

AutoMixer for Improved Multivariate Time-Series Forecasting on Business and IT Observability Data

Santosh Palaskar^{1*}, Vijay Ekambaram^{†2}, Arindam Jati^{†2}, Neelamadhav Gantayat², Avirup Saha², Seema Nagar², Nam H. Nguyen², Pankaj Dayama², Renuka Sindhgatta², Prateeti Mohapatra², Harshit Kumar², Jayant Kalagnanam², Nandyala Hemachandra¹, Narayan Rangaraj¹

¹Indian Institute of Technology Bombay

²IBM Research

Abstract

The efficiency of business processes relies on business key performance indicators (Biz-KPIs), that can be negatively impacted by IT failures. Business and IT Observability (Biz-ITObs) data fuses both Biz-KPIs and IT event channels together as multivariate time series data. Forecasting Biz-KPIs in advance can enhance efficiency and revenue through proactive corrective measures. However, BizITObs data generally exhibit both useful and noisy inter-channel interactions between Biz-KPIs and IT events that need to be effectively decoupled. This leads to suboptimal forecasting performance when existing multivariate forecasting models are employed. To address this, we introduce AutoMixer, a time-series Foundation Model (FM) approach, grounded on the novel technique of channel-compressed pretrain and finetune workflows. AutoMixer leverages an AutoEncoder for channel-compressed pretraining and integrates it with the advanced TSMixer model for multivariate time series forecasting. This fusion greatly enhances the potency of TSMixer for accurate forecasts and also generalizes well across several downstream tasks. Through detailed experiments and dashboard analytics, we show AutoMixer's capability to consistently improve the Biz-KPI's forecasting accuracy (by 11-15%) which directly translates to actionable business insights.

1 Introduction

Business and IT Observability (henceforth called *BizITObs*) is an emerging field that focuses on seamlessly integrating business and IT data for comprehensive analysis. In general, the effectiveness of business process execution is assessed using quantifiable metrics known as Business Key Performance Indicators (Biz-KPIs) (e.g. page views, search calls, number of successful or failed transactions). These are captured using Process Monitoring (BPM) tools (van der Aalst and Weijters 2004) and are gathered periodically as data streams, offering insights into business performance. Simultaneously, Application Performance Monitoring (APM) tools like Instana¹ and Dynatrace² capture

anomalous IT events, such as hardware failures and resource over-utilization as timestamped event logs. The combination of these two data streams, which encompass IT events and Biz-KPIs, constitutes what we term as BizITObs data. BizITObs data can be structured as a multivariate time-series dataset, wherein each time-point encompasses both the presence and counts of various IT events with its corresponding business KPI values. This formulation allows for a comprehensive analysis of the interplay between IT events and business performance over time.

In this ever-changing ecosystem, where anomalous IT events can swiftly affect Biz-KPIs, there is a compelling need for accurate multivariate forecasting of these Biz-KPIs. These models must rapidly and precisely forecast the Biz-KPIs by considering the joint effect of historical IT events and past Biz-KPIs. By harnessing the power of these predictive models, organizations can accurately estimate the potential business impact of various IT failures and promptly take corrective actions before they escalate into significant issues. Also, users can prioritize the remediation actions that will have the most significant positive effect on revenue generation and operational efficiency. Thus, accurate forecasting of Biz-KPIs is of utmost importance.

One of the key distinguishing features in BizITObs time series data lies in its strong correlation across channels where each channel signifies an IT event or business KPI. The channels can exhibit different correlation patterns such as IT events with business KPIs (e.g., disk failure affecting payment gateway calls), IT-event correlations (e.g., disk failure impacting database), and business KPI correlations (e.g., throughput linked to cycle time and revenue). To achieve precise forecasting in these scenarios, models that effectively target cross-channel correlations are of utmost importance. In this paper, we focus on the problem of accurate cross-channel aware multivariate forecasting of business KPIs given historical data on IT events and Biz-KPIs.

