ELIAS, ELLIOT & ELSA Theme Development Workshop on Foundation Models

July 10th, 2025
Thessaloniki, Greece (Hybrid event)

Funded by the European Union

ELIAS ELLIOT elsa European Lighthouse on Secure and Safe AI

CENTRE FOR RESEARCH & TECHNOLOGY - HELLAS
Information Technologies Institute

Tabular Foundation Models

Frank Hutter

frank@priorlabs.ai

Tabular data?

Example: Which patients have early-stage Alzheimer (based on omics blood markers)?

ID	L-Car-nitin	Crea-tinin	Homocys-tein	Beta-hydro-xybutyrat	Early-stage Alzheimer
1	45.2	85	12.1	0.4	Yes
2	38.7	72	10.5	0.3	No
3	41.0	90	13.2	0.2	Yes
4	36.5	80	9.4	0.5	No
5	44.8	78	12.0	0.6	Yes
6	39.3	88	11.1	0.4	No
7	42.1	76	13.0	0.7	Yes
8	37.5	70	8.9	0.3	No
9	40.9	92	14.1	0.5	Yes
10	36.0	75	10.2	0.4	No
⋮	⋮	⋮	⋮	⋮	⋮
5000	43.0	84	12.7	0.6	Yes
5001	41.2	81	11.5	0.4	?
5002	43.5	83	11.8	0.5	?
5003	39.9	74	10.0	0.3	?

Tabular prediction problem

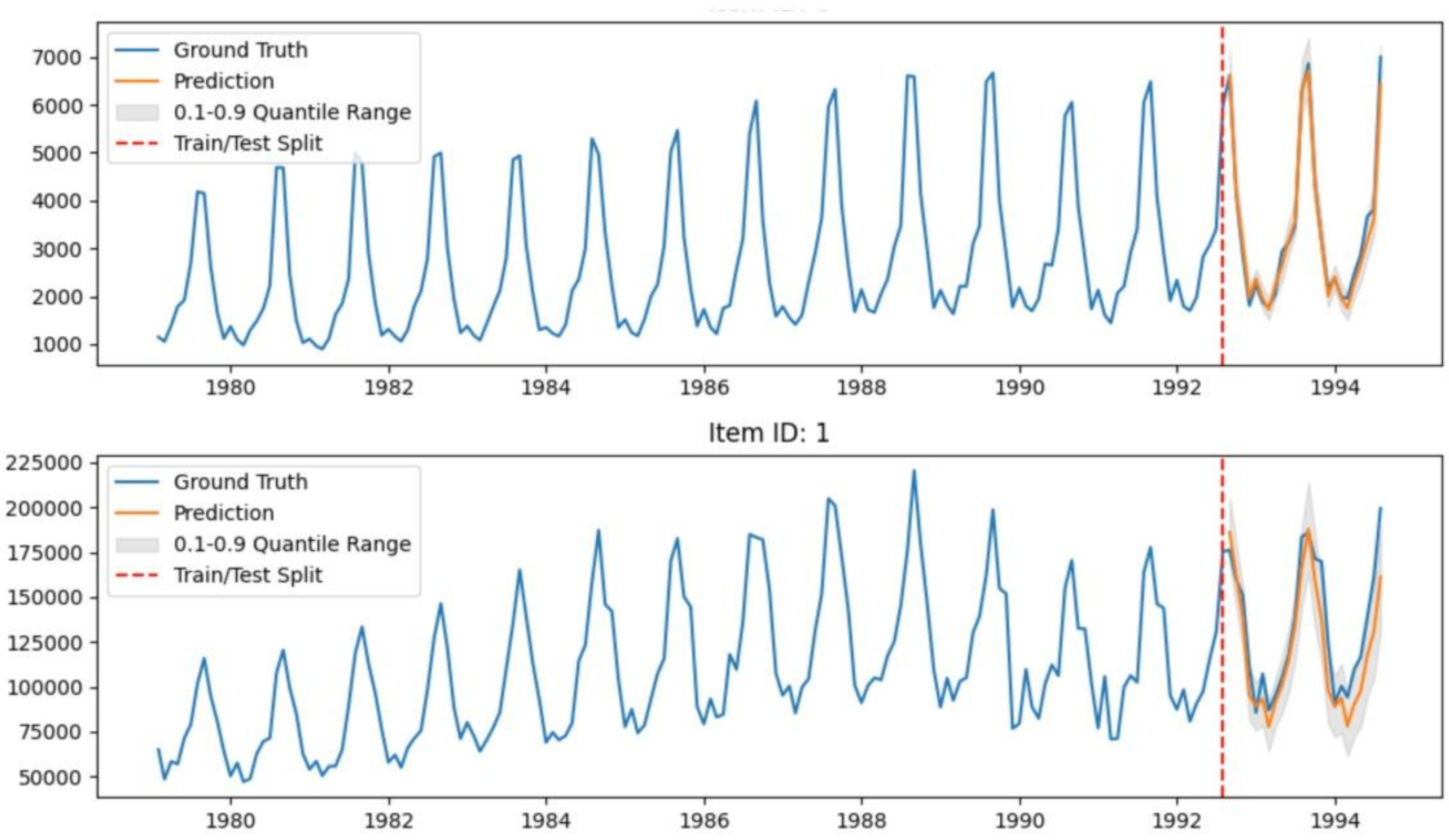
Personalized health models

- „Omics“-bloodmarkers instead of invasive diagnostics
- Better tabular models multiply value of $BN data acquisition efforts

Why tabular data?

	Healthcare		Finance		Business Analytics		Insurance	
Classification / Regression	Personalized Risk Prediction	Drug response prediction	Loan default prediction	Fraud detection	Customer Lifetime Value Prediction	Customer segmentation	Premium price prediction	Fraud detection
	Medical billing fraud detection	Sepsis detection	Credit risk assessment	AML & transaction monitoring	Pricing optimization	Resource allocation	Claim & loss prediction	Customer segmentation
Time Series	AI-Enhanced Intensive Care	Hospital readmission prediction	Trading price prediction	Currency exchange rates	Metric forecasting	Sales & inventory forecasting	Climate risk modeling	Care cost forecasting
Reco ommen der Systems	Clinical Decision Support	Clinical Trial Matching	Investment products	Personalized banking	Cross selling	Product recommender	Preventative measures	Contract recommendat ion

Why tabular data? Very related: Time Series Forecasting

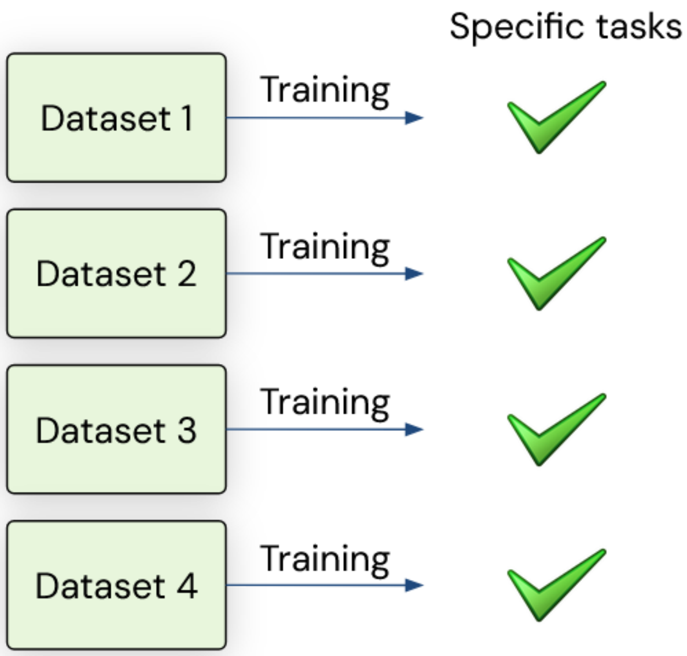


How will XYZ develop over time?

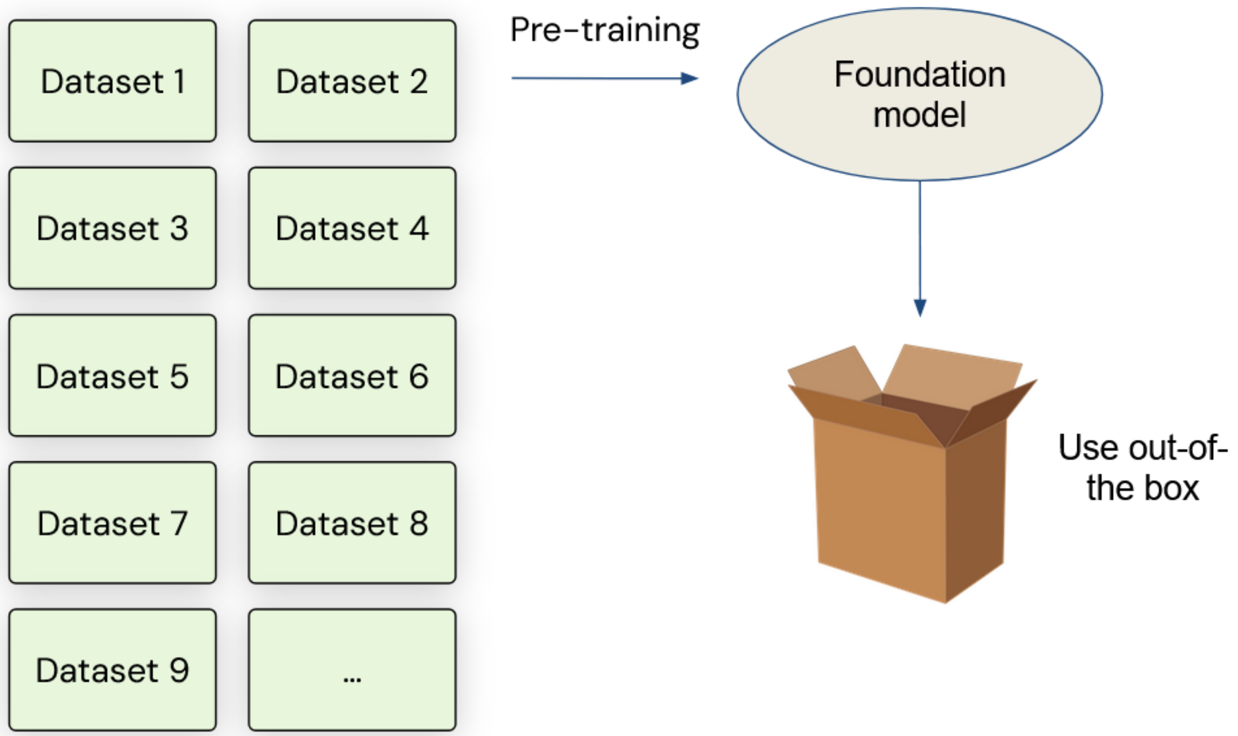
- Stock prices
- Energy price
- Supply & demand
- Temperature
- Traffic congestion
- Machine health
- …

Why Foundation Models for Tabular Data?

Traditional ML



Foundation models



- Individual siloed models
- Lengthy task-specific training

- Off-the-shelf use without retraining
- Can quickly be finetuned to new use cases

Vision

Foundation models have transformed text & images

But our most valuable data is organized in tables

Text

OpenAI

DeepMind

deepseek

Tables, Time Series & Databases

PRIOR LABS

Images

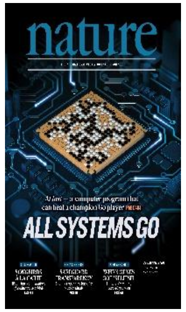
OpenAI

deepseek

black forest labs.

