RETAIL SALES FORECASTING IN THE PRESENCE OF PROMOTIONS: COMPARISON OF STATISTICAL AND MACHINE LEARNING FORECASTING METHODS

HALGAMUWE HEWAGE HARSHA RUWAN CHAMARA

208068F

Degree of Master of Science

Department of Transport and Logistics Management

University of Moratuwa

Sri Lanka

July 2022

RETAIL SALES FORECASTING IN THE PRESENCE OF PROMOTIONS: COMPARISON OF STATISTICAL AND MACHINE LEARNING FORECASTING METHODS

HALGAMUWE HEWAGE HARSHA RUWAN CHAMARA

208068F

Thesis/Dissertation submitted in partial fulfilment of the requirements for the degree of Master of Science in Supply Chain and Data Science

Department of Transport and Logistics Management

University of Moratuwa

Sri Lanka

July 2022

DECLARATION OF ORIGINALITY

I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

I hereby also grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or another medium. I retain the right to use this content in whole or part in future works (such as articles or books).

UOM Verified Signature

Signature: …………………..   Date: 08/07/2022 …………………..

(H.H.H.R. Chamara)

STATEMENT OF THE SUPERVISOR

The above candidate has carried out research for the Degree of Master of Science under my supervision.

Signature of the Supervisor: UOM Verified Signature Date: 08/07/2022…………………..

Dr. H.N. Perera,
Senior Lecturer,
Dept. of Transport & Logistics Management,
Faculty of Engineering,
University of Moratuwa.

Abstract

Retail sales forecasting is the process of estimating the number of future sales for a specific product or products. However, producing reliable and accurate sales forecasts at a product level is a very challenging task in the retail context. Many factors can influence observed sales data at the product level, such as sales promotions, weather, holidays, and special events, all of which causes demand irregularities. Sales promotions are one of the salient drivers in generating irregular sales patterns. Sales promotions confound retail operations, causing sudden demand changes not just during the promotion period, but also throughout the demand series. As a result, three types of periods are relevant for sales promotions: normal, promotional, and a post-promotional. However, previous research has mostly focused on promotional and normal (i.e., non-promotional) periods, often neglecting the post-promotional period. To address this gap, we explore the performance of comprehensive methods, namely gradient-boosted regression trees, random forests, and deep learning in all periods. Moreover, we compare proposed approaches with conventional forecasting approaches in a retail setting. Our results demonstrate that machine learning methods can deal with demand fluctuations generated by retail promotions while enhancing forecast performance throughout all time periods. The base-lift model outperformed machine learning methods, although with more effort necessary to cleanse sales data. Our findings indicate that machine learning methods can automate the forecasting process and provide significant performance even with the standard approach. Hence, our research demonstrates the way retailers can successfully apply machine learning methods in forecasting sales.

Keywords: Forecasting, Promotions, Retail supply chain, Post-promotional effect, Machine learning

ACKNOWLEDGMENTS

I would like to take this opportunity pen my appreciation and heartfelt gratitude heartfelt gratitude to each and every one of them.

The first person I am extremely grateful and indebted to is Dr. H. Niles Perera, my supervisor and mentor. During the last two years, he gave me his continued guidance and support and also provided me with lots of freedom. Your kind support has been priceless to me. I am extremely grateful to my external advisor Dr. Kasun Bandara of The University of Melbourne, Australia for his guidance and support on forecasting methodologies.

I would also wish to extend my sincere gratitude to Dr. T. Sivakumar, postgraduate research coordinator of the Department of Transport and Logistics Management, for his guidance and support during the degree program progress reviews. I also wish to express my heartiest gratitude to Dr. Charith Chitraranjan for his extensive evaluations and feedback during the progress reviews.

I am immensely grateful to Senior Prof. Amal S. Kumarage, former Head and Founder of the Department of Transport and Logistics Management at the University of Moratuwa, and Prof. A.A.D.A.J. Perera, the current Head of the Department, for providing me with the opportunity to pursue the degree program.

I am extremely grateful to Dr. Amila Thibbotuwawa for evaluating my work and providing feedback to further improve it. My special thanks are due to the Center for Supply Chain Operations, and Logistics Optimization (SCOLO) fraternity for always being there to discuss anything and everything. I would also like to extend my gratitude towards academic and non-academic staff members at the Department of Transport and Logistics Management, for their support. I am thankful to The Senate Research Committee of the University of Moratuwa and the grant bearing ID SRC/LT/2020/20 which enabled this thesis and must be mentioned with immense gratitude.

My family has been the extremely encouraging. Last but not least, I'd like to express my heartfelt gratitude to my parents and sister, whose love and wisdom are with me in whatever I do. I could not have done this without you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
BL	Base-Lift
BT	Boosted Trees
DL	Deep Learning
ETS	Exponential Smoothing
ETSX	Exponential Smoothing with Exogenous Variable
FSS	Forecasting Support System
FVA	Forecast Value Added
GBRT	Gradient-Boosted Regression Trees
LBG	LightGBM
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
ML	Machine Learning
SNAIVE	Seasonal NAIVE
NN	Neural Network
RF	Random Forest
RT	Regression Trees
SKU	Stock Keeping Unit
sMAPE	Symmetric Mean Absolute Percentage Error
SVR	Support Vector Machines
TPR	Temporary Price Reductions
XGB	xgBoost

1 INTRODUCTION

The term "retail" is derived from the Old French verb "tailer" which means "to cut off, trim, pare, or divide in terms of tailoring" (Abraham & Lodish, 1987). It was first used as a noun in 1433, and it meant "a sale of small quantities of products to customers" (Abraham & Lodish, 1987). A retail supply chain is made up of suppliers, manufacturers, intermediaries and retailers who collaborate to meet customer demand (Fildes et al., 2019). The retail industry is undergoing a dramatic shift with the competitive climate that businesses face today (Fildes et al., 2019). Thus, retailers must make operational decisions in the face of changing competitive and technological landscapes to gain operational efficiency (Fildes et al., 2019).

Many factors contribute to the difficulties and complications that retail operations confront, such as shifting customer expectations, promotional activities, partner activities, and shorter lead times (Hewage & Perera, 2022a; Ma et al., 2016; Ma & Fildes, 2021). Among those factors, retail sales promotions make retail sales forecasting challenging (Hewage et al., 2021). Generally, retail promotions raise demand for a product during promotional periods (Fildes et al., 2019). This rise in sales for the promoted product could come at the expense of other product sales or sales of the same product during other time periods (Blattberg & Briesch, 2012). Following the promotional period, sales may drop below normal levels before recovering, resulting in a post-promotional dip (Hewage et al., 2021). As a result, promotions have an impact on demand not just during the promotional period, but also across the demand period (Macé & Neslin, 2004). Figure 1-1 depicts the demand variations during a retail sales promotion.

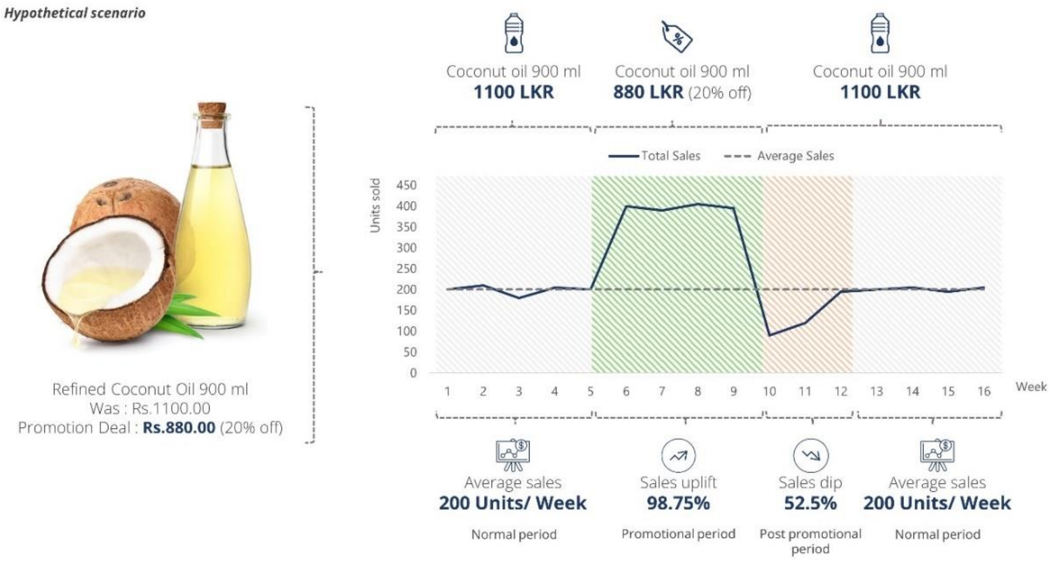

Figure 1-1 *Variations in demand in retail sales promotions*

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 to forecast sales, often with judgmental adjustments to account for promotional effects (Mou et al., 2018). These judgemental approaches are manpower-intensive because a typical retail store carries thousands of products across many store locations (Fildes et al., 2019). Hence, ML based approaches are a potential option for retailers looking to automate the retail sales 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 promotional and non-promotional periods, regardless of the post-promotional period (e.g., Abolghasemi et al., 2020; Ali et al., 2009; Cohen et al., 2020; Ali & Gürlek, 2020;

Huber & Stuckenschmidt, 2020; Ma & Fildes, 2017; Ramanathan & Muyldermans, 2011; van Steenbergen & Mes, 2020)). As a result, we explore the applicability of ML methods in the presence of promotions, considering *all* periods: normal, promotional, and post-promotional.

