UNIVERSITEIT VAN AMSTERDAM

INDE lab

Scalable Neurosymbolic Knowledge Discovery from Tabular Data

Erkan Karabulut
e.karabulut@uva.nl

23 January 2025

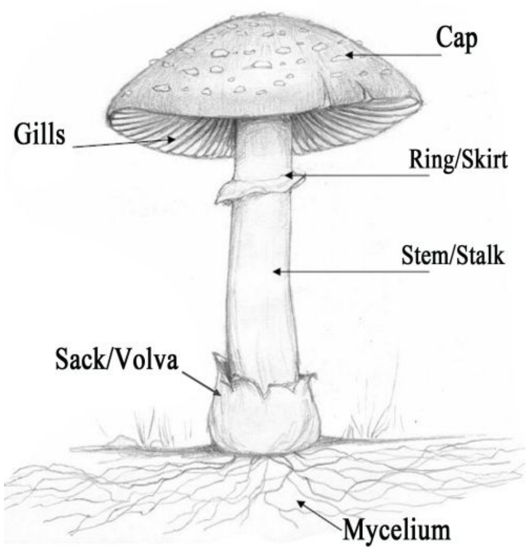
Table of Contents

Part I: Learning interpretable patterns for knowledge discovery and inference

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://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 - Positioning

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 = cap-shape(convex) → cap-color(brown)

support(R1) *= P(cap-shape=convex ∧ cap-color=brown)*
= 5/8 = 62.5%.

confidence(R1) *= P(cap-color=brown | cap-shape=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

support((cap-color=brown → habitat=woods))
= 6/8 = 75%.

Low support → anomalies, minority classes …

support(cap-shape(bell) → poisonous(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 = cap-shape(convex) → habitat(woods)
confidence(R3) = 100%!

But …
support(habitat(woods)) = ~85%.
Many other feature will imply *habitat(woods)* anyways.

cap-color(brown) → habitat(woods) 100% confidence
odor(almond) → habitat(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:

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

range [-1, 1].
-1: dissociation, 0: independence, 1: association.

Assoc_strength(habitat(woods) → cap-color(brown))
= (1 - 0) / 1 = 100%.

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!!

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

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

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 - 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 $C(20,1)\times3^1 + C(20,2)\times3^2 + C(20,3)\times3^3 \sim 1.85$ million rules!

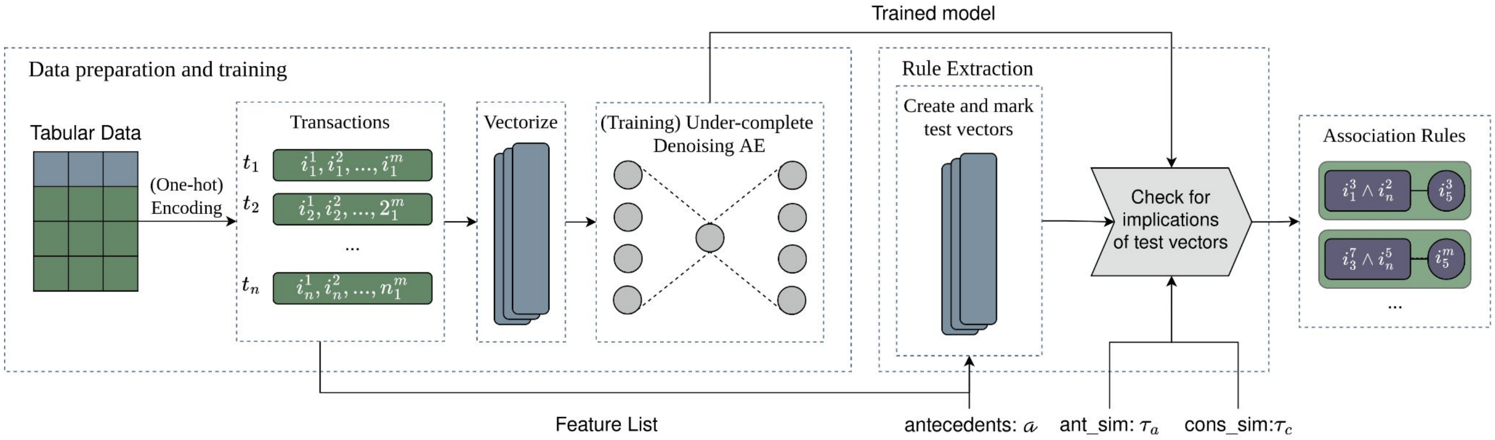
How to find a small number of interesting patterns fast?