Related work: This is a classic forecasting problem with a well-established line of research dedicated to solving it. Recent advancements in Transformer (Vaswani et al. 2017) based forecasting models have surpassed the traditional statistical and ML models (Salinas et al. 2020) (Oreshkin et al. 2019) with very high state-of-the-art accuracy. Specifically, Informer (Zhou et al. 2021), Autoformer (Wu et al. 2021),

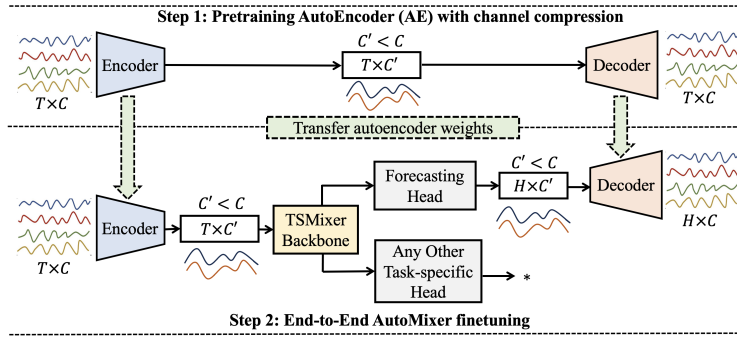
*The author was in IBM Research while the work was done.

[†]These authors contributed equally.

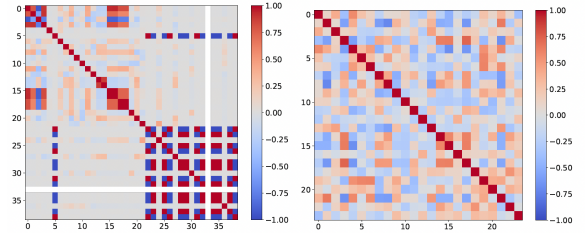
Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹instana.com

²dynatrace.com



(a) The AutoMixer framework. T : input context/history length, H : forecast horizon, C : number of input channels, C' : number of compressed channels.



(a) Raw channels (b) Channels in AE embedding

(b) Inter-channel correlation for Application data

Figure 1: AutoMixer overview

and Fedformer (Zhou et al. 2022) introduced channel mixing, wherein channels from the same timepoint are combined into an embedding and processed using self-attention networks for cross-channel modeling. These methods propose various improvements to the self-attention modules and were considered state-of-the-art in multivariate forecasting until DLinear (Zeng et al. 2023) and PatchTST (Nie et al. 2023) presented the power of the channel-independence approach. In channel-independence, authors treat multivariate modeling as a global univariate approach with shared model weights across channels, demonstrating superior forecasting accuracy compared to existing channel-mixing techniques. In contrast, CrossFormer (Zhang and Yan 2023) proposed an alternative native cross-channel modeling approach. Further advancements were made with TSMixer (Vijay et al. 2023), which introduced the cross-channel reconciliation head for explicit cross-channel modeling, leading to even better forecasting accuracy. To the best of our knowledge, TSMixer is the latest state-of-the-art for multivariate forecasting with improved cross-channel correlation support.

However, when these models were used on BizITObs data, there was evident potential to enhance cross-channel correlations for better accuracy. Toward this, we propose **AutoMixer**, a novel time-series foundation modeling approach for BizITObs data that harnesses the combined capabilities of AutoEncoder and TSMixer for improved multivariate time-series forecasting. AutoMixer follows a foundation model workflow, wherein the first step involves pretraining an AutoEncoder (AE) independent of the downstream tasks, to compactly project the input data into a compressed channel space and enable full reconstruction. Then, for every downstream task (e.g. Biz-KPI forecasting), we seamlessly interleave the pretrained AutoEncoder components at different points in the TSMixer workflow as depicted in Figure 1a and do end-to-end finetuning. The intuition behind AutoMixer is elegantly simple. Rather than feeding raw channels directly to TSMixer, we harness the power of *compressed channels* (via AutoEncoders (Murphy 2022)) where unwanted noisy channel interactions are reduced. Hence, the channel correlation task is simplified and the potency of TSMixer is significantly amplified in the compressed chan-

nel space, resulting in even more accurate forecasts.

This comprehensive two-step training process (i.e. pre-train and finetune) allows the framework to unlock the full potential of both AutoEncoder and TSMixer. It also enables the common pretrained models to be easily repurposed for various downstream tasks (such as forecasting, classification, regression, *etc.*) just by changing the final heads, and harnessing the power of foundation model workflows. Note that, in time-series, varying data dynamics mandate separate pretrained models for each dataset, capable of generalizing across multiple downstream tasks within the same dataset (Nie et al. 2023; Vijay et al. 2023). This contrasts with NLP and image domains, where pretrained models are constructed across several datasets.

