From Big Time Series Forecasting to Foundation Models for Structured Data

Xiyuan Zhang
Amazon Web Services
USA
xiyuanz@amazon.com

Abdul Fatir Ansari
Amazon Web Services
Germany
ansarnd@amazon.de

Christos Faloutsos
Carnegie Mellon University and Amazon
USA
christos@cs.cmu.edu

George Karypis
University of Minnesota and NTT Data AIVista
USA
karypis@umn.edu

Yuyang Wang
Amazon Web Services
USA
yuyawang@amazon.com

ABSTRACT

Prediction over structured data such as forecasting future values and classifying rows has long been a central concern for the database and data-mining communities. Our earlier tutorials presented in the community (VLDB, SIGMOD, KDD) traced the evolution from classical models and tensor methods to neural forecasting, probabilistic forecasting, and practical systems. Recently, foundation models have transformed machine learning, with large-scale pretraining followed by zero-shot or in-context inference now defining the state of the art. This paradigm has arrived for structured data, with foundation models emerging for both time series (Chronos, TimesFM, Moirai, Toto, etc.) and tabular tasks (TabPFN, TabICL, Mitra, Limix, etc.). This proposal continues our tutorial line and adds the modern layer of *foundation models for structured data.*

The objective of this tutorial is to provide a comprehensive survey of foundation models for structured data for a database audience. We review the state of the art in two related areas: (1) time-series foundation models, covering tokenization, zero-shot forecasting, covariates, and multivariate prediction, and (2) tabular foundation models, covering synthetic priors, in-context learning, and relational pre-training. We also discuss how these two fields are converging with shared pretraining strategies and unified architectures. As for their implication for data management systems, we provide case studies from AWS how forecasting helps improving the efficiency of the system with predictive auto-scaling and capacity planning. Further, we argue that pretrained models open the door to tighter native integration of some predictive tasks into OLAP workflows, because many workloads can now shift from task-specific training pipelines to context construction plus inference inside analytical plans.

PVLDB Reference Format:
Xiyuan Zhang, Abdul Fatir Ansari, Christos Faloutsos, George Karypis, and Yuyang Wang. From Big Time Series Forecasting to Foundation Models for Structured Data. PVLDB, 19(12): 4939 - 4943, 2026.
doi:10.14778/3827998.3828154

Proceedings of the VLDB Endowment, Vol. 19, No. 12 ISSN 2150-8097.
doi:10.14778/3827998.3828154

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://vldb-fmsd-tutorial.github.io/.

1 INTRODUCTION

Structured data, especially time series and tables, remains the dominant data format in various high-impact applications such as finance, healthcare, enterprise decision-making, and climate modeling. Forecasting product demand, planning data-center capacity, scheduling workforces, and monitoring sensor networks all depend on accurate, scalable prediction over large collections of related time series [6, 16]. Meanwhile, tabular classification and regression underpin credit scoring, clinical decision support, fraud detection, and recommendation [22, 42]. Despite decades of progress, building production-grade prediction pipelines still requires intensive feature engineering, model selection, and domain-specific tuning [12, 23].

The last two years have witnessed a paradigm shift. In natural-language processing and computer vision, foundation models, which are large pretrained models that generalize across tasks via zero-shot or few-shot inference, have transformed the landscape. The same idea is now reaching structured data. In time series, pretrained models such as Chronos [3, 4], TimesFM [10], Moirai [28, 44], Tirex [5], Toto [9, 26], and FlowState [18] can forecast unseen series without task-specific training, by learning shared temporal patterns from large and diverse corpora. In tabular learning, models such as TabPFN [20, 22], TabICL [33, 34], TabDPT [29], Limix [46], and Mitra [45], leverage synthetic priors, in-context learning, and large-scale pretraining to solve classification and regression tasks on novel tables with minimal or no adaptation. Moreover, the boundary between forecasting and broader structured prediction is rapidly dissolving, as both families share pretraining strategies, data generation pipelines, and deployment considerations.

