DATA POEM

Large-Scale Causal Foundation Models: Integrating Causal Discovery with Transformer Architectures for Multivariate Time Series Forecasting

Abstract/Executive Summary:

This research introduces FOUNT, a groundbreaking neural network-based architecture that addresses fundamental challenges in predicting future trends from complex data patterns. At its core, FOUNT's unified transformer architecture simultaneously handles multiple interconnected targets across diverse domains and time periods, representing a significant advancement in multi-dimensional forecasting capabilities. Learning multiple targets in the unified architectures enables it to comprehensively learn cross target synergies, halo effects. The system combines advanced pattern recognition with sophisticated cause-and-effect assessment to deliver accurate long-term predictions that maintain reliability even as underlying conditions evolve.

FOUNT, a large causal architecture which can handle multiple interconnected targets across different domains and time periods through a unified transformer architecture can capture complex cross-target halo effects, synergies and causal relationships across different inputs. This multitarget handling capability enables simultaneous prediction across diverse business metrics while maintaining coherent interdependencies between forecasting objectives.

The architecture processes data from multiple sources across different industries and product categories through a unified approach, discovering how factors truly influence each other over time using specialized detection methods and enabling knowledge sharing across different business areas.

Poem365, built upon the powerful Fount architecture, revolutionizes enterprise-wide analytics & planning by breaking down data silos and providing a holistic view of influencing factors. It integrates a vast array of internal data (sales history, promotions, customer demographics, investments, supply chain & logistics related data etc) and external data (economic indicators, social media trends etc) to identify complex causal and non-linear relationships, even accounting for carryover effects of past actions. The system separates systematic causal effects from random noise variations by observing relationship patterns across millions of instances for unprecedented modelling accuracy, while learning universal causal patterns and domain-specific characteristics simultaneously across diverse datasets for robust cross-domain generalization and adaptability.

FOUNT demonstrates exceptional performance across benchmark datasets including ETT, Electricity, and Weather, while excelling on the challenging M5 forecasting competition—one of the most demanding time series problems in the field that incorporates large-scale retail data, hierarchical aggregation of demands at various levels, complex demand patterns, intricate seasonality effects, and extended forecasting horizons that test the limits of predictive models.

Key Performance Achievements:

- 75% improvement on ETTh1 benchmark dataset
- 36% average superiority across standard benchmark datasets
- WMRSSE of 0.516 on M5 competition, outperforming the Kaggle winner (0.5204)
- POEM365 trained on over 15,000 real brand datasets, representing the largest causal model for enterprise forecasting

I. Introduction/Problem Statement:

Traditional time series forecasting methods often fail dealing with complex non-linear patterns, multivariate dependencies, and cross-domain generalization, especially when distinguishing correlation from causation. There have been many recent advancements in transformer-based large models like TimesFM, Moirai and Chronos. These models are pattern matchers, with limited causal inference abilities. While these methods are effective at detecting temporal patterns, these methods don't understand as to "why" those patterns occur. Also, these methods have a limitation in handling multiple inputs affecting the zero-shot multivariate forecasting and long-horizon predictions.

The key questions answered through this research are;

- How can large time series models move beyond pattern matching to understand the underlying causes that drive multiple KPIs together, effectively overcoming current limitations by learning generalized patterns that identify common factors, capture synergies between different metrics, and build unified frameworks that handle both shared and individual KPI behaviours at scale?
- How do transformer-based architectures automatically identify causal structures in multivariate time series to separate genuine temporal dependencies from spurious correlations?
- How can unified architectures simultaneously predict multiple targets while preserving cross-target dependencies, learning shared representations that capture both individual target dynamics and inter-target relationships, and enabling scalable inference without compromising prediction coherence or computational efficiency?
- How can a Large Causal foundational model enable highly accurate multivariate zero shot, and few shot long horizon forecasting

This research introduces FOUNT, a Large Causal foundation Model addressing synergistic relationships, multivariate data handling, and agile learning limitations through novel causal structure discovery. It also shows how an unified architecture which handles multiple target together can outperform the existing architectures. Additionally, POEM365 demonstrates enterprise applications in enterprise-wide planning. This work represents the World's first large causal foundational architecture, combining causal discovery with transformer processing, enabling superior cross-domain knowledge transfer and interpretable forecasting solutions for complex temporal phenomena.