Part II: Solving association rule mining with a neurosymbolic approach

Our Approach - Aerial

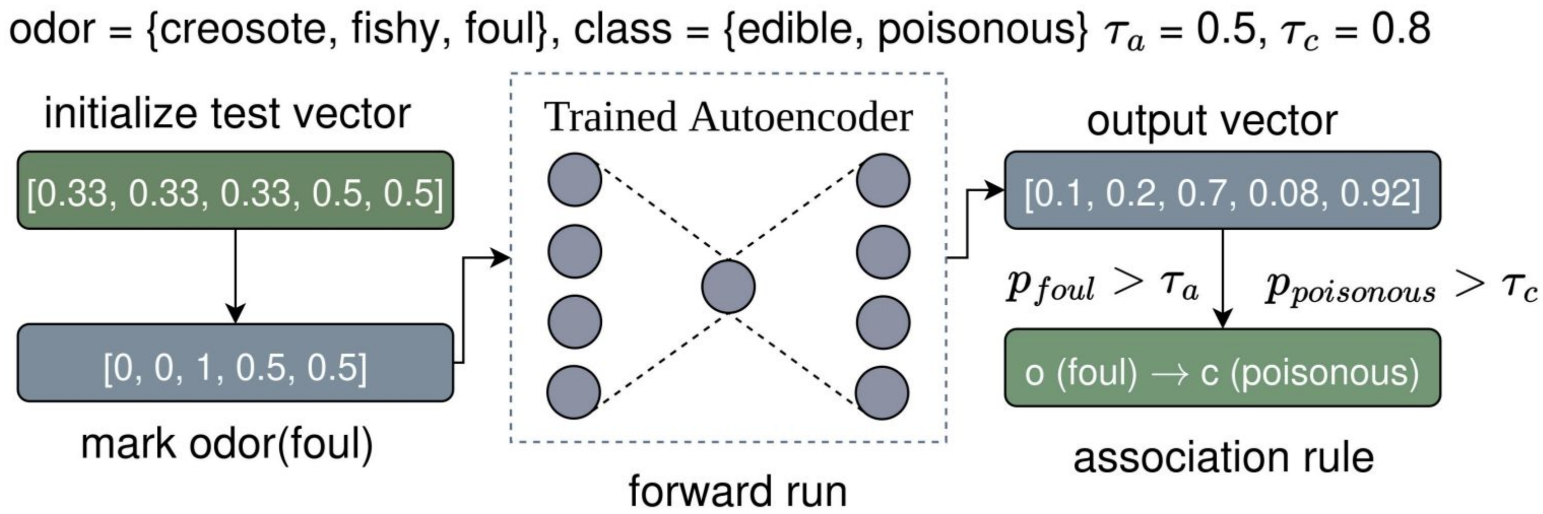
Steps:

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



E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

Aerial - Rule Extraction (Step 2)



E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

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).

E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

Scalable Rule ‘Learning’ with Aerial

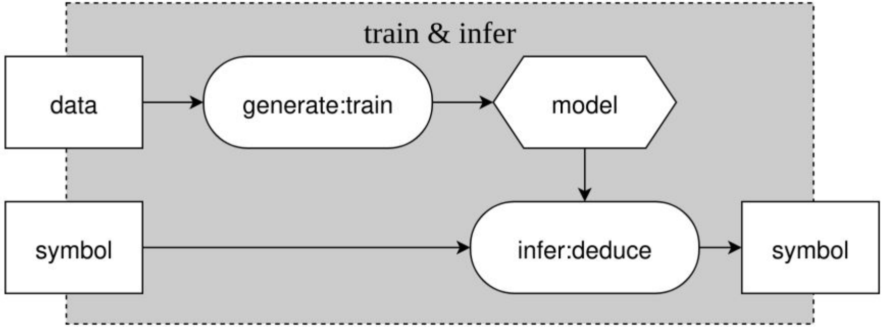


Figure 2: Boxology [van Bekkum et al., 2021] diagram of neurosymbolic ARM approaches such as Aerial+: i) a neural *model* of *data* (i.e., tabular data) is learned, ii) an algorithm (symbolic) *infers* rules (symbols) from the model using hypotheses (symbols, as in test vectors of Aerial+).

E. Karabulut, D. Daza, P. Groth and V. Degeler. "Discovering Association Rules in High-Dimensional Small Tabular Data". In ANSyA'25: 1st International Workshop on Advanced Neuro-Symbolic Applications, co-located with 28th European Conference on Artificial Intelligence (ECAI 2025).

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)* → *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!