Our proposed VLDB 2026 tutorial keeps that backbone from our earlier versions [14–16] and updates it with the most important recent development of *foundation models for structured data*, aiming at a broad database audience. We will connect the old and the new: from large-scale forecasting as a central database and data-mining problem to the emerging foundation-model view of structured data. Chronos and Mitra serve as running examples within a broad survey covering time-series foundation models, tabular foundation models, evaluation, and system considerations. Specifically, we will cover:

- **Time-series foundation models:** Pretrained models such as Chronos [3, 4], TimesFM [10], Moirai [44] and Toto [9] enable zero-shot forecasting by learning shared temporal patterns from diverse corpora. We cover tokenization strategies, covariate handling, multivariate prediction, and the progression toward universal forecasting.
- **Tabular foundation models:** Models such as TabPFN [19, 20, 22], TabICL [33, 34], TabDPT [29], Limix [46] and Mitra [45] bring in-context learning to tabular prediction. We cover synthetic priors, real-data pretraining, relational pretraining, and scaling to larger tables.
- **Data-management considerations:** efficient system management with predictive analytics with cases studies in DynamoDB Athena, and Redshift. We also argue that pre-trained predictive models enables ML as a native analytic workload, which was not possible in the traditional train-test paradigm.
- **Convergence and open problems:** Time-series and tabular foundation models increasingly share architectures, pretraining strategies, and evaluation challenges. We discuss emerging unified approaches and identify open research directions.

Who should attend. The target audience is database researchers, data-mining researchers, ML-systems researchers, and practitioners who want a principled introduction to modern prediction over structured data. The tutorial is especially relevant to attendees interested in forecasting, tabular prediction, benchmarking, warehouse-native ML, AutoML, and end-to-end deployment inside data systems. We assume familiarity with basic linear algebra, probability, and machine-learning fundamentals (supervised learning, loss functions, gradient descent), but no prior knowledge of foundation models or structured-data pretraining. Attendees will leave with: (1) a coherent map from classical forecasting to foundation models for structured data, (2) an understanding of the architectural choices, pretraining strategies, and evaluation practices that define the current state of the art, and (3) awareness of open research problems at the intersection of foundation models and data management.

Related tutorials and how the current proposal differs. This proposal continues a long-running tutorial line. *Forecasting Big Time Series: Old and New* (VLDB 2018) surveyed classical forecasting, tensor methods, neural forecasting, and lessons from large-scale systems [15]. *Classical and Contemporary Approaches to Big Time Series Forecasting* (SIGMOD 2019) broadened the treatment of classical and neural methods [16]. *Forecasting Big Time Series: Theory and Practice* (KDD 2019, WWW 2020) pushed toward operational practice and interactive materials [14]. *Modern Aspects of Big Time Series Forecasting* (IJCAI 2021) [1] and *Time Series Forecasting with Deep Probabilistic Approaches* (ECML 2023) [2] tutorials refined the deep-learning story with emphasis on probabilistic prediction, evaluation, and multivariate settings. The proposed VLDB 2026 tutorial differs in four ways. (1) It preserves the forecasting backbone but expands scope to the broader landscape of *foundation models for structured data*, covering both time-series and tabular modalities in a unified treatment. (2) It updates the modeling narrative from deep probabilistic forecasting to pretraining, zero-shot inference, in-context learning, synthetic priors, and universal multivariate forecasting, which is a qualitative shift in how prediction is done. (3) It foregrounds data-management questions that did not define the earlier tutorials: schema heterogeneity, corpus curation, contamination-aware evaluation, realistic benchmarks [1, 40], throughput and memory trade-offs, and governance. (4) It adds a new systems perspective: pretrained models create the possibility of embedding predictive operators into analytical execution plans, an implication we discuss as a forward-looking research direction.