II. Literature Review:

Time series forecasting has evolved from classical ARIMA and exponential smoothing methods to sophisticated deep learning architectures, yet systematic evaluations reveal that complex models often fail to consistently outperform simpler baselines. The transformer revolution introduced specialized architectures like Informer and Autoformer, achieving 26-29% MASE improvements in retail forecasting. Recent foundation models including TimesFM (200M parameters), Chronos, and Moirai demonstrate impressive zero-shot capabilities across diverse domains. However, causal discovery efforts from transfer entropy to NAVAR's neural networks primarily identify predictive importance rather than true causal mechanisms. Current approaches address challenges in isolation, lacking unified solutions for causal discovery, synergistic modelling, and multivariate handling. Most large models remain univariate, restricting causal learning between features, and the issue still exist after so many large model discoveries. The critical gap is the absence of comprehensive foundational architectures integrating causal mechanisms with multivariate learning for robust cross-domain generalization.

III. Methodology:

1. FOUNT a Large Causal architecture

FOUNT is a novel transformer-based architecture designed for multivariate time series forecasting with integrated causal discovery mechanisms. The system predicts future trends by understanding not just patterns in data, but how different factors cause changes in each other, distinguishing what factors cause other factors rather than simply identifying when things happen together.

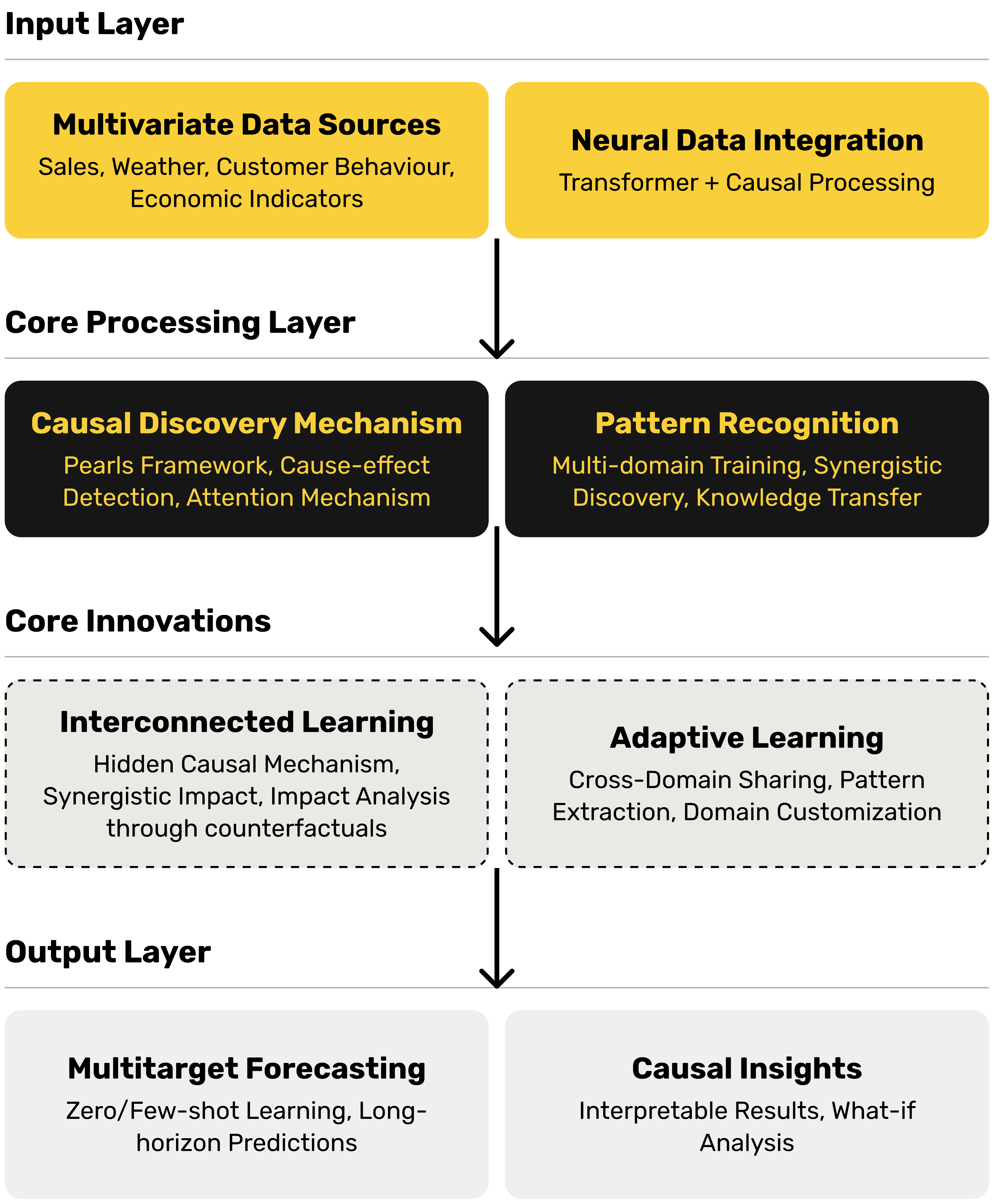

Figure.1 Foundational flow of FOUNT

FOUNT is a novel transformer-based architecture designed for multivariate time series forecasting with integrated causal discovery mechanisms. The system predicts future trends by understanding not just patterns in data, but how different factors cause changes in each other, distinguishing what factors cause other factors rather than simply identifying when things happen together. The architecture's major innovation lies in how it uses transformer's attention weights to enable causal learning for all the variables. The model first learns causality through attention weights during training, allowing natural discovery of causal patterns from data, and then estimates causality through counterfactual reasoning by processing these learned attention patterns. This approach enables the system to answer interventional questions and predict counterfactual outcomes, moving beyond traditional correlation-based forecasting. The methodology addresses key limitations in existing approaches by incorporating temporal causal relationships and enabling robust knowledge transfer across various domains. The multivariate data from diverse domains allows the architecture to understand shared learning through common patterns while preserving unique characteristics from each domain, creating a unified framework that maintains awareness of underlying causal structures governing how variables influence each other over time.

