FP-Growth	7K	2K	<b>0.84</b>	0.73	<b>0.85</b>	0.75	0.98	0.99
Aerial	73	258	<b>0.74</b>	0.65	<b>0.74</b>	0.66	1.0	0.99

Caveats: a naive approach, can not say yet that we learn ‘semantics’

# Research Questions - Part 2

1. How to measure how much semantics does a rule contain?
  - a. Semantic rule learning → Class learning?
2. Given a set of tabular medical data, and a medical KG describing the semantic relations for the columns of the table:
  - a. Can we ‘discover’ the graph from the table itself?
  - b. Can we discover unknown relations in the graph
    - i. and verify in lab experiments?
    - ii. or verify on a masked part of the graph?
3. Extending “associations by reconstruction success” over the GNNs of KGs and/or tables with auxiliary KGS?

# Aerial in Low-Data Regime

Using neural networks has its downsides!

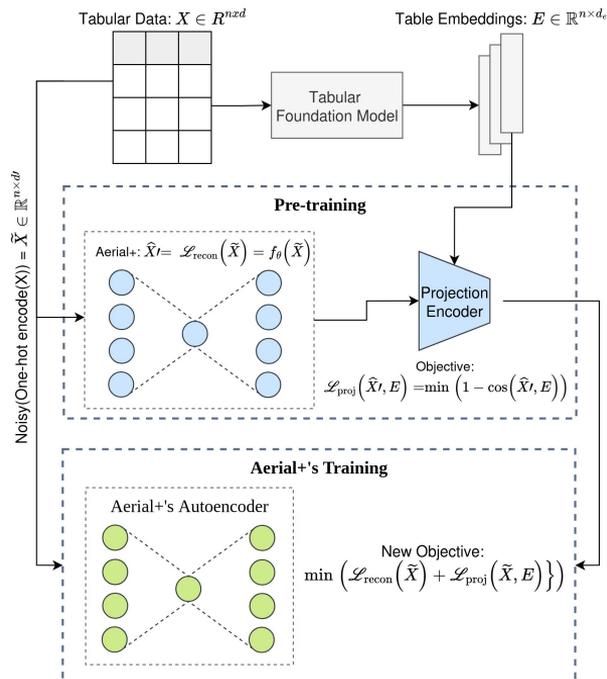
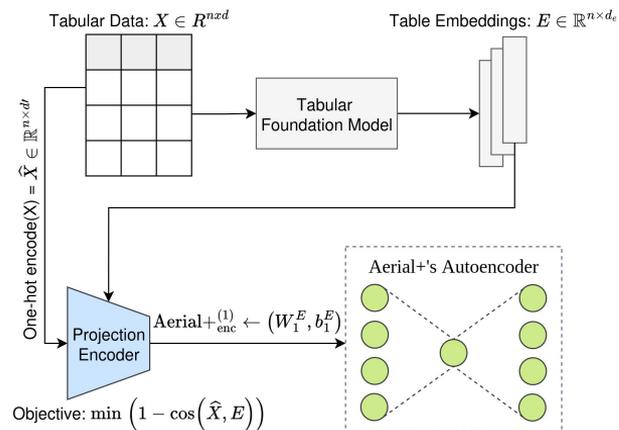
High-dimensional datasets with low number of samples (gene expression datasets, rare disease datasets etc.)

Sample / Rule	Gene_1	Gene_2	Gene_3	...	Gene_18107	Gene_18107
Sample_1	normal	normal	normal	...	normal	normal
Sample_2	normal	normal	high	...	normal	high
Sample_3	normal	normal	normal	...	normal	low
			...			
<b>Rule_1</b>	Gene2 (high) $\wedge$ Gene29 (high) $\rightarrow$ Gene14 (low)					
<b>Rule_2</b>	Gene3 (high) $\wedge$ Gene45 (high) $\rightarrow$ Gene84 (high)					

# Aerial in Low-Data Regime

Prior knowledge from tabular foundation models!

A.k.a., transfer learning for knowledge discovery.



# Aerial in Low-Data Regime

Even less rules with higher average quality!

Approach	# Rules	~Rule Coverage	~Support	~Confidence	Data Coverage	~Zhang's Metric	Exec. Time (s)
<b>Chondrosarcoma</b>							
Aerial+	200	<b>0.23</b>	0.21	0.921	0.533	0.784	2.25
Aerial+WI	75	0.217	0.206	0.945	0.524	0.813	5.80
Aerial+DL	75	0.235	0.219	<b>0.947</b>	<b>0.536</b>	<b>0.828</b>	5.36
<b>SmallCellLungCarcinoma</b>							
Aerial+	1576	0.068	0.041	0.579	<b>0.835</b>	0.476	10.58
Aerial+WI	664	<b>0.076</b>	0.052	<b>0.633</b>	0.715	<b>0.577</b>	13.48
Aerial+DL	1338	0.070	0.044	0.597	0.816	0.513	18.23
<b>NonSmallCellLungCarcinoma</b>							
Aerial+	1620	0.059	0.035	0.584	0.823	0.554	18.03
Aerial+WI	978	<b>0.078</b>	0.057	<b>0.663</b>	0.698	<b>0.639</b>	28.67
Aerial+DL	1453	0.053	0.028	0.547	<b>0.849</b>	0.501	24.27
<b>BreastCarcinoma</b>							
Aerial+	1017	0.072	0.046	0.641	<b>0.816</b>	0.575	9.64
Aerial+WI	590	0.077	0.052	<b>0.686</b>	0.686	<b>0.644</b>	12.09
Aerial+DL	535	<b>0.078</b>	0.050	0.652	0.761	0.590	15.31
<b>Melanoma</b>							
Aerial+	1220	0.067	0.035	0.545	<b>0.888</b>	0.440	13.09
Aerial+WI	773	0.070	0.038	<b>0.575</b>	0.772	<b>0.496</b>	13.19
Aerial+DL	859	<b>0.071</b>	0.038	0.566	0.860	0.461	16.49

# Conclusions

1. Rule learning for knowledge discovery and interpretable machine learning.
2. Reconstruction success implies associations.
  - a. Aerial helps learning a concise set of high-quality association rules with full data coverage.
  - b. A concise set of rules still leads to equal or better predictive performance.
  - c. Orders of magnitude better scalability on large tables
3. Including background knowledge, context, prior knowledge is possible and still an open question.



# Scalable Neurosymbolic Knowledge Discovery from Tabular Data

Erkan Karabulut  
[e.karabulut@uva.nl](mailto:e.karabulut@uva.nl)

8 December 2025

# ARM on Numerical Features