Our community is on its way to revolutionize tabular data

TabPFN v2
Publication in *Nature*



More accurate than previous ML on

>96% of use-cases

More data efficient than previous ML

Only 50% of the data needed for same accuracy

More than 1 Million Downloads

Open-source Software with 4000 stars on Github

OPEN SOURCE USED AT

Outline

- Motivation for Tabular Foundation Models
- TabPFN
- TabPFN v2
- TabPFN for time series: TabPFN-TS
- Explainability & Fairness

TabPFN: a Learned Algorithm for Small Tabular Data

- TabPFN is a **GPT-like transformer for tabular classification**
- Framed as next-word prediction: $x_1, y_1, \ldots, x_n, y_n, x_{n+1}, ?$

- To be more precise:

$$\{(x_1, y_1), \ldots, (x_n, y_n)\}, x_{n+1} \longrightarrow \boxed{\text{PFN}} \longrightarrow \hat{y}_{n+1}$$

- To be even more precise:

$$\{(x_1, y_1), \ldots, (x_n, y_n)\}, x_{n+1} \longrightarrow \boxed{\text{PFN}} \longrightarrow p(y_{n+1} \mid x_{n+1}, \{(x_1, y_1), \ldots, (x_n, y_n)\})$$

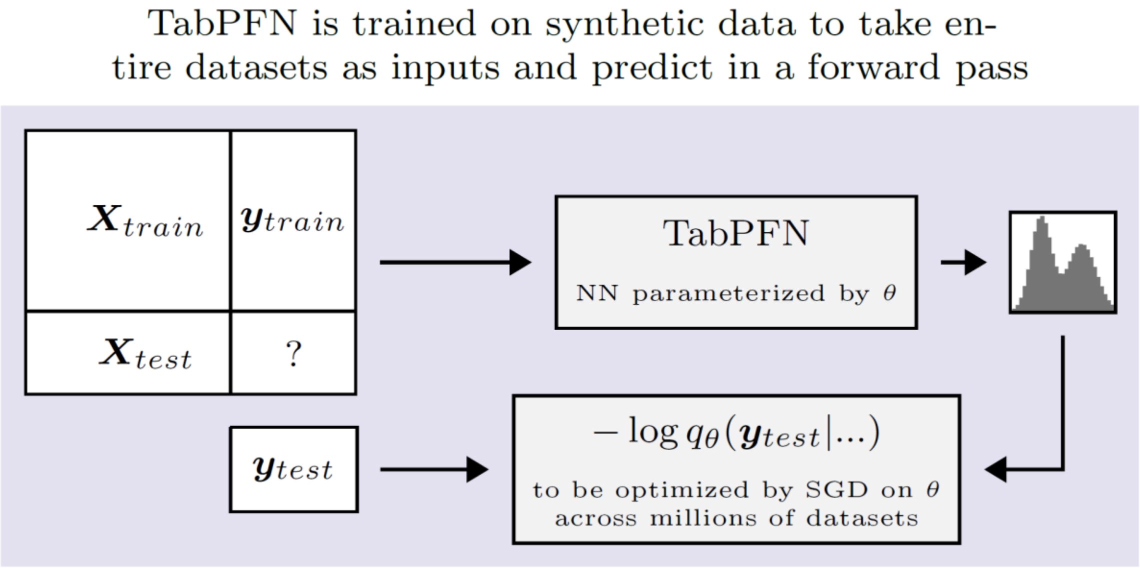
The high-level intuition

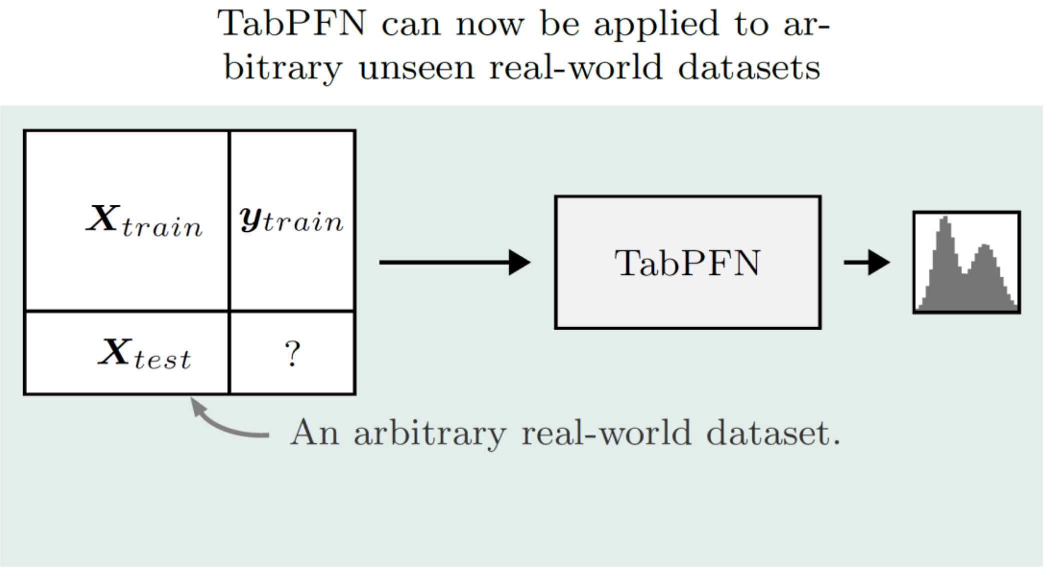
- TabPFN is a transformer with weights θ
  - A single forward pass directly approximates $p(y_{n+1} \mid x_{n+1}, \{(x_1, y_1), \ldots, (x_n, y_n)\})$

- We optimize θ to minimize average cross entropy loss across datasets
  - Across which datasets?
    - Millions of synthetically generated ones: $\{(x_1, y_1), \ldots, (x_{n+1}, y_{n+1})\}$
  - How do we train it?
    - Very standard transformer architecture (just drop the positional encoding)
    - Standard supervised learning with SGD

$$\{(x_1, y_1), \ldots, (x_n, y_n)\}, x_{n+1} \longrightarrow \boxed{\mathrm{PFN}_\theta} \longrightarrow p(y_{n+1} \mid x_{n+1}, \{(x_1, y_1), \ldots, (x_n, y_n)\})$$

TabPFN High-Level Overview of Training & Inference



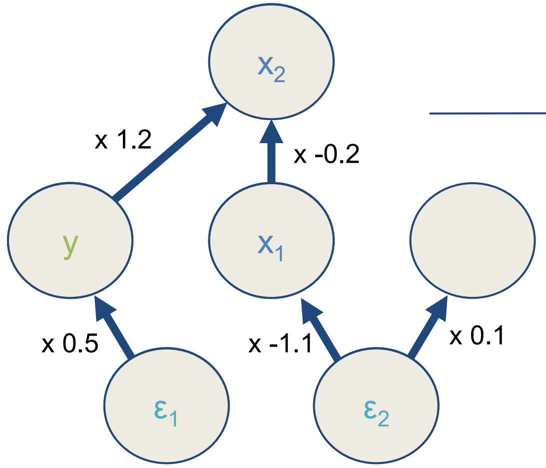


The only missing piece: a method to **generate synthetic data sets** that resemble the data sets we expect

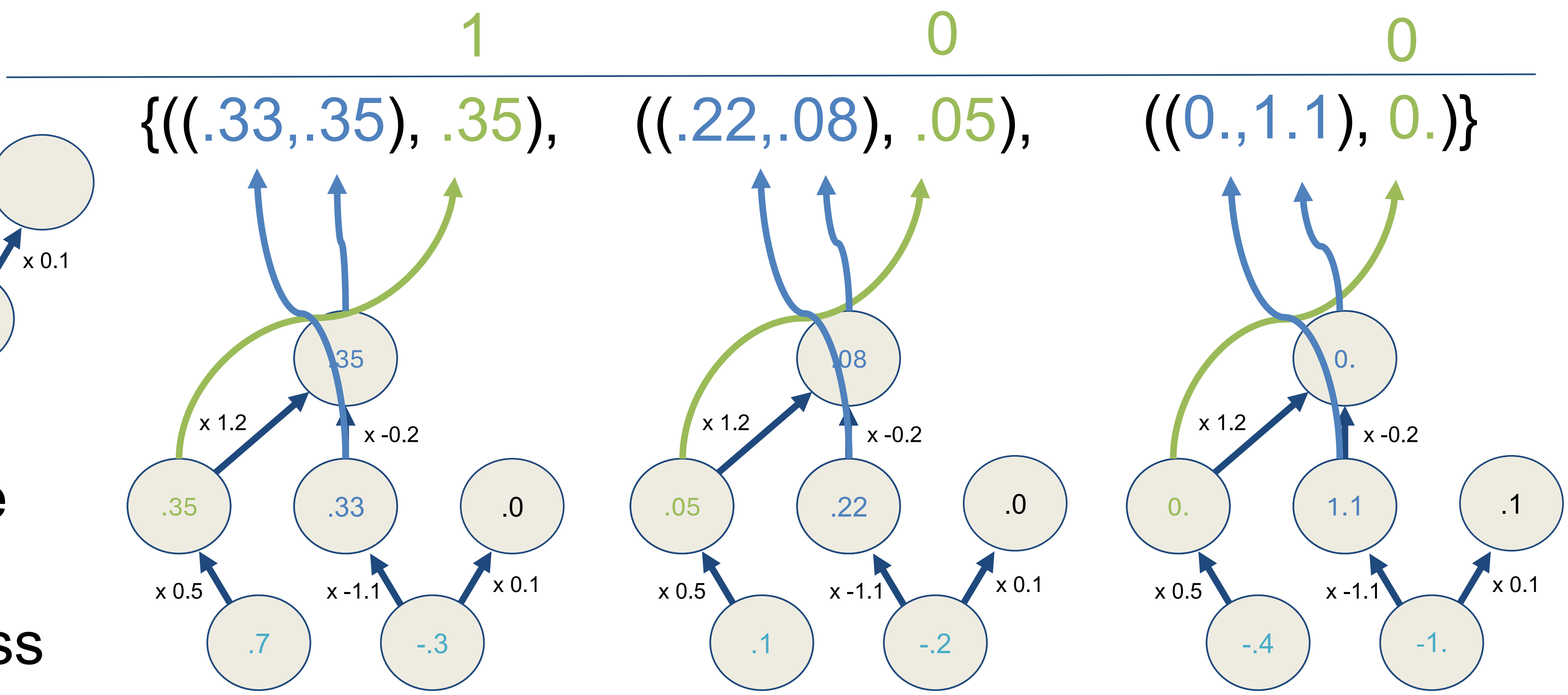
(TabPFN then approximates the Bayesian posterior for the prior we define over these datasets)

TabPFN Prior: Integrating Principles from Causality

Sample & initialize
a causal graph

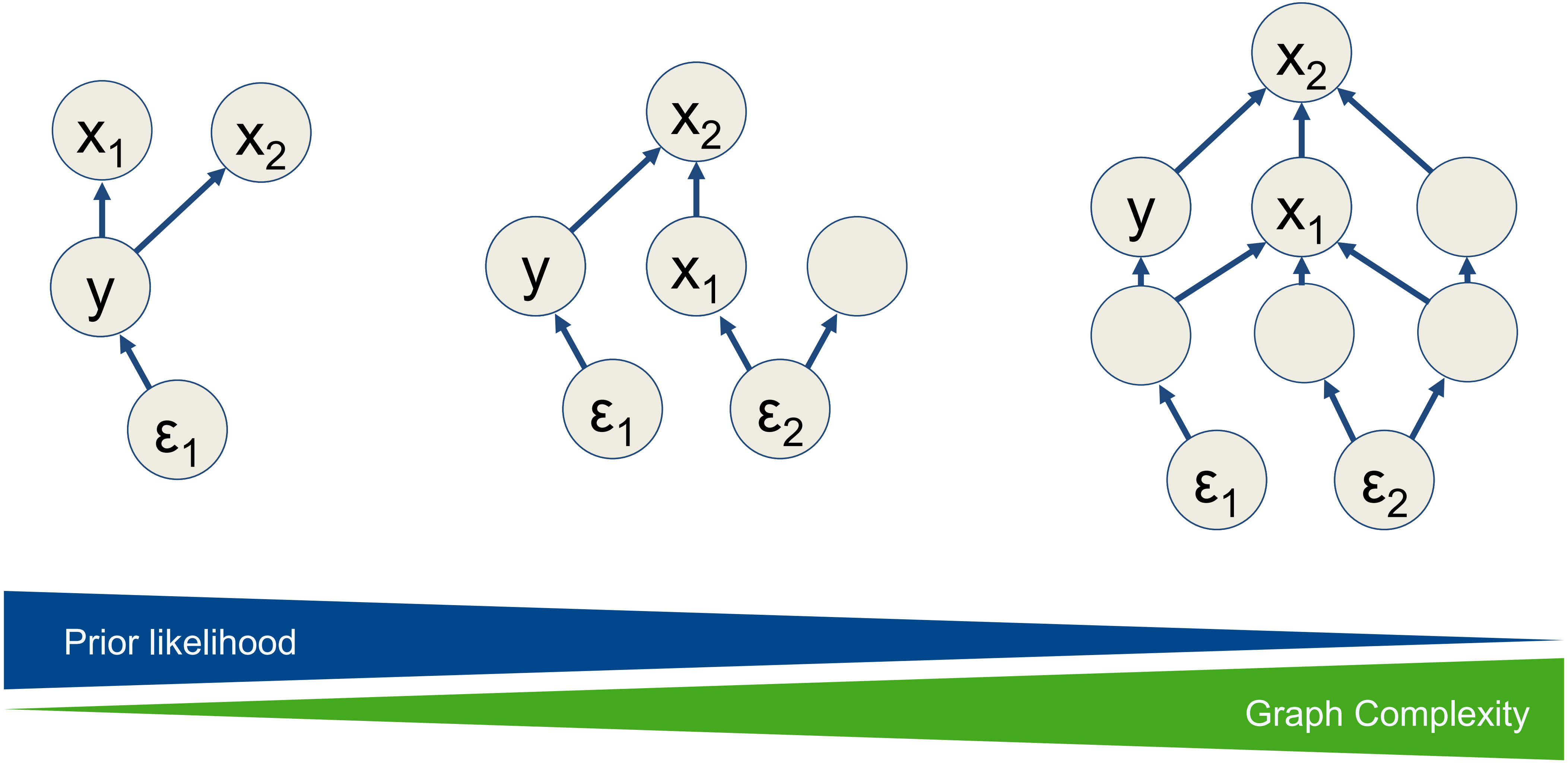


Build dataset: output >0.2?