Our study is concerned with the comparison of the forecast performances of conventional univariate methods and ML methods. Therefore, our study revolves around the following research questions which are methodically derived through the literature. Further details on the derivation are available under Section 2: Literature Review.

RQ1*: How can the retailers leverage ML techniques to predict retail promotion forecasts?*

RQ2: *Are ML techniques beneficial only when using more detailed inputs (promotion types, category effects, etc.) compared to statistical methods?*

RQ3: *Are ML-driven methods for retail sales forecasting a feasible alternative to conventional retail forecasting approaches?*

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

Our study makes the following contributions. First, we examine whether ML methods can detect the post-promotional dip automatically. The empirical results based on the data show that ML methods can detect the post-promotional period automatically. However, additional inputs need to be provided to determine the correct size of the post-promotional dip. Second, we focus on whether incorporating more causal features improve the performance of ML methods. Third, our study provides an extensive comparison of the performance between ML methods and conventional forecasting approaches in a retail context. Our findings suggest that ML approaches can automate the forecasting process whilst providing significant performance even with the standard approach.

This thesis is organized 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 focuses on the discussion and conclusion respectively.

2 LITERATURE REVIEW

Supply Chain Management (SCM) encompasses all of an organization's sourcing, procurement, manufacturing, and logistics management activities (Perera et al., 2019). Coordination and collaboration with key stakeholders like suppliers, manufacturers, third-party service providers, intermediaries and customers is a crucial component of SCM (Perera et al., 2019). The retail industry can be described as a complex supply chain comprised of numerous stakeholders (Ali et al., 2009). The retail industry is a key economic sector that creates market dynamics in various stages of the retail supply chain (Fildes et al., 2019).

2.1 Retail Supply Chain

A retail supply chain is made up of retailers, suppliers, manufacturers, and other intermediaries who collaborate to meet customer demand (Fildes et al., 2019). This entails the movement of products, money, and information along the retail supply chain (Huang et al., 2019). Among which, the role of information is crucial for retail operations to ensure the desired level of competitiveness (De Baets & Harvey, 2016; Perera & Perera, 2022). However, it is becoming challenging due to various uncertainties and challenges. These uncertainties and challenges arise as a result of shifting customer expectations, competitor actions, partner activities, promotional activities, shorter lead times, and emerging technologies (Hewage et al., 2021; Ma et al., 2016; Ma & Fildes, 2021).

All of these factors lead to a volatile retail supply chain. Even a small improvement in operational decisions allows retailers to maintain their operations at a competitive level (Hübner et al., 2018; Ma & Fildes, 2017). Demand forecasting is a salient and comprehensive operational task in retail operations (Ali & Gürlek, 2020). Because retailers must manage their demand and supply planning procedures properly in order to minimize customer service concerns, excess inventory, and excessive costs due to obsolete products (Huang et al., 2019).

2.2 Retail Sales Promotions

Retail sales promotions are a prominent form of marketing activity in the retail industry (Blattberg & Briesch, 2012). Typically, sales promotions induce the behaviour of customers to make a purchase during the promotional period (Abraham & Lodish, 1987). Retailers use various types of price promotions such as coupons, multi-buys, or temporary price reductions (TPR) coupled with non-price promotions such as displays, features, and point of sales materials (Blattberg & Briesch, 2012). Retailers generally have multiple goals in mind when it comes to sales promotions. One goal is to boost consumer traffic. The second goal is to sell extra inventory produced by overstocking. The third goal is to increase sales of a new product category (Blattberg & Briesch, 2012; DelVecchio et al., 2006; Fildes et al., 2019).

Retail sales promotions cause demand volatility not just during the promotion period, but also throughout the demand series (Abolghasemi et al., 2020). Normally, a sales uplift can be found during promotion periods. This increase in sales is usually the result of customers changing their buying patterns, either through purchase acceleration or higher consumption (Blattberg & Briesch, 2012) since customers tend to stockpile products during sales promotions for future consumption (Hewage et al., 2021). This often leads to lower sales figures than the baseline (normal) level[1] 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 & Lodish, 1987). This period of reduced sales is known as the post-promotional period (Hewage et al., 2021). Hence, a retail sales promotion has three phases: the normal period, the promotional period, and the post-promotional period (Hewage et al., 2021), creating different demand variations in each period (DelVecchio et al., 2006).

Forecasting retail sales in the context of promotions can be difficult for a variety of reasons (Fildes et al., 2018). It is common for retailers to have thousands of products across hundreds of stores being promoted simultaneously (Cohen et al., 2020).

[1] Normal sales represent the number of sales without any sales promotions (Hewage 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 (Ma et al., 2016). As a result, there is no standardized method for coping with changes in demand caused by retail promotions (Fildes et al., 2019).

2.3 Supply Chain Forecasting

2.3.1 Demand forecasting

Demand forecasting is a fundamental activity in retail since it is a primary input into many operational decisions, including sourcing, procurement, logistics, production planning, marketing, and financial decisions (Abolghasemi et al., 2020; Huber & Stuckenschmidt, 2020). The common practice of the retailers for demand forecasting is to anticipate future demand using historical sales data (Feiler et al., 2013). Overall sales, on the other hand, may not fully reflect actual customer demand (Perera et al., 2019).

In many retail stores, unsatisfied customer demand is either not observed or lost when stockout situations occur (i.e., censored demand) (Mou et al., 2018). Thus, actual customer demand is higher than the sales in stockout periods (Feiler et al., 2013). However, recent literature shows demand forecasting tools generally use censored demand data to produce retail sales forecasts (Hewage & Perera, 2022a; Tong et al., 2018).

2.3.2 Retail sales forecasting

Retail sales forecasting can be defined as the estimation of the number of future sales for a specific product or products (Hewage & Perera, 2022b). Retailers 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). However, retail sales forecasting process is highly complex since retailers need to manage a wide range of products within a limited shelf space (Mou et al., 2018).

Inaccurate sales forecasts often result in stock-outs or high stock levels that are prone to obsolescence (Huang et al., 2019). Customers may become dissatisfied if stock-outs occur frequently. Eventually, customers switching to other retail outlets (Ma et al., 2016). Thus, retailers tend to maintain a buffer stock to ensure customer satisfaction. This ultimately leads to higher inventory costs and reduced profits (Ma & Fildes, 2017; Perera et al., 2019).

However, producing reliable and accurate sales forecasts in product level is a very challenging task in the retail context (Ali & Gürlek, 2020; Trapero et al., 2015). Many factors can influence observed sales data at the product level, such as sales promotions, weather, holidays, and special events, causing demand irregularities (Fildes et al., 2019; Huang et al., 2019). Sales promotions are one of the salient factors in creating irregular sales patterns among them (Bandara et al., 2019).

2.4 Retail Sales Forecasting Methods

2.4.1 Human factor in retail sales forecasting

In practice, many retailers still use simple univariate methods supplemented by judgmental adjustments or base lift correction to cope with promotional effects (Fildes et al., 2019). Fildes & Goodwin (2007) found that 67% of cases involve judgemental adjustments in their survey. Similar results are also elucidated in past literature (e.g., Brau et al., 2019; Fildes & Petropoulos, 2015; Franses & Legerstee, 2013). Judgmental adjustments or base lift corrections have two steps: (1) determining whether statistical forecasts need to be adjusted and (2) determining the direction and magnitude of the adjustment (Arvan et al., 2018).