E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

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 τ_a, τ_c
Output: Extracted rules $\mathcal{R}$

$\mathcal{R} \leftarrow \emptyset$, $\mathcal{F} \leftarrow AE.input\_feature\_categories$;
foreach $i \leftarrow 1$ **to** a **do**
  $\mathcal{C} \leftarrow \binom{\mathcal{F}}{i}$;
  foreach $S \in \mathcal{C}$ **do**
    $\mathbf{v}_0 \leftarrow \text{UniformProbabilityVectorPerFeature}(\mathcal{F})$;
    $\mathcal{V} \leftarrow \text{MarkFeatures}(S, \mathbf{v}_0)$
    foreach $\mathbf{v} \in \mathcal{V}$ **do**
      $\mathbf{p} \leftarrow AE(\mathbf{v})$;
      if $\min_{f \in S} p_f < \tau_a$ **then**
        S.low_support $\leftarrow$ **True**;
        continue with the next $\mathbf{v}$;
      foreach $f \in \mathcal{F} \setminus S$ **do**
        if $p_f > \tau_c$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{(S \rightarrow f)\}$;
  $\mathcal{F} \leftarrow \{f \in \mathcal{F} \mid f\text{.low\_support} = \textbf{False}\}$;
Return $\mathcal{R}$;

E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

Aerial - Quantitative Insights

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

Algorithm	#Rules	Time (s)	Cov.	Support	Conf.
Congressional Voting Records					
BAT	1913	208	1	0.06	0.45
GW	2542	186	1	0.05	0.48
SC	7	186	0.46	0.01	0.43
FSS	10087	272	1	0.01	0.71
FP-G | HMine	1764	0.09 | 0.04	1	0.29	0.88
ARM-AE	347	0.21	0.03	0.23	0.45
Aerial+	149	0.25	1	0.32	**0.95**
Mushroom					
BAT	1377.2	225.57	1	0.1	0.62
GW	1924.1	184.56	1	0.11	0.63
SC	1.33	281.84	0.07	0.02	0.48
FSS	794.9	352.99	1	0.04	0.38
FP-G | HMine	1180	0.1 | 0.07	1	0.43	0.95
ARM-AE	390	0.33	0	0.22	0.23
Aerial+	321	0.38	1	0.44	**0.96**
Spambase					
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
Aerial+	43996	1.92	1	0.62	**0.97**

Algorithm	#Rules	Time (s)	Cov.	Support	Conf.
Breast Cancer					
BAT	787.1	162.18	1	0.07	0.41
GW	1584	129.18	1	0.08	0.42
SC	33.6	137.66	1	0.03	0.27
FSS	6451.6	225.71	1	0.02	0.36
FP-G | HMine	94	0.01 | 0.01	1	0.34	**0.87**
ARM-AE	131	0.09	0.01	0.19	0.27
Aerial+	50	0.19	1	0.39	0.86
Chess					
BAT	2905.9	235.34	1	0.17	0.64
GW	5605.25	255.56	1	0.31	0.65
SC	1	545.71	0	0	0.7
FSS	32.75	380.73	0.4	0	0.36
FP-G | HMine	30087	12.43 | 0.7	1	0.46	0.93
ARM-AE	22052	26.98	0.02	0.39	0.54
Aerial+	16522	0.22	1	0.45	**0.95**

E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

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	**1495**	91.91	**92.66**	0.34	**0.14**
	BRL	2547	**57**	**96.97**	**96.97**	15.37	**9.69**
	CORELS	4553	**61**	**96.97**	**96.97**	3.04	**0.17**
Mushroom	CBA	27800	**2785**	**99.82**	**99.82**	1.75	**1.30**
	BRL	5093	**493**	**99.87**	99.82	244	**167**
	CORELS	23271	**335**	90.14	**99.04**	61	**2**
Breast Cancer	CBA	695	**601**	66.42	**71.13**	**0.08**	0.28
	BRL	2047	**290**	71.13	**71.46**	16.82	**14.5**
	CORELS	2047	**369**	73.69	**75.82**	1.42	**0.40**
Chess	CBA	49775	**34490**	**94.02**	93.86	24.31	**6.24**
	BRL	19312	**1518**	**96.21**	95.93	321	**119**
	CORELS	37104	**837**	81.1	**93.71**	106	**3.87**
Spambase	CBA	125223	**33418**	84.5	**85.42**	23.87	**7.56**
	BRL	37626	**5190**	72.78	**84.93**	1169	**431**
	CORELS	275003	**1409**	85.37	**87.28**	1258	**5.23**

E. Karabulut, P. Groth, V. Degeler, Neurosymbolic association rule mining from tabular data, in: Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning, volume 284 of Proceedings of Machine Learning Research, PMLR, 2025, pp. 565–588.

PyAerial

README | MIT license

pyaerial: scalable association rule mining

python 3.9,3.10,3.11,3.12 | pypi v1.0.24 | Tests passing | license MIT | Stars 26 | last commit december 2025
Ubuntu 24.04 LTS | macOS Monterey 12.6.7 | DOI 10.5281/zenodo.17795656

📥 Install | 🚀 Quick Start | ✨ Features | 📚 Documentation | 📄 Cite | 🤝 Contribute | 🔑 License

PyAerial is a **Python implementation** of the Aerial scalable neurosymbolic association rule miner for tabular data. It utilizes an under-complete denoising Autoencoder to learn a compact representation of tabular data, and extracts a concise set of high-quality association rules with full data coverage.