[1] https://lovvge.github.io/Forecasting-Tutorial-IJCAI-2021/

[2] https://lovvge.github.io/Forecasting-Tutorial-ECML-2023/

2 OUTLINE

We propose a **1.5-hour** tutorial (five modules with their respective durations listed in Table 1). The format is *survey* with lightweight demonstrations; the emphasis is on concepts, design trade-offs, evaluation, and open research questions. The tutorial is organized into five modules. **Module 1** provides historical context, tracing the evolution for both time series (classical local and global forecasting methods) and tabular models (tree-based tabular models), setting up the paradigm shift toward foundation models. **Module 2** surveys time-series foundation models in depth: tokenization strategies, architectural choices, pretraining data and augmentation, zero-shot and in-context inference, and contamination-aware evaluation. **Module 3** covers the parallel development in tabular foundation models: synthetic priors, in-context learning, real-data pretraining, and the relationship to tree-based baselines. **Module 4** translates these developments into data-management and systems questions, using AWS case studies (DynamoDB, Athena, Redshift) to illustrate how predictive models integrate into production infrastructure, and discussing the prospect of native predictive operators in analytical systems. **Module 5** concludes with live demonstrations, open research problems, and discussion.

2.1 Module 1: From Classical Methods to Foundation Models (10 min)

This module provides the historical backbone. We revisit the progression from traditional local or global models time series models, as well as tree-based and AutoML-based tabular models that appear before foundation models, making explicit how the new models build on the existing foundations covered in the earlier tutorials.

- *Traditional local time series models.* We briefly recap exponential smoothing (ETS), ARIMA, structural time series models, and the Kalman filter as the workhorse of traditional single-series forecasting [7, 11, 21, 23].
- *Traditional global neural time series models.* We trace the shift from local to global methods: DeepAR [39], Temporal Fusion Transformer [27], N-BEATS [31], and their variants learn shared parameters across thousands of related series. We discuss how this shift motivated scalable training systems [2, 6]. We also cover deeep probabilistic forecasting models including deep state-space models [35], normalizing flows [37], and diffusion-based

forecasters [43], emphasizing probabilistic calibration and proper scoring rules [17].

- *Traditional tabular models.* Traditionally, tree-based models dominate practice in tabular domain such as gradient-boosted trees (XGBoost [8], LightGBM [25], CatBoost [32]) and tabular data has historically resisted deep-learning dominance. More recently, AutoML systems such as AutoGluon [12, 41] further automate feature preprocessing and model ensemble which leads to improved performance.
- *The paradigm shift.* We set up the central theme: pretrained foundation models move from task-specific training to large-scale pre-training and zero-shot inference.

2.2 Module 2: Time-Series Foundation Models (30 min)

This module surveys the design space and evaluation of pretrained time-series models. It compares design choices for time-series foundation models, including tokenization versus continuous representations, decoder-only versus encoder-decoder designs, multivariate and covariate-informed forecasting, and contamination-aware evaluation.

- *Tokenization and input representation.* We compare the tokenization approach of Chronos [4] (scaling and quantizing real values into discrete tokens) with patching strategies [30, 44] and continuous representations [10, 36]. We discuss trade-offs between vocabulary size, reconstruction fidelity, and sequence length.
- *Architecture families.* We survey decoder-only models such as TimesFM [10] and Lag-Llama [36], encoder–decoder designs such as Chronos [4], and hybrid architectures. We explain how attention mechanisms, in particular the group attention in Chronos-2 [3], enable multivariate and covariate-aware forecasting without explicit cross-series engineering.
- *Pretraining data and augmentation.* We discuss how large-scale corpus construction which mixes real-world data (e.g., from public repositories) with synthetic time series generated by Gaussian processes and other stochastic models drives generalization [4]. We cover data augmentation strategies such as KernelSynth [4].
- *Zero-shot, fine-tuning, and in-context learning.* We show how pretrained models can forecast unseen series out of the box, and when and how adaptation via fine-tuning or in-context conditioning improves performance. We highlight the universal forecasting capability of Chronos-2 [3] across both univariate and multivariate settings.
- *Evaluation and benchmarks.* We discuss the pitfalls of time-series evaluation: data leakage, contamination, overlapping train–test splits, and metric sensitivity. We cover GIFT-Eval [1] and fev-bench [40] as efforts toward contamination-aware, realistic evaluation, and discuss open challenges in benchmark design.