Figure 1 illustrates the foundational flow of FOUNT, organized into four distinct processing layers. The Input Layer integrates multivariate data from diverse sources including sales, weather, customer behaviour, and economic indicators, where transformers establish causal processing capabilities. The Core Processing Layer implements causal discovery mechanisms grounded in Pearl's causal framework while simultaneously extracting patterns through multi-domain training, synergistic discovery, and knowledge transfer across different data types. The Core Innovations layer focuses on interconnected learning and causal relationships through hidden causal mechanisms and synergistic impact, incorporating impact analysis through counterfactuals, while establishing adaptive learning capabilities that enable cross-domain sharing, pattern extraction, and domain customization. Finally, the Output Layer delivers the architecture's key capabilities through multitarget forecasting with zero/few-shot learning and long-horizon predictions, alongside causal insights that provide interpretable results and what-if analysis functionality.

1. Neural Data Integration and Causal Processing

FOUNT's revolutionary breakthrough in large-scale causal modelling takes data from different sources like sales numbers, weather data, and customer behaviour and combines them intelligently through massive parameter networks trained on millions of multivariate data points, employing causal embedding layers within transformer-based architecture that helps learn causal relationships purely from vast datasets while capturing the complex causal relationships that bind multiple interconnected outcomes together. The massive computational scale enables multi-layered transformer networks to implement causal attention mechanisms with specialized neural attention heads, allowing the system to learn complex causal patterns during training across hundreds of interconnected forecast targets. Through this process, the system discovers emergent causal pathways where combinations of factors create synergistic effects that exponentially amplify multi-target outcomes beyond their individual contributions, revealing non-linear interactions and emergent behaviours that arise from the complex interplay of variables. This approach leverages long previous dependencies across a large number of examples and features to learn complex temporal patterns that break down traditional forecasting silos, ensuring the network distinguishes between genuine causal relationships and spurious correlations while modelling how interventions on one target variable propagate through intricate causal webs to influence other forecast targets.

2. Finding Cause-and-Effect Relationship

FOUNT discovers which factors cause changes in other factors, not just which ones happen at the same time, by using attention weights that actively identify and learn cause-and-effect relationships rather than just statistical associations. Building on Judea Pearl's causal inference framework, our system can predict the effects of deliberate interventions on specific factors and answer counterfactual "what if" questions about alternative scenarios. The attention mechanisms achieve this by not only accounting for temporal sequences but also adapting to changes in the underlying causal structure when interventions occur. It uses special detection methods that follow logical rules about cause-and-effect, where causes must happen before effects, implementing Pearl's hierarchical approach to causal reasoning that moves from simply observing patterns to understanding interventions to predicting counterfactual outcomes. The system estimated the causal impact of each variable at its various values on the outcome considering its synergistic impact with other input variables as well as its non-linear relationship with the outcome variable.