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

PyAerial

Installation

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

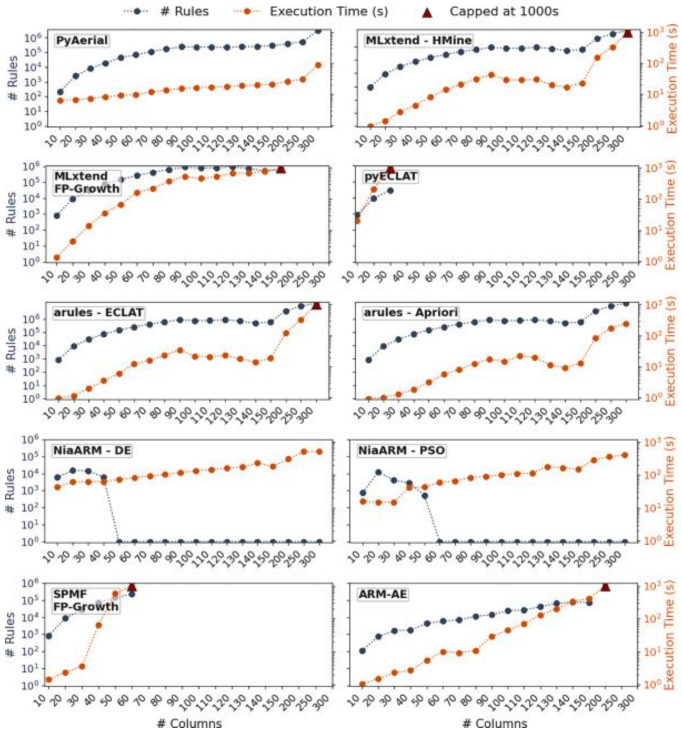
```
pip install pyaerial
```

```
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 …

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

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!

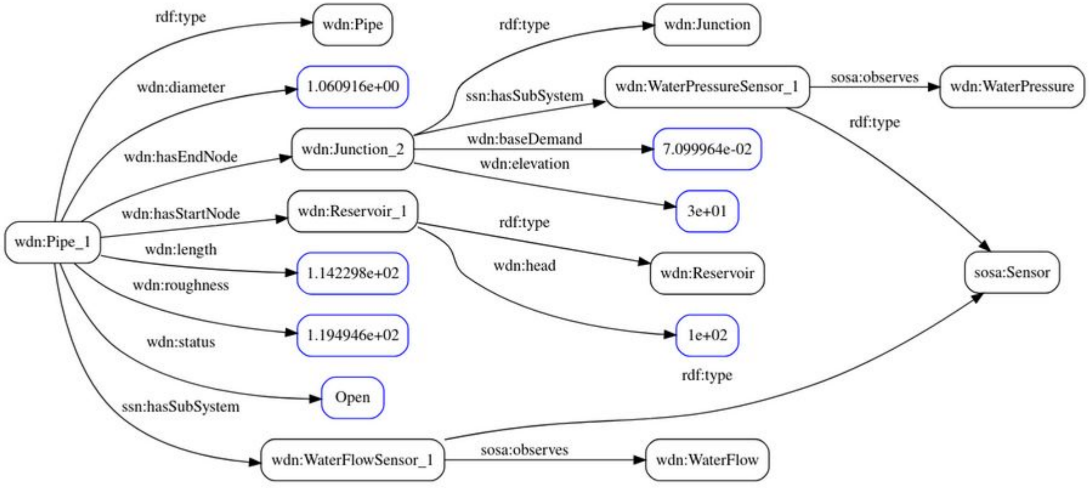
Part III: Using Prior/Background Knowledge for Knowledge Discovery

Incorporating Back. Knowledge/Context

Tables do not exists 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

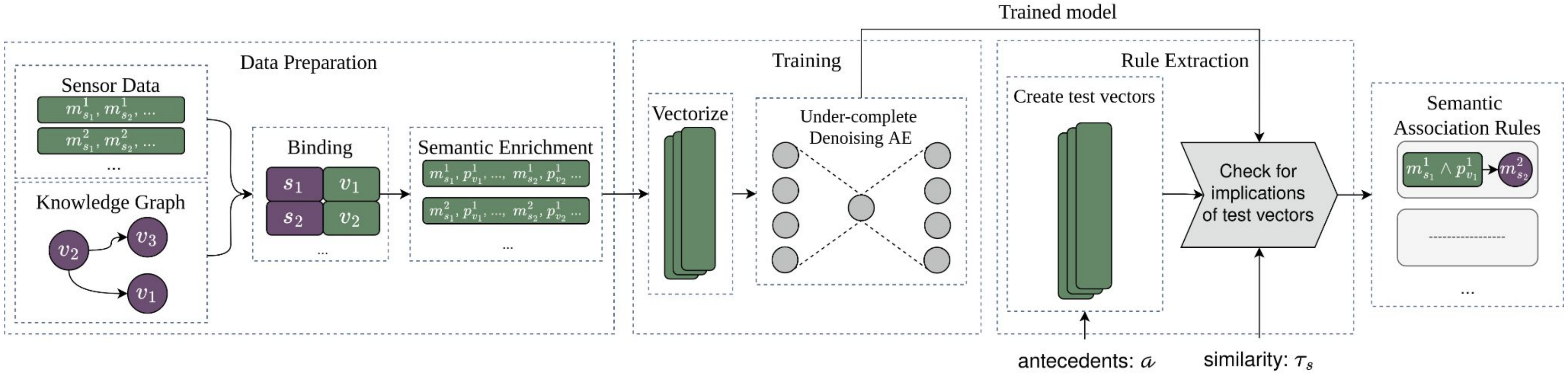
Sensor measurements table



Knowledge Graph

LeakDB Dataset: Vrachimis, Stelios G., and Marios S. Kyriakou. "LeakDB: a Benchmark Dataset for Leakage Diagnosis in Water Distribution Networks:(146)." *WDSA/CCWI Joint Conference Proceedings*. Vol. 1. 2018.

Semantic Association Rules



Numerical association rules: sensor1.value ≥ 18 ∧ sensor1.value ≤ 20→sensor2.value ≥ 20 ∧ sensor2.value ≤ 22

Semantic association rules: type(s1, WP) ∧ type(p1, Pipe) ∧ placed_in(s1, p1) ∧ p1.diameter ≥ A1 ∧ s1.value ≥ 18 ∧ s1.value ≤ 20 → type(s2, WP) ∧ type(j1, Junction) ∧ placed_in(s2, j1) ∧ connected(j1, p1) ∧ s2.value ≥ 20 ∧ s2.value ≤ 22

Erkan Karabulut, Paul Groth, and Victoria Degeler. "Learning Semantic Association Rules from Internet of Things Data". Neurosymbolic Artificial Intelligence, 2025:1. doi:10.1177/29498732251377518.

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
		LeakDB		
FP-Growth	103K | 9K	**0.41** | 0.19	**0.43** | 0.2	0.95 | 0.97
Aerial	554 | 2.5K	**0.54** | 0.25	**0.59** | 0.3	0.91 | 0.87
		L-Town		
FP-Growth	25K | 5K	**0.86** | 0.36	**0.9** | 0.38	0.96 | 0.96
Aerial	1K | 2.5K	**0.59** | 0.39	**0.65** | 0.45	0.91 | 0.88
		LBNL		
FP-Growth	7K | 2K	**0.84** | 0.73	**0.85** | 0.75	0.98 | 0.99
Aerial	73 | 258	**0.74** | 0.65	**0.74** | 0.66	1.0 | 0.99

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