Summary of Contributions

- We introduce **AutoMixer**, a **time-series foundation model approach** particularly targeted for BizITObs data for improved Biz-KPI multivariate forecasting via the proposed novel technique of channel-compressed pre-train and finetune workflows.
- AutoEncoders are traditionally used for compressing hidden features. In contrast, we employ AutoEncoders to compress the channels instead of hidden features. This simplifies the channel correlation task which in turn amplifies the potency of TSMixer as it works in the compressed less noisy channel space, resulting in even more accurate and insightful forecasts.
- We publicly release 4 new BizITObs datasets and provide benchmarks for various state-of-the-art (SOTA) forecasting models. Our comprehensive experiments emphasize AutoMixer’s effectiveness, surpassing SOTAs in the Biz-KPI forecasting task with a substantial 11-15% MSE reduction and 1-2X enhancements in PCC.
- We also demonstrate the effectiveness and ease of porting these channel-compressed pretrained models across several BizITObs downstream tasks, harnessing the power of foundation model workflows.
- Through an actionable business insight dashboard and real-time examples, we demonstrate how these forecast-

ing improvements practically empower business and IT analytics for effective decision-making.

2 AutoMixer: A Foundation Model Approach for BizITObs

2.1 Problem Formulation

We denote a multivariate time series as $\mathbf{x} \in \mathbb{R}^{T \times C}$, where T is the length of the time series and C is the number of channels (or variates). Multivariate time series data in the BizITObs domain blend channels from both business KPIs and IT events together. The multivariate forecasting problem can be formulated as a sequence-to-sequence prediction problem: given a history sequence $\mathbf{x} \in \mathbb{R}^{T \times C}$ we have to predict the subsequent future sequence $\mathbf{y} \in \mathbb{R}^{H \times C}$ with a trainable model $f(\cdot)$. The model's forecast is denoted as $\hat{\mathbf{y}} = f(\mathbf{x}) \in \mathbb{R}^{H \times C}$. H is the forecast horizon.

2.2 TSMixer Background

Given AutoMixer is built on top of TSMixer, we provide an overview of TSMixer here. TSMixer employs MLP-Mixer blocks and segments input time series into fixed windows, followed by gated MLP transformations and permutations for enhanced accuracy. TSMixer adopts a channel-independent strategy, modeling all channels globally in a univariate manner, with shared model weights across channels to implicitly account for channel correlations. To explicitly capture channel correlations in a pure multivariate fashion, the TSMixer-CC variant is further introduced, which blends a channel-independent backbone with explicit channel-reconciliation heads. This approach achieves the latest state-of-the-art accuracy. Next, we explain how its potency can be further improved using AutoMixer's channel-compressed pretrain and finetune workflows.

2.3 Channel Compression as a Pretraining Task

Foundation models and self-supervised representation learning have been found to be effective in natural language and vision domains. Pretraining through a self-supervised objective can learn a generic representation of the data, and the pretrained model weights can be subsequently finetuned for various downstream tasks like forecasting or classification.

BizITObs data in general exhibit a strong presence of intricate and potentially lagged correlation patterns within and across business KPIs and IT events. Some of these correlations hold significant importance for effective modeling, while others merely contribute noise and hinder the modeling process. APM tools like Instana can potentially capture hundreds of IT events, some of which may be rare. It is difficult to determine a priori which IT events are responsible for influencing the business KPIs. The IT events which do not affect the business KPIs only add noise to the data.

To tackle this, we propose to project the raw channels to a compressed channel space where unimportant and irrelevant channel correlations are pruned away and only the important correlations are compactly retained. We pose the channel compression as a pretraining task, and achieve it with an AutoEncoder (AE) model. Specifically, the AutoEncoder