3. Large-Scale Interconnected Pattern Recognition & Multitarget forecasting

FOUNT's pattern recognition system represents a revolutionary breakthrough in large-scale causal modelling that revolutionizes multi-target forecasting by leveraging massive parameter networks trained on millions of multivariate data points to simultaneously predict multiple interconnected outcomes while capturing the complex causal relationships that bind them together. The system's significant computational scale enables it to model how interventions on one target variable propagate through intricate causal webs to influence hundreds of other forecast targets, discovering emergent causal pathways where combinations of factors create synergistic effects that exponentially amplify outcomes beyond their individual contributions. By processing vast datasets across diverse domains, the architecture breaks down traditional forecasting silos to learn holistically from the interconnected impact of influencing factors, enabling it to distinguish systematic causal effects from random noise through observing how relationships manifest across a large number of examples, ultimately creating a unified predictive framework where multiple forecast targets are modelled as components of a single dynamic causal system.

This revolutionary large-scale causal modelling capability transforms multi-target forecasting from independent prediction tasks into integrated ecosystem modelling, where the system simultaneously forecasts multiple target variables such as demands of different regions, demand of different retailers etc. while accounting for their complex interdependencies, temporal lag effects, and feedback mechanisms that create either reinforcing or dampening cycles across the entire target landscape. The architecture's breakthrough ability to capture subtle yet powerful synergistic relationships between variables enables multi-target forecasting scenarios where predictions become significantly more accurate because the system leverages causal knowledge about how achieving one forecast target influences the probability and magnitude of achieving others, fundamentally changing how organizations approach strategic planning by providing integrated predictions that reveal the causal consequences of decisions across all critical outcome dimensions simultaneously.

Core Innovations:

- Orchestrates multiple interconnected targets across different domains and time periods through unified transformer architecture that captures complex cross-target dependencies and causal relationships with advanced multitarget handling capabilities
- Reveals emergent variable interactions and interconnected effects through million data points multivariate training that uncover hidden causal mechanisms invisible to smaller-scale approaches while managing multiple prediction targets simultaneously
- Separates systematic causal effects from random noise variations by observing relationship patterns across millions of instances for unprecedented modelling accuracy across diverse target variables
- Learns universal causal patterns and domain-specific characteristics simultaneously across diverse datasets for robust cross-domain generalization and adaptability with coherent multitarget prediction frameworks.

FOUNT understands deeper relationships between factors rather than just surface patterns. This deeper understanding makes predictions more accurate and reliable, especially when conditions change from what the system has seen before.

4. Cross-Domain Shared Learning and Adaptive Fine-Tuning

The system employs a shared learning framework that simultaneously trains across multiple domains to identify and leverage common patterns, enabling efficient knowledge transfer and robust generalization capabilities. This multi-domain approach allows the model to extract universal principles that transcend individual domains while maintaining sensitivity to domain-specific characteristics and requirements. FOUNT's hierarchical fine-tuning capability provides exceptional flexibility, operating from broad multi-domain understanding down to highly granular individual domain customization, where the system can be precisely adapted for specific temporal patterns, unique business contexts, and specialized forecasting requirements while preserving the foundational knowledge gained from shared learning across diverse operational environments.

2.POEM 365, a pretrained model built upon FOUNT

POEM365 is a revolutionary pre-trained model built upon the FOUNT architecture, leveraging FOUNT's breakthrough causal modelling capabilities to create an unprecedented enterprise forecasting solution. By harnessing the power of FOUNT's large-scale pattern recognition and causal inference mechanisms, POEM365 is developed using comprehensive multi-domain datasets with unified characteristics that span diverse business contexts and industries.

- **Trained of Large dataset :** The model represents the world's first large-scale model specifically designed for enterprise planning, trained on over 15,000 brand datasets across multiple domains, establishing an unparalleled foundation for business forecasting and strategic decision-making.
- **Fine Tuning Capabilities:** The model incorporates advanced fine-tuning capabilities that enable customization for any domain-specific granularity by considering sophisticated spatio-temporal expansion, allowing precise adaptation to unique business contexts, geographical requirements, and temporal patterns.

