Journal of Forecasting

WILEY

RESEARCH ARTICLE OPEN ACCESS

Enhancing Demand Forecasting in Retail: A Comprehensive Analysis of Sales Promotional Effects on the Entire Demand Life Cycle

Harsha Chamara Hewage[1,2] | H. Niles Perera[2,3] | Kasun Bandara[4]

[1]Data Lab for Social Good, Cardiff Business School, Cardiff University, Cardiff, UK | [2]Center for Supply Chain, Operations and Logistics Optimization, University of Moratuwa, Moratuwa, Sri Lanka | [3]Department of Transport Management and Logistics Engineering, Faculty of Engineering, University of Moratuwa, Moratuwa, Sri Lanka | [4]School of Computing and Information Systems, Melbourne Centre for Data Science, The University of Melbourne, Melbourne, Victoria, Australia

Correspondence: Harsha Chamara Hewage (halgamuwehewagehr@cardiff.ac.uk)

Received: 21 October 2023 | **Revised:** 11 November 2024 | **Accepted:** 24 September 2025

Funding: This research was supported by Grant ID SRC/LT/2020/20 provided by the Senate Research Committee of the University of Moratuwa, Sri Lanka.

Keywords: machine learning | postpromotional effect | promotions | retail supply chain | sales forecasting

ABSTRACT

Sales promotions pose challenges to retail operations by causing sudden fluctuations in demand, not only during the promotional period but also across the entire sales promotional life cycle. Previous research has predominantly focused on promotional and nonpromotional periods, often overlooking the postpromotional phase, where demand decreases due to consumer stockpiling during promotions. To address this research gap, we investigate both traditional statistical forecasting methods and contemporary approaches, such as global models, implemented using gradient boosting and deep learning techniques. We assess their performance throughout the entire demand life cycle. We employ the base-lift approach as our benchmark model, commonly used in the retail sector. Our study results confirm that machine learning methods effectively manage demand volatility induced by retail promotions while enhancing forecast accuracy across the demand life cycle. The base-lift model performs comparably to alternative machine learning methods, albeit with the additional effort required for data cleansing. Our proposed forecasting framework possesses the capability to automate the retail forecasting process in the presence of sales promotions, facilitating efficient retail planning. Thus, this research introduces a novel demand forecasting framework that considers the complete demand life cycle for generating forecasts, and we rigorously evaluate it using real-world data.

1 | Introduction

Retail operations encounter a variety of difficulties and complexities due to many factors such as shifting customer expectations, promotional activities, partner activities, and shorter lead times (Hewage and Perera 2022b; Ma et al. 2016; Ma and Fildes 2021). Of which, retail sales promotions make retail sales forecasting a difficult task (Hewage et al. 2021). Generally, retail promotions often increase product sales during promotional periods (Fildes et al. 2019). This increase in sales for the promoted product may occur at the expense of sales of other products or sales of the same product during other time periods (Blattberg and Briesch 2012). Following the promotional period, sales may fall below normal levels and then recover, creating a postpromotional dip (Hewage et al. 2021). Thus, promotions cause demand changes not only during the promotional time but also throughout the demand life cycle (Macé and Neslin 2004). Figure 1 depicts the demand variations during a retail sales promotion.

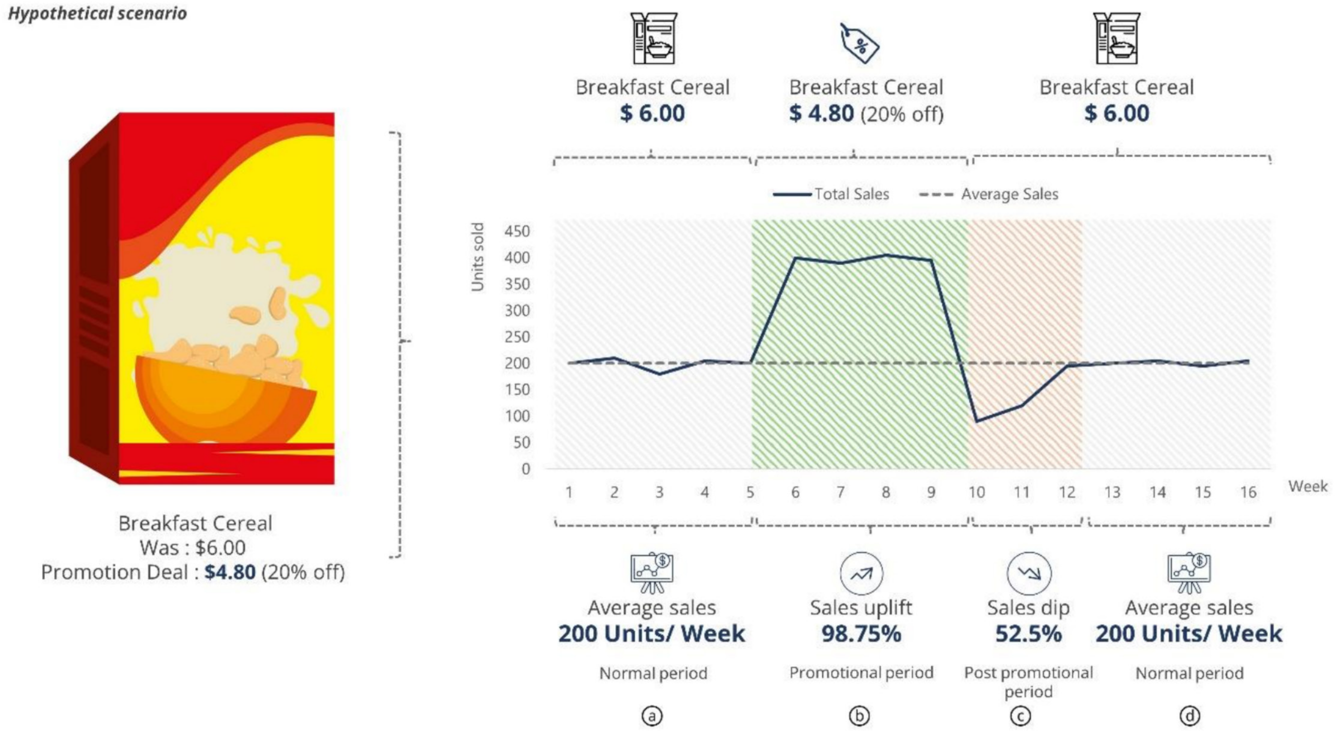

FIGURE 1 | Variations in demand in retail sales promotions: (a) normal sales represent the number of sales without any sales promotions; (b) a sales uplift can be found during promotional periods; (c) the postpromotional period has lower sales figures than the normal period; and (d) sales figures gradually recover to a normal level.

This necessitates the development of more comprehensive approaches to deal with these issues in retail forecasting (Hewage et al. 2021) such as regression-based models (e.g., Cooper et al. 1999; Divakar et al. 2005; Leeflang et al. 2005) and machine learning (ML) models (Spiliotis et al. 2020). Though these methods can incorporate causal features like promotional information (Perera et al. 2019), retailers continue to use simple methods like exponential smoothing (ETS) to forecast sales, often with judgmental adjustments to account for promotional effects (Mou et al. 2018). These judgmental approaches are manpower-intensive because a typical retail store carries thousands of products across many store locations (Fildes et al. 2019). Therefore, ML methods are a viable alternative that enables retailers to automate the forecasting process (Ali et al. 2009).

Importantly, with enhancements to the current technology, utilizing ML methods does not pose a technical challenge for retailers (Fildes et al. 2019). Nevertheless, past literature highlights only a few studies focused on stock keeping unit (SKU) level sales forecasting using ML methods and all the studies have focused only on the promotional and nonpromotional periods, regardless of the postpromotional period (e.g., Abolghasemi, Beh, et al. 2020; Ali and Gürlek 2020; Ali et al. 2009; Huber and Stuckenschmidt 2020; Ma et al. 2016; Ma and Fildes 2017). Therefore, we explore the applicability of ML methods in the presence of promotions, considering the whole demand life cycle: normal, promotional, and postpromotional.

In reality, people are sensitive to retail sales promotions and try to get an advantage out of them (Hewage et al. 2021). Often, traditional forecasting methods are unable to cope with such complexities. Thus, our study is concerned with the comparison of the forecast performances of conventional univariate methods and ML methods. Therefore, we aim to explore the potential of ML methods in retail operations to support better decision-making using available data. Specifically, we focus on investigating the use of ML methods that incorporate exogenous features, such as the type of promotional period, to improve the accuracy of retail promotional forecasting and to compare their performance with univariate methods in each period.

This investigation is relevant from both a theoretical and practical point of view, as retailers have access to an increasing array of data and need to understand how data can improve decision-making in retail operations. The primary goal of this research is to determine whether ML approaches are viable for forecasting sales in the context of retail sales promotions.

Therefore, our study makes the following contributions:

- We explore whether ML methods have the capacity to automatically identify the promotional effects.
- We investigate the impact of incorporating more exogenous features on the performance of ML methods, using real-world sales data from https://www.dunnhumby.com/source-files/.
- Our study provides an extensive comparison of the performance between the proposed forecasting framework and statistical univariate forecasting methods.

Our paper is structured as follows. Section 2 discusses the relevant literature as well as the theoretical foundation for the hypothesis development. The methodology is presented under

Section 3. Section 4 includes a full analysis and the findings of the empirical study. Sections 5 and 6 of the paper focus on the discussion and conclusion respectively.

2 | Literature Review

A retail supply chain is made up of retailers, suppliers, other intermediaries, and manufacturers who collaborate to meet customer demand (Fildes et al. 2019; Perera et al. 2019). Retailing is a highly competitive industry due to the various complexities and uncertainties (Ma et al. 2016). These challenges and uncertainties arise as a result of shifting customer expectations, competitor actions, partner activities, promotional activities, shorter lead times, and emerging technologies (Hewage and Perera 2022b; Ma et al. 2016; Ma and Fildes 2021). All these factors contribute to a volatile retail supply chain affected by demand volatility. Therefore, even a small improvement in operational decisions allows retailers to maintain their operations at a competitive level (Hübner et al. 2018; Ma and Fildes 2017).

Importantly, retailers need to effectively manage their supply chain in order to successfully meet customer demand (Fildes et al. 2019). Despite this, retail planning tasks are highly complex since retailers need to manage a wide range of products within limited shelf space (Mou et al. 2018). Therefore, sales forecasting[1] is a critical task in retail planning (Guo et al. 2013; Hewage et al. 2021; Hewage and Perera 2022a). Retailers, in particular, must generate proper forecasts for individual products in order to manage all logistics services while avoiding stock imbalances and ensuring consumer satisfaction (Ali et al. 2009). Thus, sales forecasting is a fundamental input in retail operations (Brinke et al. 2023) as it is required for making various operational decisions such as sourcing, procurement, production planning, logistics, marketing, and financial decisions (Hanssens 1998; Huber and Stuckenschmidt 2020).

Inaccurate forecasts often result in stock-outs or high stock levels that are prone to obsolescence (Huang et al. 2019). If stock-outs occur on a frequent basis, they can lead to customer dissatisfaction and, eventually, customers switching to other retail outlets (Huang et al. 2019). Thus, retailers tend to maintain a buffer stock to ensure customer satisfaction. This ultimately leads to higher inventory costs and reduced profits (Ma et al. 2016; Perera et al. 2019). However, producing reliable and accurate sales forecasts is a very challenging task in the retail context (Ali and Gürlek 2020; Trapero et al. 2015). Many factors, such as sales promotions, weather, holidays, and special events, can influence observed sales data at the product level, causing demand irregularities (Fildes et al. 2019; Perera et al. 2019). Sales promotions are one of the salient factors in creating irregular sales patterns among them (Baek 2019; Bandara et al. 2019).

Sales promotions cause demand volatility not just during the promotional period, but also throughout the demand life cycle (Abolghasemi, Beh, et al. 2020). Normally, a sales uplift can be found during promotional periods. This increase in sales is usually the result of customers changing their buying patterns, either through purchase acceleration or higher consumption (Blattberg and Briesch 2012) because customers tend to stockpile products during sales promotions for future consumption (Perera et al. 2019). This often leads to lower sales figures than the baseline (normal) level[2] for a short period of time in the immediate aftermath of a promotion. The sales figures then recover to a normal level again with time (Abraham and Lodish 1987). This period where there is a dip in demand is identified as the postpromotional period (Hewage et al. 2021). Hence, a retail sales promotion has three phases: the normal period, the promotional period, and the postpromotional period (Hewage et al. 2021), creating different demand variations in each period (DelVecchio et al. 2006).