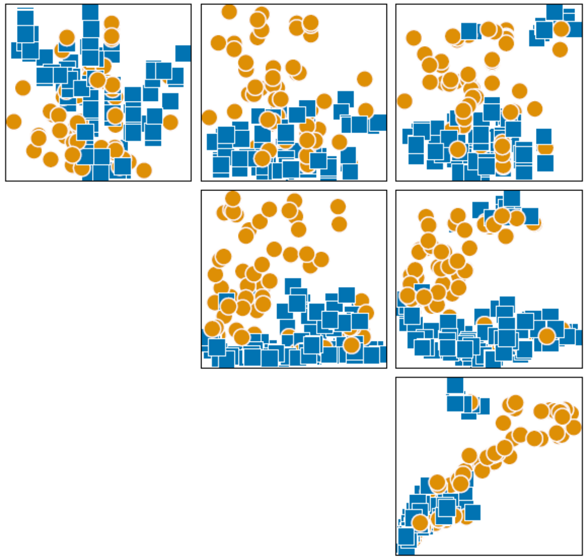
Sample noise
per example
& forward pass

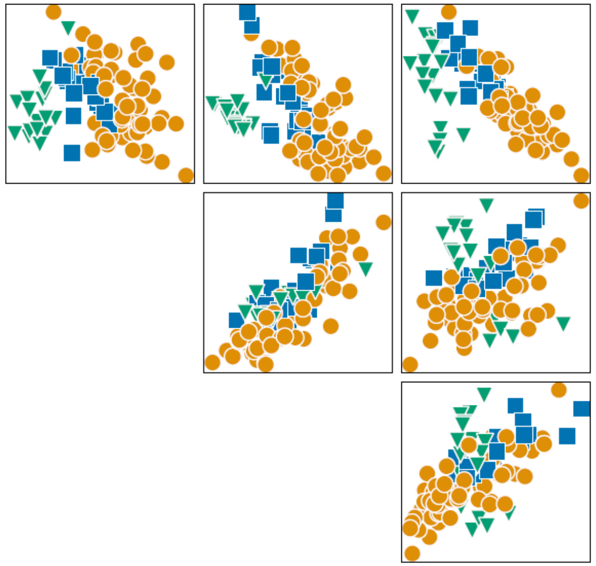
TabPFN Prior: Simplicity Principle



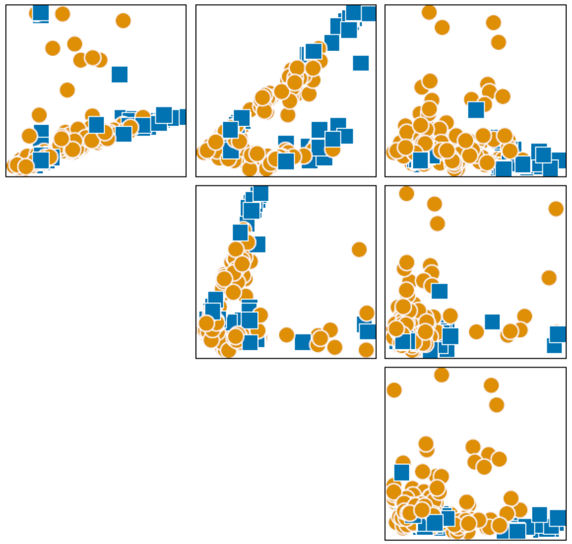
The generated datasets look similar to real datasets

Synthetic datasets

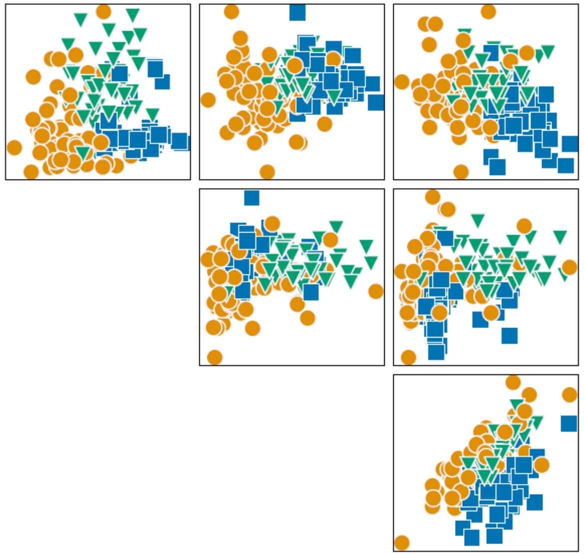




Parkinsons dataset

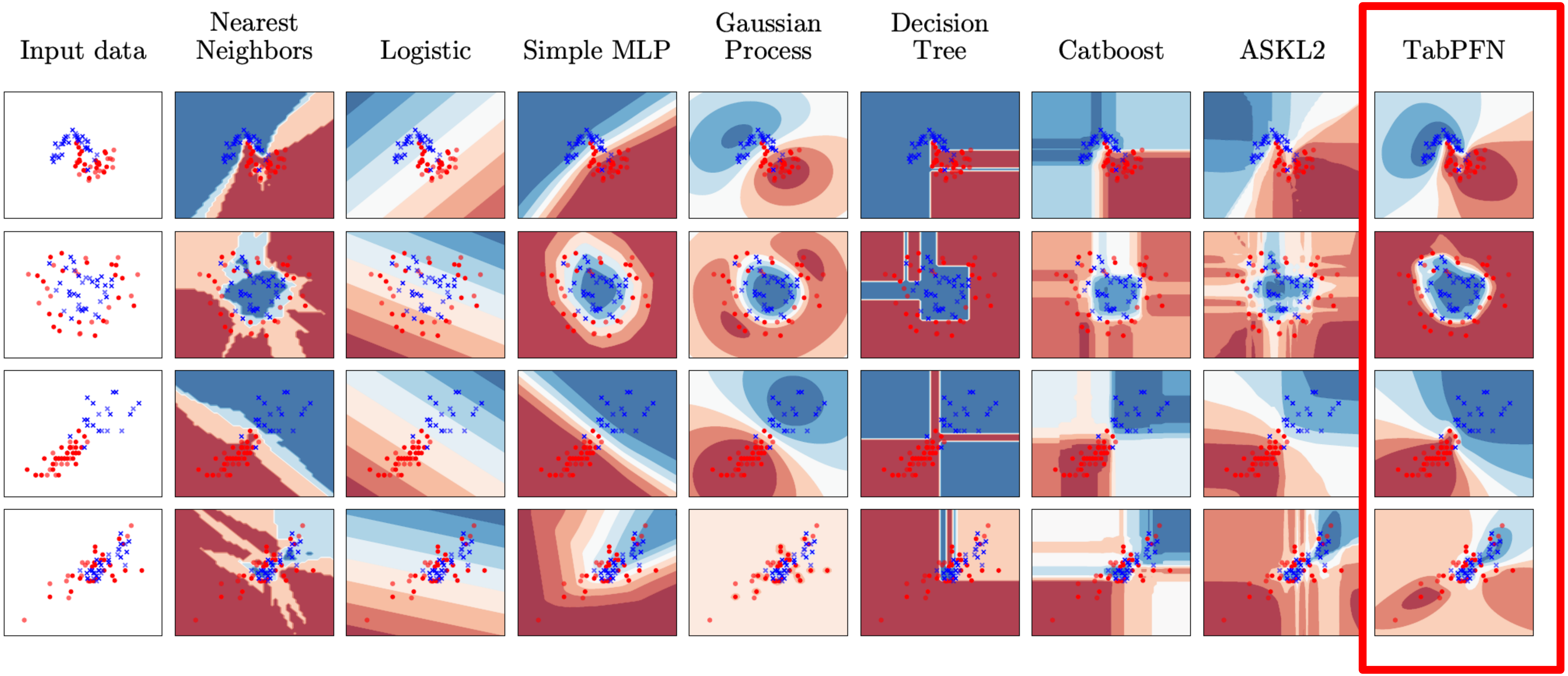


Wine dataset

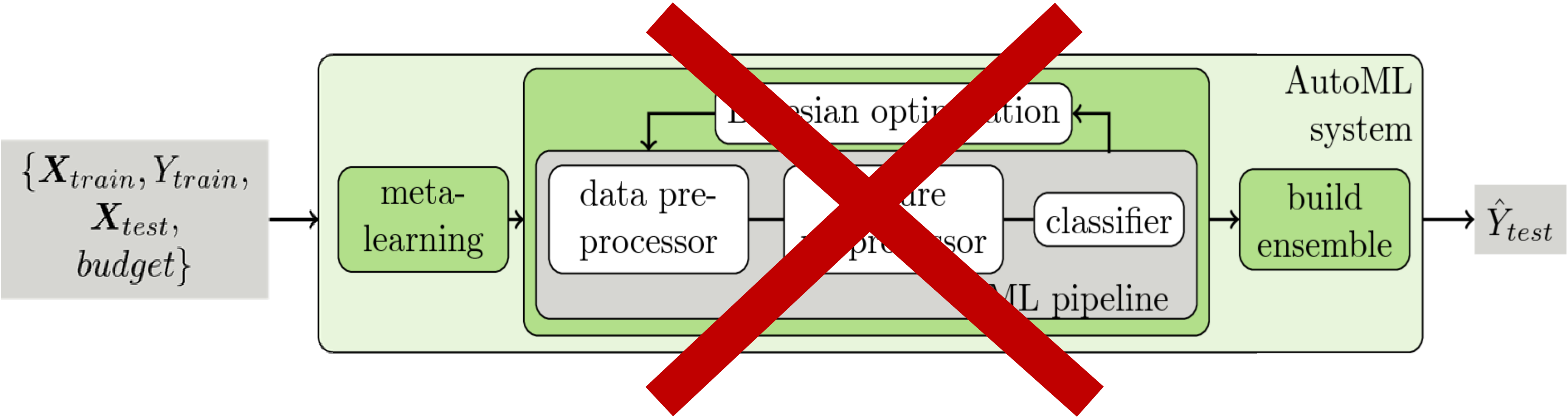


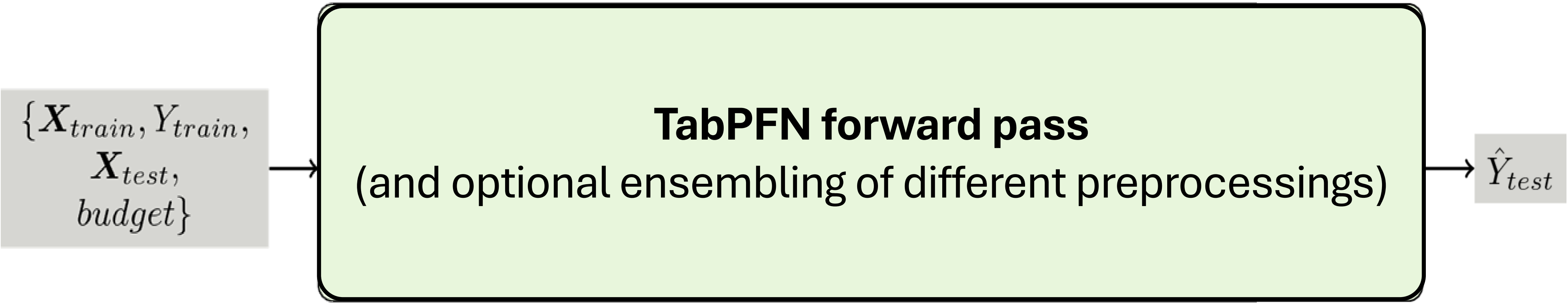
Qualitative result: smooth & well-calibrated predictions