This multi-domain shared learning approach significantly enhances the system's capabilities, enabling the fine-tuned model to deliver exceptionally accurate predictions, understand causal relationships with precision, and maintain robust reliability when operating conditions change across different business environments, all while preserving the causal structure through the transformer's modified attention mechanisms.

Figure 2 presents POEM365 built upon the comprehensive foundation architecture of FOUNT, structured across four interconnected layers that demonstrate the end-to-end enterprise deployment framework. The Foundation Architecture establishes FOUNT as a Large Causal Model with Transformer Architecture is utilised, serving as the core foundation for all subsequent processing layers for POEM365. The Training Infrastructure layer encompasses two critical components: Dataset Integration, which incorporates over 15,000 brand datasets with multi-domain coverage, and Enterprise Data integration including sales, supply chain, customer, and investment data sources. The Advanced Processing Layer implements the model's learning capabilities through Shared Learning mechanisms that enable cross-domain transfer and pattern recognition, alongside Fine-Tuning processes for domain-specific adaptation and business context customization. Finally, the Enterprise Applications layer showcases the practical deployment scenarios including Sales Forecasting with multi-horizon and strategic planning capabilities, Supply Chain optimization for inventory and demand forecasting, and Decision Support systems for financial planning and market analysis. The architecture emphasizes three key features: being the first large-scale enterprise model, enabling continuous adaptation, and providing robust reliability across diverse business applications.