For various reasons, sales forecasting in the presence of promotions can be challenging (Fildes et al. 2018). It is common for retailers to have thousands of products across hundreds of stores being promoted simultaneously (Cohen et al. 2021). However, the relative infrequency of such promotions, as well as the varying sales uplift achieved, makes the forecasting process challenging (Fildes et al. 2018). On the other hand, when a product is promoted, it not only affects the demand for that product but also the demand for other items, resulting in cross-item effects (Cohen et al. 2021). As a result, there is no standardized method for coping with changes in demand caused by retail promotions (Fildes et al. 2019).

In practice, many retailers still use simple univariate methods supplemented by judgmental adjustments or base-lift (BL) correction to cope with promotional effects (Fildes et al. 2019). Simple moving averages, ETS and its extensions, or autoregressive integrated moving average (ARIMA) approaches to state space models, are the most common univariate methods used in the retail sector (Fildes et al. 2019; Hewage et al. 2021; Ma and Fildes 2021; Perera et al. 2019; Hyndman and Khandakar 2008). Although univariate methods are extensively used due to their simplicity and robustness, using univariate models with judgmental adjustments to forecast sales in promotions might result in systemic errors (Hewage et al. 2021). Thus, these forecasts can be inaccurate, costly, and inconsistent due to bias (Fildes et al. 2009; Trapero et al. 2013).

In contrast, causal methods are capable of incorporating sales promotions into forecasts without any judgmental interference (Fildes et al. 2008). These models are often based on multiple regression, incorporating causal effects of promotions into the forecasts (Trapero et al. 2015). Some of the known implementations of these methods are SCAN*PRO (Leeflang et al. 2005), PromoCast (Cooper et al. 1999), CHAN4CAST (Divakar et al. 2005), and Driver Moderator (Gür Ali 2013; Huang et al. 2014; Ma et al. 2016). These methods, however, are quite sophisticated and have stringent data requirements (Lee et al. 2007; Trapero et al. 2013). Thus, these models are not widely employed in the industry (Fildes et al. 2019).

Unstructured methods, on the other hand, can use past sales and causal variables with lags as input to provide forecasts during promotional periods (Ali and Gürlek 2020). Thus, ML methods are gaining traction as a viable option for forecasting retail sales (Fildes et al. 2019). Some of the popular implementations include support vector machines (SVMs), regression trees (RTs), artificial neural networks (ANN), and boosted trees (BTs) (Fildes et al. 2019; Loureiro et al. 2018; Perera et al. 2019). With ever-increasing volumes of data created by both retailers and customers, ML is expected to have a significant impact on retail (Wang et al. 2020). ML models, despite being computationally expensive, give flexibility and high predicted accuracy when there is a large amount of data (Ali and Gürlek 2020). Furthermore, the results of a recent M5 competition on Kaggle show the potential of ML in retail forecasting tasks (Spiliotis et al. 2020).

Past literature shows that ML methods often improve forecast accuracy compared to linear models in the presence of promotions for retail products (Fildes et al. 2019). As an example, Ali et al. (2009) proposed a RT-based method incorporating a range of causal variables such as promotion and price, along with past sales at the SKU level. They found that the proposed model with causal features substantially improved the forecast accuracy in promotional periods. Also, Huber and Stuckenschmidt (2020) report that ML methods including ANN and BT provide more accurate forecasts suitable for large-scale demand forecasting scenarios. Further, Abolghasemi, Beh, et al. (2020) show that the SVR model generates robust forecasts in the presence of promotions.

Most ML methods in the past have been employed as univariate techniques, which were not successful and resulted in overfitted models that often give poor forecast accuracy (Godahewa et al. 2020). With access to massive amounts of retail store-generated data, ML-based forecasting methods that are fitted globally across multiple time series outperform forecasting models trained on isolated series (Godahewa et al. 2020). Nevertheless, there are only a few studies focused on SKU-level sales forecasting using ML methods fitted globally using multiple time series (Hewage et al. 2021). Table 1 provides a comprehensive comparison of the past literature with our proposed model. However, all the previous studies have only focused on the promotional and nonpromotional (i.e., normal) periods, without considering the postpromotional period. This leaves the actual benefits and challenges of integrating *all* types of promotional periods with ML methods unexplored.

Therefore, our proposed model offers a more comprehensive approach. It includes not only traditional methods used in existing frameworks but also recent methods like BTs and deep learning (DL) models. By incorporating *all* types of promotional periods, we can gain a deeper understanding of the effectiveness of promotional strategies and how they affect sales in the long term. Thus, we aim to investigate whether ML methods are a viable alternative for forecasting retail sales in the presence of promotions. To evaluate their relative performance, we compare the ML methods with widely applied univariate methods.

2.1 | Hypothesis Development

Promotions are the main reason for incorporating judgmental adjustments into retail sales forecasting (Aruchunarasa and Perera 2022; Perera et al. 2019). However, practitioners tend to ignore quantitative forecasts altogether when making adjustments to tackle promotional effects (Perera et al. 2019). Furthermore, Goodwin (2000) and Hewage et al. (2021) report that practitioners often fail to identify the promotional periods correctly or ignore the postpromotional period and treat it as a normal period. Thus, they frequently make inappropriate adjustments that impair forecast accuracy during promotional periods (De Baets and Harvey 2018). In contrast, Trapero et al. (2015) found that the dynamic regression model can automatically identify the postpromotional dip. Ali and Gürlek (2020) also state that the FAIR model identifies the postpromotional dip. However, the postpromotional effect was not incorporated into these models. Hence, we hypothesize:

H1. *ML methods require minimal feature engineering in order to recognize the postpromotional period and accurately predict the magnitude of the postpromotional dip.*

Interestingly, previous literature shows RT with explicit features improves accuracy significantly during promotional periods (Ali et al. 2009). Huber and Stuckenschmidt (2020) further suggest expanding the feature space with exogenous features such as features of a product or information on the store to allow ML methods to implicitly cluster time series while reducing the loss function. Thus, incorporating more sophisticated variables benefits ML methods as they have the capability to take advantage of them effectively (Ali et al. 2009). As a result, we hypothesize:

H2. *ML methods can improve forecast performance during all periods, including normal, promotional, and postpromotional periods, when promotional periods are provided as an exogenous variable.*

Specifically, Huber and Stuckenschmidt (2020) points out that it is unclear whether ML methods can outperform conventional methods in retail sales forecasting. Furthermore, previous literature emphasizes the need of research in retail sales forecasting due to limited availability of objective evidence on performance comparisons (Fildes et al. 2019; Makridakis et al. 2018). Hence:

H3. *In the retail industry, ML methods outperform statistical univariate methods across the demand life cycle.*

3 | Methodology

This section summarizes the proposed forecasting framework, which consists of three components: input data/preprocessing, forecast engine, and postprocessing/final forecasts. We employ multiple forecasting techniques, including both statistical and ML techniques, as the main core of our forecast engine in the proposed framework. Figure 2 depicts the overview of the proposed forecasting framework.

3.1 | Input Data

Retailers require SKU level sales forecasting as it is the primary operational unit for managing daily stock replenishment (Fildes et al. 2019). Thus, we focus on SKU level sales forecasting in this study. We used a publicly available dataset[3] from a leading US-based retailer. The dataset used in our study consists of four product categories (cereal, frozen pizza, oral hygiene products, and snacks) carrying 55 SKUs across 75 stores, resulting in 3364 unique time series. The dataset spans over 156 weeks. Table 2 shows the descriptive statistics of the collected dataset.

TABLE 1 | A comparison of relevant previous studies with our proposed model.

Paper	Methods	Time series features	Cleansed sales	Product features	Store features	Promotion features	Demand life cycle: Normal period	Demand life cycle: Promotion period	Demand life cycle: Postpromotional period
Cooper et al. (1999)	Base-lift model Historical averages	✓	✓	✓	✓	✓	✓	✓	✖
Ali et al. (2009)	Regression Base-lift model SVR RT	✓	✓	✓	✓	✓	✓	✓	✖
Guo et al. (2013)	IELM GLM MID	✓	✓	✓	✓	✓	✓	✓	✖
Trapero et al. (2015)	NAÏVE Simple exponential smoothing Base-lift model Dynamic regression Pooled regression	✓	✓	✓	✖	✓	✓	✓	✖
Ma et al. (2016)	ETS Base-lift model ADL	✓	✓	✓	✓	✓	✓	✓	✖
Ma and Fildes (2017)	ADL	✓	✖	✓	✓	✓	✓	✓	✖
Huang et al. (2019)	Base-lift model ADL	✓	✓	✓	✓	✓	✓	✓	✖
Abolghasemi, Hurley, et al. (2020)	Theta ARIMA ARIMAX ETS ETSX DLR ANN SVR	✓	✖	✖	✖	✖	✓	✓	✖

(Continues)

TABLE 1 | (Continued)

Paper	Methods	Time series features	Cleansed sales	Product features	Store features	Promotion features	Demand life cycle		
							Normal period	Promotion period	Postpromotional period
Abolghasemi, Beh, et al. (2020)	ARIMA ARIMAX SVR RT FSE	✓	✖	✓	✖	✓	✓	✓	✖
Ali and Gürlek (2020)	ARIMA ETS VAR FAIR Boosted trees STL	✓	✖	✓	✓	✓	✓	✓	✖
Huber and Stuckenschmidt (2020)	sNAÏVE Regression ETS FNN LSTM GBRT	✓	✖	✓	✓	✓	✓	✓	✖
van Steenbergen and Mes (2020)	RF Quantile regression forest	✓	✖	✓	✓	✖	✓	✖	✖
Ma and Fildes (2021)	ETS ADL ARIMAX SVR ELM ELMP RF GBRT	✓	✓	✓	✓	✓	✓	✓	✖

(Continues)

TABLE 1 | (Continued)

Paper	Methods	Time series features	Cleansed sales	Product features	Store features	Promotion features	Demand life cycle: Normal period	Promotion period	Postpromotional period
Our proposed framework	NAÏVE Base-lift model ARIMA ETS ETSX GBRT RF DeepAR WaveNet	✓	✓	✓	✓	✓	✓	✓	✓

Abbreviations: ADL: autoregressive distributed lag, ANN: artificial neural network, ARIMA: autoregressive integrated moving average, ARIMAX: ARIMA with exogenous variable, DLR: dynamic linear regression, ELM: extreme learning machine, ELMP: ELM with data pooling, ETS: exponential smoothing, ETSX: ETS with exogenous variable, FAIR: fully automatic interpretable retail forecasting, FNN: feedforward neural network, FSE: forecasting systematic events, GBRT: gradient-boosted regression tree, GLM: generalized linear model, IELM: improved ELM, MID: multivariate intelligent decision-making, RF: random forest, RT: regression tree, STL: seasonal trend decomposition, SVR: support vector regression, VAR: vector autoregressive.

Next, we examine the time series plot at different cross-sectional aggregation levels to observe the time series features such as trend, seasonality, and noise. Figure 3 shows that at higher aggregate levels, we can observe seasonality patterns with low trend. However, at the product level, the time series becomes more volatile in some cases due to low sales volume.

However, due to the high number of series, it is not visually feasible to observe time series features properly. Therefore, we extracted the time series features of all 3364 series using the "STL" (Seasonal and Trend Decomposition using Loess) decomposition method (Cleveland et al. 1990). Figure 4 illustrates the strengths of trend and seasonality of each time series, with both measures on a scale of [0, 1]. The majority of time series in the cereal, frozen pizza, and oral hygiene categories show a low and moderate strength of trend and low seasonality. Conversely, in snacks, we can see a moderate and high seasonality as well. However, it was evident that even within the same product category (e.g., Snacks), different patterns of trends and seasonality were observed, making the forecasting process challenging.