Erkan Karabulut, Paul Groth, and Victoria Degeler. "Learning Semantic Association Rules from Internet of Things Data". Neurosymbolic Artificial Intelligence, 2025:1. doi:10.1177/29498732251377518.

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.)

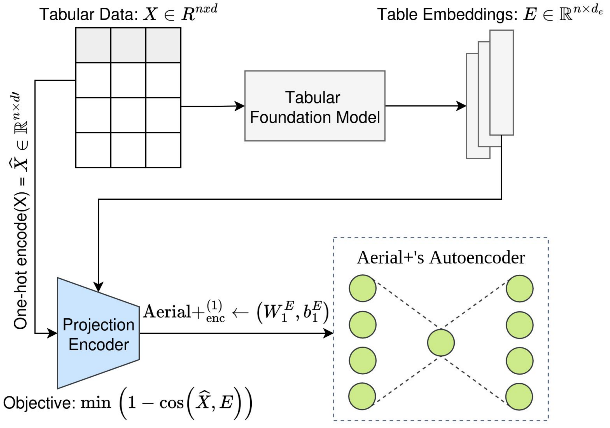
Sample / Rule	**Gene_1**	**Gene_2**	**Gene_3**	⋯	**Gene_18107**	**Gene_18107**
Sample_1	normal	normal	normal	⋯	normal	normal
Sample_2	normal	normal	high	⋯	normal	high
Sample_3	normal	normal	normal	⋯	normal	low
			⋯			
Rule_1	Gene2 (high) $\wedge$ Gene29 (high) $\rightarrow$ Gene14 (low)					
Rule_2	Gene3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

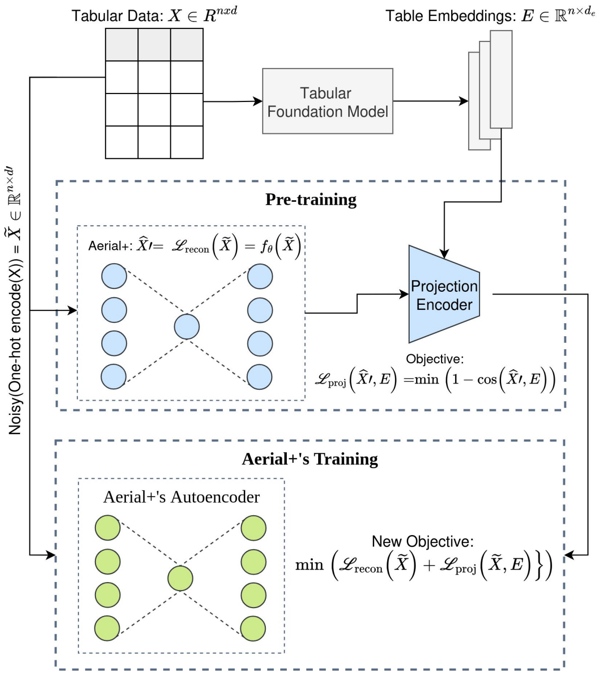
E. Karabulut, D. Daza, P. Groth and V. Degeler. "Discovering Association Rules in High-Dimensional Small Tabular Data". In ANSyA'25: 1st International Workshop on Advanced Neuro-Symbolic Applications, co-located with 28th European Conference on Artificial Intelligence (ECAI 2025).

Aerial in Low-Data Regime

Prior knowledge from tabular foundation models!

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





E. Karabulut, D. Daza, P. Groth and V. Degeler. "Discovering Association Rules in High-Dimensional Small Tabular Data". In ANSyA'25: 1st International Workshop on Advanced Neuro-Symbolic Applications, co-located with 28th European Conference on Artificial Intelligence (ECAI 2025).

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)
				Chondrosarcoma			
Aerial+	200	**0.23**	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	**0.947**	**0.536**	**0.828**	5.36
				SmallCellLungCarcinoma			
Aerial+	1576	0.068	0.041	0.579	**0.835**	0.476	10.58
Aerial+WI	664	**0.076**	0.052	**0.633**	0.715	**0.577**	13.48
Aerial+DL	1338	0.070	0.044	0.597	0.816	0.513	18.23
				NonSmallCellLungCarcinoma			
Aerial+	1620	0.059	0.035	0.584	0.823	0.554	18.03
Aerial+WI	978	**0.078**	0.057	**0.663**	0.698	**0.639**	28.67
Aerial+DL	1453	0.053	0.028	0.547	**0.849**	0.501	24.27
				BreastCarcinoma			