Learning on synthetic datasets yields strong performance on new datasets



Simplicity: it's just a forward pass

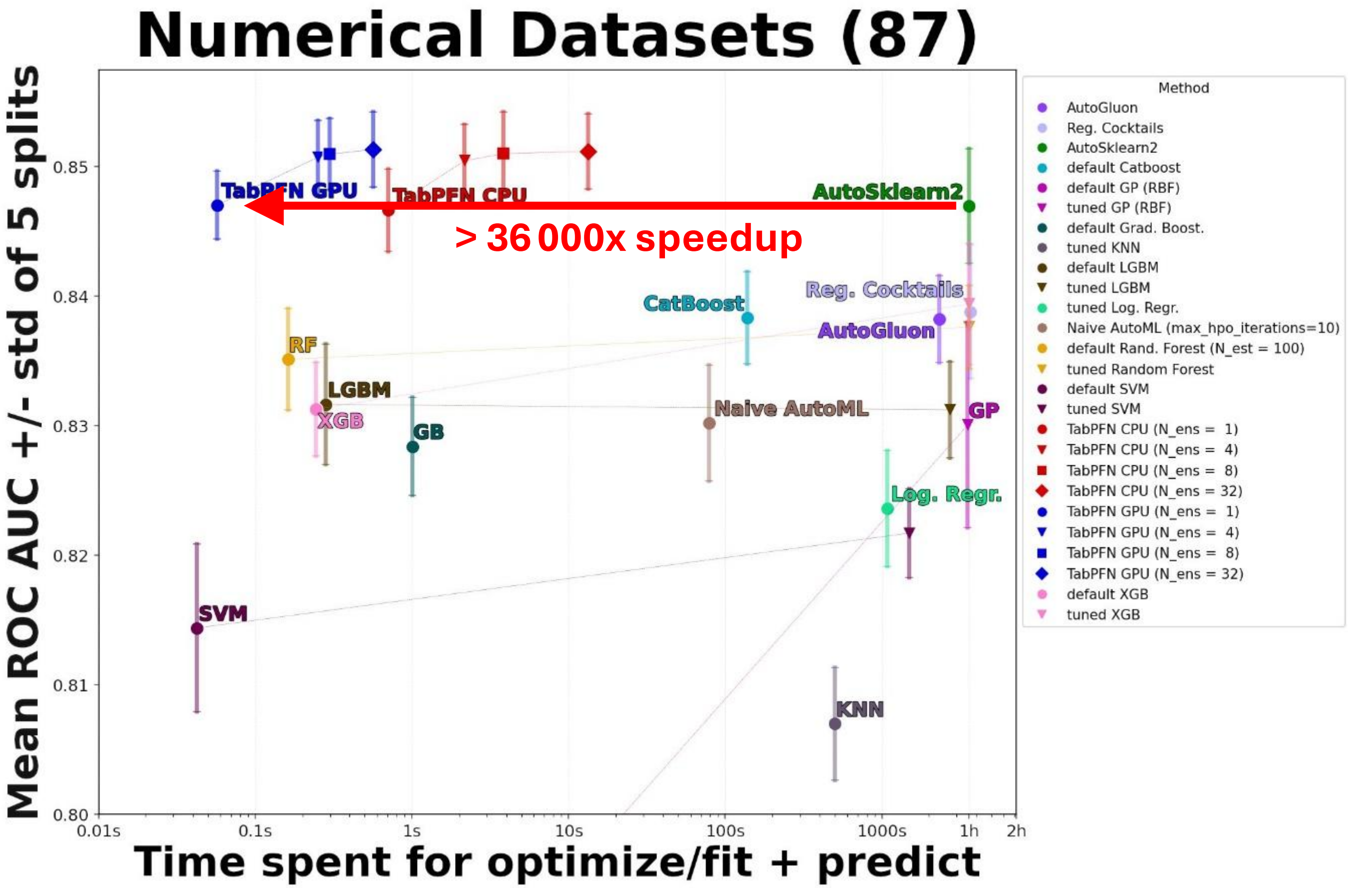




Quantitative Result (87 numerical datasets, no missing values)

- **Better performance in 1s than than any other ML / AutoML method in 1h**
  - Disclaimer: these are average results; TabPFN is not the best on every single dataset

- **Limitations** (in 2022)
  - Size: up to 1000 data points, 100 features, 10 classes
  - Not (yet) designed for: categorical features, missing values, uninformative features
  - Only classification
  - High inference time



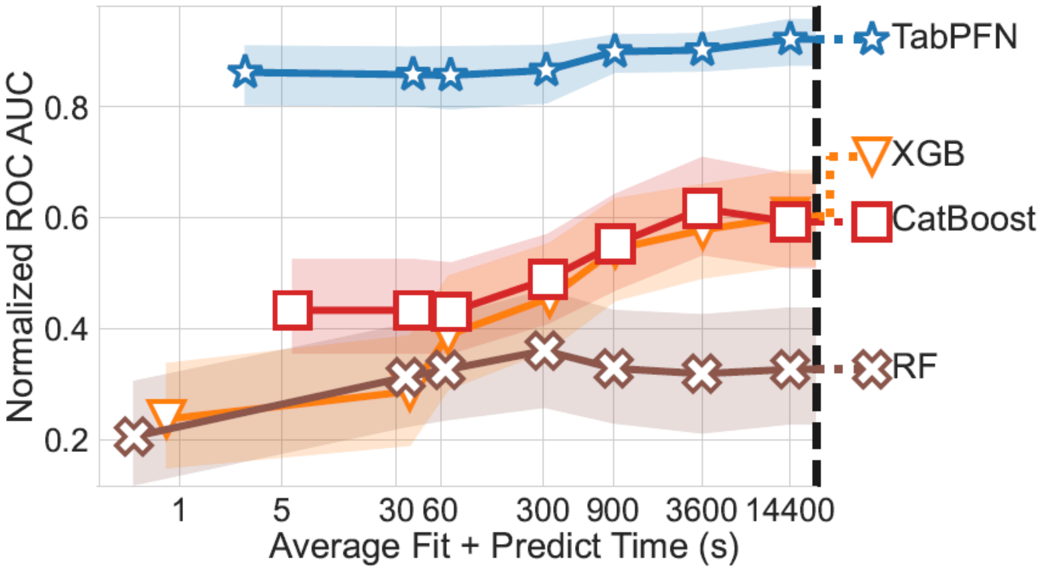
Outline

- Motivation
- TabPFN
- TabPFN v2
- TabPFN for time series: TabPFN-TS
- Explainability & Fairness

Improvements since TabPFN v1

- Now best tabular ML algorithm for <= 10000 data points, 500 features
  - Better in 5 seconds than any other method in 4 hours

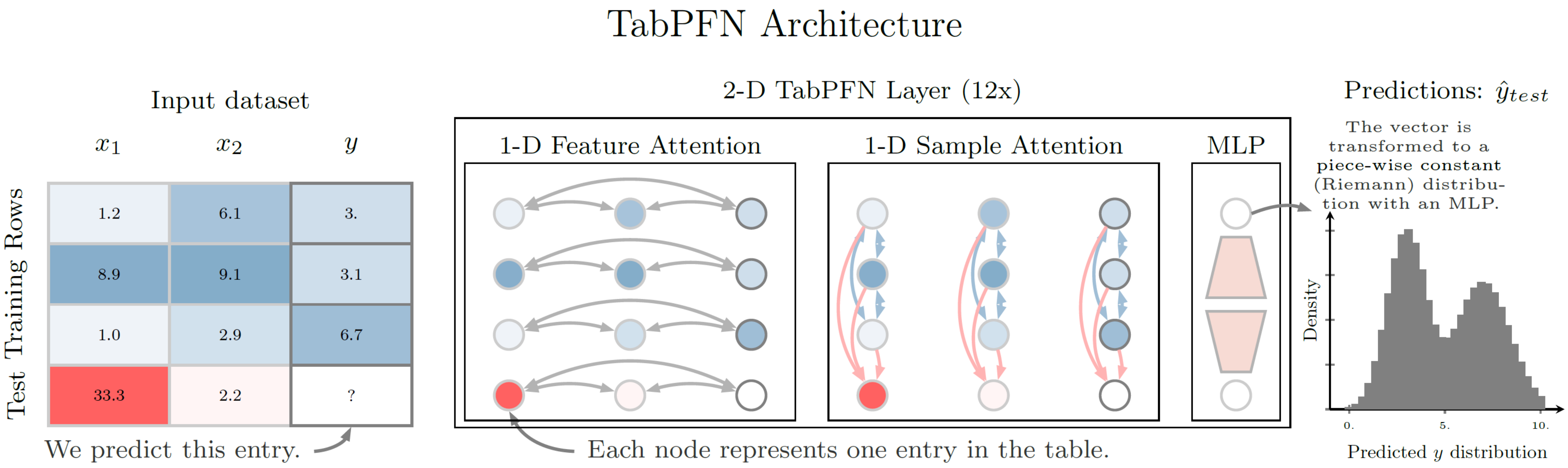
- **Limitations resolved**
  - Size: up to ~~1000~~ 10000 data points, ~~100~~ 500 features, 10 classes
  - ~~Not (yet)~~ Now also designed for: categorical features, missing values, uninformative features
  - Classification & regression
  - ~~High~~ Moderate inference time



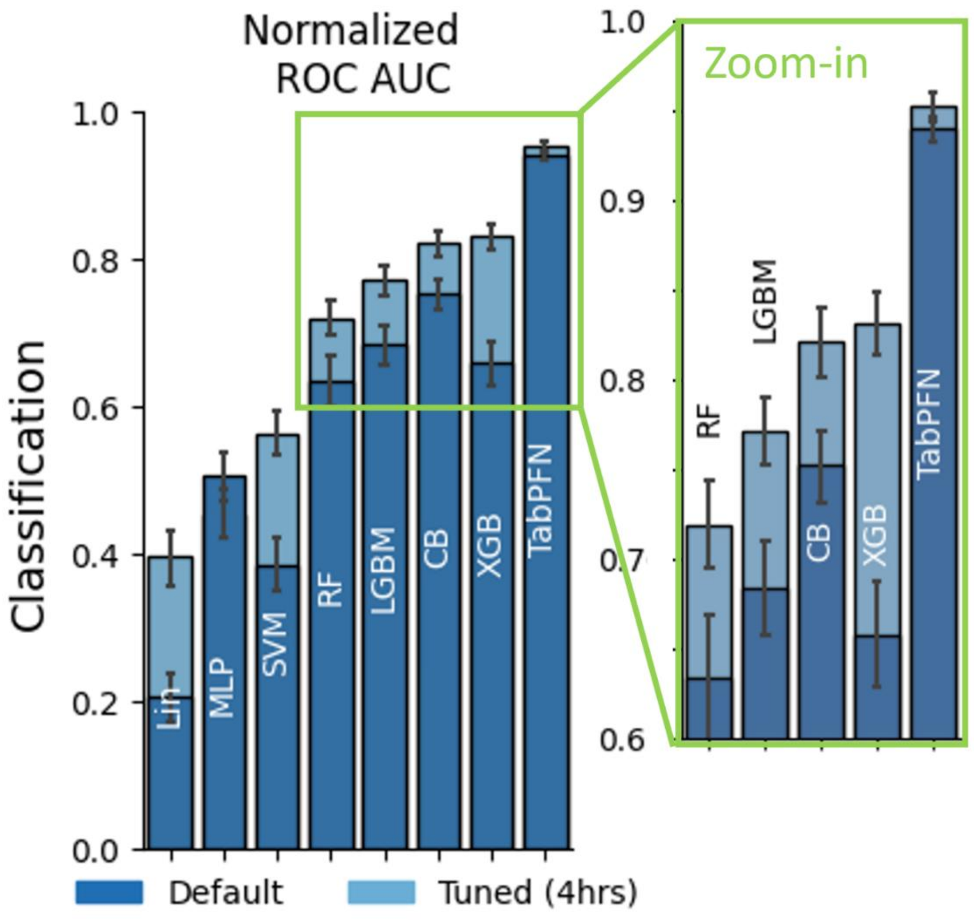
Extensions since TabPFN v1

- **Scaling up**
  - More efficient attention to support more data points
  - Change in architecture to support arbitrary #features
  - Inference speedups
- **Improving the prior**
  - Trees in the structural causal models
  - Supporting more activation functions (sine, log, exponentials, ..)
  - Discretizing categoricals in the prior already
  - A lot of engineering …
- **Demonstrating foundation model capabilities**