3.2 | Input Features

In our study, we utilize a combination of time series and causal data, encompassing both static and dynamic features. We defined a total of 14 features, which are depicted in Table 3. To identify the most significant lag predictors, we explored the correlation between the target variable and its lag predictors using lag scatterplots. This initial exploration suggests that the first three lag predictors are the most suitable for modeling. Moreover, previous literature has emphasized the significance of these variables in the retail domain (Ali and Gürlek 2020; Huber and Stuckenschmidt 2020; Ma and Fildes 2021). Furthermore, it is worth noting that these features are typically accessible to most retailers.

3.3 | Data Preprocessing

We started by looking for missing values in the dataset, but there were none. Estimating baseline demand is fundamental to classifying the promotional periods as normal, promotional, and postpromotional. Baseline demand represents the sales level without taking promotional effects into consideration. For this, we employed the ETS model using the sales levels in normal periods. The ETS model has several advantages for our study, including simplicity and robustness (Hyndman and Khandakar 2008). We used the *ets()* function in the R *forecast* package (Hyndman and Khandakar 2008) to implement the ETS model. We elaborate on the model further under Section 3.4.

Following that, the promotional calendar was used to classify the normal and promotional periods. We employed Equation (1) to identify the postpromotional periods.

$$D_{it} = B_{it} - A_{it}, \tag{1}$$

where A_{it}: actual sales for SKU i at the t^{th} week, B_{it}: baseline demand for SKU i at the t^{th} week, and D_{it}: difference between baseline demand and actual sales at the t^{th} week for SKU i. If

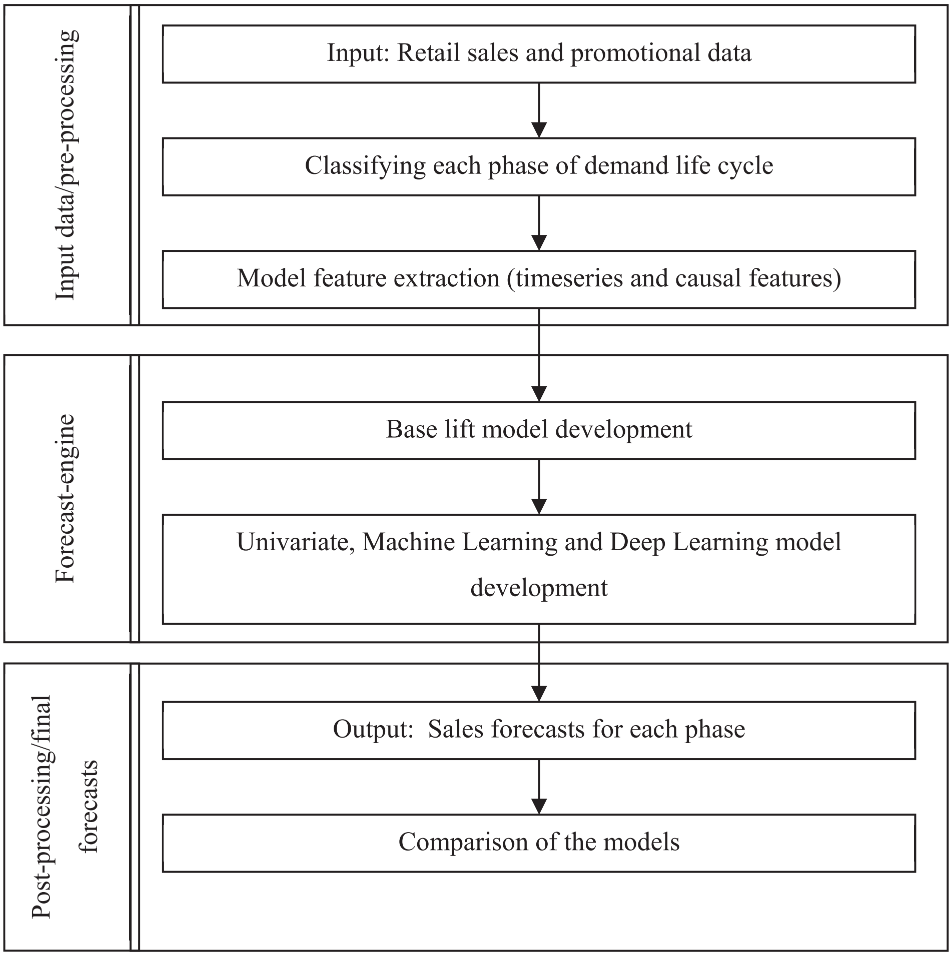

FIGURE 2 | The proposed framework consists of three components, namely, input data/preprocessing, forecast engine, and postprocessing/final forecasts.

TABLE 2 | Descriptive summary of the dataset.

		Weekly sales					
		Normal		**Promotional**		**Postpromotional**	
	# of SKUs	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Cereal	15	41.6	25.3	103.2	63.2	37.5	26.7
Frozen pizza	12	18.3	9.64	60.5	35.4	14.0	9.48
Oral hygiene products	13	12.9	6.07	30.5	14.0	8.10	4.22
Snacks	15	29.8	26.2	62.6	43.5	22.4	18.9

D_{it} is negative immediately following a promotion, the t^{th} period ($t \geq 0$) is classified as a postpromotional period. However, the postpromotional effect is most noticeable only within the first 1 or 2 weeks following a promotion (Macé and Neslin 2004).

We further divided the dataset into training and test sets. The training data were utilized to estimate the parameters of each forecasting method, while the test data were used to evaluate the forecast accuracy. The training set encompassed the initial 130 weeks, totaling 439,400 observations, whereas the test set comprised the subsequent 26 weeks, totaling 87,880 observations. Additionally, when developing the forecasting models, we set the frequency of each series to 52 weeks. This frequency was selected to synchronize with the weekly data, ensuring that the models adequately captured the inherent seasonality and other temporal patterns.

3.4 | Benchmark Model

As a benchmark method, we implemented a BL model using the baseline demand estimation described in Section 3.2. This model, which is commonly used by retailers for forecasting (Ma and Fildes 2021) and is also implemented in commercial

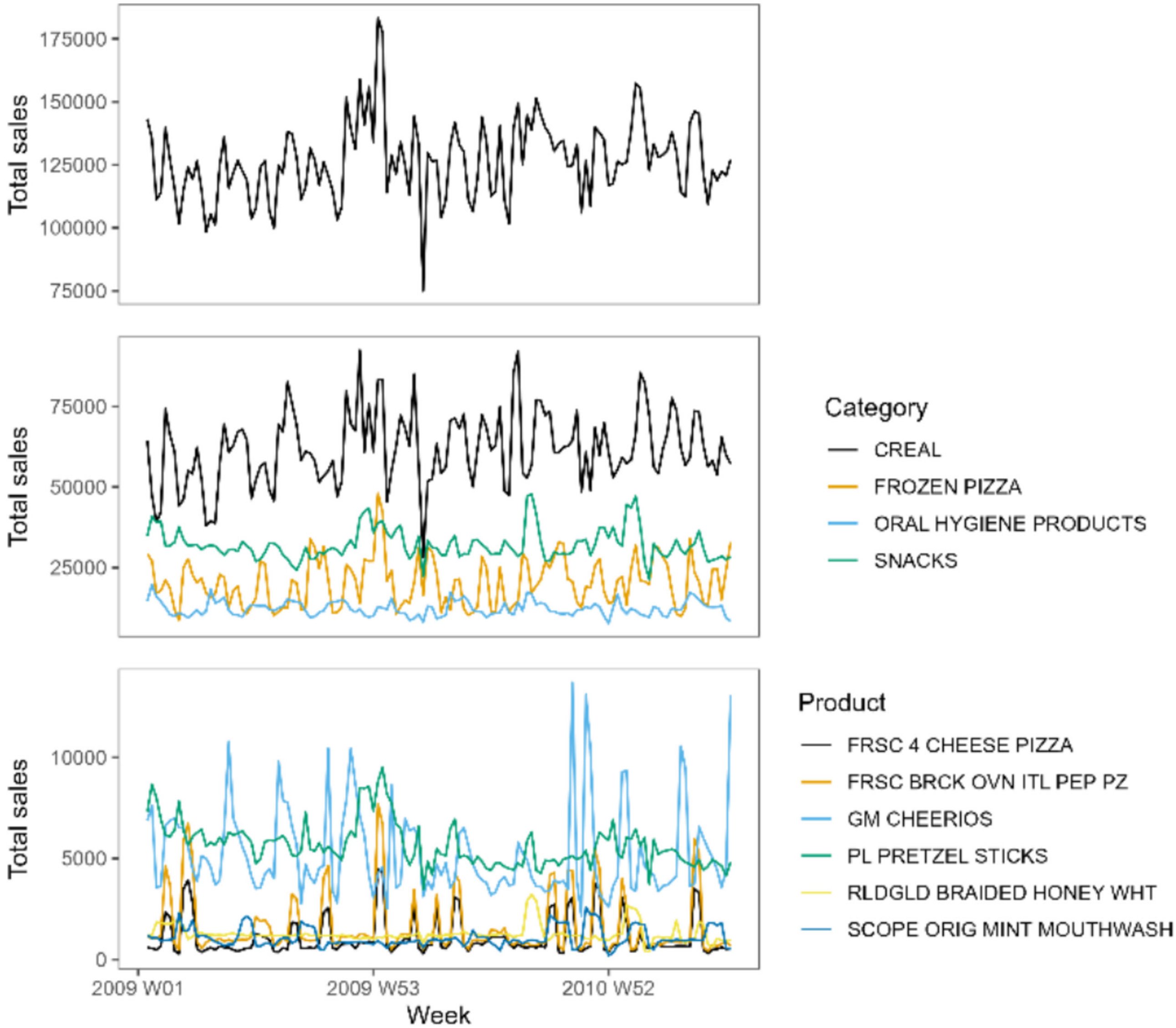

FIGURE 3 | Time series of total sales (January 14, 2009–July 06, 2011) at cross-sectional aggregate levels. The *x* axis displays the week, and the *y* axis indicates the total sales in terms of units. The top panel shows total sales for the entire company, the second panel shows total sales for each category, and the bottom panel shows total sales for each product. To ensure clarity and prevent overplotting, only five of the time series at the product level are displayed.

applications (Ali et al. 2009), adjusts the forecast based on promotional or postpromotional effects. If a promotion is planned for the coming week, the average promotional lift is added to the forecast. If the week is identified as a postpromotional period, the average postpromotional dip is added to the forecast. Otherwise, the forecast value is left unchanged for the normal period, as shown in Equation (2).

$$BL_{it}\begin{cases} B_{it} & ;(t = normal\ period), \\ B_{it} + (Average\ promotional\ uplift)_i & ;(t = promotional\ period), \\ B_{it} - (Average\ postpromotional\ dip)_i & ;(postpromotional\ period). \end{cases} \tag{2}$$

where i: selected SKU, B_{it}: baseline demand for SKU i at t^{th} week, and BL_{it}: final forecast for SKU i at t^{th} week.

3.5 | Forecasting Methods

We consider three groups of methods in our study, namely, (1) univariate methods, (2) ML-based methods, and (3) DL-based methods. As univariate methods, we use ARIMA and ETS models because these are widely applied in both the retail industry and academia (Fildes et al. 2019; Perera et al. 2019; Hyndman and Khandakar 2008). Furthermore, we implement the sNAÏVE model in our study for comparison purposes. We also use ETS with exogenous variable (ETSX), an extension of the ETS model (Abolghasemi, Beh, et al. 2020). For ML-based methods, we use LightGBM (LGB), xgBoost (XGB), and random forest (RF) methods. Finally, within the DL family of methods, we use DeepAR and WaveNet in our study. Next, we detail the methods and specific implementations we used in our study.