Aerial+	1017	0.072	0.046	0.641	**0.816**	0.575	9.64
Aerial+WI	590	0.077	0.052	**0.686**	0.686	**0.644**	12.09
Aerial+DL	535	**0.078**	0.050	0.652	0.761	0.590	15.31
				Melanoma			
Aerial+	1220	0.067	0.035	0.545	**0.888**	0.440	13.09
Aerial+WI	773	0.070	0.038	**0.575**	0.772	**0.496**	13.19
Aerial+DL	859	**0.071**	0.038	0.566	0.860	0.461	16.49

E. Karabulut, D. Daza, P. Groth and V. Degeler. "Discovering Association Rules in High-Dimensional Small Tabular Data". In ANSyA'25: 1st International Workshop on Advanced Neuro-Symbolic Applications, co-located with 28th European Conference on Artificial Intelligence (ECAI 2025).

Part IV: Tabular Foundation Models Can Learn Association Rules

TFMs Can Learn Rules Out-Of-The-box

Definition 2 (Conditional Probabilistic Model). *Let $\mathcal{M}_\theta$ be a trained model that allows computing conditional queries of the form*

$$P_\theta(i \mid X), \quad i \in I,\ X \subseteq I, \tag{2}$$

interpreted as the probability estimated by the model that item i holds given that all items in X are observed to hold.

Requirement 1 (Antecedent Validation). *A model $\mathcal{M}_\theta$ provides a scoring function $s_\theta : 2^I \to [0,1]$, where $s_\theta(X)$ quantifies how plausible the partial configuration X is under the learned data distribution. An antecedent X is considered valid if $s_\theta(X) \geq \tau_a$, for a user-specified threshold $\tau_a \in [0,1]$.*

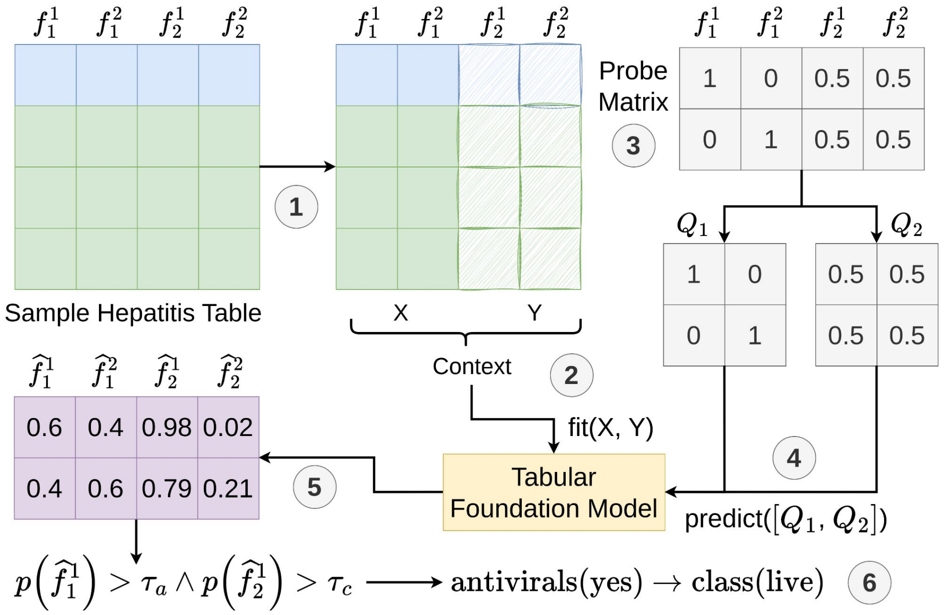
Intuitively, this prevents the extraction of rules based on highly unlikely or spurious combinations of feature values (e.g., rare co-occurrences in the clinical data).

Requirement 2 (Consequent Extraction). *Given a valid antecedent X, the model allows computing conditional probabilities $P_\theta(i \mid X)$ for all $i \in I \setminus X$. An item i is accepted as a consequent of X if $P_\theta(i \mid X) \geq \tau_c$, for a user-specified threshold $\tau_c \in [0,1]$.*