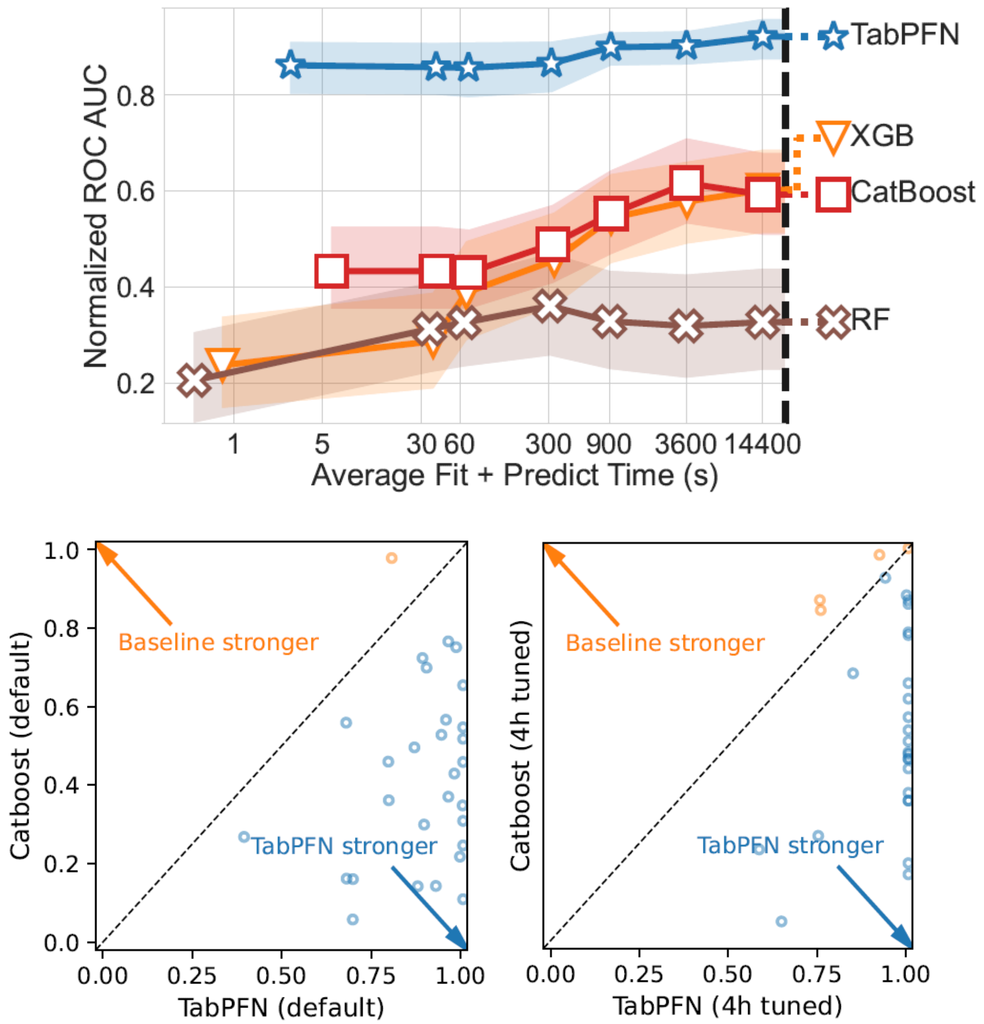
New TabPFN v2 Architecture



Results for Classification

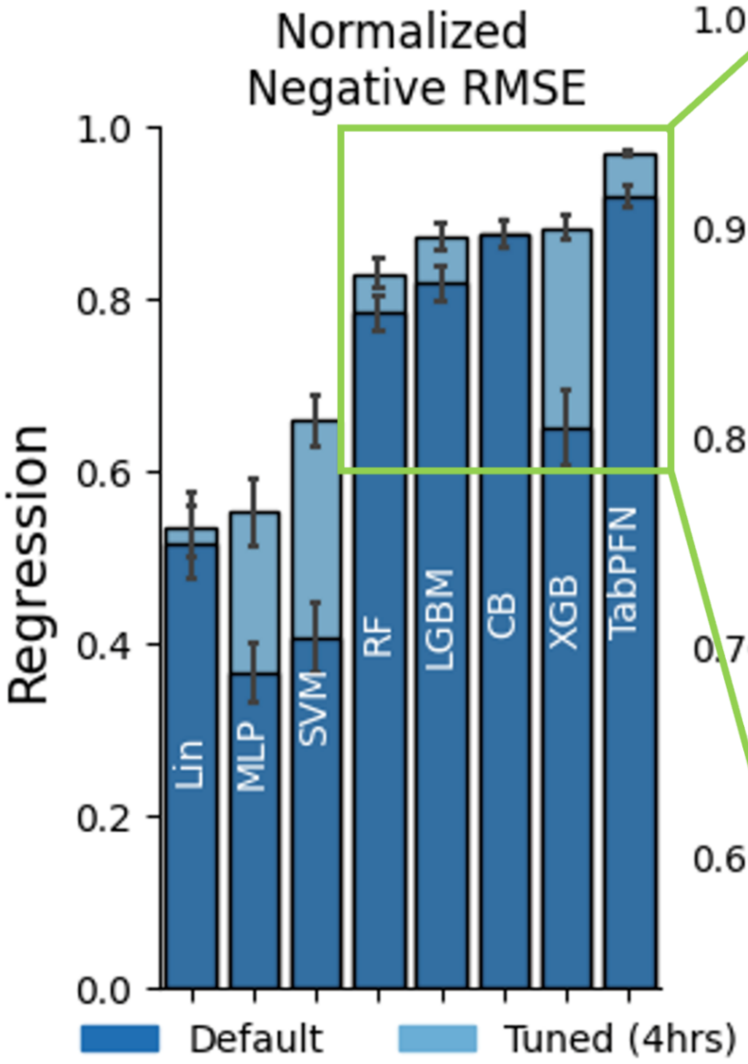


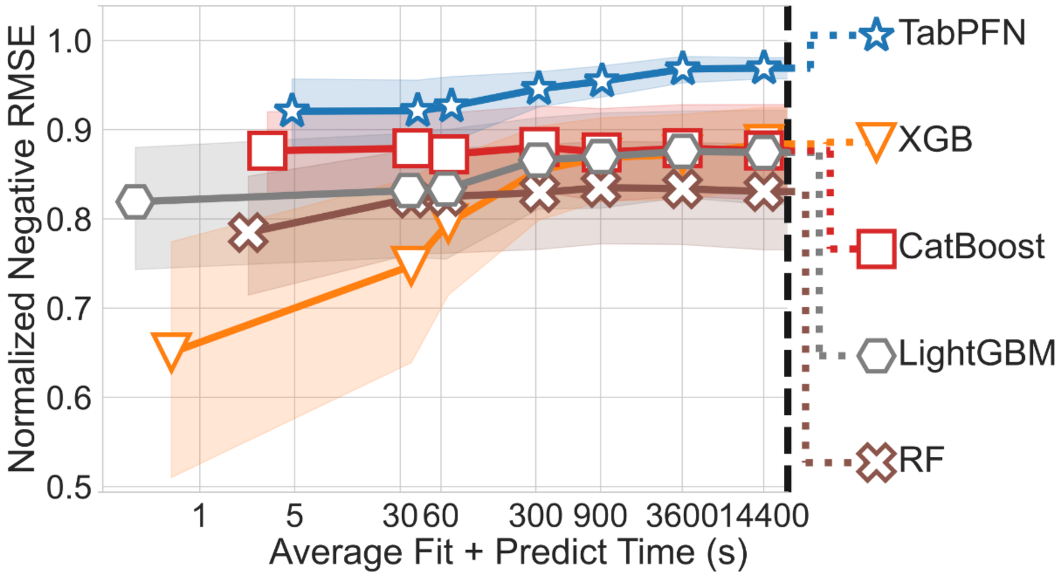
Result across 29 datasets:
better in 5s than other methods in 4h