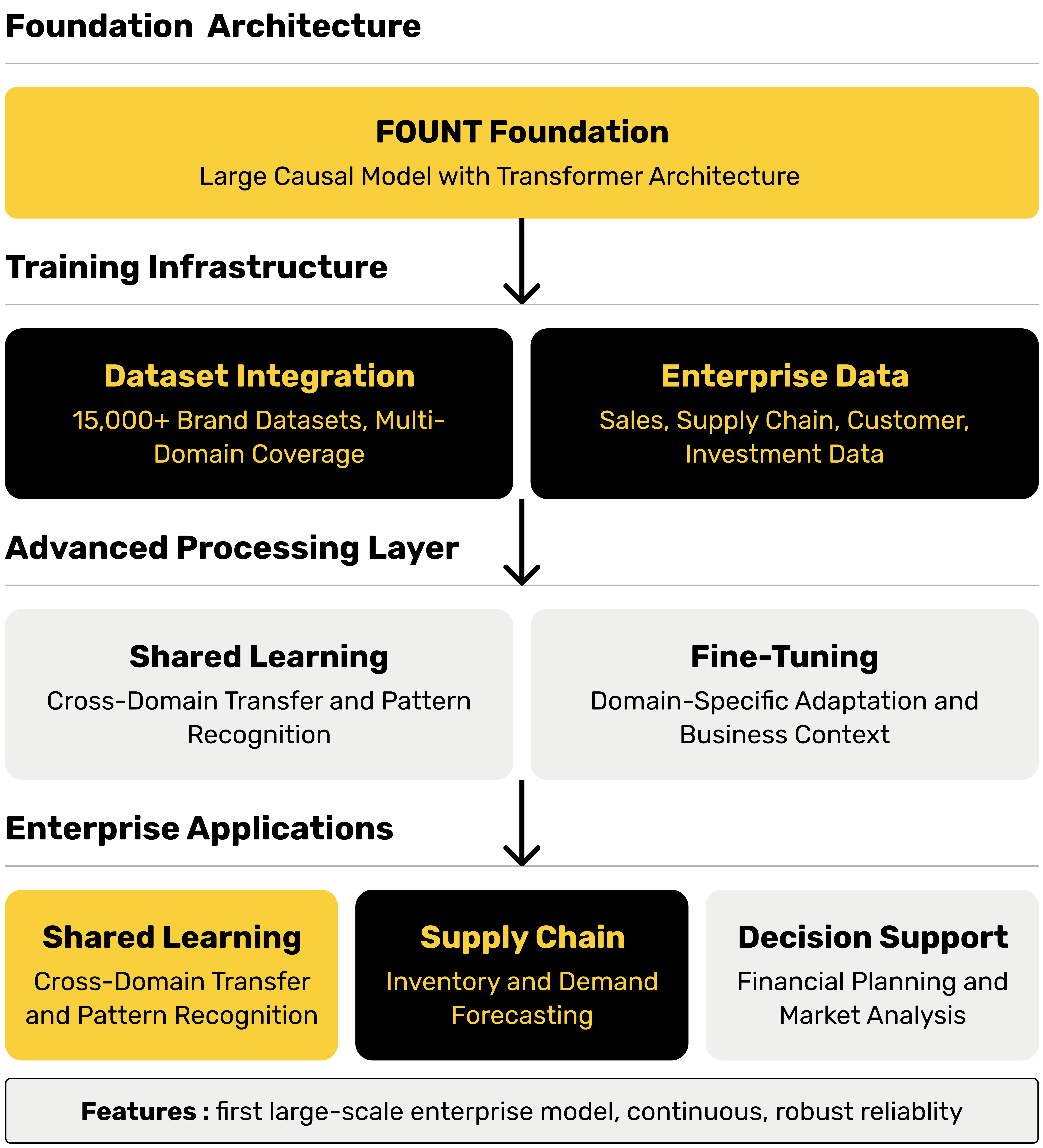

Figure.2 POEM365 built upon FOUNT

Groundbreaking Use Cases and Applications: Considering the comprehensive capabilities of POEM365, the following represent the key capabilities where it delivers transformative business solutions:

- This enables precise forecasting across multiple time horizons, seamlessly addressing short-term operational needs while supporting long-term strategic planning initiatives through leveraging incremental model training capabilities.
- POEM365 continuously learns and adapts to evolving market shifts and changing business dynamics, ensuring its predictions remain robust, resilient, and highly relevant in dynamic commercial environments.
- Its capabilities extend across crucial business functions, including advanced sales and revenue forecasting, intelligent inventory and supply chain optimization, sophisticated financial planning and budgeting support, and strategic decision-making assistance for initiatives such as new product launches and comprehensive market trend analysis, ultimately transforming demand forecasting into a core strategic capability that drives enterprise-wide competitive advantage.

FOUNT model performance is analyzed in two of the widely adopted benchmarking categories, i.e. Standard benchmark datasets and M5 forecasting competition dataset.

Standard benchmark Dataset performance comparison

Dataset	Domain	Frequency	Description
ETTh1/h2	Energy Infrastructure	Hourly	Transformer Temperature
ETTm1/m2	Energy Infrastructure	15-min	Transformer Temperature
Electricity	Energy Consumption	Hourly	321 clients consumption
Weather	Meteorological	Hourly	1,600 locations indicators

Table.1 Standard benchmark datasets information

Background and Dataset Information

- ETTh1 and ETTm1 datasets contain Electricity Transformer Temperature data from Electric Power Systems with oil temperature measurements and load data at hourly and 15-minute intervals, while Electricity dataset includes consumption data from 321 clients and Weather dataset contains meteorological measurements from weather stations
- These datasets collectively span different temporal granularities and domains including energy, infrastructure, and environmental monitoring, providing comprehensive multivariate time series forecasting challenges across various real-world applications
- The datasets provide diverse temporal patterns, multivariate relationships, and varying prediction horizons that comprehensively test forecasting model capabilities. They represent real-world scenarios with practical applications in energy management, infrastructure monitoring, and environmental forecasting.

Significance

- These benchmark datasets have become foundational evaluation standards in the time series forecasting research community due to their diverse characteristics and real-world applicability
- These datasets are widely adopted by leading technology companies and research institutions, with Google's TS-Mixer, Salesforce's MOIRAI, Intel's WPMixer, and TIMER-XL representing state-of-the-art approaches.

Performance Analysis

- FOUNT demonstrates exceptional performance superiority across multiple benchmark datasets and temporal horizons, achieving overall ~75% performance improvement for ETTh1, ~20% improvement in electricity dataset, and ~25% improvement in Weather datasets, with average superiority of 36% across all benchmark datasets.