Evidence suggests that retailers use simple univariate methods to produce base forecasts using only past sales history (Fildes et al., 2019; Petropoulos et al., 2022). Therefore, univariate methods might not be able to incorporate promotional periods into the sales forecasts (Abolghasemi et al., 2020). The most widely used univariate methods in the retail industry are simple moving averages, exponential smoothing and its extensions, or Auto Regressive Integrated Moving Average (ARIMA) approaches (Fildes et al., 2019; Perera et al., 2019; Hyndman & Khandakar, 2008). Thereafter, retailers use their experience and domain knowledge to adjust the statistical forecast to include promotional effects (Trapero et al., 2015).

Past literature shows judgmental adjustments can improve the statistical forecasts in the presence of sales promotions (Fildes et al., 2009; Hewage et al., 2021). However, Alvarado-Valencia et al. (2017) argue that the success of these methods depends on the experience of the retailers. Evidence also reveals that forecasters make unwarranted adjustments even when there is insufficient information available (Fildes et al., 2009; Lawrence et al., 2006). Therefore, forecasting retail sales in promotions using univariate models with judgmental adjustments may result in systemic errors (Hewage et al., 2021). Thus, these forecasts can be inaccurate, costly, and inconsistent due to bias (Baecke et al., 2017).

2.4.2 Incorporating sales promotions in retail sales forecasting

Retailers are frequently required to generate a large number of sales forecasts for multiple products across multiple stores at the same time (Ali, 2013). On the other hand, retailers need to generate sales forecasts at the SKU level, incorporating factors such as historical sales, product attributes, promotional attributes, and store information (Ramanathan & Muyldermans, 2010). Furthermore, these sales forecasts often need to be made daily, weekly, and sometimes monthly (Fildes et al., 2006). Thus, the univariate methods based on human judgment may restrict the scale of retail sales forecasts (Hewage & Perera, 2022a).

In contrast, causal methods are capable of incorporating sales promotions into forecasts without any judgmental interference (Trapero et al., 2015). 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 PromoCast (Cooper et al., 1999), SCAN*PRO (Leeflang et al., 2005), CHAN4CAST (Divakar et al., 2005) and Driver Moderator (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).

On the other hand, unstructured methods such as ML can use past sales and causal variables with lags as input to provide forecasts during promotional periods (Ali & 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 (SVR), Regression Trees (RT), Artificial Neural Networks (ANN), and Boosted Trees (BT) (Perera et al., 2019; Petropoulos et al., 2022). 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 high predicted accuracy and flexibility when there is a large amount of data (Ali & Gürlek, 2020). Furthermore, the results of a recent M5 competition on Kaggle shows the potential of ML in retail forecasting tasks (Spiliotis et al., 2020).

Previous literature reveals that ML approaches often enhance forecast accuracy when compared to linear models in the context of retail sales promotions (Fildes et al., 2019). As 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 & Stuckenschmidt (2020) reports that ML methods including ANN and BT provide more accurate forecasts suitable for large scale demand forecasting scenarios. Abolghasemi et al. (2020) shows that the SVR model generates robust forecasts in the presence of promotions. Aburto & Weber (2007) proposed a hybrid method combining the ARIMA model with a Neural Network (NN) model. They use NN models to estimate the promotional uplift and combine it with ARIMA model forecasts. Further, Ma & Fildes (2021) developed a meta-learning framework based on deep convolutional neural networks to produce SKU level forecasting in retail setting.

2.5 Problem Description

2.5.1 Research problem derivation

Nevertheless, there are only a few studies focused on SKU level sales forecasting using ML methods in the area of sales forecasting. Majority of those employed NN methods

(Spiliotis et al., 2020). Moreover, all the studies have focused only on the promotional and non-promotional (i.e., normal) periods, regardless of the post-promotional period (e.g., Abolghasemi et al., 2020; Cohen et al., 2020; Ali & Gürlek, 2020; Huber & Stuckenschmidt, 2020; Ma & Fildes, 2017; Ramanathan & Muyldermans, 2011; van Steenbergen & Mes, 2020). As a result, the true benefits, and limitations of incorporating different types of promotional periods with ML approaches are yet to be explored. Hence, we aim to see if ML approaches are a feasible option for forecasting retail sales in the context of promotions. In addition, we compare the ML approaches to widely used univariate methods to assess their relative performance.

2.5.2 Hypothesis development

Promotions are the main reason for incorporating judgmental adjustments into retail sales forecasting (Aruchunarasa & Perera, 2022; Perera et al., 2019). However, practitioners tend to ignore quantitative forecasts altogether when making adjustments to tackle promotion 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 post-promotional period and treat it as a normal period. Thus, they frequently make inappropriate adjustments that impair forecast accuracy during promotional periods (De Baets & Harvey, 2018). In contrary, Trapero et al. (2015) found that the Dynamic Regression Model is capable of detecting the post-promotional period automatically. Ali and Gürlek (2020) also state that the FAIR model identifies the post-promotional dip. Yet, the post-promotional period was not incorporated into these models as an input feature. Therefore, we hypothesize:

Hypothesis 1. *ML methods can automatically recognize the post-promotional period and the correct magnitude of the post-promotional 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:

Hypothesis 2. *ML methods improve forecast performance in all periods when promotional periods are included as an additional variable.*

Specifically, Huber and Stuckenschmidt (2020) state that it is uncertain whether ML methods can outperform conventional approaches in retail sales forecasting. Also, 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:

Hypothesis 3*. In the retail setting, ML methods outperform conventional forecasting methods across all periods.*

3 METHODOLOGY

We discussed some of the major retail sales forecasting methods in the previous section, as well as how sales promotions can substantially influence customer purchasing behaviour. We realise that retailers require accurate and reliable SKU level sales forecasting (Ramanathan & Muyldermans, 2011), as it is the primary operational unit for retail operations (Fildes et al., 2019). Thus, we focus on SKU level sales forecasting in this study. We increase the model scope to incorporate features from multiple SKU-store combinations along with the promotional periods[2].

3.1 Data and Input Features

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. The dataset spans over 156 weeks and was collected from a leading US-based retailer. Table 3-1 shows the descriptive statistics of the collected dataset. Figure 3-1 shows the category distribution of the dataset and Figure 3-2 shows the weekly sales by category of the collected dataset.

Table 3-1 *Descriptive summary of the dataset*

	# of SKUs	Weekly sales					
		Normal		Promotional		Post-promotional	
		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

[2] The promotional period term refers to the three promotional periods; the normal period, the promotional period, and the post-promotional period (Hewage et al., 2021).