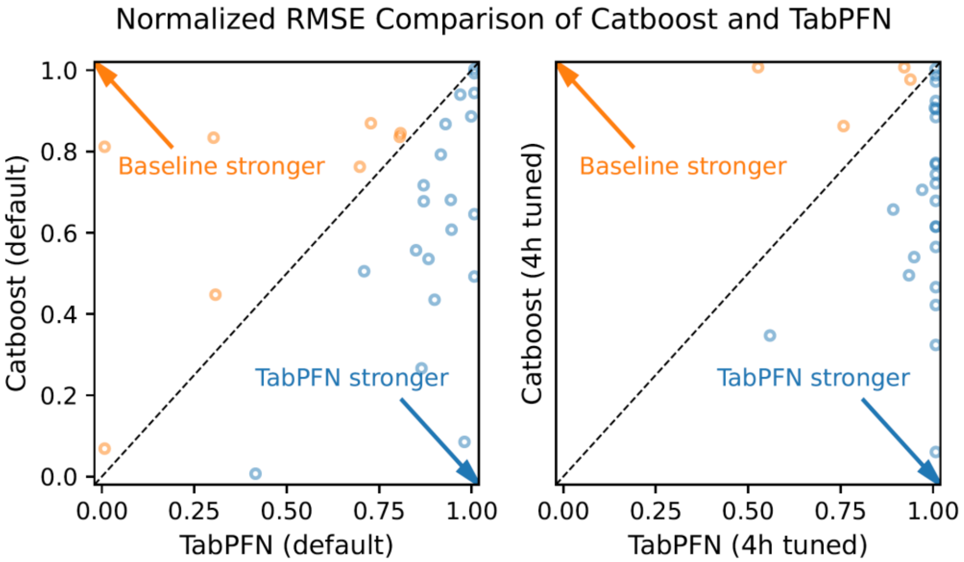


Improvements are quite stable across datasets,
for both default & tuned

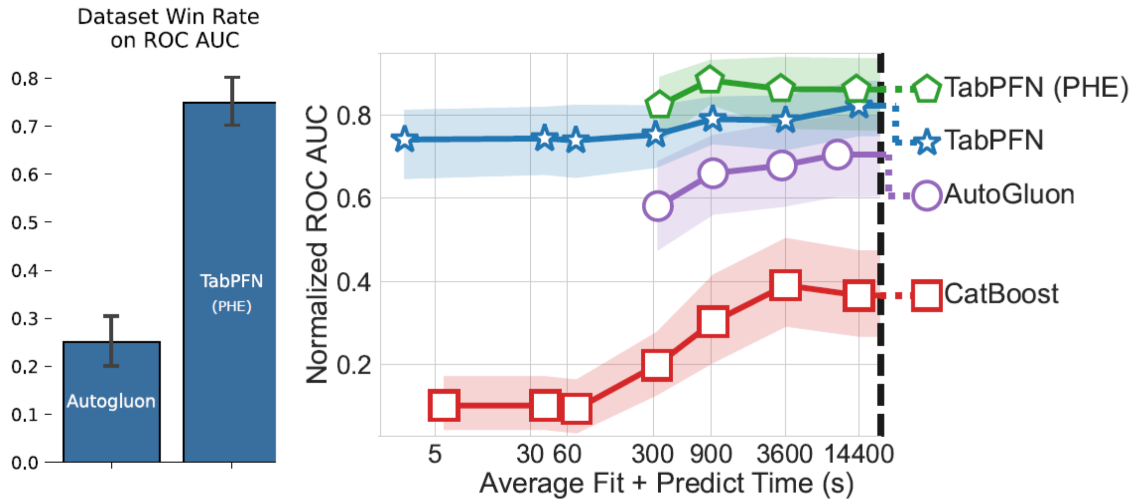
Results for Regression



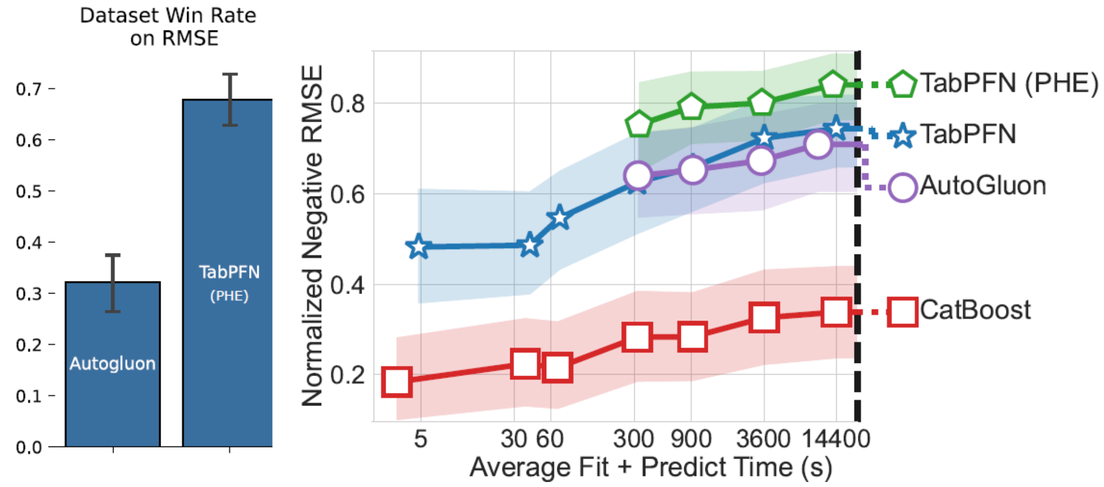




Comparison to the Leading AutoML Method AutoGluon

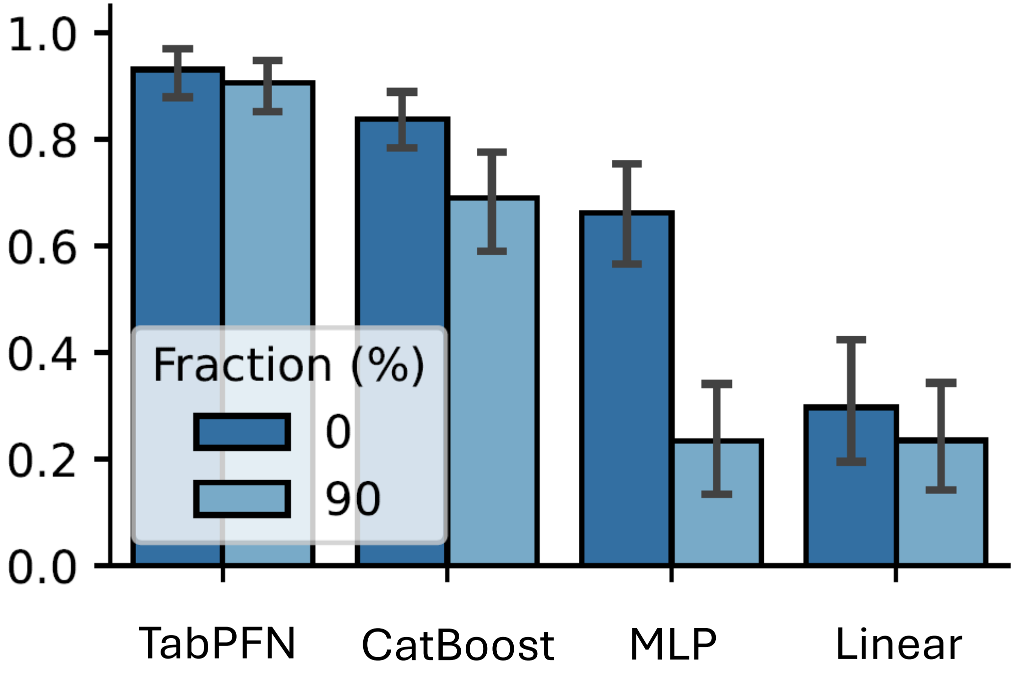


- Classification:
  - Even 5s of native TabPFN is better than AutoGluon (4h)
  - TabPFN (PFE) better yet



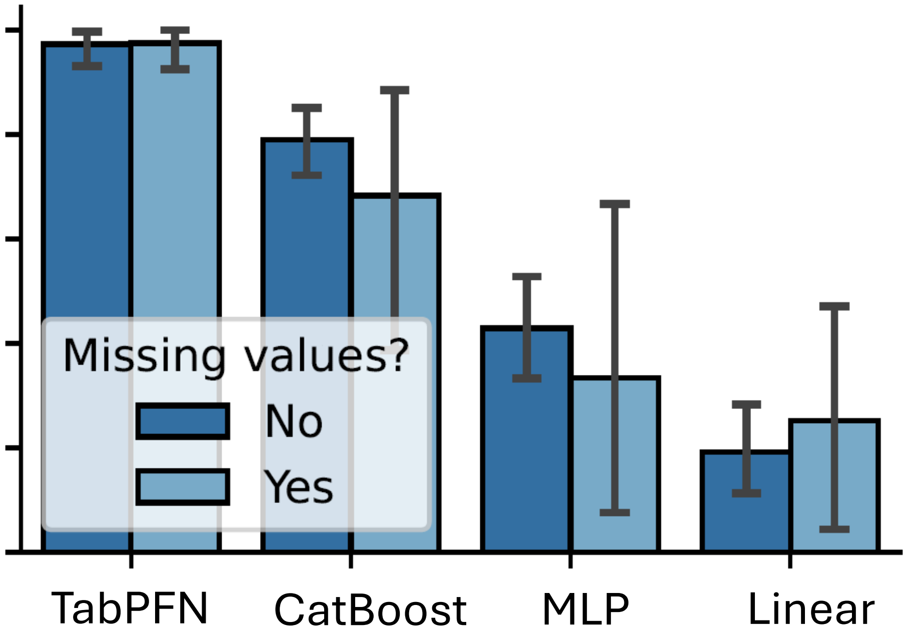
- Regression:
  - TabPFN similar to AutoGluon
  - TabPFN (PHE) still better
    - 5s matches AutoGluon 4h

TabPFN is now robust against uninformative features



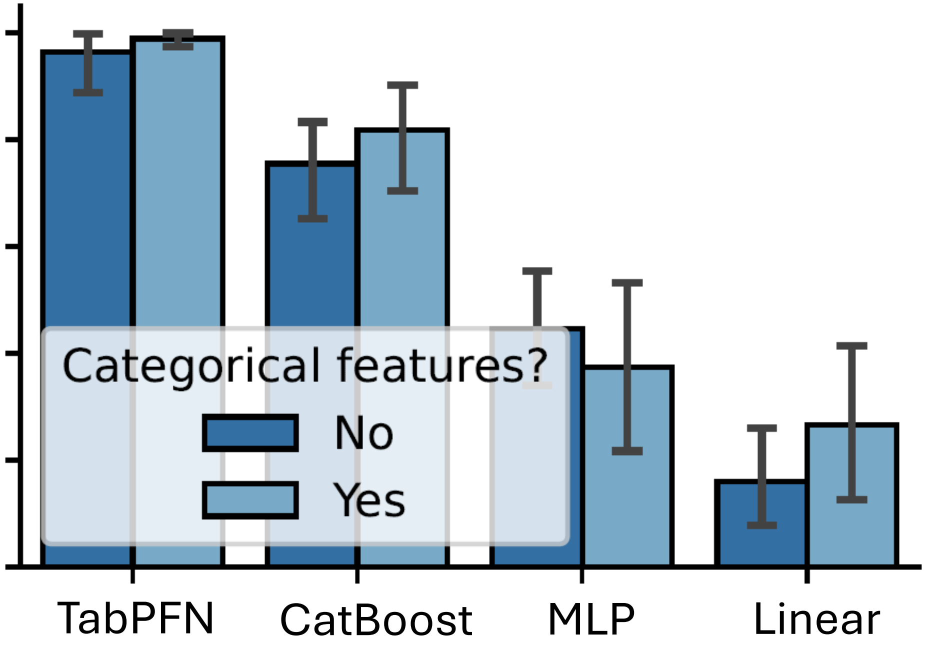
- Setup: Add 9x uninformative features to actual features
- TabPFN v1 had big problems with uninformative features
  - Neural networks are notoriously bad at handling uninformative features, see MLP performance
- Including the possibility of uninformative features in the prior fixed this

TabPFN is now robust against missing values



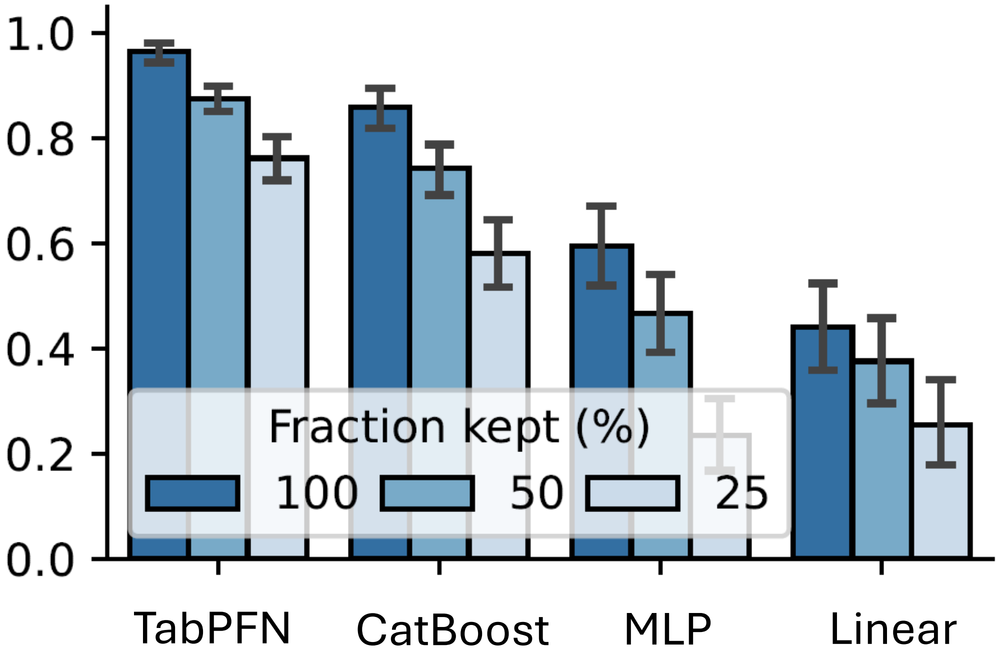
- Setup: subset of datasets with / without missing values
- TabPFN v1 had some problems with missing values
- Including the possibility of missing features in the prior fixed this

TabPFN is now robust for categorical features



- Setup: subset of datasets with / without categorical features
- TabPFN v1 had problems with categorical features
- Including the possibility of categorical features in the prior fixed this