Karabulut, Erkan, Daniel Daza, Paul Groth, Martijn C. Schut, and Victoria Degeler. "Tabular Foundation Models Can Learn Association Rules" (pre-print out soon).

TFMs Can Learn Rules Out-Of-The-box

$$\{f_1^1, f_1^2, f_2^1, f_2^2\} = \{\text{antivirals(yes)}, \text{antivirals(no)}, \text{class(live)}, \text{class(die)}\}$$
$$\tau_a = 0.5, \tau_c = 0.8$$



Karabulut, Erkan, Daniel Daza, Paul Groth, Martijn C. Schut, and Victoria Degeler. "Tabular Foundation Models Can Learn Association Rules" (pre-print out soon).

TFMs Can Learn Rules Out-Of-The-box

Algorithm	#Rules	Supp.	Conf.	Zhang	Inter.	Cov.
Small Tabular Datasets						
TabICL	378	38	84.3	**31.8**	42.6	98.1
TabPFNv2.5	**286**	38.9	86.3	30.3	43.2	98.1
TabDPT	294	36.6	82.9	26.2	40.9	**1**
Aerial+	313	28.7	89.3	23.5	34.4	88.1
FP-G (0.3)	430	**42.4**	89.8	23.6	**48.6**	99.7
FP-G (0.2)	680	34.5	90.9	24.4	41.1	**1**
FP-G (0.1)	1462	22.9	**91.7**	26	29.2	**1**
Larger Tabular Datasets						
TabICL	11069	50.5	88.5	**38.9**	**58.4**	98.2
TabPFNv2.5	12238	48.8	86.6	34.3	55.8	99.0
TabDPT	**5177**	45.8	85	36.4	52.9	98.3
Aerial+	7214	50.9	**93.6**	22	57.5	97.6
FP-G (0.3)	26538	**51**	91.1	25.9	58	98.8
FP-G (0.2)	29211	43.5	91.4	25.2	50.4	99.6
FP-G (0.1)	35376	35.6	91.6	24.4	41.6	**1**

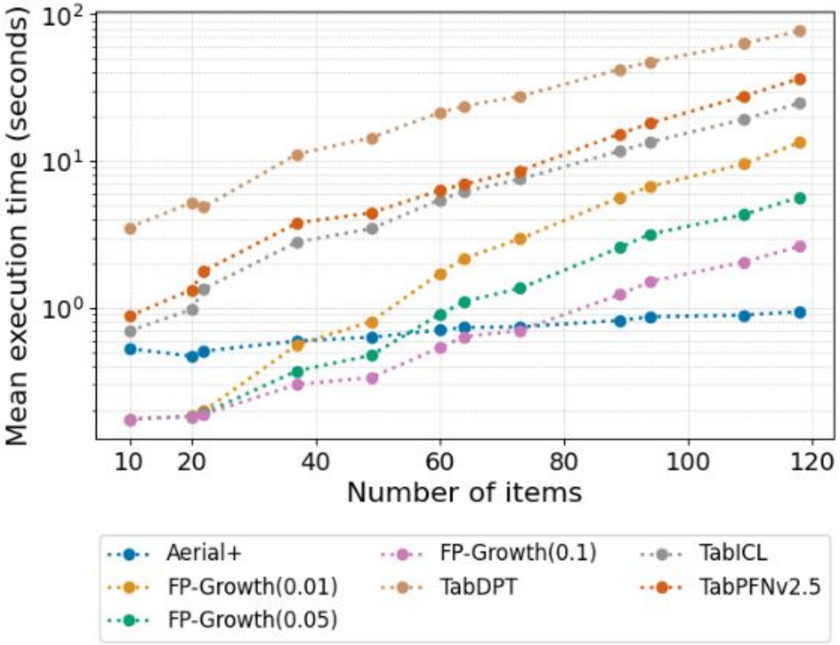
Table 2: Averages of rule quality metrics in percentage (%) across 5 small and 5 larger datasets. Tabular foundation models can learn concise number association rules with higher association strength, interestingness on average, and full data coverage, with slightly lower confidence scores (FP-G (x): refers to FP-Growth with x being min. support threshold).

Method	#Rules	Accuracy	F1	Precision	Recall
Small Tabular Datasets					
TabICL	616	82.93	81.07	81.13	82.93
TabDPT	641	82.22	80.35	80.58	82.22
TabPFNv2.5	657	**84.47**	**82.75**	**83.02**	**84.47**
Aerial+	579	80.69	77.35	76.51	80.69
FP-G (0.3)	266	79.72	76.45	75.84	79.72
FP-G (0.2)	422	79.95	77.08	76.86	79.95
FP-G (0.1)	934	83.54	81.37	82.49	83.54
FP-G (0.05)	2066	83.08	80.63	81.42	83.08