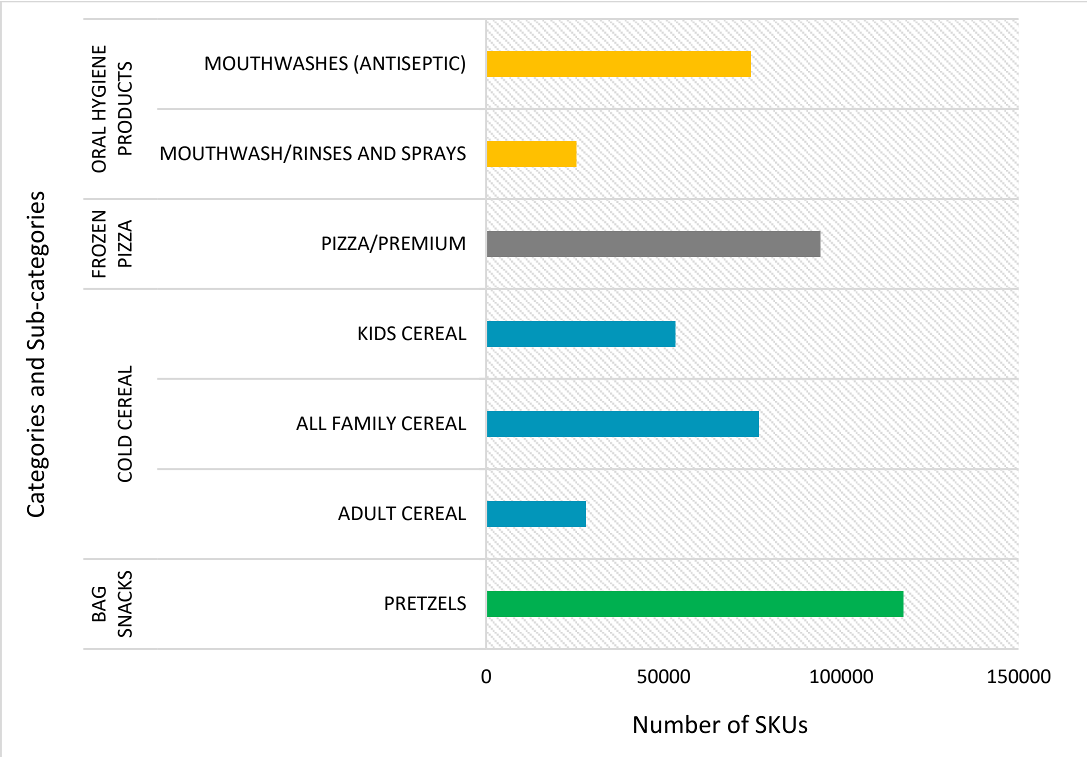

Figure 3-1 *Category distribution of the SKUs*

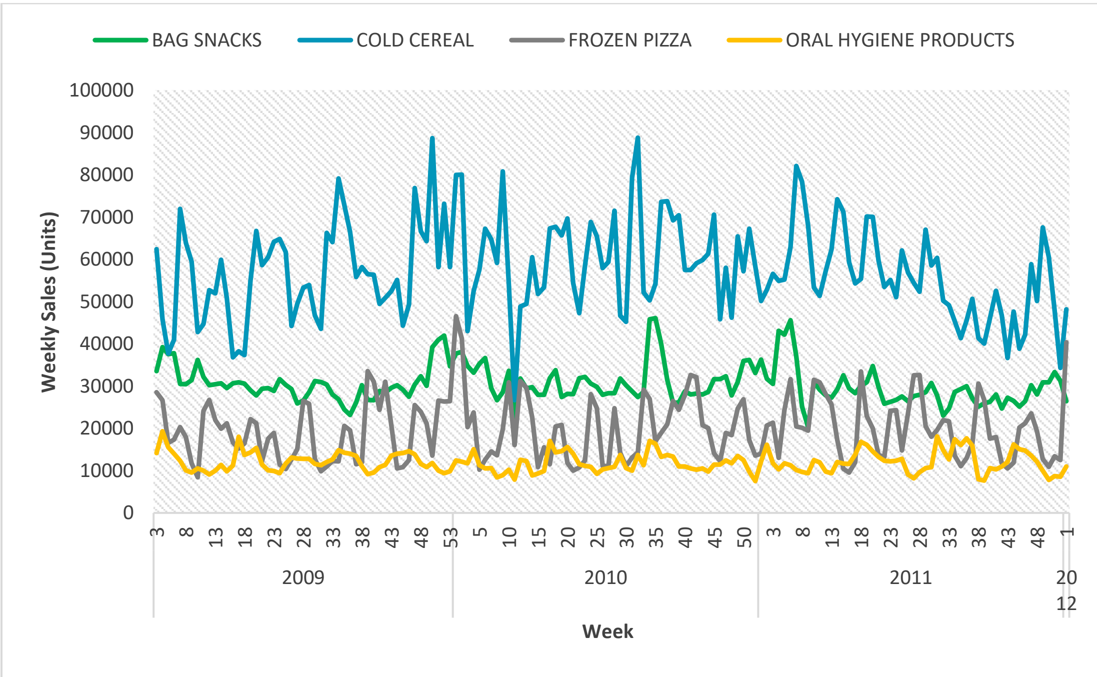

Figure 3-2 *Weekly sales by category*

Figure 3-3 shows the average weekly sales for each period for each category. At a glance, it shows that the average sales level during the post-promotional period is lower than the average sales level during the normal period.

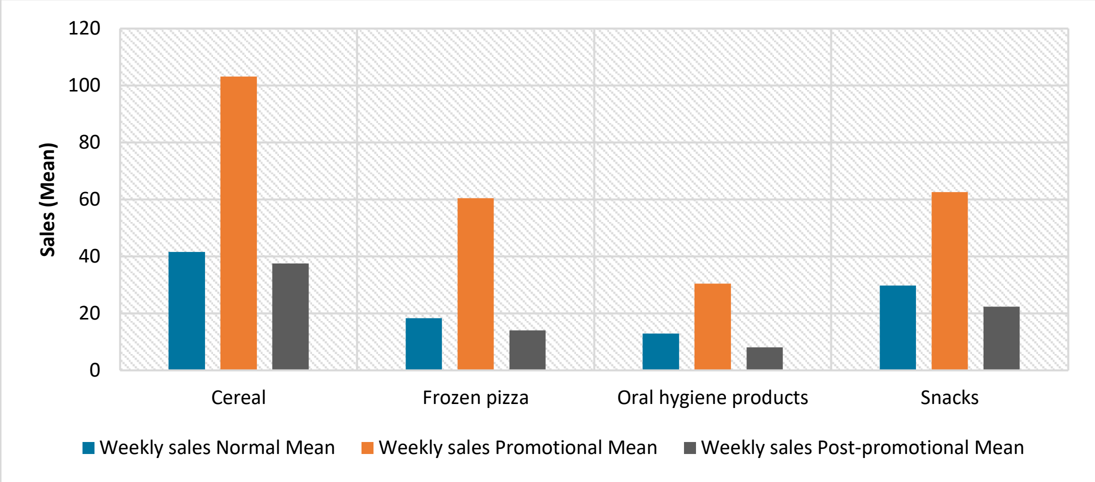

Figure 3-3 *Average weekly sales by promotional period*

In our study, we use a combination of both time series and causal data, including static and dynamic features. Furthermore, we defined lag features based on the maximum correlation. As Figure 3-4 depicts, we selected lag sales for 3 weeks for the ML models as lag features. Altogether, we defined 14 features as depicted in Table 3-2.

Table 3-2 *Selected input features*

Feature type	Feature
Raw data	Weekly Sales
Time series features	Lagged sales for 1 week, 2 weeks and 3 weeks
Dynamic features	Calendar features, promotion types (i.e., TPR, display and feature), magnitude of discounts and selling price
Static features	Store ID, Category and sub-category ID, SKU
Additional feature	Promotional period (i.e., normal, promotional, or post-promotional)