$g(\cdot)$ takes a time series segment $\mathbf{x} \in \mathbb{R}^{T \times C}$ and the encoder part compresses *only* the channel dimension i.e., $\mathbf{z} = g(\mathbf{x}) \in \mathbb{R}^{T \times C'}$, where $C' < C$ is the number of channels in the compressed space. The decoder of the AE tries to reconstruct the original \mathbf{x} from the compressed representation \mathbf{z} . The decoder's output $\hat{\mathbf{x}} = h(\mathbf{z}) \in \mathbb{R}^{T \times C}$ is compared with the original input \mathbf{x} , and the mean squared error (MSE) objective $\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2$ is employed to train the AE. Intuitively, if \mathbf{x} can be reconstructed with low MSE, then, the original input data possibly has latent correlations between the channels which was essentially leveraged by the AE for a successful channel compression task. Figure 1b shows the inter-channel correlations for raw channels and channels in the embedding space on Application dataset (details in Section 3.1). We can see that the raw channels have strong absolute correlations between many channels, which is compressed and evenly distributed by the AE. As we observe, these compressed channels demonstrate diminished inter-channel correlation, signifying that important and meaningful correlations are already amalgamated within separate channels in the compressed space. This streamlined structure simplifies the channel correlation task and the potency of TSMixer is significantly amplified in the compressed channel space, resulting in even more accurate and insightful forecasts.

2.4 End-to-End Forecasting as a Finetuning Task

The pretrained AE is used in the final stage of training the AutoMixer end-to-end for a downstream Biz-KPI forecasting task. The input \mathbf{x} is first compressed into \mathbf{z} through the encoder part of the AE. The backbone of the TSMixer takes the compressed representation $\mathbf{z} \in \mathbb{R}^{T \times C'}$ and converts it into $\mathbf{z}' \in \mathbb{R}^{C' \times n \times l}$ by a series of transformations. Here, n denotes the number of patches (TSMixer performs the patching internally), and l is the length of every patch. Hence, $T = n \times l$. The output of the TSMixer backbone can be fed to *task-specific heads*. For forecasting, the head converts \mathbf{z}' into $\mathbf{y}' \in \mathbb{R}^{H \times C'}$ which is fed into the pretrained decoder of the AE to produce the final output, $\hat{\mathbf{y}} \in \mathbb{R}^{H \times C}$. The entire model is trained (hence, the AE weights are finetuned and the TSMixer weights are trained from scratch) end-to-end with a mean squared objective, $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$. For any other downstream task like classification or regression, the decoder part of the AE might be removed, and the output of the TSMixer backbone \mathbf{z}' can be directly fed to the task-specific head to generate the prediction (see Figure 1a).

2.5 Discussion on Design Choices

In this section, we delve into the intricacies of the AutoMixer architectures, shedding light on the multitude of design choices that have been considered and thoughtfully incorporated. Each component of the AutoMixer has been crafted with a specific purpose in mind, guided by a coherent set of reasons that harmoniously converge to yield an efficient and effective prediction system for BizITObs data.

Choice of AutoEncoders Autoencoders can be constructed using diverse architectures like CNNs and RNNs. Our forecasting pipeline requires sequence-invariant Autoencoders, allowing flexible handling of different input

Dataset	#Obs	#Channels	#Biz-KPIs	#IT Events
Service	8835	107	72	35
Application	8834	39	4	35
L2C	31968	9	7	2
SO	20475	45	13	32

Table 1: Dataset statistics

and output sequence lengths. Since CNNs lack sequence-invariance, we opted for RNN-based Autoencoders. We will explore the contrast between GRU and LSTM in subsequent experiments.

Choice of Backbone models Our approach is adaptable to various deep learning prediction backbones. While options like Informer, Fedformer, Patchtst, and TSMixer were available, TSMixer stood out with its high state-of-the-art accuracy complimented with 2-3X speed and memory improvement. Thus, we chose TSMixer as our preferred backbone among the alternatives.

Choice of training workflows In training workflows, we face two distinct choices. The first approach trains a single AutoMixer end-to-end by initializing the AE weights at random for channel compression. Alternatively, we can also follow a two-step process (as done in foundation models (Bommasani et al. 2021)): commencing with AE weight learning in a pretraining phase and then transferring this knowledge to enhance the AutoMixer through finetuning. Based on empirical analysis (Section 3), we choose foundation modeling over direct end-to-end training due to its significant performance improvements.

3 Experiments and Results

We comprehensively evaluate AutoMixer’s accuracy against established benchmarks and perform ablation studies to emphasize the importance of different model configurations.

3.1 Data

For experimental purposes, we used four datasets that are synthesized based on various realistic business process workflows³.

