TabTransformer++: Residualized, Calibrated Transformer Architecture for Tabular Data

Author: L. Elaine Dazzio

Abstract

We present TabTransformer++, an extension of transformer architectures for tabular prediction tasks. Unlike prior work, TabTransformer++ is residualized against strong ensemble baselines, incorporates fold-wise isotonic calibration, and introduces value-tower fusion of standardized numerical features into embedding space. These innovations yield consistent improvements in out-of-fold RMSE and more reliable uncertainty quantification. Our results demonstrate that residualized calibration enables transformers to match or exceed gradient-boosted trees while maintaining the flexibility of deep learning.

1. Introduction

Tabular data remains a frontier for deep learning. Boosted decision trees (BDTs) dominate in practice, while transformers have shown promise with contextual embeddings and attention mechanisms. Recent approaches such as cross-table pretraining and LLM-based embeddings highlight growing momentum. However, reproducible and calibrated improvements remain elusive.

2. Related Work

Gradient Boosted Trees (GBDTs): XGBoost, LightGBM, CatBoost. Transformer baselines: FT-Transformer, TabTransformer. Enhanced tabular transformers: SAINT, XTab. Emerging LLM approaches: leveraging pretrained embeddings for tabular tasks.

3. TabTransformer++ Architecture

Residualization: train transformer against predictions of strong BDT ensembles. Leak-free tokenization: per-fold quantile binning of continuous features. Value towers: gated fusion of standardized numerical inputs into embeddings. Calibration: isotonic regression applied per-fold to correct predictive bias.

4. Experiments

Benchmarks: UCI datasets, financial tabular tasks, synthetic residualized baselines. Metrics: RMSE, AUC, reliability diagrams. Result: consistent improvements over TabTransformer and FT-Transformer.

5. Conclusion

TabTransformer++ demonstrates that carefully residualized and calibrated transformers can match or surpass tree-based ensembles, while enabling integration with multimodal pipelines. This work establishes a new baseline for reproducible, calibrated tabular deep learning.

References (IEEE style)

[1] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, “Revisiting deep learning models for tabular data,” Advances in Neural Information Processing Systems, vol. 34, 2021.

[2] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, “TabTransformer: Tabular data modeling using contextual embeddings,” arXiv preprint arXiv:2012.06678, 2020.

[3] G. Somepalli, M. Goldblum, A. Schwarzschild, C. B. Bruss, and T. Goldstein, “SAINT: Improved neural networks for tabular data via row attention and contrastive pre-training,” arXiv preprint arXiv:2106.01342, 2021.

[4] B. Zhu, X. Shi, N. Erickson, M. Li, and G. Karypis, “XTab: Cross-table pretraining for tabular transformers,” in Proc. ICML, 2023.

[5] B. Koloski, A. Margeloiu, X. Jiang, B. Škrlj, N. Simidjievski, and M. Jamnik, “LLM embeddings for deep learning on tabular data,” arXiv preprint arXiv:2502.11596, 2025.