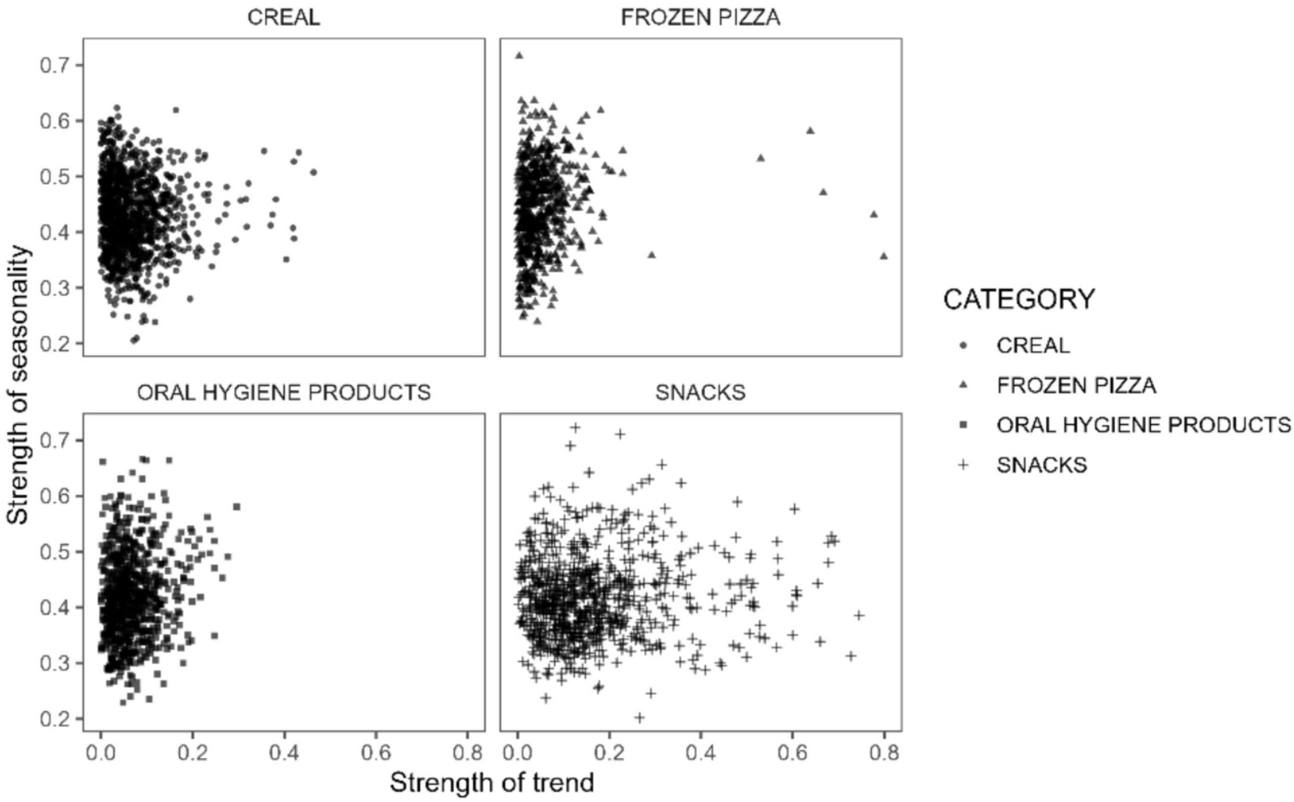

FIGURE 4 | The strengths of trend and seasonality in the time series of sales. The scatter plot comprises 3364 data points, with each point representing a specific time series.

TABLE 3 | Selected input features.

Feature type	Feature
Target variable	Weekly sales
Time series features	Lagged sales for 1, 2, and 3 weeks
Dynamic features	Calendar features, promotional types (i.e., temporary price reduction, display, and feature), magnitude of discounts, and selling price
Static features	Store ID, category, and subcategory ID, SKU
Additional feature	Promotional period (i.e., normal, promotional, or postpromotional)

3.5.1 | sNAÏVE

This simple forecasting model involves generating forecasts using the last known observation from the previous same period. To implement this model, we utilized the SNAIVE() function in the fable package in R (Hyndman and Athanasopoulos 2021).

3.5.2 | ARIMA

ARIMA model is a widely used approach in practice since it can take into consideration trend, seasonality, and error, as well as the nonstationarity of a time series (Hewamalage et al. 2021). In our study, we used the AutoARIMA algorithm (Hyndman and Khandakar 2008), which finds the best ARIMA model automatically (Hyndman and Khandakar 2008). First, it finds the appropriate order of difference (d) by using the Kwiatkowski–Phillips–Schmidt–Shin unit root test. Second, it determines the appropriate order of the autoregressive component (p) and the moving average component (q) values by fitting different models and selecting the model with the lowest Akaike information criterion (AICc) (Hyndman and Khandakar 2008). Moreover, the AutoARIMA algorithm is capable of fitting seasonal ARIMA models by identifying the number of seasonal differences and other model parameters by minimizing the AIC, similar to nonseasonal ARIMA models. Thus, it determines the best model for each time series by selecting the one with the lowest AIC after comparing both nonseasonal and seasonal ARIMA models simultaneously (Hyndman and Athanasopoulos 2021). To implement the ARIMA model, we used the *auto.arima()* function in the R *forecast* package (Hyndman and Khandakar 2008).

3.5.3 | ETS and ETSX

ETS is a univariate method based on ETS in a state space framework that takes seasonality, trend, and error into account (Petropoulos and Svetunkov 2020; Hyndman and Khandakar 2008). This automatically determines the best model by minimization of a prespecified information criterion from the underlying 15 ETS models (Hyndman and Khandakar 2008). Moreover, the ETS model can be extended using a regressor variable when additional information is available to construct the ETSX model. This allows us to incorporate causal features into the ETS model (Petropoulos and Svetunkov 2020; Hyndman and Khandakar 2008). We used the *ets()* function in the R *forecast* package (Hyndman and Khandakar 2008) and the *es()* function in the R *smooth* package (Petropoulos and Svetunkov 2020) to implement ETS and ETSX models, respectively.

3.5.4 | Gradient-Boosted RTs (GBRTs)

GBRTs have gained popularity as a potential approach in time series forecasting (Ma and Fildes 2021) and as a viable alternative to ANNs (Huber and Stuckenschmidt 2020). LGB and XGB are the most widely used implementations among these (Huber and Stuckenschmidt 2020). They train a series of decision trees, one at a time. It is based on the accumulated errors of the last tree, similar to boosting approaches. Therefore, the final forecast is the aggregate of all trees trained (Hewage and Perera 2022b; Huber and Stuckenschmidt 2020). We used the *LightGBM* Python Package (Microsoft Corporation 2022) and *XGBoost* Python Package (xgboost Developers 2021) to implement LGB and XGB models, respectively.

3.5.5 | RF

RF is a collection of RTs, each of which is based on the values of a random vector with the same distribution that is sampled independently (Breiman 2001). The accuracy of the RF is determined by the correlation and strength of the individual trees, as well as the size of the forest. RF averages the forecasts of multiple RTs to produce the final forecast (Breiman 2001). Therefore, it is more resistant to noise and is less prone to overfit the training data (Breiman 2001). Further, past literature states RF is a promising approach in the retail context (Spiliotis et al. 2020). We used the *RandomForestRegressor* Python Package (scikit-learn Developers 2022) to implement the RF model.

3.5.6 | DeepAR and WaveNet

DeepAR model is based on an autoregressive recurrent neural network framework and trains a large number of related time series simultaneously (Salinas et al. 2020). On the other hand, WaveNet is made up of detailed causal convolutional layers. Thus, it can produce real-valued data sequences in response to some conditional inputs (Sprangers et al. 2022). Though these models were introduced recently, they have been identified as potential approaches for sales forecasting (Vallés-Pérez et al. 2022). Moreover, the WaveNet model finished second in the Kaggle competition that featured the Corporaci Favorita data (Vallés-Pérez et al. 2022). We used the *GluonTS toolkit* in Python (Amazon Web Service 2022) to implement both DeepAR and WaveNet models.

4 | Experimental Study

This section highlights the forecasting methods and the error measures utilized in our study.

4.1 | Overview of the Candidate Models

In our study, we developed 14 candidate models using different combinations of input features, as shown in Table 4. For the ML and DL methods, we created two forecasting groups for each method by varying the availability of the promotional period as an input feature. Models denoted by 1 included the promotional period as an input feature, while models denoted by 2 did not. This experimental setup was used to investigate whether providing additional feature variables improved the performance of the ML and DL models. We used the default parameters to train all of the forecasting models and did not perform any hyperparameter tuning for the ML and DL methods in order to keep the models simple.

4.2 | Error and Performance Measures

We focused on four major areas in our analysis. First, we evaluated the magnitude and sign of the postpromotional effect identified by each candidate model. We measured the magnitude and the sign of the postpromotional effect identified using Equation (3).

$$PM_{it} = \left(F_{it} - B_{it}\right) / B_{it}, \tag{3}$$

where F_{it}: forecasted sales for SKU i at t^{th} week, B_{it}: baseline demand for SKU i at t^{th} week, and PM_{it}: magnitude of the postpromotional effect at t^{th} week for SKU i.

Second, we evaluate the forecast accuracy of the models using symmetric MAPE (sMAPE) (Bandara et al. 2020) and mean absolute scaled error (MASE) (Hyndman and Koehler 2006) using Equations (4) and (5), respectively.

$$sMAPE = \frac{\sum_{t=1}^{n} \frac{|F_t - A_t|}{(|F_t| + |A_t|)/2}}{n} \times 100, \tag{4}$$

$$MASE = \frac{\frac{1}{(n-1)} \sum_{t=2}^{n} |A_t - A_{t-1}|}{\frac{1}{(n)} \sum_{t=1}^{n} |F_t - A_t|} \times 100, \tag{5}$$

where A_t: actual sales at t^{th} week, F_t: forecasted sales at t^{th} week, and n: number of series. Both of these error metrics are widely used in the field of time series forecasting (Huang et al. 2019). However, sMAPE has some shortcomings, such as a lack of interpretability, a lack of robustness, and being unstable with values close to zero (Bandara et al. 2019). To mitigate some of these issues, we use MASE as our second error metric, as it is scale independent (Hyndman and Koehler 2006).

Thirdly, we compare the value addition from the additional variable to the ML and DL methods using the forecast value added (FVA) model: Equation (6). Chybalski (2017) explains that FVA compares the forecast improvement of a model with another. Error metrics such as the mean absolute percentage

TABLE 4 | Overview of the candidate models.

	BL	sNAÏVE	ETS	ETSX	ARIMA	LGB		XGB		RF		DeepAR		WaveNet	
						1	2	1	2	1	2	1	2	1	2
Week	✓	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Raw sales	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cleansed sales	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Lagged sales	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Store ID	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SKU	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product category	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Product subcategory	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Discount rate	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
TPR (binary)	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Display (binary)	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Feature (binary)	—	—	—	—	—	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Promotion period	✓	—	—	✓	—	✓	—	✓	—	✓	—	✓	—	✓	—
Average promotional uplift	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Average postpromotional dip	✓	—	—	—	—	—	—	—	—	—	—	—	—	—	—

error (MAPE), MASE, or any other measure can be used during the analysis (Chybalski 2017). In our study, we used MASE to produce the FVA calculation. FVA > 0 indicates that there is an improvement in the forecast performance in comparison to the benchmark. On the contrary, FVA < 0 shows that there is no forecast improvement against the selected method.

$$FVA_{i,k} = |MASE_k| - |MASE_i| \tag{6}$$

where $FVA_{i,t}$: FVA for model i compared to model k, $MASE_k$: MASE for model k, and $MASE_i$: MASE for model i.

Finally, we used the Friedman test to examine the statistical significance of the differences in these methods (Friedman 1940). To further explore these differences with respect to each method, we used the Wilcoxon signed-rank test (Wilcoxon and Wilcox 1964). We used the *friedman_test()* function in the R *rstatix package* to employ the Friedman test and *pairwise.wilcox.test()* in R.

Finally, we compare the efficiency of each model by analyzing the run time (computational time) required for each one. This will give us a more comprehensive understanding of the realistic nature of the proposed framework.

5 | Analysis and Results

5.1 | Comparison of Magnitude and Sign of Postpromotional Effect

Figure 5 shows the distribution of the postpromotional dip identified by each forecasting model.

Table 5 shows the descriptive summary of the postpromotional effects identified by the forecasting models. A first exploration indicates that ML methods are able to identify the postpromotional period compared to univariate and DL methods. Friedman test results indicate that there are significant differences ($\chi^2(235) = 390.4$, $p < 0.000$) in identified postpromotional dips by forecasting methods.