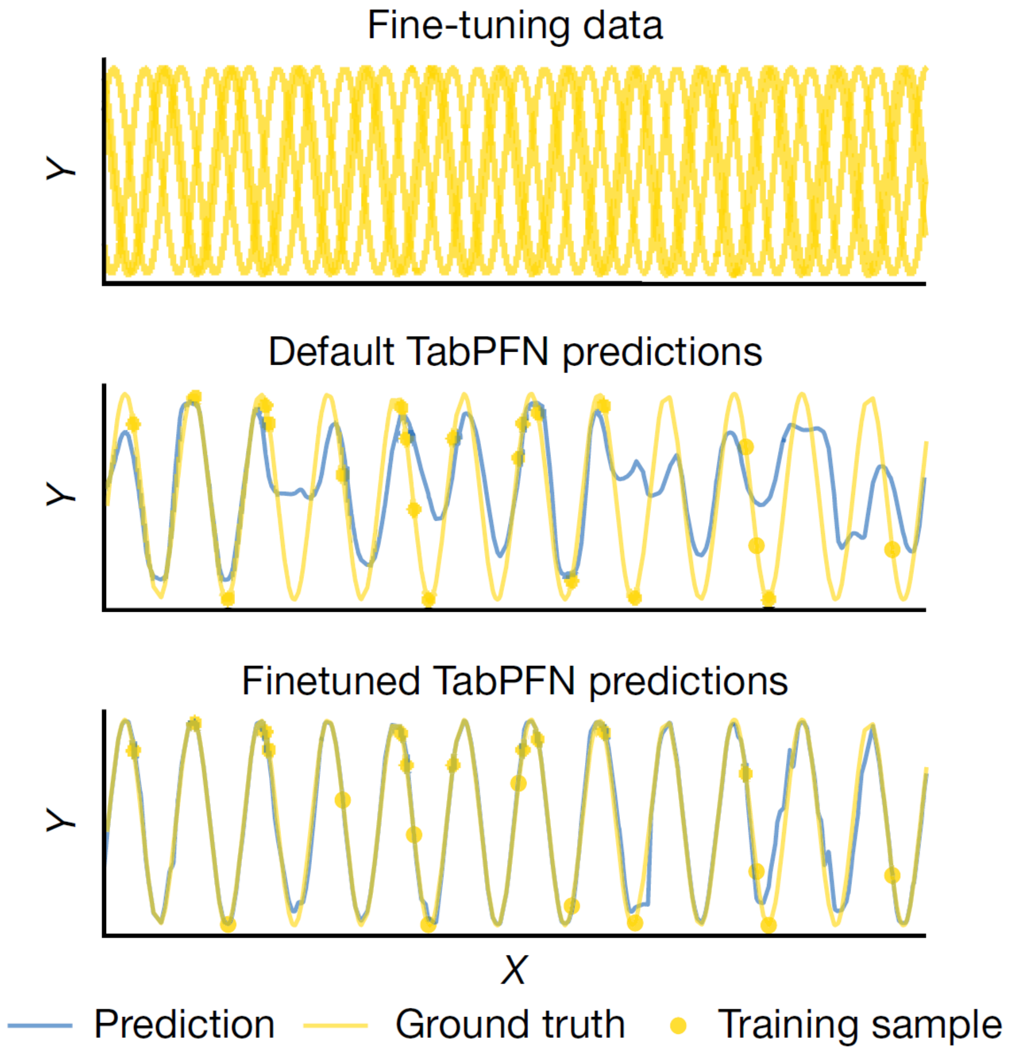
TabPFN works well with less samples



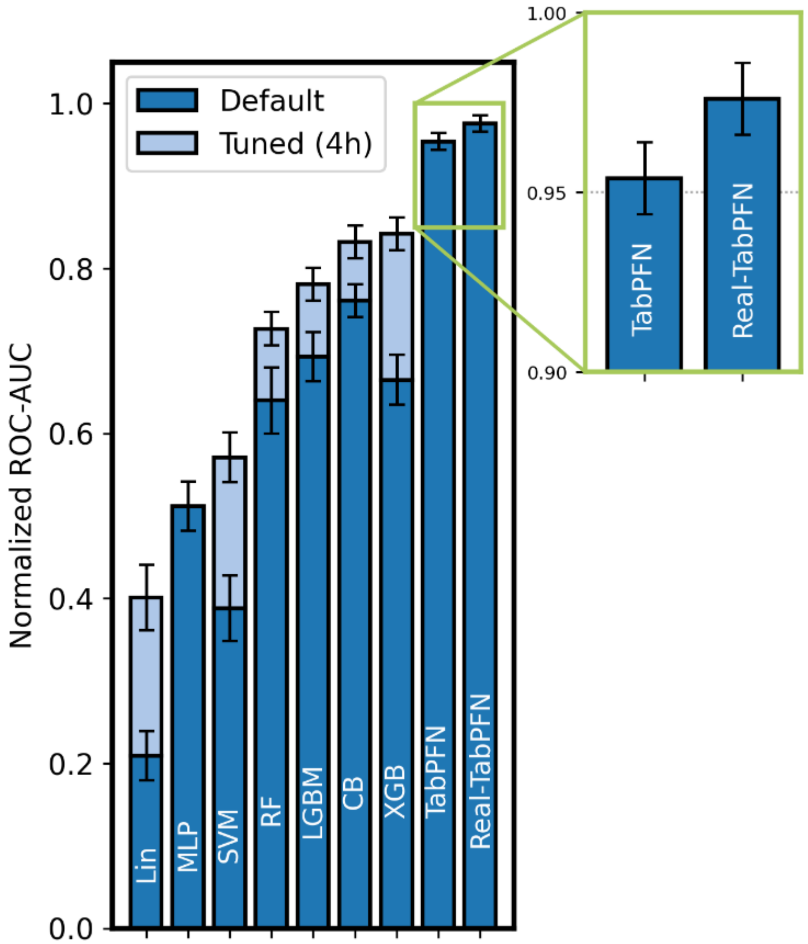
- TabPFN using 50% of the data ties with CatBoost using 100% of the data

Finetuning: customizing the model (just like an LLM)

Fine-tuning to toy (sine) functions



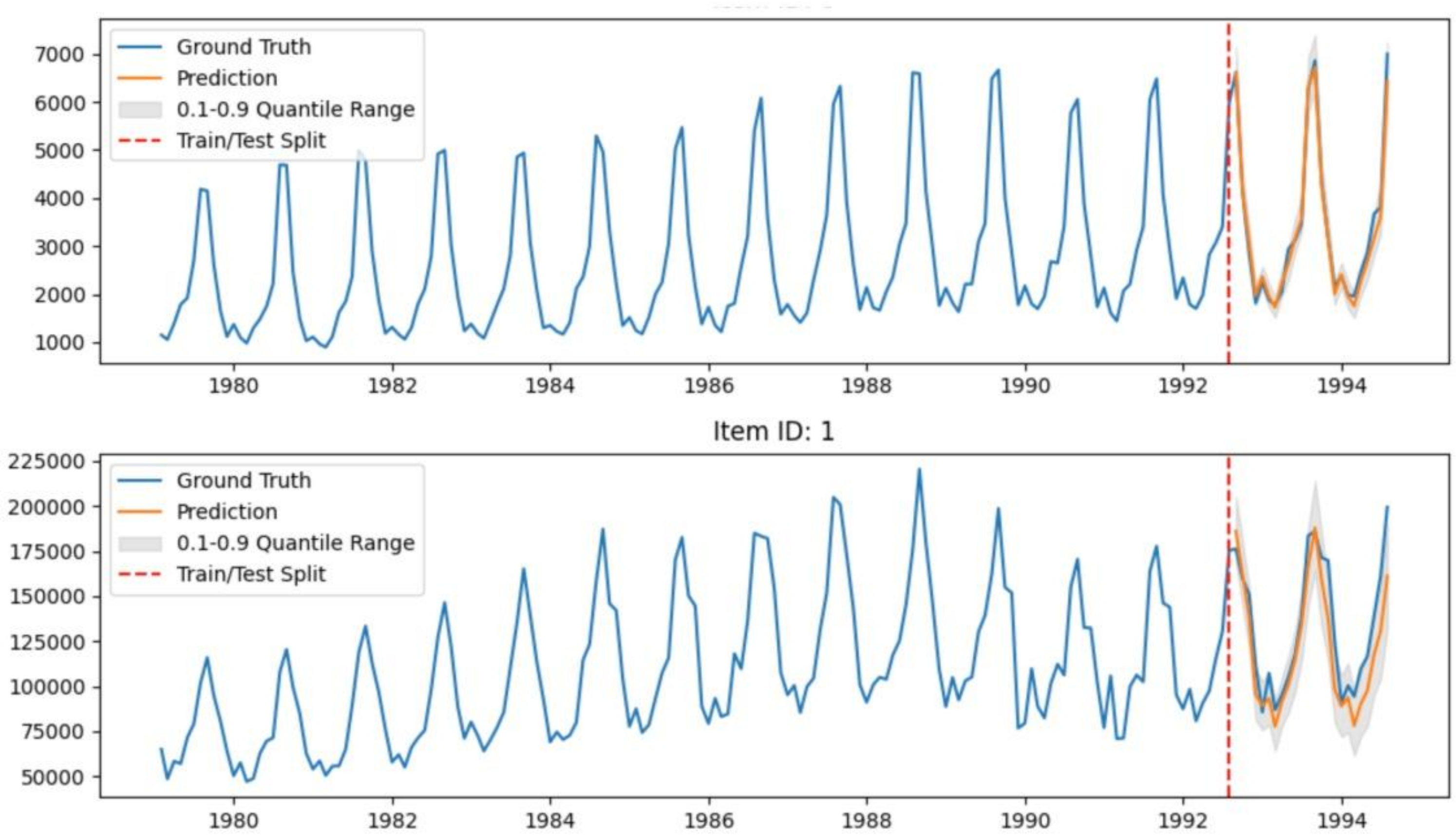
Fine-tuning to (broad collection of) real datasets



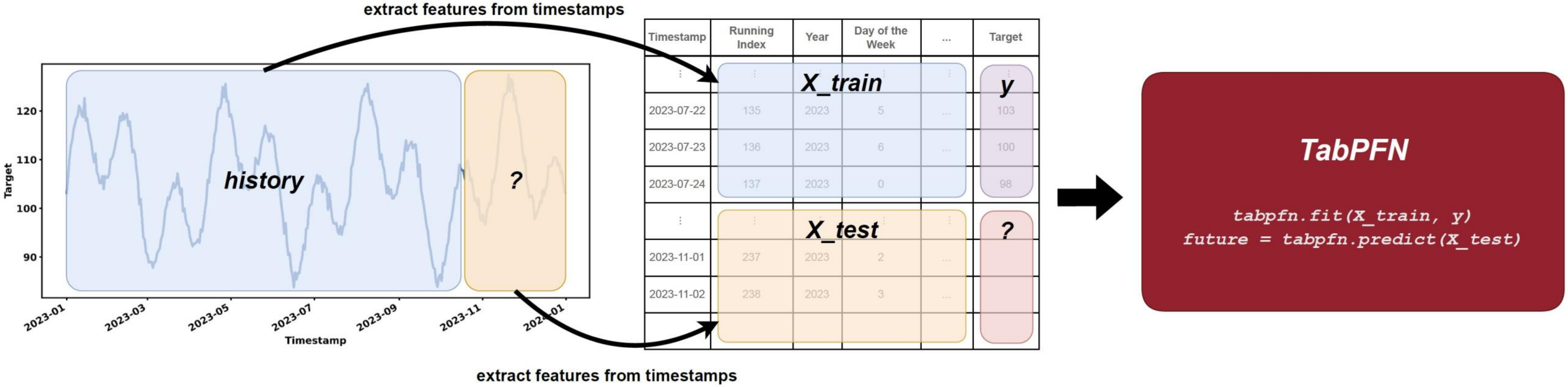
Outline

- Motivation
- TabPFN
- TabPFN v2
- TabPFN for time series: TabPFN-TS
- Explainability & Fairness

TabPFN v2 also excels on time series data: TabPFN-TS



Casting time series forecasting as tabular regression



January 2025: This simple extension achieves SOTA on GIFT-Eval

	Point Forecast	Probabilistic Forecast	Rank of CRPS

T	model	MASE	CRPS	Rank
●	TabPFN-TS	0.748 #4	0.48 #2	6.649 #1
●	chronos_bolt_base	0.725	0.485	6.856
●	timesfm_2_0_500m	0.680	0.465	6.897
●	chronos_bolt_small	0.738	0.487	7.392
◆	PatchTST	0.762	0.496	8.258
●	Moirai_large	0.785	0.506	8.381
●	Moirai_base	0.809	0.515	8.454
◆	TFT	0.822	0.511	9.505
●	Moirai_small	0.849	0.549	11.227

January 2025: This simple extension achieves SOTA on GIFT-Eval

	T	model	MASE	CRPS	Rank
			Point Forecast	*Probabilistic Forecast*	*Rank of CRPS*
Prior Labs, 11M		TabPFN-TS	0.748 **#4**	0.48 **#2**	6.649 **#1**
Amazon, 205M		chronos_bolt_base	0.725	0.485	6.856
Google, 500M		timesfm_2_0_500m	0.680	0.465	6.897
		chronos_bolt_small	0.738	0.487	7.392
		PatchTST	0.762	0.496	8.258
		Moirai_large	0.785	0.506	8.381
		Moirai_base	0.809	0.515	8.454
		TFT	0.822	0.511	9.505
		Moirai_small	0.849	0.549	11.227