2.3 Module 3: Tabular Foundation Models (30 min)

Tabular data presents unique challenges from time series and tree-based models have long dominated tabular benchmarks precisely because they handle heterogeneous features, missing values, and irregular distributions without extensive preprocessing. The recent emergence of tabular foundation models challenges this by reframing tabular prediction as in-context learning: given a table as context, the model predicts labels for query rows without gradient updates. This formulation enables zero-shot transfer across datasets with different schemas. We examine how these models achieve cross-table generalization and when they complement or outperform tree-based pipelines.

This module surveys the emerging landscape of pretrained models for tabular classification and regression. It covers tabular foundation models through the lens of synthetic priors, in-context learning, real-data pretraining, scaling to larger tables, and the relationship to tree-based baselines.

- *Synthetic priors and TabPFN.* We explain how TabPFN [19, 20, 22] uses prior-data fitted networks to pre-train on millions of synthetic classification tasks sampled from structural casual models and perform in-context learning at inference time without gradient updates.
- *Scaling in-context learning.* We cover TabICL [33, 34] and its techniques for scaling in-context tabular learning to larger datasets, and TabDPT [29], which pretrains on real-world tabular corpora to improve transfer.
- *Mixed priors and Mitra.* We present Mitra [45], which combines synthetic and real-data priors to enhance robustness and generalization across heterogeneous table schemas.
- *Ensembling and AutoML integration.* We discuss how foundation models fit into multi-layer stack ensembles alongside tree-based and neural baselines [12], and when zero-shot tabular models outperform or complement traditional pipelines.
- *Evaluation and benchmarks.* Similarly as time series, we discuss the pitfalls of tabular evaluation including benchmark contamination from web-scraped pretraining data, inconsistent preprocessing, and unfair comparison against baselines. We cover TabRepo [38] and TabArena [13] as efforts toward standardized evaluation, and discuss open challenges in benchmark design.

2.4 Module 4: Data Management and Systems Implications (15 min)

This module translates the modeling developments into questions central to the VLDB community, efficient system management with predictive analytics and the possibility of native predictive operators in analytical systems.

- *DynamoDB: long-horizon capacity planning*. We use the case study of DynamoDB from the AWS forecasting stack paper to show how predictive models support annual service-planning decisions for highly seasonal, heteroscedastic demand. We discuss why deep state-space models were chosen in that setting: they combine data-driven learning with explicit control over recurring patterns, helping avoid long-horizon error accumulation [24].
- *Athena: forecasting for warm-pool sizing*. We present Athena as a concrete example of forecasting directly improving system efficiency. Replacing hand-tuned heuristics with probabilistic DeepAR forecasts for serverless instance demand reduced warm-pool sizes by 30–45% for selected instance types and availability zones while preserving customer experience, and increased cluster utilization by roughly 4× [24].
- *Redshift: cache-pool provisioning under intermittent demand.* We discuss Redshift as a case where intermittent demand and many

Table 1: The tutorial spans 1.5 hours and is organized into five modules.

Section	Minutes
Historical view of classical, global, deep probabilistic, and traditional tabular models	10
Time-series foundation models: tokenization, patching, zero-shot forecasting, covariates, universal forecasting	30
Tabular foundation models: synthetic priors, ICL, real-data pretraining, efficiency, robustness	30
Data-management and system implications: efficient system management with predictive analytics and the possibility of native predictive operators in analytical systems	15
Demos, failure modes, open problems, future outlook and discussion	5

configuration-specific series make probabilistic forecasting operationally valuable. The team used Amazon Forecast's AutoML capabilities to choose models suited to sparse demand, making the cost–latency trade-off explicit through forecast quantiles and yielding utilization gains of 35–91% with millions of dollars in annual savings [24].

- *Toward predictive operators in analytical systems.* We argue that pretrained models, by decoupling training from inference, make it feasible to express some predictive workloads as context construction plus model invocation inside analytical query plans. We survey early work and discuss open challenges for native integration, including cost modeling, plan optimization, result caching, and efficient execution across heterogeneous workloads.