The Wilcoxon signed-rank test results show that all univariate models, including ETSX, are significantly different from the test dataset's real mean postpromotional dip ($p < 0.000$). Though ML models are able to identify the postpromotional period, we see that they only identify the correct magnitude of the postpromotional dip when the additional feature is incorporated; LGB1 versus LGB2 ($p < 0.000$), XGB1 versus XGB2 ($p < 0.000$), and RF1 versus RF2 ($p < 0.000$) fail to provide support for H1. Noticeably, DL methods are unable to identify the postpromotional period even with the additional variable and are significantly different from the actual postpromotional dip ($p < 0.000$). Furthermore,

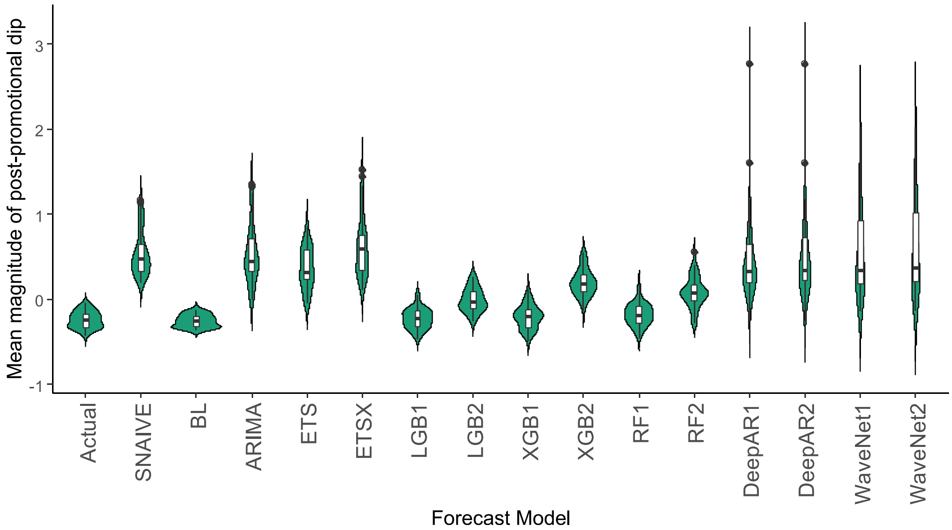

FIGURE 5 | Distribution of the magnitude of postpromotional dip.

TABLE 5 | Descriptive summary of the sign and magnitude of the postpromotional dip, the top performing models are highlighted in boldface.

	Postpromotional dip		
Forecasting model	**Mean**	**Median**	**SD**
Postpromotional dip in test dataset	−23.86%	−24.66%	9.81
BL	**−25.50%**	**−25.66%**	**7.31**
sNAÏVE	54.77%	47.43%	26.82
ARIMA	52.54%	44.37%	34.49
ETS	36.99%	30.99%	24.67
ETSX	63.25%	59.97%	36.68
LGB1	**−22.65%**	**−22.20%**	**11.92**
LGB2	−0.67%	−3.36%	13.88
XGB1	**−20.81%**	**−20.39%**	**13.82**
XGB2	19.73%	18.60%	15.84
RF1	**−18.88%**	**−18.25%**	**13.02**
RF2	8.31%	7.81%	18.49
DeepAR1	47.38%	33.12%	51.90
DeepAR2	50.30%	33.63%	53.22
WaveNet1	54.86%	34.48%	55.83
WaveNet2	59.70%	36.38%	58.55

LGB1 ($p = 0.628$), XGB1 ($p = 0.361$), and RF1 ($p = 0.054$) show no significant differences from the actual postpromotional dip, providing partial support for H2.

5.2 | Comparison of Forecast Performance

We separately compare model performance in each promotional period using sMAPE and MASE. Table 6 summarizes the descriptive statistics of sMAPE and MASE across forecasting methods.

5.2.1 | Forecast Performance During the Normal Period

Figure 6 shows the distribution of sMAPE and MASE values in the normal period. A comparison of sMAPE and MASE of the normal period was conducted using the Friedman test. The results show that there are significant differences (*sMAPE*: $\chi^2(436) = 450.43$, $p = 0.036$: *MASE*: $\chi^2(436) = 495.57$, $p = 0.042$) between forecasting methods in the normal period.

The Wilcoxon signed-rank test results show significant differences between univariate methods and other forecasting methods ($p < 0.000$), except in two cases: DeepAR2 and WaveNet2. The results further show no significant differences between ML methods (*sMAPE*: $p > 0.05$; *MASE*: $p > 0.05$) in the normal period irrespective of providing the additional variable. However, DL methods show a significant improvement in terms of sMAPE (*DeepAR*: $p < 0.000$; *WaveNet*: $p < 0.000$) when the promotional period is provided as an additional variable. H2, therefore, is only partially supported in the normal period.

5.2.2 | Forecast Performance During the Promotional Period

Figure 7 shows the distribution of sMAPE and MASE values in the promotional period. Results of the Friedman test indicate that there are significant differences (*sMAPE*: $\chi^2(436) = 527.14$, $p = 0.001$; *MASE*: $\chi^2(436) = 466.05$, $p = 0.037$) between forecasting models in the promotional period.

Wilcoxon signed-rank test reveals significant differences ($p < 0.000$) between univariate methods and ML methods during the promotional period, providing evidence for H3. However, all the ML models show no significant differences among themselves, even with the additional variable (*sMAPE*: *LGB*: $p = 0.933$; *XGB*: $p = 0.533$; *RF*: $p = 0.973$|*MASE*: *LGB*: $p = 0.830$; *XGB*: $p = 0.695$; *RF*: $p = 0.898$). Only LGB1 (*sMAPE*:

TABLE 6 | Forecast accuracy for each forecasting method, the top performing model(s) in each column are highlighted in boldface.

Forecasting method	sMAPE						MASE					
	Normal		Promotional		Postpromotional		Normal		Promotional		Postpromotional	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
BL	0.27	0.06	0.33	0.18	**0.27**	**0.12**	**0.71**	**0.08**	0.98	0.85	**0.26**	**0.13**
sNAÏVE	0.33	0.09	0.42	0.11	0.67	0.16	1.05	0.34	1.02	0.14	0.79	0.17
ARIMA	0.41	0.16	0.49	0.15	0.70	0.28	1.33	0.62	1.17	0.37	0.97	0.51
ETS	0.37	0.10	0.58	0.20	0.61	0.25	1.15	0.36	1.41	1.00	0.84	0.47
ETSX	0.38	0.15	0.45	0.14	0.74	0.26	1.15	0.57	1.03	0.22	1.06	0.38
LGB1	0.28	0.04	**0.25**	**0.10**	0.29	0.10	0.75	0.08	**0.69**	**0.42**	0.28	0.14
LGB2	0.29	0.04	0.27	0.11	0.38	0.15	0.78	0.10	0.70	0.43	0.41	0.21
XGB1	0.30	0.05	0.31	0.12	0.35	0.12	0.82	0.09	0.89	0.56	0.34	0.16
XGB2	0.31	0.05	0.30	0.13	0.51	0.18	0.85	0.14	0.88	0.58	0.59	0.23
RF1	0.29	0.05	0.28	0.12	0.31	0.12	0.80	0.09	0.81	0.52	0.31	0.15
RF2	0.30	0.05	0.29	0.12	0.44	0.17	0.82	0.11	0.82	0.55	0.52	0.26
DeepAR1	**0.24**	**0.05**	0.58	0.22	0.56	0.20	1.30	0.47	1.24	0.55	0.74	0.42
DeepAR2	0.43	0.11	0.64	0.27	0.33	0.12	1.62	0.55	1.31	0.51	0.88	0.50
WaveNet1	**0.24**	**0.05**	0.55	0.25	0.33	0.12	1.61	0.59	1.17	0.56	0.92	0.51
WaveNet2	0.41	0.10	0.67	0.22	0.59	0.26	1.75	0.83	1.28	0.47	1.04	0.54

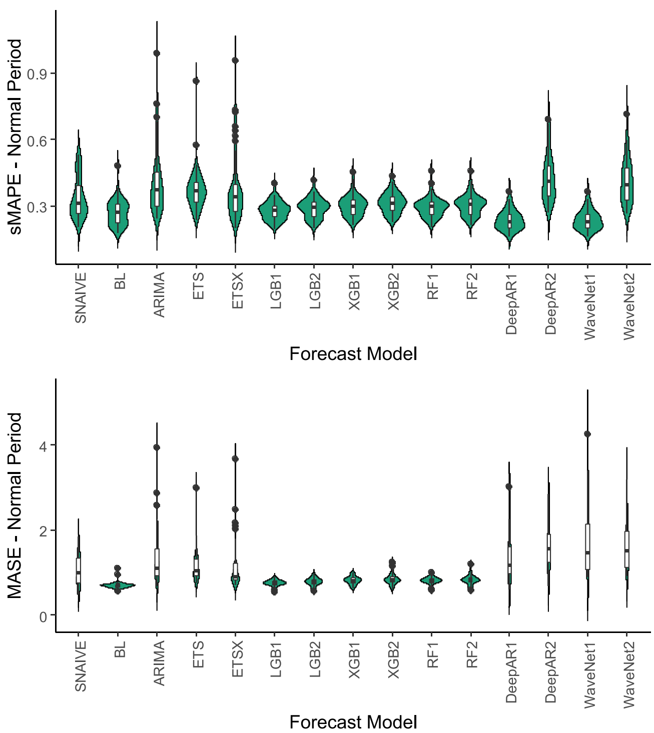

FIGURE 6 | First panel shows sMAPE values in normal period; second panel shows MASE values in normal period.

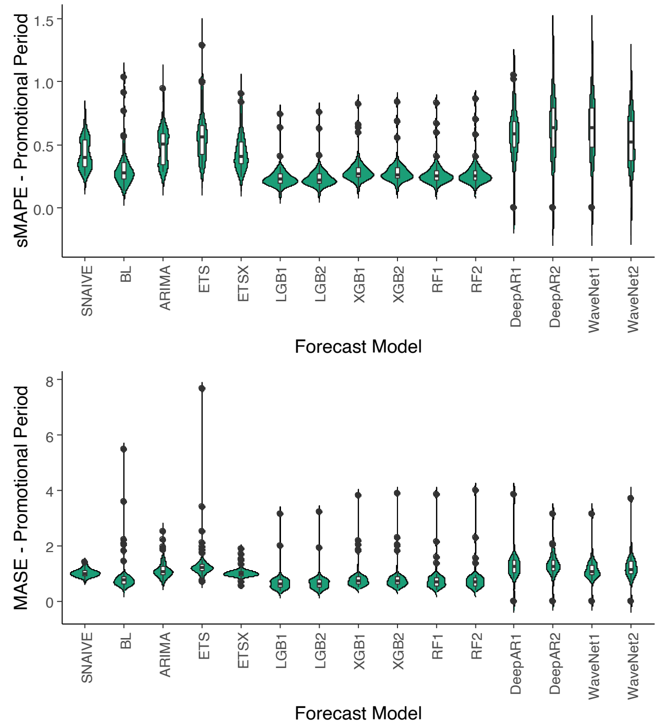

FIGURE 7 | First panel shows sMAPE values in promotional period; second panel shows MASE values in promotional period.

$p = 0.000$; *MASE*: $p = 0.000$) and LGB2 (*sMAPE*: $p = 0.000$; *MASE*: $p = 0.001$) models outperform the BL method in the promotional period. This lends some credence to H3. All other models (i.e., XGB and RF) perform similarly ($p > 0.05$) to the BL method. Surprisingly, all the DL methods show no significant differences ($p > 0.05$) with univariate methods. This provides no evidence for H2 in the promotional period. Furthermore, as expected, the ETSX model outperformed the ETS model in the promotional period (*sMAPE*: $p = 0.000$; *MASE*: $p < 0.000$).

5.2.3 | Forecast Performance During the Postpromotional Period

Figure 8 shows the distribution of sMAPE and MASE values in the postpromotional period.

Friedman test results (*sMAPE*: $\chi^2(436) = 510.43, p = 0.007$; *MASE*: $\chi^2(436) = 460.46$, $p = 0.043$) demonstrate that there are significant differences in forecasting models in the postpromotional period. The Wilcoxon signed-rank test reveals that ML methods significantly differ from univariate methods ($p < 0.000$), providing partial support for H3. Notably, the pairwise comparison shows that incorporating the additional variable significantly improves the performance of ML methods (*sMAPE*: $p < 0.000$; *MASE*: $p < 0.000$). This provides support for H2. However, even with support for the additional variable, ML methods fail to outperform the BL method. Only the LGB1 model performs similarly to the BL method (*sMAPE*: *LGB1*: $p = 0.055$; *XGB1*: $p < 0.000$; *RF1*: $p = 0.000$|*MASE*: *LGB1*: $p = 0.490$; *XGB1*: $p = 0.001$; *RF1*: $p = 0.048$). This provides no support for H3. On the other hand, DL methods show a significant improvement only in terms of sMAPE (*DeepAR*: $p < 0.000$; *WaveNet*: $p < 0.000$), providing evidence for H2. Surprisingly, the ETS model outperforms the ETSX model in the postpromotional period (*sMAPE*: $p = 0.008$; *MASE*: $p = 0.000$).