January 2025: This simple extension achieves SOTA on GIFT-Eval

Synthetic tabular data
Synthetic + real TS data
Real TS data

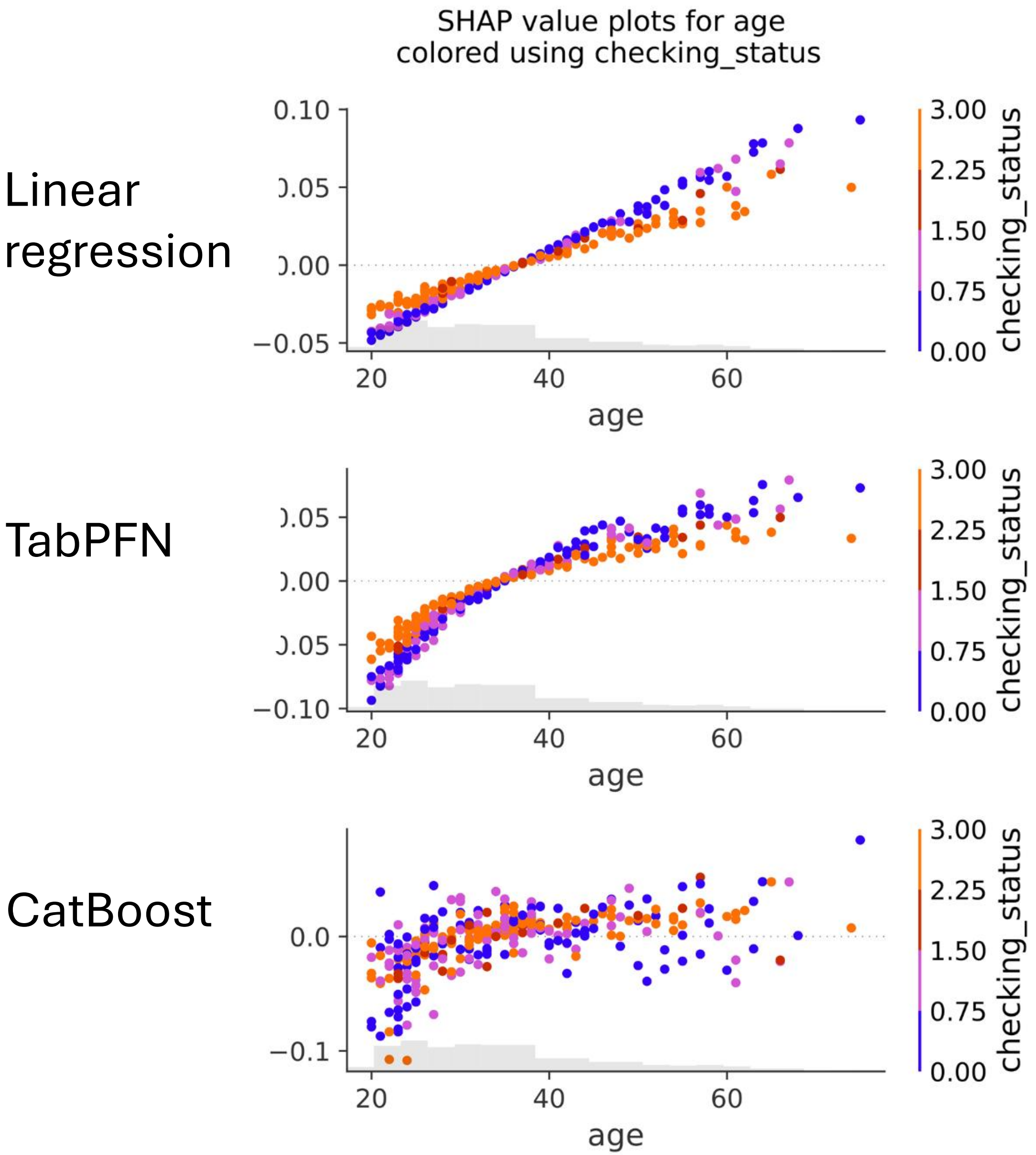
		Point Forecast	Probabilistic Forecast	Rank of CRPS
T	model	MASE	CRPS	Rank
●	TabPFN-TS	0.748 #4	0.48 #2	6.649 #1
●	chronos_bolt_base	0.725	0.485	6.856
●	timesfm_2_0_500m	0.680	0.465	6.897
●	chronos_bolt_small	0.738	0.487	7.392
◆	PatchTST	0.762	0.496	8.258
●	Moirai_large	0.785	0.506	8.381
●	Moirai_base	0.809	0.515	8.454
◆	TFT	0.822	0.511	9.505
●	Moirai_small	0.849	0.549	11.227

Outline

- Motivation
- TabPFN
- TabPFN v2
- TabPFN for time series: TabPFN-TS
- Explainability & Fairness

Explainability: what effect does each feature have?

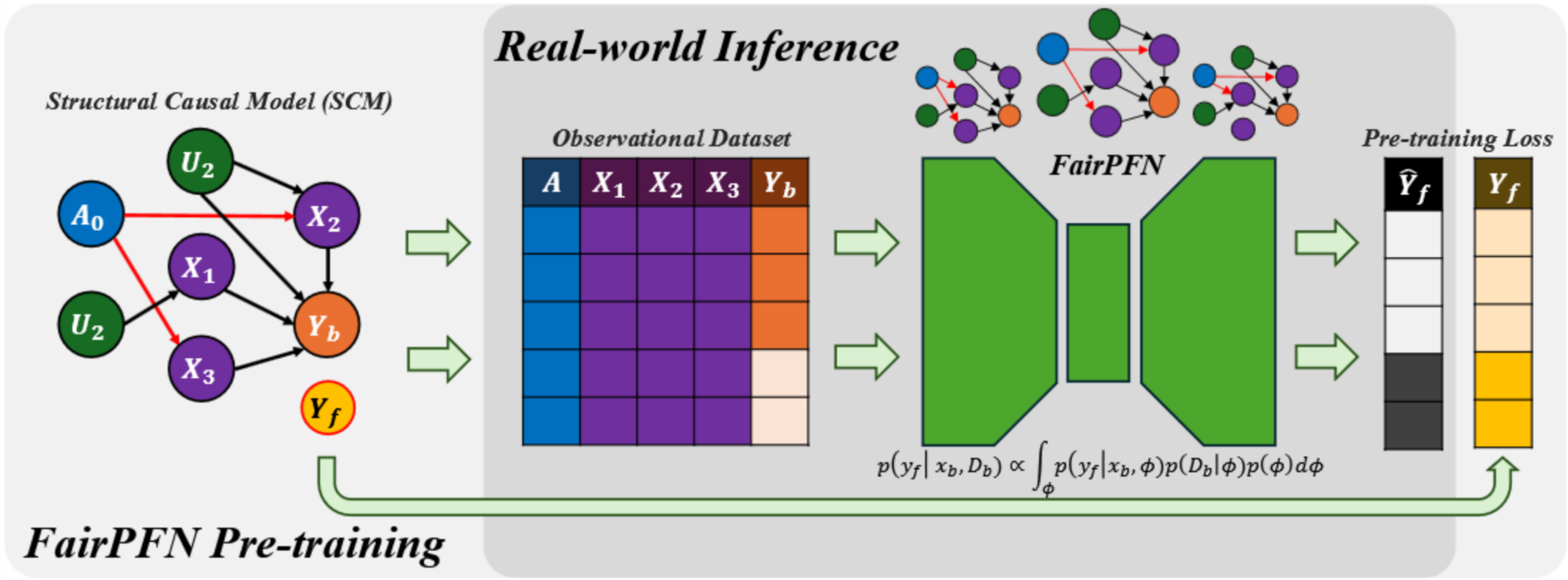
- SHAP analysis yields more reliable results for TabPFN
  - Much better predictions than linear regression
    $\rightarrow$ captures nonlinear effects
  - Much smoother predictions than boosted trees
    $\rightarrow$ clearer SHAP patterns



Counterfactual Fairness with TabPFN

[Robertson et al, ICML 2025]

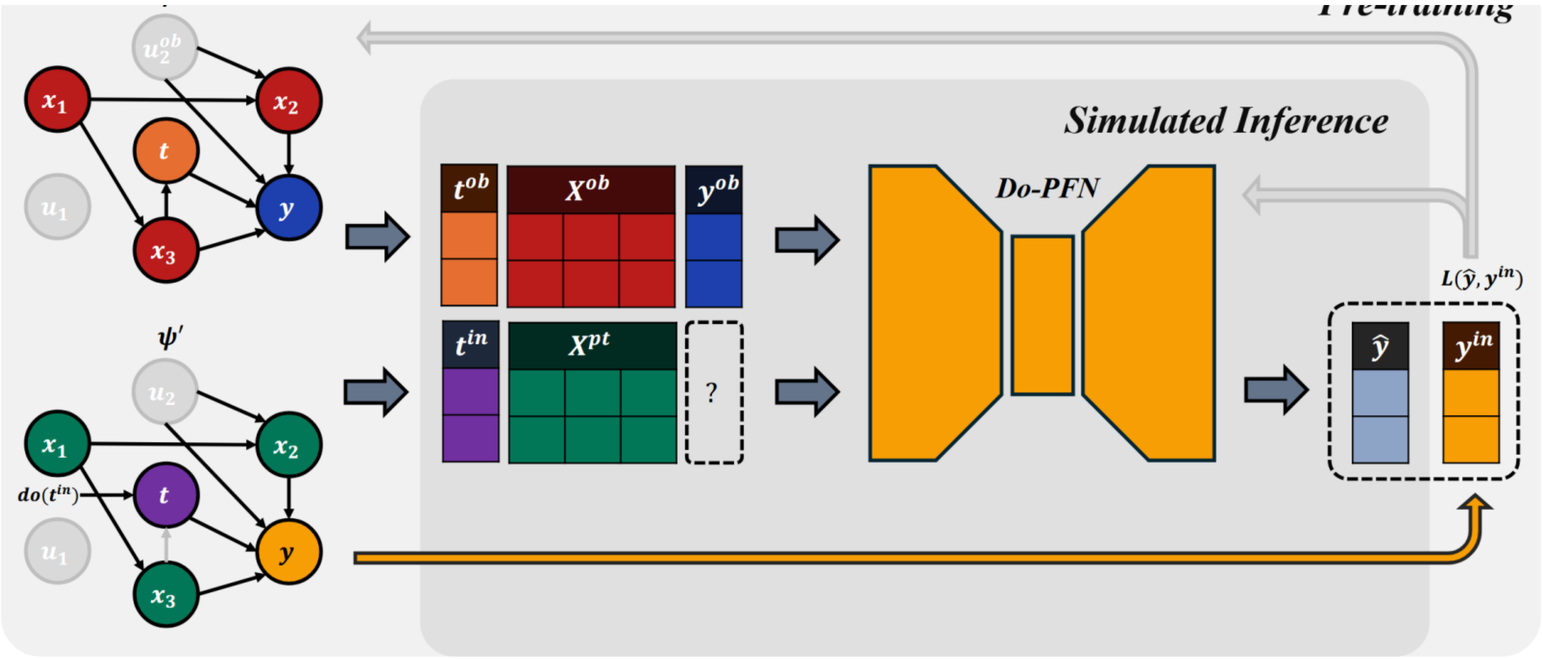
- Counterfactual reasoning: what would the result be IF the protected feature changed?
  - "Holy Grail": **remove the protected feature's causal effect on other features**
- Solution with TabPFN's prior sampling:
  - Generate standard X_{biased} and y_{biased}, and remove causal effect to generate X_{fair}, Y_{fair}
  - Learn to map from $X_{train, biased}$, $y_{train, biased}$, $X_{test, biased}$ to $y_{test, fair}$
- Substantially outperforms standard methods



Interventional predictions with Do-PFN

[Robertson et al, arXiv 2025]

- Interventional reasoning: what will happen to y if I change t?
- Solution with TabPFN's prior sampling:
  - Generate standard observational data tob ,Xob and yob, and interventional t, Xpt, yin
  - Learn to map from Xob, yob, Xpt, t, to yin
- Substantially outperforms standard methods



Take-aways

- **TabPFN is the new default for small tabular ML**
  - Currently: up to 10k data points, 500 features; scaling up further
  - Unique features compared to previous methods
    - **Faster** (no HPO needed, more interactive data science)
    - Better **peak performance**
    - Works well with **less data**
  - More interpretable

- **Finetuning clearly improves performance**
  - Customization to various use cases

Open source

Thank you for your attention, and to my fantastic team!

- ✓ AI Scientists
- ✓ Engineers
- ✓ Data Scientists
- ✓ Developer relations
- ✓ Internships
- ✓ Founder Associate
- ✓ Product Manager

$5000 USD thank-you
if we hire your referral
for a fulltime position!

http://priorlabs.ai

Email: frank@priorlabs.ai