Service This dataset relates to the “Stan’s Robot Shop” (Instana 2023) app on the cloud, overseen by Instana. It models a user’s e-commerce journey, from site access to shipping, using a load generator. Intermittent fault injection induces various IT events. It offers business KPIs for services (e.g. payment, catalog) and Instana-tracked IT events. It is sampled every 10 seconds due to high traffic and events.

Application This dataset is similar to the above dataset but contains KPIs at the level of the whole application.

L2C This dataset is related to the Lead-to-Cash (L2C) process ranging from customers’ intent or interest to buy a product, to a company’s realization of revenue based on product sales (SAP 2023a). It covers steps like order updates, approvals, and acceptance. It is sampled every 5 minutes due to moderate IT event frequency.

SO This dataset pertains to the Sales Order (SO) process, encompassing order creation, outbound deliveries, goods issues, and invoicing (SAP 2023b). Generated in a controlled setting based on the existing sales order process with additional order validation, and order correction, the data is sampled every 1 hour due to infrequent IT events. It also includes environmental metrics from Environmental Intelligence Suite (EIS 2023), designed to potentially impact outbound shipping in adverse weather. These metrics are also treated as auxiliary IT events in our experiments.

Table 1 provides the total number of observations, time series variables, KPIs, and IT events for each dataset.

3.2 Experimental Setting

Model Variants We present four AutoMixer variants: (i) AutoMixer LSTM, (ii) AutoMixer LSTM CC, (iii) AutoMixer GRU, and (iv) AutoMixer GRU CC. These options arise from different underlying model (TSMixer, TSMixer-CC) and AutoEncoder (LSTM, GRU) selections.

Forecasting Benchmarks We compare the AutoMixer variants with various established multivariate time series forecasting benchmarks, primarily grouped as follows: (i) **Mixer Models:** TSMixer and TSMixer-CC (Latest SOTAs, refer to Section 2.2), (ii) **Other Deep Learning Models:** PatchTST (Nai et al. 2022), DLinear (Zeng et al. 2023), FEDFormer (Zhou et al. 2022), Recurrent-GRU (Cho et al. 2014), DecoderMLP and TFT (Lim et al. 2021) from the PyTorch forecasting library (PyTorch 2023), AGCRN (Bai et al. 2020) (iii) **Other ML and Statistical Models:** XGBoost Regressor (Chen and Guestrin 2016) and VAR (Lütkepohl 2005).

Model Configuration First, we trained the autoencoder with a context/history length, $sl = T = 24$, and a forecast length, $fl = H = 24$. Subsequently, we utilized these weights in the AutoMixer for the end-to-end training for forecasting. Throughout all the datasets and experiments, the following model configurations from TSMixer paper is used: patch length (pl) = 8, batch size (b) = 8, feature scalar (fs) = 2, hidden feature size (hf) = $fs \times pl = 16$, and expansion feature size (ef) = $fs \times hf = 32$. For SO and Application data, we used the number of Mixer layers (nl) = 8 and dropout (do) = 0.4, and for the Service and L2C data, $nl = 3$ and $do = 0.3$. The training, validation, and test sets are chronologically split in a ratio of 0.6 : 0.2 : 0.2 using temporal cross-validation (Jati et al. 2023). For all training, we incorporated early stopping based on the validation performance. For each dataset, we choose the optimal channel compression ratio (cr) and model variant by assessing their effects on the validation set.

Metrics We use the mean squared error (MSE) to measure the model performance and also employ the Pearson Correlation Coefficient (PCC) to evaluate the correlation between predicted and actual value. PCC is particularly crucial for BizITObs data, as we expect a good model to not only capture mean predictions but also capture accurate trends and patterns. We independently normalize channels by subtracting means and dividing by standard deviations. Both loss