5.3 | Comparison of Forecast Improvement

Tables 7–9 provide a summary of FVA values for forecasting methods in the normal period. Notably, ML methods outperform all the univariate methods across demand life cycle (Tables 6–8; $FVA > 0$). However, they did not improve the forecast compared to the BL method and performed similarly in the normal period ($p < 0.000$). In the promotional period, ML methods outperform the BL method ($p < 0.000$). On the contrary, only the LGB1 model shows no significant differences from the BL method in the postpromotional period (*sMAPE*: *LGB1*: $p = 0.055$; *MASE*: *LGB1*: $p = 0.490$). Thus, this only provides partial support to H3 as ML methods only outperform conventional univariate methods. Surprisingly, DL methods rarely outperform univariate methods and were unable to outperform the BL methods across demand life

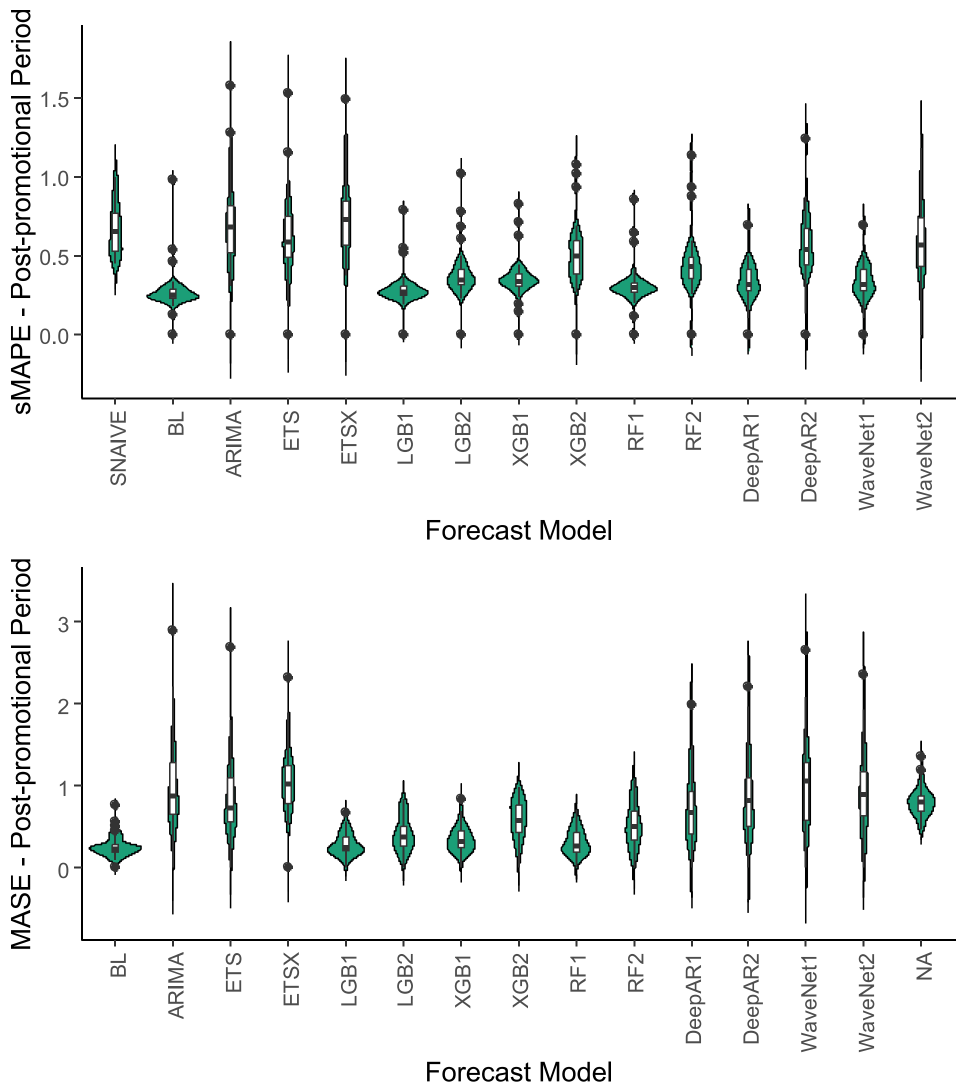

FIGURE 8 | First panel shows sMAPE values in postpromotional period; second panel shows MASE values in postpromotional period.

TABLE 7 | FVA value comparison for normal period.

	MASE	BL	sNAÏVE	ARIMA	ETS	ETSX	LGB1	LGB2	XGB1	XGB2	RF1	RF2	DeepAR1	DeepAR2	WaveNet1
BL	0.71	—	—	—	—	—	—	—	—	—	—	—	—	—	—
sNAÏVE	1.05	−0.35	—	—	—	—	—	—	—	—	—	—	—	—	—
ARIMA	1.33	−0.63	−0.28	—	—	—	—	—	—	—	—	—	—	—	—
ETS	1.15	−0.44	−0.10	0.18	—	—	—	—	—	—	—	—	—	—	—
ETSX	1.05	−0.35	0.00	0.28	0.10	—	—	—	—	—	—	—	—	—	—
LGB1	0.75	−0.04	0.31	0.59	0.40	0.31	—	—	—	—	—	—	—	—	—
LGB2	0.78	−0.07	0.28	0.56	0.37	0.27	−0.03	—	—	—	—	—	—	—	—
XGB1	0.82	−0.11	0.24	0.52	0.33	0.23	−0.07	−0.04	—	—	—	—	—	—	—
XGB2	0.85	−0.14	0.20	0.48	0.30	0.20	−0.11	−0.07	−0.03	—	—	—	—	—	—
RF1	0.80	−0.09	0.25	0.54	0.35	0.25	−0.05	−0.02	0.02	0.05	—	—	—	—	—
RF2	0.82	−0.12	0.23	0.51	0.33	0.23	−0.08	−0.05	0.00	0.03	−0.02	—	—	—	—
DeepAR1	1.30	−0.60	−0.25	0.03	−0.15	−0.25	−0.56	−0.53	−0.49	−0.45	−0.51	−0.48	—	—	—
DeepAR2	1.62	−0.91	−0.57	−0.29	−0.47	−0.57	−0.87	−0.84	−0.80	−0.77	−0.82	−0.80	−0.32	—	—
WaveNet1	1.62	−0.9	−0.56	−0.28	−0.5	−0.6	−0.86	−0.84	−0.8	−0.76	−0.82	−0.8	−0.31	0.01	—
WaveNet2	1.75	−1.05	−0.70	−0.42	−0.6	−0.7	−1.01	−0.98	−0.94	−0.9	−0.96	−0.9	−0.45	−0.13	−0.13

TABLE 8 | FVA value comparison for promotional period.

	MASE	BL	sNAÏVE	ARIMA	ETS	ETSX	LGB1	LGB2	XGB1	XGB2	RF1	RF2	DeepAR1	DeepAR2	WaveNet1
BL	0.98	—	—	—	—	—	—	—	—	—	—	—	—	—	—
sNAÏVE	1.02	−0.04	—	—	—	—	—	—	—	—	—	—	—	—	—
ARIMA	1.17	−0.19	−0.15	—	—	—	—	—	—	—	—	—	—	—	—
ETS	1.41	−0.43	−0.39	−0.24	—	—	—	—	—	—	—	—	—	—	—
ETSX	1.03	−0.05	−0.01	0.14	0.38	—	—	—	—	—	—	—	—	—	—
LGB1	0.69	0.29	0.33	0.48	0.72	0.34	—	—	—	—	—	—	—	—	—
LGB2	0.70	0.28	0.32	0.47	0.71	0.33	0.00	—	—	—	—	—	—	—	—
XGB1	0.89	0.09	0.14	0.29	0.53	0.15	−0.19	−0.19	—	—	—	—	—	—	—
XGB2	0.88	0.10	0.15	0.29	0.53	0.16	−0.18	−0.18	0.01	—	—	—	—	—	—
RF1	0.81	0.17	0.22	0.37	0.61	0.23	−0.11	−0.11	0.08	0.07	—	—	—	—	—
RF2	0.82	0.16	0.21	0.36	0.60	0.22	−0.12	−0.12	0.07	0.06	−0.01	—	—	—	—
DeepAR1	1.24	−0.26	−0.22	−0.07	0.17	−0.21	−0.55	−0.54	−0.36	−0.36	−0.43	−0.43	—	—	—
DeepAR2	1.31	−0.33	−0.28	−0.14	0.10	−0.27	−0.61	−0.61	−0.42	−0.43	−0.50	−0.49	−0.07	—	—
WaveNet1	1.12	−0.14	−0.09	0.06	0.30	−0.08	−0.42	−0.42	−0.23	−0.24	−0.31	−0.30	0.12	0.19	—
WaveNet2	1.22	−0.24	−0.20	−0.05	0.19	−0.19	−0.53	−0.52	−0.34	−0.34	−0.41	−0.41	0.02	0.09	−0.10

TABLE 9 | FVA value comparison for postpromotional period.

	MASE	BL	sNAÏVE	ARIMA	ETS	ETSX	LGB1	LGB2	XGB1	XGB2	RF1	RF2	DeepAR1	DeepAR2	s
BL	0.26	—	—	—	—	—	—	—	—	—	—	—	—	—	—
sNAÏVE	0.80	−0.54	—	—	—	—	—	—	—	—	—	—	—	—	—
ARIMA	0.97	−0.72	−0.17	—	—	—	—	—	—	—	—	—	—	—	—
ETS	0.84	−0.58	−0.04	0.14	—	—	—	—	—	—	—	—	—	—	—
ETSX	1.06	−0.81	−0.26	−0.09	−0.23	—	—	—	—	—	—	—	—	—	—
LGB1	0.28	−0.02	0.52	0.70	0.56	0.78	—	—	—	—	—	—	—	—	—
LGB2	0.41	−0.15	0.39	0.57	0.43	0.65	−0.13	—	—	—	—	—	—	—	—
XGB1	0.34	−0.09	0.46	0.63	0.49	0.72	−0.07	0.06	—	—	—	—	—	—	—
XGB2	0.59	−0.33	0.21	0.38	0.25	0.47	−0.31	−0.18	−0.25	—	—	—	—	—	—
RF1	0.31	−0.05	0.49	0.66	0.53	0.75	−0.03	0.10	0.03	0.28	—	—	—	—	—
RF2	0.52	−0.26	0.28	0.46	0.32	0.55	−0.24	−0.11	−0.17	0.07	−0.21	—	—	—	—
DeepAR1	0.74	−0.49	0.06	0.23	0.09	0.32	−0.47	−0.34	−0.40	−0.15	−0.43	−0.23	—	—	—
DeepAR2	0.88	−0.62	−0.08	0.09	−0.04	0.18	−0.60	−0.47	−0.54	−0.29	−0.57	−0.36	−0.14	—	—
WaveNet1	0.92	−0.66	−0.12	0.05	−0.08	0.14	−0.64	−0.51	−0.58	−0.33	−0.61	−0.40	−0.18	−0.04	—
WaveNet2	1.04	−0.78	−0.24	−0.07	−0.20	0.02	−0.76	−0.63	−0.70	−0.45	−0.73	−0.52	−0.30	−0.16	−0.12

cycle. Furthermore, univariate methods are unable to improve the forecast performance across demand life cycle compared to the BL method ($p < 0.000$).

5.4 | Comparison of Forecasting Models Run Times

To understand forecasting efficiency, we first briefly evaluate the run time of each model. Table 10 clearly indicates that all the ML methods have significantly less run times compared to ARIMA, ETS, and ETSX. Importantly, BL is also not efficient compared to ML models. These findings are supported by similar results discussed by Makridakis et al. (2020).