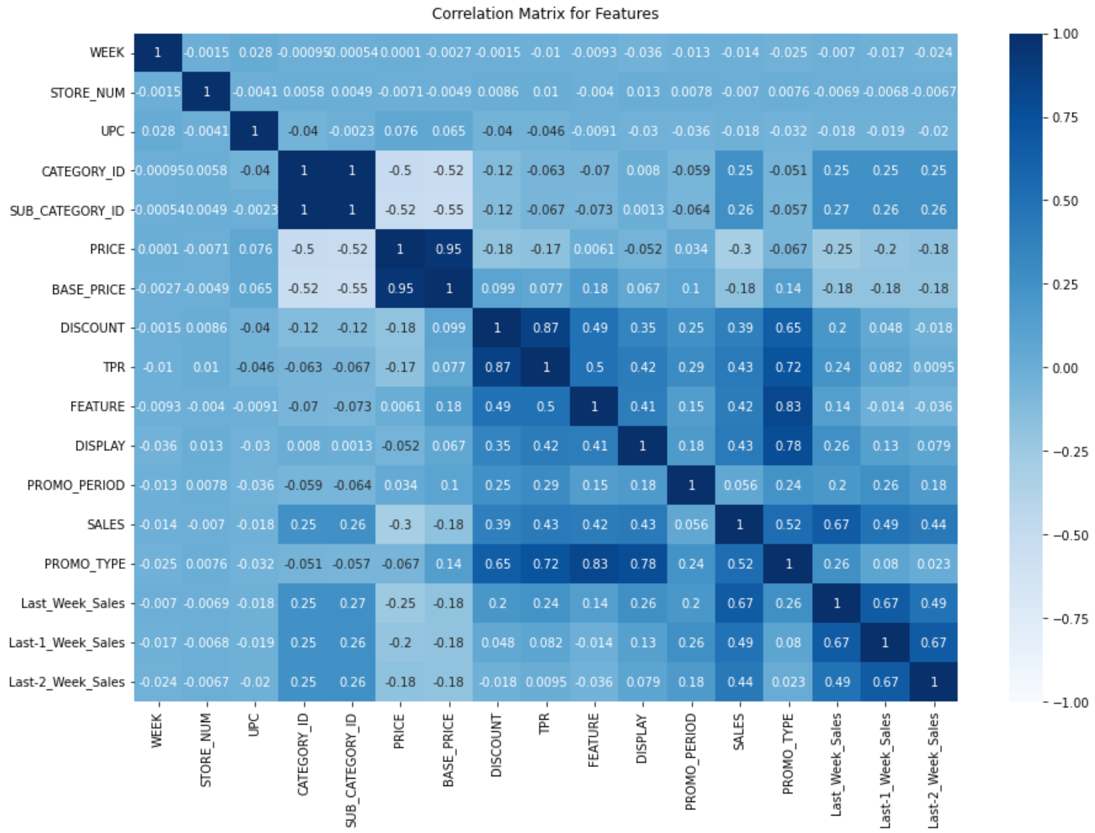

Figure 3-4 *Correlation matrix for features*

3.2 Data Pre-processing

We started by looking for missing values in the dataset, but there were none. Estimating baseline demand is critical for characterizing promotional periods as normal, promotional, or post-promotional. We used the ETS (ExponenTial Smoothing) model with normal sales levels to accomplish this. The ETS model has several advantages for our study, including simplicity and robustness (Hyndman & Khandakar, 2008). Furthermore, non-expert users can readily grasp and use the ETS model. We elaborate the model further under Section 3.4.

The normal and promotional periods were then classified using the promotional calendar. Following that, we used Eq. (3-1) to determine the post-promotional periods.

Equation 3-1 *Post-promotional effect calculation*

$$D_{it} = B_{it} - A_{it}$$

Where, A_{it}: actual sales for SKU i at t^{th} week, B_{it}: baseline demand for SKU i at t^{th} week and D_{it}: difference between baseline demand and actual sales at t^{th} week for SKU i. If D_{it} is negative directly after a promotion, the t^{th} period ($t \geq 0$) is labelled as a post-promotional period.

3.3 Benchmark Model

We employed a base-lift (BL) model with baseline demand estimation and promotional or post-promotional effect correction as the benchmark approach (see Section 2.4). It is a common approach that many retailers use to produce retail forecasts (Ma & Fildes, 2021) and is also implemented in commercial applications (Ali et al., 2009). If a promotion is planned for the next week, the average promotional lift is added to the forecast; if the week is designated as a post-promotional period, the average post-promotional dip is added to the forecast; otherwise, the forecast value is taken as is for the normal period, as shown in Eq (3-2).

Equation 3-2 *Base-lift estimation calculation*

$$BL_{it} = \begin{cases} B_{it} & ;\ (t = \text{normal period}) \\ B_{it} + (\text{Average promotional uplift})_i & ;\ (t = \text{promotional period}) \\ B_{it} - (\text{Average post-promotional dip})_i & ;\ (t = \text{post-promotional period}) \end{cases}$$

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.

3.4 Forecasting methods

We consider three groups of methods in our study, namely (1) univariate methods, (2) ML-based methods and (3) Deep Learning (DL) based methods. As univariate methods, we use ARIMA and ETS models since these are widely applied in both retail industry and academia (Fildes et al., 2019; Hyndman & Khandakar, 2008). Furthermore, we implement NAIVE and Seasonal NAIVE (SNAIVE) methods in our study for the comparison purposes. We also use Exponential Smoothing with exogenous variable (ETSX), an extension of ETS model (Abolghasemi 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.

3.4.1 NAIVE and SNAIVE

These methods are treated as simple forecasting methods. We simply use the last observation value as the forecast value in the NAIVE method. Whereas in SNAIVE, each forecast is set to the most recent observation from the same season (Hyndman & Athanasopoulos, 2021).

3.4.2 ARIMA

ARIMA model is a widely used approach in practice since it can take into consideration trend, seasonality, and error, as well as the non-stationarity of a time series (Hewamalage et al., 2021). The ARIMA parameters (p, d, q) indicate the following: p: auto-regressive (AR) component order; d: difference order; q: moving average (MA) component order (Pankratz, 1989). Eq. (3-3) demonstrates the ARIMA (p,d,q) model.

Equation 3-3 *ARIMA (p,d,q) model*

$$y_t = \emptyset_1 y_{t-1} + \ldots + \emptyset_p y_{t-p} + e_t + \theta_1 y_{t-1} + \ldots + \theta_q y_{t-q}$$

In our study, we used the AutoARIMA model (Hyndman & Khandakar, 2008), which finds the best ARIMA model automatically (Hyndman & Khandakar, 2008). First, it finds the appropriate order of difference (*d*) by using the Kwiatkowski-Phillips–Schmidt–Shin unit root test. Second, it calculates the optimal p and q values by fitting various models and choosing the model with the lowest Akaike Information Criterion (AIC) (Hyndman & Khandakar, 2008). We used the *auto.arima()* function in the R *forecast* package to implement the ARIMA model (Hyndman & Khandakar, 2008).

3.4.3 ETS and ETSX

ETS is a univariate approach that takes seasonality, trend, and error into consideration and is based on exponential smoothing in a state space framework (Petropoulos &

Svetunkov, 2020; Hyndman & Khandakar, 2008). This automatically determines the best model by minimization of a prespecified information criterion from the underlying fifteen exponential smoothing models (Hyndman & Khandakar, 2008). The following equations elaborate the ETS model: Eq. (3-4.a) depicts the observations and Eq. (3-4.b) and (3-4.c) elaborate the state equations of the model. Moreover, ε_t is a white noise process.

Equation 3-4 *ETS model*

$$y_t = l_{t-1} + \emptyset b_{t-1} + \varepsilon_t \qquad \text{(a)}$$

$$l_t = l_{t-1} + \emptyset b_{t-1} + \propto \varepsilon_t \qquad \text{(b)}$$

$$b_t = \emptyset b_{t-1} + \beta \varepsilon_t \qquad \text{(c)}$$

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 & Svetunkov, 2020; Hyndman & Khandakar, 2008). Eq. (3-4.a) can be modified by incorporating promotional period p_t as the covariate and providing it with a time-invariant coefficient of c. Eq. (3-5.a) depicts the observations with promotional period at t^{th} time and Eq. (3-5.b) and (3-5.c) elaborate the state equations of the model. Moreover, ε_t is a white noise process.

Equation 3-5 *ETSX model*

$$y_t = l_{t-1} + \emptyset b_{t-1} + c p_t + \varepsilon_t \qquad \text{(a)}$$

$$l_t = l_{t-1} + \emptyset b_{t-1} + \propto \varepsilon_t \qquad \text{(b)}$$

$$b_t = \emptyset b_{t-1} + \beta \varepsilon_t \qquad \text{(c)}$$

We used the *ets()* function in the R *forecast* package (Hyndman & Khandakar, 2008) and the *es()* function in the R *smooth* package (Petropoulos & Svetunkov, 2020) to implement ETS and ETSX models, respectively. We obtained the forecasts from the ETS model by estimating the model parameters using the *ets()* function and selecting the appropriate model by default.

3.4.4 Gradient-boosted Regression Trees

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

3.4.5 Random Forest

RF is a collection of RTs, each of which is based on the values of a random vector with the same distribution which 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). it is more resistant to noise and less likely to overfit the training data (Breiman, 2001). Further, past literature states RF is a promising approach in retail context (Spiliotis et al., 2020). We used *RandomForestRegressor* Python Package (scikit-learn Developers, 2022) to implement RF model.

3.4.6 DeepAR and WaveNet

DeepAR is built on an autoregressive recurrent neural network framework and simultaneously trains a large number of related time series (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, WaveNet model finished second in the Kaggle competition that featured the Corporaci Favorita data (Vallés-Pérez et al., 2022). We used *GluonTS* *toolkit* in Python (Amazon Web Service, 2022) to implement both DeepAR and WaveNet models.

3.4.7 Overview of the Candidate Models

In our study, we developed 13 candidate models using different combinations of input features as shown in Table 3-3. We set 52 as the frequency of each time series when developing the models, as these are weekly time series. We separated the data set into training and test sets using an 80:20 ratio, where we used training data to estimate the parameters of each forecasting method and test data to evaluate the forecast accuracy. Thus, the first 130 weeks are fed to each of the candidate models as the training set. The test data comprises of the subsequent 26 weeks. This results in 414,482 observations in the training dataset and 55,487 observations in the testing dataset. We used the default parameters to train all of the forecasting models. Specifically, we did not perform any hyperparameter tuning for the ML and DL methods to ensure the simplicity of the model.