³<https://github.com/BizITObs/BizITObservabilityData>

Category	Datasets	Application		SO		Service		L2C	
	Models	MSE ↓	PCC ↑	MSE ↓	PCC ↑	MSE ↓	PCC ↑	MSE ↓	PCC ↑
AutoMixer variants (ours)	AutoMixer GRU	0.1535	0.9086	0.5627	<u>0.2198</u>	0.148	0.8684	0.0774	0.9719
	AutoMixer GRU CC	0.0937	0.9354	0.5697	0.213	0.1494	0.862	0.0772	0.9719
	AutoMixer LSTM	0.0788	0.9407	0.5694	0.2131	0.1593	0.8569	0.0771	0.9719
	AutoMixer LSTM CC	0.1012	0.9324	0.5657	0.2194	0.1602	0.8576	0.0772	0.9719
SOTAs	TSMixer	0.2194	0.8352	0.6042	0.0096	0.2094	0.8576	0.0723	0.9734
	TSMixer CC	0.1313	0.9148	<u>0.5636</u>	0.1641	0.1654	0.8542	0.0716	0.9737
	PatchTST	0.3484	0.5735	<u>0.8572</u>	0.0647	0.2392	0.7873	0.1086	0.9574
	FEDFormer	0.4643	0.4855	0.8971	0.0570	0.2845	0.7524	0.1342	0.9475
	DLinear	0.3256	0.5983	0.9089	0.0050	0.2536	0.7875	0.1046	0.9578
Baselines	VAR	NA		0.6505	0.1603	NA		1.0100	0.1131
	XGBoost	2.7192	0.3923	0.8331	0.1804	0.2955	0.2065	0.1545	0.0574
	AGCRN	0.2653	0.6458	0.5839	0.3159	0.1515	0.2491	0.0671	0.1851
	DecoderMLP	0.5591	-0.4585	0.6943	0.0015	0.5892	-0.0070	0.1233	0.0585
	Recurrent GRU	0.4873	0.0763	0.7118	0.1203	0.3326	0.1128	0.092	0.0645
	TFT	0.3843	0.1698	0.7283	0.0901	0.3435	0.1295	<u>0.0694</u>	0.2588
Avg Imp. over best SOTA		MSE: 11%, PCC: 1X							
Avg Imp. over best Baseline		MSE: 15%, PCC: 2X							

Table 2: Comparing AutoMixer with benchmarks for Biz-KPIs forecasting task. Comparison is across AutoMixer best variant with the benchmark’s best variant in SOTA and Baselines categories. Average Improvement (Imp.) across all datasets *w.r.t* MSE is given in % and improvement *w.r.t* PCC is in X’s (number of times). ↓ indicates lower is better. ↑ indicates higher is better. The best and second-best results are in bold and underlined respectively. VAR models failed in few datasets due to constant channel values.

Data	PT	NPT
Application	0.0937	0.0956
SO	0.5697	0.5551
Service	0.1494	0.1618
L2C	0.0772	0.079
Avg % Imp	2.3%	

Table 3: MSE Improvements with pretraining (PT) vs without pretraining (NPT).

<i>cr</i>	Application	SO	Service	L2C
20%	0.0935	0.5697	0.1677	-
40%	0.1074	0.5715	0.1463	0.0819
60%	0.0937	0.5708	0.1494	0.0772
80%	0.0731	0.5598	0.1485	-

Table 4: Comparing across channel compression ratios (*cr*). “-” indicates compression infeasibility due to fewer channels

and reported metrics are then computed on this normalized data, eliminating bias from the original data scale.

3.3 Biz-KPI Forecasting

In this section, we delve into various empirical studies highlighting the potential of AutoMixer for Biz-KPI forecasting. All results in this section report the MSE or PCC averaged across all the normalized Biz-KPI channels at a dataset level.

Accuracy Improvements We compare AutoMixer to existing benchmarks for multivariate forecasting of Biz-KPIs ($fl = 24$), using past IT events and Biz-KPIs ($sl = 24$). The results presented in Table 2 reveal AutoMixer’s superior performance over Baselines, displaying a 15% average

Datasets	Model	Biz-KPI Forecast (MSE ↓)	IT Event Forecast (MSE ↓)	IT Event Clf (Acc ↑)
Application	AutoMixer	0.0937	0.3035	77.82%
	TSMixer CC	0.1313	0.32	33.92%
Service	AutoMixer	0.1494	0.3255	83.21%
	TSMixer CC	0.1654	0.3294	44.75%
Avg improvement		19.2%	3.2%	2X

Table 5: Comparing downstream tasks, KPI forecast, IT event forecast and IT event classification (Clf). Avg improvements are *w.r.t* AutoMixer GRU CC over TSMixer CC on MSE and Accuracy (Acc).

reduction in MSE and a twofold increase in PCC. When compared to TSMixer and other SOTAs, AutoMixer demonstrates an average MSE improvement of 11%, though there are no significant PCC improvements. This can be attributed to TSMixer’s strong trend-capturing abilities, which limit the potential for notable PCC improvements, despite notable enhancements in MSE. These insights highlight the potency of AutoMixer to capture both mean and trends more accurately in most of the considered datasets, a challenge in existing benchmarks. Among the considered AutoMixer variants, each variant performed well in different scenarios, and hence in practice, we would need to select the best variant using the validation metric for customer deployments.