6 | Discussion

6.1 | Findings

Retailers depend on reliable and accurate sales forecasts to manage their supply chain. However, the presence of sales promotions makes sales forecasting more challenging and complex. Yet, many retailers still use simple univariate methods supplemented by judgmental adjustments or BL correction to cope with promotional effects. It is typical for retailers to run various promotions for thousands of products across hundreds of stores at the same time. Therefore, retailers need an automated sales forecasting process to gain a competitive advantage.

Our study explores the applicability of ML methods in retail sales forecasts in the presence of promotions. We specifically focused on incorporating promotional periods into the models as this is a topic that has received little attention in the literature. Thus, the primary goal of our research is to evaluate the forecast performance of ML algorithms against existing methodologies in the retail setting across the demand life cycle.

TABLE 10 | Model run time in minutes.

Candidate model	Software	Run time in minutes
BL	R	516.5 + data cleansing time
sNAÏVE	R	7.2
ARIMA	R	31.1
ETS	R	642.2
ETSX	R	24.8
LGB1	Python	1.2
LGB2	Python	1.1
XGB1	Python	0.55
XGB2	Python	0.51
RF1	Python	2.3
RF2	Python	2.1
DeepAR1	Python	16.1
DeepAR2	Python	13.2
WaveNet1	Python	46.7
WaveNet2	Python	42.1

Note: Total number of unique time series: 3380. CPU: 2.3-GHz AMD Ryzen 5 4500U (six-core, 8-MB cache) and 8-GB RAM.

First, our findings reinforce previous research findings (Ali and Gürlek 2020; Huber and Stuckenschmidt 2020; Trapero et al. 2015) on the ability of multivariate models to automatically detect the postpromotional period. However, ML models require the additional variable as an input feature to determine the correct sign and magnitude of the postpromotional dip. Notably, DL methods did not identify the correct postpromotional dip even with the additional variable as an input.

Second, in normal periods, ML and DL models (with an additional variable) were able to outperform conventional univariate methods in normal periods. However, this finding is notably different from the previous literature. Ali et al. (2009) report that simple univariate methods perform similarly to advanced methods in the period without promotions. On the other hand, the BL method outperformed all the univariate methods in the normal period. This reinforces previous findings that when univariate algorithms are used with uncleansed sales data,[4] they frequently overestimate during normal periods (De Baets and Harvey 2018). Thus, our results suggest that ML methods can provide better results compared to univariate methods based on uncleansed sales data.

In terms of ETS and ETSX, we find that ETSX outperforms ETS throughout the promotional period due to the inclusion of the additional variable. However, this is not the case in the post-promotional period. This is interesting given that the inclusion of the additional variable should enhance ETSX. On the other hand, results show that ML methods improve the forecasting performance remarkably in both promotional and postpromotional periods compared to conventional univariate methods. Furthermore, adding the additional variable enhances the forecast performance of ML models only during the postpromotional period. Although the DL methods did not perform as expected, the inclusion of the additional variable improved forecast performance across the demand life cycle. This aligns with the previous findings that when advanced methods are used, more detailed inputs can improve the performance (Ali et al. 2009).

Third, our study compares all the forecasting methods with the BL method, a well-established retail implementation. Importantly, ML methods perform similarly to the BL method in *all* periods even though ML methods benefit from the additional variable. On the other hand, while the BL method generates significantly better forecasts compared to conventional univariate methods in all periods, it requires additional effort and time for data cleansing. This process can be time-consuming and prone to bias, as noted by Hewage et al. (2021) and Perera et al. (2019). Moreover, our brief encounter with the run time of each model provides evidence that although the BL method provides effective forecasts, it is not as efficient as ML methods. Additionally, it is important to note that the BL method is unable to identify multiple seasonalities and use information from multiple series since the base model is a univariate model. In contrast, ML methods can exploit information from multiple series and provide both efficient and effective forecasts.

Furthermore, Hewage et al. (2021) state that forecasters tend to apply an initial anchor when making adjustments to incorporate promotional effects to the base forecasts, even with the support of information guidance. Importantly, retailers are often required to generate sales forecasts for multiple products across multiple stores simultaneously, making it a manpower-intensive process (Fildes et al. 2019). This stresses the importance of an automated approach for retail sales forecasting. Therefore, we believe that ML approaches are a viable solution for retail sales forecasting because they can handle any SKU-store combination at the same time.

6.2 | Managerial Implications

Forecasting retail sales is essential for most managerial decisions made across the supply chain. In today's competitive market, many factors influence demand, making it volatile and unpredictable. However, most retailers still use simple methods supplemented by judgmental adjustments. Thus, managers need to invest a considerable amount of effort into retail sales forecasting in the presence of demand volatility. Our research suggests that using ML approaches can help automate the retail sales forecasting process. As a result, managers are no longer required to forecast future demand despite being informed of the underlying model and its implications. This will save both money and time for managers. They can use the time saved for other operational tasks. Furthermore, ML models coupled with a forecasting support system (FSS) can improve the quality of the decision-making process. Importantly, improvements in forecast performances will lead to increased operational profitability for retail stores.

6.3 | Limitations and Future Works

Clearly, our study is limited to the domain of our analysis, which comprises data from a US-based retailer for four product categories. Thus, it may not be generalizable to other product categories without appropriate customization. Our study only includes three types of promotions (i.e., temporary price reductions, display, and feature) instead of incorporating a variety of promotions. The dataset in our study primarily consists of weekly granularity rather than daily granularity. This limitation posed challenges in precisely aligning promotions with calendar events. Thus, we did not explore the impact of incorporating special days and holidays into our study. However, promotions are often associated with holiday events. Thus, it is important to understand how these major seasonal events affect the promotional life cycle of products. Therefore, how to incorporate other causal factors such as multiple promotional types, special days and events, and holidays might be an interesting future research avenue. Additionally, we did not see any intermittent demand patterns in our dataset. Thus, the proposed methodology may not work similarly in the presence of intermittent demand. We also did not consider the hierarchical structure of the sales forecasting problem. Thus, leveraging the hierarchical structure (e.g., store vs. category vs. product) and exploring hierarchical reconciliation of sales forecasts is a potential avenue for future work.

Retailers tend to apply human judgment in retail sales forecasting in the real world. Therefore, further research efforts are required to identify how to incorporate human judgment with advanced methods in the retail context. With enhancements to the current technology, this does not create a technical challenge for retailers. Subsequently, this outlines unexplored research avenues: (1) the ability of users to comprehend the implications of the various variables incorporated into ML methods and (2) their ability and capacity to make judgmental adjustments to forecasts in order to add value.

We did not employ any hyperparameter tuning or combination of methods. Our study also shows that no single model performs well for *all* periods. Thus, investigating how to identify appropriate forecast models in each period and how to combine them to create an integrated approach would be worthy of further investigation. Moreover, our study shows that sophisticated methods like DL methods can improve their forecasting performances by incorporating more detailed inputs. Thus, determining how and what feature inputs improve the performance of DL methods in the retail industry could be an interesting research question.

7 | Conclusion

Retail promotions create demand irregularities for products, making it difficult to generate accurate forecasts. Nonetheless, retailers generally forecast sales during promotional periods using either the BL method or human judgment. Retailers need to handle thousands of SKUs across multiple stores at any given time, underscoring the need for automated forecasting since the sheer volume of SKUs makes it redundant to use BL or judgmental approaches. Therefore, more advanced approaches are becoming relevant in retail sales forecasting due to these complexities. Furthermore, the need to improve decision-making in retail operations and the increasing availability of data has paved the way for such advanced methods.

In the context of promotions, our research reveals that ML methods are a robust alternative for retail sales forecasting. Our empirical study shows that ML methods have the capacity to incorporate causal factors with the sales history. Also, the inclusion of additional variables provides an additional improvement in the performance of ML methods. Unlike ML methods, the BL method necessitates more time and effort to cleanse the sales data. As a result, ML methods would enable retailers to reduce the time and effort required for sales forecasting and concentrate more time on other pain points in the supply chain.

Furthermore, with the availability of more data, advanced methods such as GBRT, RF, and DL methods continuously improve performance. This also provides the flexibility to process larger datasets with no restrictions on inputs. Thus, ML methods have the capacity to exploit similarities in time series across products and stores, increasing their effectiveness in the retail context dramatically. In sum, ML methods can deal with demand volatility caused by retail sales promotions while enhancing forecasting performance across the demand life cycle.

Acknowledgments

This research was supported by Grant ID SRC/LT/2020/20 provided by the Senate Research Committee of the University of Moratuwa, Sri Lanka.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Endnotes

[1] Sales forecasting is the process of estimating the number of future sales for a specific product or products (Hewage and Perera 2022b).

[2] Normal sales represent the number of sales without any sales promotions (Abraham and Lodish 1987).

[3] Dunnhumby source files: https://www.dunnhumby.com/source-files/.

[4] Raw sales data, which has not been treated to remove promotional effects to normalize the sales data.

References

Abolghasemi, M., E. Beh, G. Tarr, and R. Gerlach. 2020. "Demand Forecasting in Supply Chain: The Impact of Demand Volatility in the Presence of Promotion." *Computers and Industrial Engineering* 142: 106380. https://doi.org/10.1016/j.cie.2020.106380.

Abolghasemi, M., J. Hurley, A. Eshragh, and B. Fahimnia. 2020. "Demand Forecasting in the Presence of Systematic Events: Cases in Capturing Sales Promotions." *International Journal of Production Economics* 230: 107892. https://doi.org/10.1016/j.ijpe.2020.107892.

Abraham, M. M., and L. M. Lodish. 1987. "Promoter: An Automated Promotion Evaluation System." *Marketing Science* 6, no. 2: 101–123.

Ali, G. Ö., and R. Gürlek. 2020. "Automatic Interpretable Retail Forecasting With Promotional Scenarios." *International Journal of Forecasting* 36, no. 4: 1389–1406. https://doi.org/10.1016/j.ijforecast.2020.02.003.

Ali, Ö. G., S. Sayin, T. van Woensel, and J. Fransoo. 2009. "SKU Demand Forecasting in the Presence of Promotions." *Expert Systems With Applications* 36, no. 10: 12340–12348. https://doi.org/10.1016/j.eswa.2009.04.052.

Amazon Web Service. 2022. "GluonTS—Probabilistic Time Series Modeling." https://ts.gluon.ai/.

Aruchunarasa, B., and H. N. Perera. 2022. "Mitigating the Proclivity Towards Multiple Adjustments Through Innovative Forecasting Support Systems." In *Innovation Analytics Tools for Competitive Advantage*, edited by N. Subramanian, S. G. Ponnambalam, and M. Janardhanan, 1–15. World Scientific Publishing Co Pte Ltd. https://doi.org/10.1142/q0293.

Baek, Y. 2019. "Forecasting in Long Horizons Using Smoothed Direct Forecast." *Journal of Forecasting* 38, no. 4: 277–292. https://doi.org/10.1002/for.2572.

Bandara, K., C. Bergmeir, and S. Smyl. 2020. "Forecasting Across Time Series Databases Using Recurrent Neural Networks on Groups of Similar Series: A Clustering Approach." *Expert Systems With Applications* 140: 112896. https://doi.org/10.1016/j.eswa.2019.112896.

Bandara, K., P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, and B. Seaman. 2019. "Sales Demand Forecast in E-Commerce Using a Long Short-Term Memory Neural Network Methodology." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11955, 462–474. Springer. https://doi.org/10.1007/978-3-030-36718-3_39.

Blattberg, R. C., and R. A. Briesch. 2012. "Sales Promotions." In *The Oxford Handbook of Pricing Management*. Oxford Academic. https://doi.org/10.1093/oxfordhb/9780199543175.013.0024.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45, no. 1: 5–32. https://doi.org/10.1023/A:1010933404324.

Brinke, A., L. Fadejeva, B. Siliverstovs, and K. Vilerts. 2023. "Assessing the Informational Content of Card Transactions for Nowcasting Retail Trade: Evidence for Latvia." *Journal of Forecasting* 42, no. 3: 566–577. https://doi.org/10.1002/for.2945.

Chybalski, F. 2017. "Forecast Value Added (FVA) Analysis as a Means to Improve the Efficiency of a Forecasting Process." *Operations Research and Decisions* 27, no. 1: 5–19. https://doi.org/10.5277/ord170101.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. J. Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess." *Journal of Official Statistics* 6: 3–73.