Table 3-3 *Overview of the candidate models*

	BL	**NAIVE**	**SNAIVE**	**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 Post-promotional Dip	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

4 ANALYSIS AND RESULTS

In our analysis, we concentrated on three key areas. First, we evaluate the magnitude and sign of the post-promotional effect identified by each candidate model. We measured the magnitude and the sign of the post-promotional effect identified using Eq. (4-1).

Equation 4-1 *Post-promotional effect calculation*

$$PM_{it} = (F_{it} - B_{it}) / B_{it}$$

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 post-promotional effect at t^{th} week for SKU i.

Second, we evaluate the forecast accuracy of the models using Symmetric Mean Absolute Percentage Error (sMAPE) (Bandara et al., 2020) and Mean Absolute Scaled Error (MASE) (Hyndman & Koehler, 2006) using eq. (4-2) and eq. (4-3), respectively.

Equation 4-2 *sMAPE calculation*

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

Equation 4-3 *MASE calculation*

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

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). sMAPE has desirable properties such as simplicity, ease of communication (Huang et al., 2019). However, sMAPE has some drawbacks, including a lack of robustness, interpretability, and instability with values close to zero (Bandara et al., 2019). We use MASE as our second error metric, as it is scale-independent to mitigate some of these issues (Hyndman & Koehler, 2006).

Third, we compare the value addition from the additional variable to the ML and DL methods using Forecast Value Added (FVA) model; Eq. (4-4). Chybalski (2017) explains that FVA compares the forecast improvement of a model with another. Error metrics such as Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (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.

Equation 4-4 *Forecast value added calculation*

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

Where: $FVA_{i,t}$: forecast value added for model i compared to model k, $MASE_k$: MASE for model k and $MASE_i$: MASE for model i.

Finally, we used the non-parametric Friedman test to examine the statistical significance of the differences in these methods (Friedman, 1940). We utilized the non-parametric Wilcoxon signed-rank test to investigate these differences further with respect to each approach (Wilcoxon & Wilcox, 1964).

4.1 Magnitude and Sign of Post-promotional Effect

Figure 4-1 depicts the distribution of the post-promotional dip identified by each forecasting model. Table 4-1 shows the descriptive summary of the post-promotional effects identified by the forecasting models. A preliminary investigation shows that ML methods outperformed univariate and DL methods in identifying the post-promotional period. Friedman test results indicate that there are significant differences ($\chi^2(235) = 390.4$, $p < .000$) in identified post-promotional dips by forecasting methods.

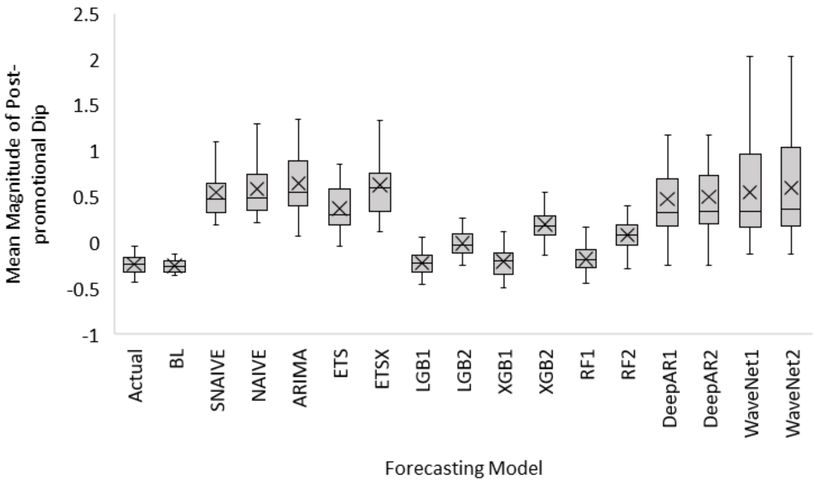

Figure 4-1 *Distribution of the magnitude of post-promotional dip*

Table 4-1 *Descriptive summary of the magnitude and sign of the post-promotional dip*

Forecasting model	Post-promotional dip	
	Mean	SD
Post-promotional dip in test dataset	-23.86%	9.81
BL	-25.50%	7.31
NAIVE	58.53%	33.19
SNAIVE	54.77%	26.82
ARIMA	52.54%	34.49
ETS	36.99%	24.67
ETSX	63.25%	36.68
LGB1	-22.65%	11.92
LGB2	-0.67%	13.88
XGB1	-20.81%	13.82
XGB2	19.73%	15.84
RF1	-18.88%	13.02
RF2	8.31%	18.49
DeepAR1	47.38%	51.90
DeepAR2	50.30%	53.22
WaveNet1	54.86%	55.83
WaveNet2	59.70%	58.55

The Wilcoxon signed-rank test results show that all univariate models, including ETSX, are significantly different from the test dataset's real mean post-promotional dip ($p < .000$). Though ML models were able to identify the post-promotional period, we see that they only identified the correct magnitude of the post-promotional dip when the additional feature was incorporated; LGB1 vs LGB2 ($p < .000$), XGB1 vs XGB2 ($p < .000$) and RF1 vs RF2 ($p < .000$), failed to provide support for *Hypothesis 1*. Noticeably, DL methods were unable to identify the post-promotional period even with the additional variable and were significantly different from the actual post-promotional dip ($p < .000$). Furthermore, LGB1 ($p = 0.628$), XGB 1 ($p = 0.361$), and RF1 ($p = 0.054$) show no significant differences from the actual post-promotional dip, providing partial support for *Hypothesis 2*.

4.2 Comparison of Forecast Performances

We separately compare model performance in each promotional period using sMAPE and MASE. Table 4-2 summarises the descriptive statistics of sMAPE and MASE across forecasting methods.

Table 4-2 *Forecast accuracy for each forecasting method*

Forecasting method	sMAPE						MASE					
	Normal		Promotional		Post-promotional		Normal		Promotional		Post-promotional	
	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
NAIVE	0.39	0.08	0.44	0.13	0.64	0.11	1	0	1	0	1	0
SNAIVE	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

4.2.1 *Forecast performance during the normal period*

Figure 4-2 indicates 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.

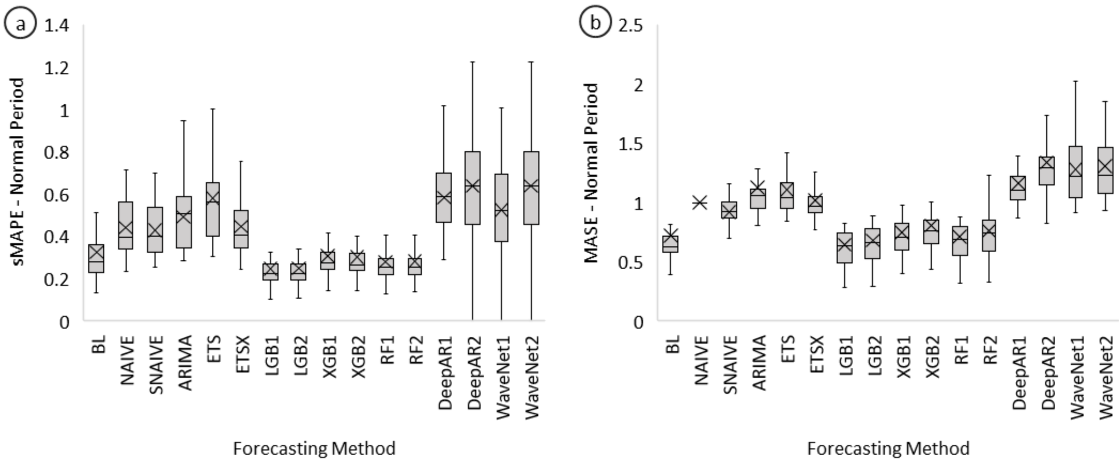

Figure 4-2 *(a) sMAPE values in normal period; (b) MASE values in normal period*

The Wilcoxon signed-rank test results show significant differences between univariate methods and other forecasting methods ($p < .000$), 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*: $p < .000$; *WaveNet*: $p < .000$) when the promotional period is provided as an additional variable. *Hypothesis 2*, therefore, is only partially supported in the normal period.