FP-G (0.01)	13991	83.28	80.58	81.16	83.28
Larger Tabular Datasets					
TabPFNv2.5	8719	86.53	85.74	86.15	86.53
TabICL	8100	86.51	85.77	86.16	86.51
TabDPT	4341	85.37	84.67	85.22	85.37
Aerial+	9095	87.63	86.20	86.41	87.63
FP-G (0.3)	13454	84.90	84.36	84.53	84.90
FP-G (0.2)	14882	88.10	**87.30**	**87.30**	88.10
FP-G (0.1)	19588	87.87	86.50	86.33	87.87
FP-G (0.05)	44424	**88.13**	86.61	86.34	**88.13**
FP-G (0.01)	59603	88.04	86.60	86.41	88.04

Table 3: Predictive performance in percentage and rule (CBA) or itemset (CORELS) size given in column two. TFMs achieve better predictive performance with a concise number of rules in small tabular data, while being behind FP-G by only -1.5% despite FP-G's rule explosion (FP-G (x) denotes FP-Growth with minimum support threshold x).

Karabulut, Erkan, Daniel Daza, Paul Groth, Martijn C. Schut, and Victoria Degeler. "Tabular Foundation Models Can Learn Association Rules" (pre-print out soon).

TFMs Can Learn Rules Out-Of-The-box



Karabulut, Erkan, Daniel Daza, Paul Groth, Martijn C. Schut, and Victoria Degeler. "Tabular Foundation Models Can Learn Association Rules" (pre-print out soon).

TFMs Can Learn Rules Out-Of-The-box

```python
from ucimlrepo import fetch_ucirepo
from src.wrapper import TabProbe

# Load breast cancer dataset from UCI ML repository
dataset = fetch_ucirepo(id=14).data.features

# Mine rules with TabPFN
miner = TabProbe(method='tabicl', ant_similarity=0.5, cons_similarity=0.8)
rules = miner.mine_rules(dataset, metrics=["support", "confidence"])
```

Karabulut, Erkan, Daniel Daza, Paul Groth, Martijn C. Schut, and Victoria Degeler. "Tabular Foundation Models Can Learn Association Rules" (pre-print out soon).

Prediction success implies associations!

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.
4. TFMs can learn association rules out-of-the-box.
   a. Prediction success implies associations.

Scalable Neurosymbolic Knowledge Discovery from Tabular Data

Erkan Karabulut
e.karabulut@uva.nl

23 January 2026