2.5 Module 5: Demos, Open Problems, and Discussion (5 min)

We conclude with short live demonstrations showing zero-shot forecasting with a pretrained time-series model and in-context tabular prediction, illustrating how a pretrained model can be dropped into an inference-first pipeline without task-specific training. We then summarize open research problems, including multi-modal structured-data models, continual pretraining, active data acquisition, privacy-preserving pretraining, and tighter database integration, and open the floor for discussion.

We expect the tutorial to leave the audience with three takeaways: (1) a coherent map from classical forecasting and deep probabilistic forecasting to foundation models for structured data; (2) a concrete understanding of the statistical, evaluation, and systems questions that arise in this new regime; and (3) a forward-looking view of why pretrained structured-data models may enable tighter native database integration for some predictive workloads.

3 PRESENTERS' SHORT BIOGRAPHIES

Xiyuan Zhang is an Applied Scientist at Amazon Web Services working on structured foundation models, especially on pre-training and multimodal analysis. She is the lead author of Mitra, one of the most downloaded tabular foundation models on Huggingface, and co-author of Chronos and Chronos-2. Xiyuan earned her PhD in Computer Science from the University of California, San Diego. She has published and received Top Reviewer Award at top machine learning conferences such as NeurIPS, and has co-organized the "Foundation Models for Structured Data" workshop at ICML 2025 and ICML 2026, as well as the "Time Series in the Age of Large Models" at ICLR 2026. Xiyuan is a recipient of the Qualcomm Innovation Fellowship and has been recognized as a Cyber-Physical-System (CPS) Rising Star.

Abdul Fatir Ansari is a Senior Applied Scientist at Amazon Web Services, focusing on time series forecasting. He is the lead developer of the Chronos family of time series foundation models, which ranked among the most downloaded models on HuggingFace in 2024. He received his PhD in 2022 from the National University of Singapore, where he worked on deep generative models for images and time series, receiving the Dean's Award for his research. He has published in and reviewed for top machine learning conferences including ICML, ICLR, and NeurIPS, and was twice recognized as a top reviewer.

Christos Faloutsos is a Professor at Carnegie Mellon University and an Amazon Scholar. He is the recipient of the Fredkin Professorship in Artificial Intelligence (2020); he has received the Presidential Young Investigator Award by the National Science Foundation (1989), the Research Contributions Award in ICDM 2006, the SIGKDD Innovations Award (2010), the PAKDD Distinguished Contributions Award (2018), 32 "best paper" awards (including 8 "test of time" awards), and four teaching awards. He is an ACM Fellow, he has served as a member of the executive committee of SIGKDD; he has published over 500 refereed articles, 17 book chapters and three monographs. He holds 12 patents (and several more are pending), and he has given over 50 tutorials and over 25 invited distinguished lectures. His research interests include large-scale data mining with emphasis on graphs and time sequences; anomaly detection, tensors, and fractals.

George Karypis is a Distinguished McKnight University Professor and a William Norris Chair in Large Scale Computing in the Department of Computer Science and Engineering at the University of Minnesota, and the chief scientist at NTT Data AIVista. His academic career spans decades of influential research across machine learning, artificial intelligence, data mining, high performance computing, and recommender systems. His contributions have earned him the 2025 ACM SIGKDD Innovation Award for pioneering work in graph partitioning, recommendation systems, and scalable algorithms for data mining and machine learning, recognition as an IEEE Fellow, several test-of-time and research contributions awards, and numerous best-paper awards.

Yuyang (Bernie) Wang is a Member of Technical Staff at NTT DATA AIVista. Previously, he was a Principal Scientist at Amazon, where he led research in automated machine learning, time series forecasting, and foundation models. His work focuses on making advanced® AI/ML systems reliable, scalable, and broadly accessible. He has helped build widely used open-source systems including AutoGluon, GluonTS, and Chronos, and contributed to production AI services such as Amazon Forecast and SageMaker. He has led science teams across multiple AWS products, built scalable algorithms and forecasting systems for Amazon retail and AWS platforms, and delivered more than six tutorials on time series forecasting at major venues including VLDB, SIGMOD, and KDD.