Ablation Study In this section, we study how a couple of crucial design settings, specifically pretraining and channel compression ratios, influence the accuracy of AutoMixer, taking the AutoMixer GRU CC variant as an example.

Effect of pretraining: Here, we analyze the impact of

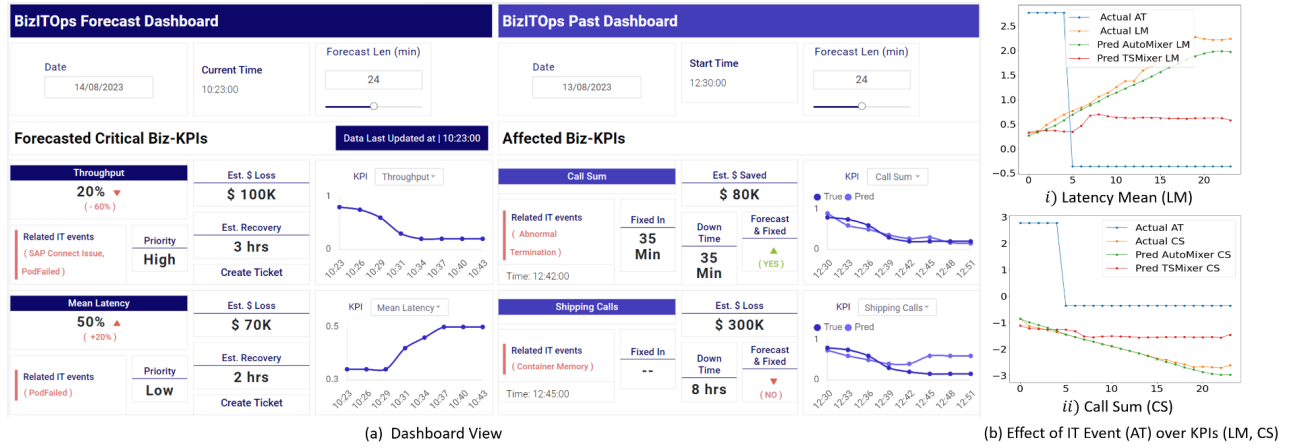


Figure 2: Actionable Business Insight Dashboard and real-time prediction examples

pretraining on AutoMixer for Biz-KPI forecasting. Table 3 highlights that utilizing pretraining (PT) of AE reduces the MSE by 2.3% as compared to direct AutoMixer training with randomly initialized AE weights (NPT). This reiterates our initial hypothesis that employing the foundation modeling workflow can provide better generalization and improved forecasts. Moreover, these pretrained models can easily cater to multiple downstream tasks as explained in Section 3.4

Effect of Channel Compression: Channel compression ratio is an important hyperparameter for improving AutoMixer’s accuracy, and it varies across datasets as shown in the Table 4. Hence based on the validation metric, we choose the right compression ratio for every dataset and model variant. This validation-based evaluation is important to get the right trade-off between channel-compression and information loss for a given model and data. However, in general, the best compression ratios mostly lie in the range between 40-80% leading to improved model performance. This compression helps us eliminate channels that contribute noise to the model.

3.4 Applicability to Multiple Downstream Tasks

In this section, we show the applicability of our pretrained model across 3 downstream tasks *viz.* (i) Biz-KPI forecasting: Temporal forecast of Biz-KPIs for the next 24 points (same as in Section 3.3), (ii) IT event forecasting: Temporal forecast of IT events for the next 24 points (iii) IT event classification: Multi-label classification of various IT events to occur in the next 24 points. As shown in Table 5, the channel compressed encoder from the pretrained model when used in conjunction with TSMixer consistently boosts the performance of TSMixer across all the considered downstream tasks. Thus, the one-time pretraining step facilitates improved generalization across multiple downstream tasks, highlighting the potential of time-series foundation models.