Model	MSE ETTh1	MSE ETTm1	MSE Electricity	MSE Weather	MAE ETTh1	MAE ETTm1	MAE Electricity	MAE Weather
FOUNT	**0.111**	**0.237**	**0.122**	**0.179**	**0.127**	**0.127**	**0.218**	**0.235**
TS-Mixer (Google)	0.412	0.347	0.135	0.225	0.428	0.380	0.257	0.264
MOIRAI (Salesforce)	0.510	0.390	0.188	0.260	0.469	0.389	0.273	0.275
WPMixer (Intel)	0.379	0.336	0.173	0.218	0.409	0.371	0.252	0.267
TIMER-XL	0.409	0.350	0.155	0.238	0.430	0.378	0.241	0.294

Table.2 Detailed results comparison of FOUNT on standard datasets with existing large architecture

- Unlike competing models that focus primarily on pattern recognition, FOUNT's causal embedding approach with separate "cause" and "effect" variable spaces enables deeper understanding of underlying temporal relationships
- The architecture's integration of causal discovery mechanisms and counterfactual learning provides more robust predictions that maintain accuracy even under distribution shifts, addressing a critical limitation of traditional forecasting approaches

M5 Competition Performance

Background and Dataset Information

- The M5 dataset contains 42,840 hierarchical time series representing unit sales of Walmart products across three states spanning 1,941 days from 2011 to 2016, with data at multiple aggregation levels from individual item-store combinations to state-level aggregations

- The dataset includes comprehensive features such as calendar information, price data, promotional events, and special occasions that influence sales patterns, enabling evaluation across different business planning levels from operational to strategic decision making.

Significance

- M5 represents the largest and most comprehensive retail forecasting competition, attracting nearly 6,400 participating teams and establishing new standards for industry-relevant forecasting challenges
- M5 provides real-world complexity with hierarchical sales data from the world's largest retailer, ensuring practical relevance and business applicability
- The dataset's intermittent demand patterns (zero sales periods) create additional forecasting challenges that mirror real.

Model	WMRSSE
FOUNT	0.516
YeonJun IN_STU (Kaggle Winner)	0.520
Matthias (1st Runner-up)	0.528
TS Mixer (Google)	0.568
TFT	0.579
DeepAR (Amazon)	611

Table.3 Results comparison or FOUNT for M5 competition datasets with existing methodologies

Performance Analysis

- FOUNT demonstrates exceptional performance on the M5 competition dataset with a WMRSSE of 0.516, establishing itself as the top performer among all compared models and outperforming the original Kaggle competition winner.
- The outstanding performance represents significant practical value: FOUNT shows 15.5% better performance than DeepAR (0.611 vs 0.516), 11.0% better than TFT (0.579 vs 0.516), and 9.2% better than TS Mixer (0.568 vs 0.516).

Critically, while the Kaggle winner used 200+ models as an ensemble, FOUNT achieves superior results using only one model establishing the true significance of a Large causal model