Cohen, M. C., J. J. Kalas, and G. Perakis. 2021. "Promotion Optimization for Multiple Items in Supermarkets." *Management Science* 67, no. 4: 2340–2364. https://doi.org/10.1287/mnsc.2020.3641.

Cooper, L. G., P. Baron, W. Levy, M. Swisher, and P. Gogos. 1999. "PromoCast: A New Forecasting Method for Promotion Planning." *Marketing Science* 18, no. 3: 301–316. https://doi.org/10.1287/mksc.18.3.301.

De Baets, S., and N. Harvey. 2018. "Forecasting From Time Series Subject to Sporadic Perturbations: Effectiveness of Different Types of Forecasting Support." *International Journal of Forecasting* 34, no. 2: 163–180. https://doi.org/10.1016/j.ijforecast.2017.09.007.

DelVecchio, D., D. H. Henard, and T. H. Freling. 2006. "The Effect of Sales Promotion on Post-Promotion Brand Preference: A Meta-Analysis." *Journal of Retailing* 82, no. 3: 203–213. https://doi.org/10.1016/j.jretai.2005.10.001.

Divakar, S., B. T. Ratchford, and V. Shankar. 2005. "Practice Prize Article—CHAN4CAST: A Multichannel, Multiregion Sales Forecasting Model and Decision Support System for Consumer Packaged Goods." *Marketing Science* 24, no. 3: 334–350. https://doi.org/10.1287/mksc.1050.0135.

Fildes, R., P. Goodwin, M. Lawrence, and K. Nikolopoulos. 2009. "Effective Forecasting and Judgmental Adjustments: An Empirical Evaluation and Strategies for Improvement in Supply-Chain Planning." *International Journal of Forecasting* 25, no. 1: 3–23. https://doi.org/10.1016/j.ijforecast.2008.11.010.

Fildes, R., P. Goodwin, and D. Önkal. 2018. "Use and Misuse of Information in Supply Chain Forecasting of Promotion Effects." *International Journal of Forecasting* 35, no. 1: 144–156. https://doi.org/10.1016/j.ijforecast.2017.12.006.

Fildes, R., S. Ma, and S. Kolassa. 2019. "Retail Forecasting: Research and Practice." *International Journal of Forecasting* 38, no. 4: 1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004.

Fildes, R., K. Nikolopoulos, S. F. Crone, and A. A. Syntetos. 2008. "Forecasting and Operational Research: A Review." *Journal of the Operational Research Society* 59, no. 9: 1150–1172. https://doi.org/10.1057/palgrave.jors.2602597.

Friedman, M. 1940. "A Comparison of Alternative Tests of Significance for the Problem of *m* Rankings." *Annals of Mathematical Statistics* 11: 86–92.

Godahewa, R., K. Bandara, G. I. Webb, S. Smyl, and C. Bergmeir. 2020. "Ensembles of Localized Models for Time Series Forecasting." Preprint, arXiv, September 21. http://arxiv.org/abs/2012.15059.

Goodwin, P. 2000. "Improving the Voluntary Integration of Statistical Forecasts and Judgment." *International Journal of Forecasting* 16, no. 1: 85–99. https://doi.org/10.1016/S0169-2070(99)00026-6.

Guo, Z. X., W. K. Wong, and M. Li. 2013. “A Multivariate Intelligent Decision-Making Model for Retail Sales Forecasting.” *Decision Support Systems* 55, no. 1: 247–255. https://doi.org/10.1016/j.dss.2013.01.026.

Gür Ali, Ö. 2013. “Driver Moderator Method for Retail Sales Prediction.” *International Journal of Information Technology and Decision Making* 12, no. 6: 1261–1286. https://doi.org/10.1142/S0219622013500363.

Hanssens, D. M. 1998. “Order Forecasts, Retail Sales, and the Marketing Mix for Consumer Durables.” *Journal of Forecasting* 17, no. 3–4: 327–346. https://doi.org/10.1002/(sici)1099-131x(199806/07)17:3/4<327::aid-for699>3.0.co;2-q.

Hewage, H. C., and H. N. Perera. 2022a. “Comparing Statistical and Machine Learning Methods for Sales Forecasting During the Post-Promotional Period.” In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 462–466. IEEE. https://doi.org/10.1109/ieem50564.2021.9672954.

Hewage, H. C., and H. N. Perera. 2022b. “Retail Sales Forecasting in the Presence of Promotional Periods.” In *Lecture Notes in Networks and Systems*, vol. 364. Springer International Publishing. https://doi.org/10.1007/978-3-030-92604-5_10.

Hewage, H. C., H. N. Perera, and S. De Baets. 2021. “Forecast Adjustments During Post-Promotional Periods.” *European Journal of Operational Research* 300, no. 2: 461–472. https://doi.org/10.1016/J.EJOR.2021.07.057.

Hewamalage, H., C. Bergmeir, and K. Bandara. 2021. “Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions.” *International Journal of Forecasting* 37, no. 1: 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008.

Huang, T., R. Fildes, and D. Soopramanien. 2014. “The Value of Competitive Information in Forecasting FMCG Retail Product Sales and the Variable Selection Problem.” *European Journal of Operational Research* 237, no. 2: 738–748. https://doi.org/10.1016/j.ejor.2014.02.022.

Huang, T., R. Fildes, and D. Soopramanien. 2019. “Forecasting Retailer Product Sales in the Presence of Structural Change.” *European Journal of Operational Research* 279, no. 2: 459–470. https://doi.org/10.1016/j.ejor.2019.06.011.

Huber, J., and H. Stuckenschmidt. 2020. “Daily Retail Demand Forecasting Using Machine Learning With Emphasis on Calendric Special Days.” *International Journal of Forecasting* 36, no. 4: 1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005.

Hübner, A., P. Amorim, H. Kuhn, S. Minner, and T. Van Woensel. 2018. “Retail Operations.” *OR Spectrum* 40, no. 4: 831–835. https://doi.org/10.1007/s00291-018-0535-1.

Hyndman, R. J., and G. Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3rd ed. OTexts. OTexts.com/fpp3.

Hyndman, R. J., and Y. Khandakar. 2008. “Automatic Time Series Forecasting: The Forecast Package for R.” *Journal of Statistical Software* 26, no. 3: 1–22. https://doi.org/10.18637/jss.v027.i03.

Hyndman, R. J., and A. B. Koehler. 2006. “Another Look at Measures of Forecast Accuracy.” *International Journal of Forecasting* 22, no. 4: 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Lee, W. Y., P. Goodwin, R. Fildes, K. Nikolopoulos, and M. Lawrence. 2007. “Providing Support for the Use of Analogies in Demand Forecasting Tasks.” *International Journal of Forecasting* 23, no. 3: 377–390. https://doi.org/10.1016/j.ijforecast.2007.02.006.

Leeflang, P. S. H., H. J. van Heerde, and D. R. Wittink. 2005. “How Promotions Work: SCAN*PRO-Based Evolutionary Model Building.” *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.321003.

Loureiro, A. L. D., V. L. Miguéis, and L. F. M. da Silva. 2018. “Exploring the Use of Deep Neural Networks for Sales Forecasting in Fashion Retail.” *Decision Support Systems* 114: 81–93. https://doi.org/10.1016/j.dss.2018.08.010.

Ma, S., and R. Fildes. 2017. “A Retail Store SKU Promotions Optimization Model for Category Multi-Period Profit Maximization.” *European Journal of Operational Research* 260, no. 2: 680–692. https://doi.org/10.1016/j.ejor.2016.12.032.

Ma, S., and R. Fildes. 2021. “Retail Sales Forecasting With Meta-Learning.” *European Journal of Operational Research* 288, no. 1: 111–128. https://doi.org/10.1016/j.ejor.2020.05.038.

Ma, S., R. Fildes, and T. Huang. 2016. “Demand Forecasting With High Dimensional Data: The Case of SKU Retail Sales Forecasting With Intra- and Inter-Category Promotional Information.” *European Journal of Operational Research* 249, no. 1: 245–257. https://doi.org/10.1016/j.ejor.2015.08.029.

Macé, S., and S. A. Neslin. 2004. “The Determinants of Pre- and Postpromotion Dips in Sales of Frequently Purchased Goods.” *Journal of Marketing Research* 41, no. 3: 339–350. https://doi.org/10.1509/jmkr.41.3.339.35992.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos. 2018. “Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward.” *PLoS ONE* 13, no. 3: 1–26. https://doi.org/10.1371/journal.pone.0194889.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos. 2020. “The M4 Competition: 100,000 Time Series and 61 Forecasting Methods.” *International Journal of Forecasting* 36, no. 1: 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014.

Microsoft Corporation. 2022. “LightGBM (v3.3.3).” https://github.com/Microsoft/LightGBM.

Mou, S., D. J. Robb, and N. DeHoratius. 2018. “Retail Store Operations: Literature Review and Research Directions.” *European Journal of Operational Research* 265, no. 2: 399–422. https://doi.org/10.1016/j.ejor.2017.07.003.

Perera, H. N., J. Hurley, B. Fahimnia, and M. Reisi. 2019. “The Human Factor in Supply Chain Forecasting: A Systematic Review.” *European Journal of Operational Research* 274, no. 2: 574–600. https://doi.org/10.1016/j.ejor.2018.10.028.

Petropoulos, F., and I. Svetunkov. 2020. “A Simple Combination of Univariate Models.” *International Journal of Forecasting* 36, no. 1: 110–115. https://doi.org/10.1016/j.ijforecast.2019.01.006.

Salinas, D., V. Flunkert, J. Gasthaus, and T. Januschowski. 2020. “DeepAR: Probabilistic Forecasting With Autoregressive Recurrent Networks.” *International Journal of Forecasting* 36, no. 3: 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001.

scikit-learn Developers. 2022. “3.2.4.3.2. sklearn.ensemble.RandomForestRegressor (3.2.4.3.2.).” https://scikit-learn.org/0.16/modules/generated/sklearn.ensemble.RandomForestRegressor.html.

Spiliotis, E., S. Makridakis, A. A. Semenoglou, and V. Assimakopoulos. 2020. “Comparison of Statistical and Machine Learning Methods for Daily SKU Demand Forecasting.” *Operational Research* 22: 3037–3061. https://doi.org/10.1007/s12351-020-00605-2.

Sprangers, O., S. Schelter, and M. de Rijke. 2022. “Parameter-Efficient Deep Probabilistic Forecasting.” *International Journal of Forecasting* 39, no. 1: 332–345. https://doi.org/10.1016/j.ijforecast.2021.11.011.

Trapero, J. R., N. Kourentzes, and R. Fildes. 2015. “On the Identification of Sales Forecasting Models in the Presence of Promotions.” *Journal of the Operational Research Society* 66, no. 2: 299–307. https://doi.org/10.1057/jors.2013.174.

Trapero, J. R., D. J. Pedregal, R. Fildes, and N. Kourentzes. 2013. “Analysis of Judgmental Adjustments in the Presence of Promotions.” *International Journal of Forecasting* 29, no. 2: 234–243. https://doi.org/10.1016/j.ijforecast.2012.10.002.

Vallés-Pérez, I., E. Soria-Olivas, M. Martínez-Sober, A. J. Serrano-López, J. Gómez-Sanchís, and F. Mateo. 2022. “Approaching Sales Forecasting Using Recurrent Neural Networks and Transformers.”

Expert Systems With Applications 201: 116993. https://doi.org/10.1016/j.eswa.2022.116993.

van Steenbergen, R. M., and M. R. K. Mes. 2020. "Forecasting Demand Profiles of New Products." *Decision Support Systems* 139: 113401. https://doi.org/10.1016/j.dss.2020.113401.

Wang, X. S., J. H. Ryoo, N. Bendle, and P. K. Kopalle. 2020. "The Role of Machine Learning Analytics and Metrics in Retailing Research." *Journal of Retailing* 97, no. 4: 658–675. https://doi.org/10.1016/j.jretai.2020.12.001.

Wilcoxon, F., and R. A. Wilcox. 1964. *Some Rapid Approximate Statistical Procedures*. Lederle Laboratories.

xgboost Developers. 2021. "XGBoost Python Package (1.5.0)." https://xgboost.readthedocs.io/en/stable/python/index.html.