REFERENCES

[1] Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. 2024. Gift-eval: A benchmark for general time series forecasting model evaluation. *arXiv preprint arXiv:2410.10393* (2024).

[2] Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Danielle C Maddix, Syama Rangapuram, David Salinas, Jasper Schulz, et al. 2020. Gluonts: Probabilistic and neural time series modeling in python. *Journal of Machine Learning Research* 21, 116 (2020), 1–6.

[3] Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, et al. 2025. Chronos-2: From univariate to universal forecasting. *arXiv preprint arXiv:2510.15821* (2025).

[4] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. 2024. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815* (2024).

[5] Andreas Auer, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. 2025. Tirex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. *arXiv preprint arXiv:2505.23719* (2025).

[6] Joos-Hendrik Böse, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger, and Yuyang Bernie Wang. 2017. Probabilistic demand forecasting at scale. (2017).

[7] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control.* John Wiley & Sons.

[8] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.* 785–794.

[9] Ben Cohen, Emaad Khwaja, Kan Wang, Charles Masson, Elise Ramé, Youssef Doubli, and Othmane Abou-Amal. 2024. Toto: Time series optimized transformer for observability. *arXiv preprint arXiv:2407.07874* (2024).

[10] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. 2023. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688* (2023).

[11] James Durbin and Siem Jan Koopman. 2012. *Time series analysis by state space methods.* Oxford University Press (UK).

[12] Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. 2020. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505* (2020).

[13] Nick Erickson, Lennart Purucker, Andrej Tschalzev, David Holzmüller, Prateek Mutalik Desai, David Salinas, and Frank Hutter. 2025. Tabarena: A living benchmark for machine learning on tabular data. *arXiv preprint arXiv:2506.16791* (2025).

[14] Christos Faloutsos, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. 2019. Forecasting big time series: Theory and practice. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 3209–3210.

[15] Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. 2018. Forecasting Big Time Series: Old and New. *Proc. VLDB Endow.* 11, 12 (2018), 2102–2105.

[16] Christos Faloutsos, Jan Gasthaus, Tim Januschowski, and Yuyang Wang. 2019. Classical and contemporary approaches to big time series forecasting. In *Proceedings of the 2019 international conference on management of data.* 2042–2047.

[17] Tilmann Gneiting and Adrian E Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association* 102, 477 (2007), 359–378.

[18] Lars Graf, Thomas Ortner, Stanisław Woźniak, and Angeliki Pantazi. 2026. FlowState: Sampling-Rate-Equivariant Time-Series Forecasting. In *Forty-third International Conference on Machine Learning.*

[19] Léo Grinsztajn, Klemens Flöge, Oscar Key, Felix Birkel, Philipp Jund, Brendan Roof, Benjamin Jäger, Dominik Safaric, Simone Alessi, Adrian Hayler, et al. 2025. Tabpfn-2.5: Advancing the state of the art in tabular foundation models. *arXiv preprint arXiv:2511.08667* (2025).

[20] Léo Grinsztajn, Klemens Flöge, Oscar Key, Felix Birkel, Philipp Jund, Brendan Roof, Mihir Manium, Shi Bin Hoo, Magnus Bühler, Anurag Garg, et al. 2026. TabPFN-3: Technical Report. *arXiv preprint arXiv:2605.13986* (2026).

[21] Andrew C Harvey and Andrew Harvey. 1990. Forecasting, structural time series models and the Kalman filter. (1990).

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

[23] Rob J Hyndman and George Athanasopoulos. 2018. *Forecasting: principles and practice.* OTexts.

[24] Tim Januschowski, Jan Gasthaus, Yuyang Bernie Wang, Syama Rangapuram, Caner Turkmen, Jasper Zschiegner, Lorenzo Stella, Michael Bohlke-Schneider, Danielle Maddix Robinson, Konstantinos Benidis, et al. 2024. A flexible forecasting stack. (2024).

[25] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017).