4.2.2 *Forecast performance during the promotional period*

Figure 4-3 depicts 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.

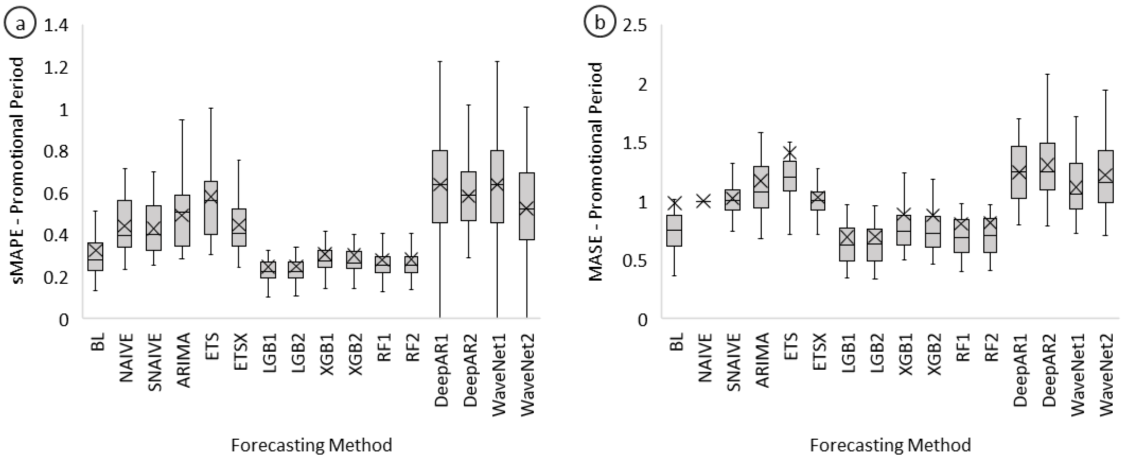

Figure 4-3 *(a) sMAPE values in promotional period; (b) MASE values in promotional period*

Wilcoxon signed-rank test reveals significant differences ($p < .000$) between univariate methods and ML methods during the promotional period, providing evidence for *Hypothesis 3*. 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*: $p = .000$; *MASE:* $p = .000$) and LGB2 (*sMAPE*: $p = .000$; *MASE:* $p = 0.001$) models outperform the BL model in the promotional period. This lends some credence to *Hypothesis 3*. 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 *Hypothesis 2* in the promotional period. Furthermore, as expected, the ETSX model outperformed the ETS model in the promotional period (*sMAPE*: $p = .000$; *MASE:* $p < .000$).

4.2.3 Forecast performance during the post-promotional period

Figure 4-4 shows the distribution of sMAPE and MASE values in the post-promotional 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 post-promotional period.

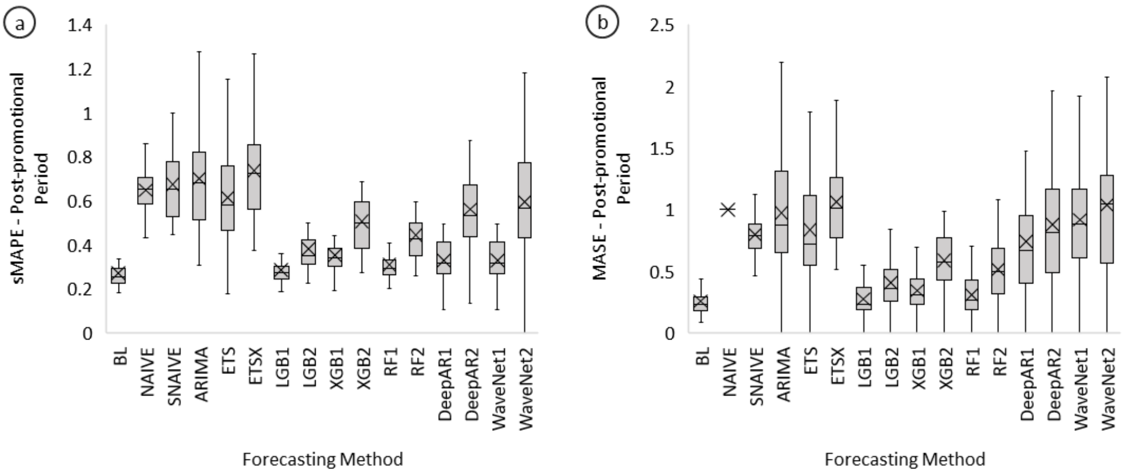

Figure 4-4 *(a) sMAPE values in post-promotional period; (b) MASE values in post-promotional period*

The Wilcoxon signed-rank test reveals that ML methods significantly differ from univariate methods ($p < .000$), providing partial support for *Hypothesis 3*. Notably, the pairwise comparison shows that incorporating the additional variable significantly improves the performance of ML methods (*sMAPE*: $p < .000$; *MASE*: $p < .000$). This provides support for *Hypothesis 2*. However, even with support of the additional variable, ML methods failed to outperform the BL method. Only the LGB1 model performed similar to the BL method (*sMAPE*: *LGB1:* $p = 0.055$; *XGB1:* $p < .000$; *RF1:* $p = .000$ | *MASE*: *LGB1:* $p = 0.490$; *XGB1:* $p = 001$; *RF1:* $p = 0.048$). This provides no support for *Hypothesis 3*. On the other hand, DL methods show a significant improvement only in terms of sMAPE (*DeepAR*: $p < .000$; *WaveNet*: $p < .000$), providing evidence for *Hypothesis 2*. Surprisingly, the ETS model outperformed the ETSX model in the post-promotional period (*sMAPE*: $p = 0.008$; *MASE:* $p = .000$).

4.3 Forecast improvement under compared methods

Table 4-3, 4-4 and 4-5 provide a summary of FVA values for forecasting methods in the normal period. Notably, ML methods outperform all the univariate methods in *all* periods (*Table 4-3, Table 4-4 and Table 4-5*; *FVA* > 0). However, they did not improve the forecast compared to the BL method and performed similarly in the normal period ($p < .000$). In the promotional period, ML methods outperform the BL method ($p < .000$). On the contrary, only the LGB1 model shows no significant differences from the BL method in the post-promotional period (*sMAPE*: *LGB1:* p = 0.055; *MASE*: *LGB1:* p = 0.490). Thus, this only provides partial support to *Hypothesis 3* as ML methods only outperform conventional univariate methods. Surprisingly, DL methods rarely outperformed univariate methods and were unable to outperform the BL methods in *all* periods. Furthermore, univariate methods were unable to improve the forecast performance in all periods compared to the BL method ($p < .000$).

Table 4-3 *FVA value comparison for normal period*

	MASE	BL	NAIVE	SNAIVE	ARIMA	ETS	ETSX	LGB1	LGB2	XGB1	XGB2	RF1	RF2	DeepAR1	DeepAR2	WaveNet1
BL	0.71	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAIVE	1.00	-0.29	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SNAIVE	1.05	-0.35	-0.05	-	-	-	-	-	-	-	-	-	-	-	-	-
ARIMA	1.33	-0.63	-0.33	-0.28	-	-	-	-	-	-	-	-	-	-	-	-
ETS	1.15	-0.44	-0.15	-0.10	0.18	-	-	-	-	-	-	-	-	-	-	-
ETSX	1.05	-0.35	-0.05	0.00	0.28	0.10	-	-	-	-	-	-	-	-	-	-
LGB1	0.75	-0.04	0.25	0.31	0.59	0.40	0.31	-	-	-	-	-	-	-	-	-
LGB2	0.78	-0.07	0.22	0.28	0.56	0.37	0.27	-0.03	-	-	-	-	-	-	-	-
XGB1	0.82	-0.11	0.18	0.24	0.52	0.33	0.23	-0.07	-0.04	-	-	-	-	-	-	-
XGB2	0.85	-0.14	0.15	0.20	0.48	0.30	0.20	-0.11	-0.07	-0.03	-	-	-	-	-	-
RF1	0.80	-0.09	0.20	0.25	0.54	0.35	0.25	-0.05	-0.02	0.02	0.05	-	-	-	-	-
RF2	0.82	-0.12	0.18	0.23	0.51	0.33	0.23	-0.08	-0.05	0.00	0.03	-0.02	-	-	-	-
DeepAR1	1.30	-0.60	-0.30	-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	-0.62	-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	-0.62	-0.563	-0.281	-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	-0.75	-0.703	-0.421	-0.6	-0.7	-1.01	-0.98	-0.94	-0.9	-0.96	-0.9	-0.45	-0.13	-0.13

