# NanoTabPFN Speedrun

Salih Bora Öztürk [1] Alexander Pfefferle Frank Hutter [1,2]

[1]University of Freiburg [2]Prior Labs

## TL;DR

**modded-nanoTabPFN**: This repository hosts the nanoTabPFN speedrun, in which we search for the fastest way to train a tabular foundation model that beats Random Forest on TabArena datasets.

## Introduction

### Motivation

- TabPFN have shown that pretraining on synthetic datasets can lead to strong performance... [1] [2]
- However, training these models takes long, and no one has time to wait...

### Background

- **PFN** prior fitted network... [3]
- **TabPFN** for tabular data... [1] [2]
- **nanoTabPFN** a lightweight and educational reimplementation of TabPFN [4]
- **TabArena** a living benchmark for machine learning on tabular data [5]
- **modded-nanogpt** OG speedrunning repo... [6]

### Goal

- Pretrain a neural network to ≤0.8068462330697953 validation average ROC AUC on subsampled TabArena using 1 NVIDIA L40S.

**This repo now contains a training algorithm which attains the target performance in:**

- **10.10 minutes** on 1xL40S (baseline needed 74.32)
- **13184 synthetic datasets** (baseline needed 80576)

## Rules

New records must:

1. Not modify the evaluation pipeline.
2. Not load any pretrained weights.
3. Run faster than prior record when baselined on the same hardware.

Other than that, anything and everything is fair game!

## Evaluation details

Evaluation is on all of 38 TabArena classification datasets.

- if >100 features, randomly select 100
- if >1000 rows, randomly select 1000 (stratified by class labels)
- 5-fold StratifiedKFold with shuffling, class labels are encoded with integers per fold
- average binary or one-vs-rest ROC AUC over all datasets

## Improvement techniques

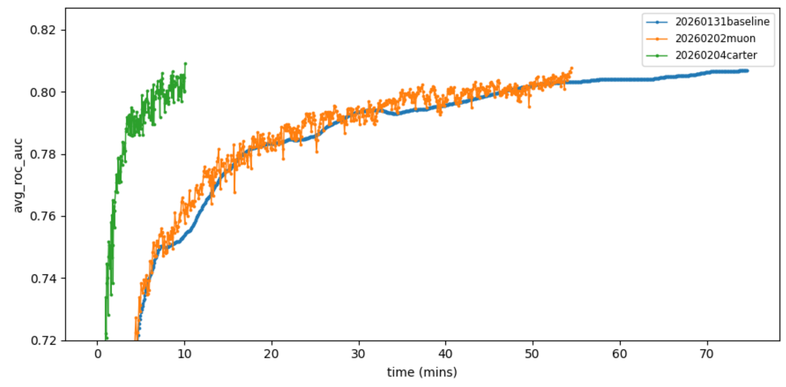
**With the help of the following techniques:**

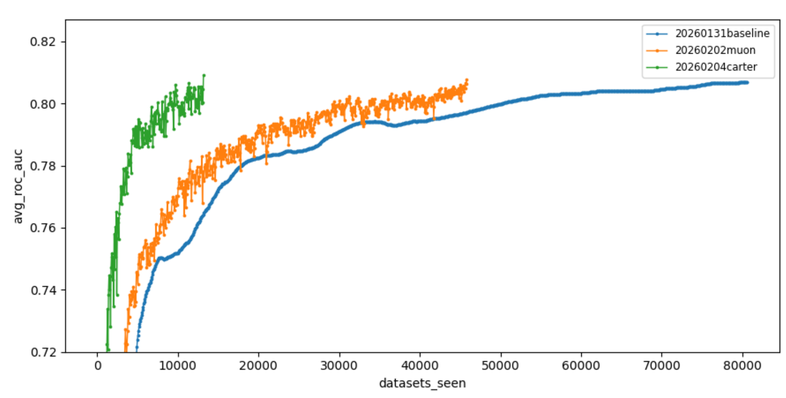
- Muon optimizer [7]
- ...
- ...
- ...

**Also these were tried but did not lead to improvements:**

- Xavier initialization

## Record history




| Record time | Description | Contributors |
|---|---|---|
| 74.32 mins | Baseline | @borawhocodess, nano-tabpfn contributors |
| 54.41 mins | Muon optimizer | @borawhocodess |
| 10.10 mins | SDPA, bf16, higher LR, wider embeddings, fewer heads | @carterprince |
| ... | ... | ... |
| ... | ... | ... |
| ... | ... | ... |

## Join the race!

Figure 1. github.com/borawhocodess/modded-nanotabpfn

## References

[1] Noah Hollmann, Samuel Müller, Katharina Eggensperger, and Frank Hutter. TabPFN: A transformer that solves small tabular classification problems in a second. In *International Conference on Learning Representations (ICLR)*, 2023.

[2] Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045):319–326, 2025.

[3] Samuel Müller, Noah Hollmann, Sebastian Pineda Arango, Josif Grabocka, and Frank Hutter. Transformers can do Bayesian inference. In *International Conference on Learning Representations (ICLR)*, 2022.

[4] Alexander Pfefferle, Johannes Hog, Lennart Purucker, and Frank Hutter. nanotabpfn: A lightweight and educational reimplementation of tabpfn, 2025.

[5] Nick Erickson, Lennart Purucker, Andrej Tschalzev, David Holzmüller, Prateek Mutalik Desai, David Salinas, and Frank Hutter. Tabarena: A living benchmark for machine learning on tabular data, 2025.

[6] Keller Jordan, Jeremy Bernstein, Brendan Rappazzo, @fernbear.bsky.social, Boza Vlado, You Jiacheng, Franz Cesista, Braden Koszarsky, and @Grad62304977. modded-nanogpt: Speedrunning the nanogpt baseline, 2024.

[7] Keller Jordan, Yuchen Jin, Vlado Boza, Jiacheng You, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks, 2024.