4 Actionable Business Insight Dashboard

In this section, we present the prototype dashboard view that conceptualizes how AutoMixer’s forecasting benefits

enhance actionable business insights and decision-making. Figure 2a showcases the BizITObs Forecast and Past Dashboards. In the Forecast Dashboard (left view), we display Top-k biz-KPIs with crucial forecasts requiring preemptive attention to avert downtime or revenue loss. The system also estimates potential downtime and revenue loss based on similar historical ticket data, providing these estimates in the UI for informed decisions. It also assigns a priority score for each insight for effective ranking during resource constraints. In the Past Dashboard (right view), we exhibit the top-k affected Biz-KPIs within a defined timeframe and indicate whether AutoMixer has forecasted it correctly or missed it. This dashboard also highlights the impact of a forecast hit or miss by visually representing the downtime prevented/incurred and revenue saved/lost estimated from similar historical past tickets. Figure 2b (i, ii) illustrates an actual forecasting instance involving 2 Biz-KPIs [Calls Sum (CS), Latency Mean (LM)] in response to an IT event [Abnormal Termination (AT)]. Notably, AutoMixer demonstrates superior accuracy compared to TSMixer in both scenarios, offering valuable preemptive business insights and informed data-driven decisions when presented via actionable business insight dashboards.

5 Conclusion

In summary, we introduce AutoMixer, a novel time-series foundational model designed specifically for BizITObs data. Through comprehensive empirical analysis, we demonstrate the efficacy of our proposed channel compression based pre-train and finetune workflows in capturing the combined impact of historical IT events and past Biz-KPIs. This technique leads to improved Biz-KPI forecasting accuracy (by 11-15%), and also generalizes well to multiple downstream tasks. We then showcase how these forecasting insights translate into valuable business insights through actionable dashboards, facilitating swift and informed decision-making. Our next steps involve deploying AutoMixer in select customer ecosystems, conducting real-time user studies, and analyzing deployment outcomes.

References

- Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33: 17804–17815.
- Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.
- Cho, K.; Van Merriënboer, B.; Bahdanau, D.; and Bengio, Y. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.
- EIS. 2023. Environmental Intelligence Suite. <https://www.ibm.com/in-en/products/environmental-intelligence-suite>. Accessed: 2023-07-15.
- Instana. 2023. Stan’s Robot Shop. <https://github.com/instana/robot-shop>. Accessed: 2023-07-15.
- Jati, A.; Ekambaram, V.; Pal, S.; Quanz, B.; Gifford, W. M.; Harsha, P.; Siegel, S.; Mukherjee, S.; and Narayanaswami, C. 2023. Hierarchical Proxy Modeling for Improved HPO in Time Series Forecasting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD ’23, 891–900. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701030.
- Lim, B.; Arik, S. Ö.; Loeff, N.; and Pfister, T. 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4): 1748–1764.
- Lütkepohl, H. 2005. *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Murphy, K. P. 2022. *Probabilistic machine learning: an introduction*. MIT press.
- Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2023. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *The Eleventh International Conference on Learning Representations*.
- Oreshkin, B. N.; Carpo, D.; Chapados, N.; and Bengio, Y. 2019. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.
- PyTorch. 2023. PyTorch Forecasting. <https://pytorch-forecasting.readthedocs.io/>. Accessed: 2023-07-15.
- Salinas, D.; Flunkert, V.; Gasthaus, J.; and Januschowski, T. 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3): 1181–1191.
- SAP. 2023a. Lead-to-Cash process. https://help.sap.com/docs/SAP_CX/fe1b9a180649472eb58843ec01732ed9/03559e96735249e7bd37929f1f387b85.html. Accessed: 2023-07-15.
- SAP. 2023b. Sales Order process. https://help.sap.com/docs/SAP_BUSINESS_BYDESIGN/2754875d2d2a403f95e58a41a9c7d6de/2cee28f1722d101488a1b9611201520c.html?version=2202. Accessed: 2023-07-15.
- van der Aalst, W. M. P.; and Weijters, A. J. M. M. 2004. Process mining: a research agenda. *Computers in Industry*, 53(3): 231–244.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Vijay, E.; Jati, A.; Nguyen, N.; Sinthong, G.; and Kalagnanam, J. 2023. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *Advances in Neural Information Processing Systems*.
- Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11121–11128.
- Zhang, Y.; and Yan, J. 2023. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In *The Eleventh International Conference on Learning Representations*.
- Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence*, volume 35, 11106–11115.
- Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning*.