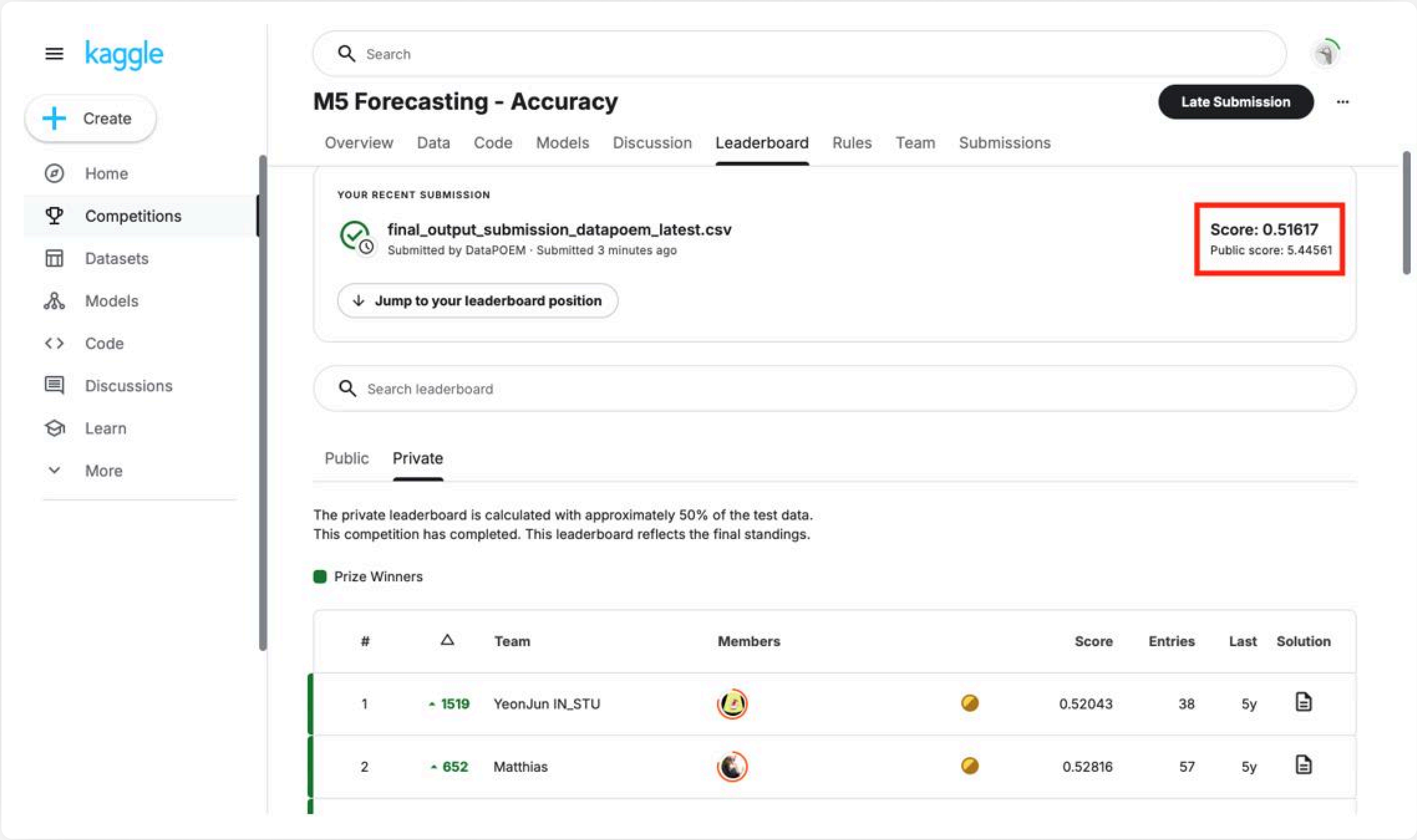

Figure.3 DataPOEM's FOUNT score after live submission to Kaggle M5 Forecasting accuracy check

Business Impact of Multi-Granularity Forecasting

FOUNT's strong WMRSSE performance across different aggregation levels enables comprehensive enterprise planning:

- **Operational Efficiency**: Fine-grained forecasts (individual item sales per store per day) support immediate operational decisions like inventory replenishment, shelf stocking, and daily staff scheduling.
- **Strategic Planning:** Higher-level forecasts (total sales by product, category, or state) enable strategic business planning including budgeting, procurement planning, media investment planning, and market trend assessment

V. Conclusion:

This research introduces FOUNT as a novel Large Causal Model architecture integrating causal discovery with transformer processing for superior time series forecasting. Key contributions include unified multi-domain forecasting, causal-learning integration, exceptional zero-shot/few-shot performance, and specialized enterprise applications.

FOUNT outperforms existing foundational models including Kaggle competition winners, demonstrating robust long-horizon forecasting under distribution shifts. POEM365 represents a transformative advancement in enterprise forecasting, enabling organizations to move beyond pattern recognition to true causal understanding, fundamentally changing how businesses approach predictive analytics and strategic planning.

The architecture's ability to distinguish genuine causal relationships from spurious correlations while maintaining exceptional predictive performance across diverse domains establishes a new paradigm for foundation models in time series forecasting, with profound implications for enterprise decision-making and strategic planning across industries.

References:

1. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.
2. Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Basic Books.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.
4. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106-11115.
5. Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems, 34, 22419-22430.
6. Das, A., Kong, W., Leach, A., Mathur, S., Sen, R., & Yu, R. (2023). Long-term forecasting with TiDE: Time-series dense encoder. Proceedings of the 40th International Conference on Machine Learning, 202, 6966-6981.
7. Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Pineda Arango, S., Kapoor, S., Zschiegner, J., Maddix, D. C., Wang, H., Mahoney, M. W., Torkkola, K., Gordon Wilson, A., Bohlke-Schneider, M., & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.
8. Woo, G., Liu, C., Sahoo, D., Kumar, A., & Hoi, S. (2022). ETSformer: Exponential smoothing transformers for time-series forecasting. arXiv preprint arXiv:2202.01381.
9. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346-1364.
10. Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent neural networks. International Journal of Forecasting, 36(3), 1181-1191.
11. Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), 1748-1764.
12. Spirtes, P., Glymour, C., & Scheines, R. (2000). Causation, prediction, and search (2nd ed.). MIT Press.