[26] Emaad Khwaja, Chris Lettieri, Gerald Woo, Eden Belouadah, Marc Cenac, Guillaume Jarry, Enguerrand Paquin, Xunyi Zhao, Viktoriya Zhukov, Othmane Abou-Amal, et al. 2026. Toto 2.0: Time series forecasting enters the scaling era. *arXiv preprint arXiv:2605.20119* (2026).

[27] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting* 37, 4 (2021), 1748–1764.

[28] Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. 2025. Moirai 2.0: When less is more for time series forecasting. *arXiv preprint arXiv:2511.11698* (2025).

[29] Junwei Ma, Valentin Thomas, Rasa Hosseinzadeh, Alex Labach, Hamidreza Kamkari, Jesse C Cresswell, Keyvan Golestan, Guangwei Yu, Anthony L Caterini, and Maksims Volkovs. 2024. Tabdpt: Scaling tabular foundation models on real data. *arXiv preprint arXiv:2410.18164* (2024).

[30] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730* (2022).

[31] Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. 2019. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437* (2019).

[32] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems* 31 (2018).

[33] Jingang Qu, David Holzmüller, Gaël Varoquaux, and Marine Le Morvan. 2025. TabICL: A tabular foundation model for in-context learning on large data. *arXiv preprint arXiv:2502.05564* (2025).

[34] Jingang Qu, David Holzmüller, Gaël Varoquaux, and Marine Le Morvan. 2026. TabICLv2: A better, faster, scalable, and open tabular foundation model. *arXiv preprint arXiv:2602.11139* (2026).

[35] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. 2018. Deep state space models for time series forecasting. *Advances in neural information processing systems* 31 (2018).

[36] Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, et al. 2023. Lag-llama: Towards foundation models for probabilistic time series forecasting. *arXiv preprint arXiv:2310.08278* (2023).

[37] Kashif Rasul, Abdul-Saboor Sheikh, Ingmar Schuster, Urs Bergmann, and Roland Vollgraf. 2020. Multivariate probabilistic time series forecasting via conditioned normalizing flows. *arXiv preprint arXiv:2002.06103* (2020).

[38] David Salinas and Nick Erickson. 2023. Tabrepo: A large scale repository of tabular model evaluations and its automl applications. *arXiv preprint arXiv:2311.02971* (2023).

[39] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting* 36, 3 (2020), 1181–1191.

[40] Oleksandr Shchur, Abdul Fatir Ansari, Caner Turkmen, Lorenzo Stella, Nick Erickson, Pablo Guerron, Michael Bohlke-Schneider, and Yuyang Wang. 2025. fev-bench: A realistic benchmark for time series forecasting. *arXiv preprint arXiv:2509.26468* (2025).

[41] Oleksandr Shchur, Ali Caner Turkmen, Nick Erickson, Huibin Shen, Alexander Shirkov, Tony Hu, and Bernie Wang. 2023. AutoGluon–TimeSeries: AutoML for probabilistic time series forecasting. In *International Conference on Automated Machine Learning*. PMLR, 9–1.

[42] Ravid Shwartz-Ziv and Amitai Armon. 2022. Tabular data: Deep learning is not all you need. *Information fusion* 81 (2022), 84–90.

[43] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. 2021. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems* 34 (2021), 24804–24816.

[44] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. 2024. Unified training of universal time series forecasting transformers. In *Forty-first International Conference on Machine Learning.*

[45] Xiyuan Zhang, Danielle C Maddix, Junming Yin, Nick Erickson, Abdul Fatir Ansari, Boran Han, Shuai Zhang, Leman Akoglu, Christos Faloutsos, Michael W Mahoney, et al. 2025. Mitra: Mixed synthetic priors for enhancing tabular foundation models. *arXiv preprint arXiv:2510.21204* (2025).

[46] Xingxuan Zhang, Gang Ren, Han Yu, Hao Yuan, Hui Wang, Jiansheng Li, Jiayun Wu, Lang Mo, Li Mao, Mingchao Hao, et al. 2025. Limix: Unleashing structured-data modeling capability for generalist intelligence. *arXiv preprint arXiv:2509.03505* (2025).