***Table 4-4** FVA value comparison for promotional period*

	MASE	BL	NAIVE	SNAIVE	ARIMA	ETS	ETSX	LGB1	LGB2	XGB1	XGB2	RF1	RF2	DeepAR1	DeepAR2	WaveNet1
BL	0.98	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAIVE	1.00	-0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SNAIVE	1.02	-0.04	-0.02	-	-	-	-	-	-	-	-	-	-	-	-	-
ARIMA	1.17	-0.19	-0.17	-0.15	-	-	-	-	-	-	-	-	-	-	-	-
ETS	1.41	-0.43	-0.41	-0.39	-0.24	-	-	-	-	-	-	-	-	-	-	-
ETSX	1.03	-0.05	-0.03	-0.01	0.14	0.38	-	-	-	-	-	-	-	-	-	-
LGB1	0.69	0.29	0.31	0.33	0.48	0.72	0.34	-	-	-	-	-	-	-	-	-
LGB2	0.70	0.28	0.30	0.32	0.47	0.71	0.33	0.00	-	-	-	-	-	-	-	-
XGB1	0.89	0.09	0.11	0.14	0.29	0.53	0.15	-0.19	-0.19	-	-	-	-	-	-	-
XGB2	0.88	0.10	0.12	0.15	0.29	0.53	0.16	-0.18	-0.18	0.01	-	-	-	-	-	-
RF1	0.81	0.17	0.19	0.22	0.37	0.61	0.23	-0.11	-0.11	0.08	0.07	-	-	-	-	-
RF2	0.82	0.16	0.18	0.21	0.36	0.60	0.22	-0.12	-0.12	0.07	0.06	-0.01	-	-	-	-
DeepAR1	1.24	-0.26	-0.24	-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	-0.31	-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	-0.12	-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	-0.22	-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 4-5 *FVA value comparison for post-promotional period*

	MASE	BL	NAIVE	SNAIVE	ARIMA	ETS	ETSX	LGB1	LGB2	XGB1	XGB2	RF1	RF2	DeepAR1	DeepAR2	WaveNet1
BL	0.26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NAIVE	1.00	-0.74	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SNAIVE	0.80	-0.54	0.20	-	-	-	-	-	-	-	-	-	-	-	-	-
ARIMA	0.97	-0.72	0.03	-0.17	-	-	-	-	-	-	-	-	-	-	-	-
ETS	0.84	-0.58	0.16	-0.04	0.14	-	-	-	-	-	-	-	-	-	-	-
ETSX	1.06	-0.81	-0.06	-0.26	-0.09	-0.23	-	-	-	-	-	-	-	-	-	-
LGB1	0.28	-0.02	0.72	0.52	0.70	0.56	0.78	-	-	-	-	-	-	-	-	-
LGB2	0.41	-0.15	0.59	0.39	0.57	0.43	0.65	-0.13	-	-	-	-	-	-	-	-
XGB1	0.34	-0.09	0.66	0.46	0.63	0.49	0.72	-0.07	0.06	-	-	-	-	-	-	-
XGB2	0.59	-0.33	0.41	0.21	0.38	0.25	0.47	-0.31	-0.18	-0.25	-	-	-	-	-	-
RF1	0.31	-0.05	0.69	0.49	0.66	0.53	0.75	-0.03	0.10	0.03	0.28	-	-	-	-	-
RF2	0.52	-0.26	0.48	0.28	0.46	0.32	0.55	-0.24	-0.11	-0.17	0.07	-0.21	-	-	-	-
DeepAR1	0.74	-0.49	0.26	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	0.12	-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	0.08	-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	-0.04	-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

5 DISCUSSION

5.1 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 base lift 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 context of sales 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 *all* periods.

Firstly, our findings reinforce previous research findings (Ali & Gürlek, 2020; Huber & Stuckenschmidt, 2020; Trapero et al., 2015) on the ability of multivariate models to automatically detect the post-promotional period. In order to estimate the right sign and magnitude of the post-promotional dip, ML models require the additional variable as an input feature. Notably, DL methods did not identify the correct post-promotional dip even with the additional variable as an input.

Secondly, 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 similar 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[3], they frequently overestimate during normal periods (De Baets & 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 post-promotional periods compared to conventional univariate methods. Furthermore, adding the additional variable enhances the forecast performance of ML models only during the post-promotional period. Although the DL methods did not perform as expected the inclusion of the additional variable improved forecast performance in *all* periods. This aligns with the previous findings that when advanced methods are used, more detailed inputs can improve the performance (Ali et al., 2009).

Thirdly, our study compares all the forecasting methods with the base-lift model, a well-established retail implementation. Importantly, ML methods perform similar to the BL method in all periods even though ML methods benefit from the additional variable. On the other hand, the BL method generates significantly better forecasts compared to conventional univariate methods in *all* periods. Although the BL method performs effectively across *all* time periods, data cleansing takes additional effort and time. Yet, this process can be time-consuming and prone to bias (Hewage et al., 2021; Perera et al., 2019). 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 of 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

---

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

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.

5.2 Managerial and Practical 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 put in 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.

5.3 Limitations and Future Directions

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. Our study only includes three types of promotions (i.e., temporary price reductions, display, and feature) instead of incorporating a variety of promotions. Furthermore, we did not explore the impact of incorporating special days and holidays into our study. Therefore, how to incorporate other causal factors such as multiple promotion types, special days and events, and holidays might be an interesting future research avenue. Furthermore, we did not see any intermittent demand patterns in our data set. 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 highlights 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 only used the standard versions of the ML and DL methods. We did not employ any hyper-parameter 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.

6 CONCLUSION

Retail promotions create demand irregularities for products, making the generation of accurate forecasts more difficult. Nonetheless, retailers generally forecast sales during promotional periods using either the base-lift model 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 base-lift 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 feasible alternative in retail sales forecasting. Our empirical study shows that ML methods have the capacity to incorporate causal factors with the sales history. ML methods perform as well as the BL method. 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.

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 over *all* periods.

REFERENCES

Abolghasemi, M., Beh, E., Tarr, G., & Gerlach, R. (2020). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers and Industrial Engineering*, *142*(February), 106380. https://doi.org/10.1016/j.cie.2020.106380

Abolghasemi, M., Hurley, J., Eshragh, A., & Fahimnia, B. (2020). Demand forecasting in the presence of systematic events: Cases in capturing sales promotions. *International Journal of Production Economics*, *230*(December 2018), 107892. https://doi.org/10.1016/j.ijpe.2020.107892

Abraham, M. M., & Lodish, L. M. (1987). Promoter : An Automated Promotion Evaluation System. *Marketing Science*, *6*(2), 101–123.

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing Journal*, *7*(1), 136–144. https://doi.org/10.1016/j.asoc.2005.06.001

Ali, Ö. G., Sayin, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, *36*(10), 12340–12348. https://doi.org/10.1016/j.eswa.2009.04.052

Alvarado-Valencia, J., Barrero, L. H., Önkal, D., & Dennerlein, J. T. (2017). Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting. *International Journal of Forecasting*, *33*(1), 298–313. https://doi.org/10.1016/j.ijforecast.2015.12.010

Amazon Web Service. (2022). *GluonTS - Probabilistic Time Series Modeling*. https://ts.gluon.ai/

Aruchunarasa, B., & Perera, H. N. (2022). Mitigating the Proclivity towards Multiple Adjustments through Innovative Forecasting Support Systems. In N. Subramanian, S. G. Ponnambalam, & M. Janardhanan (Eds.), *Innovation Analytics Tools for Competitive Advantage* (pp. 1–15). World Scientific Publishing Co Pte Ltd. https://doi.org/https://doi.org/10.1142/q0293

Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2018). Integrating human judgement into quantitative forecasting methods: A review. *Omega (United Kingdom)*. https://doi.org/10.1016/j.omega.2018.07.012

Baecke, P., De Baets, S., & Vanderheyden, K. (2017). Investigating the added value of integrating human judgement into statistical demand forecasting systems. *International Journal of Production Economics*, *191*(June), 85–96. https://doi.org/10.1016/j.ijpe.2017.05.016

Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Systems with Applications*, *140*. https://doi.org/10.1016/j.eswa.2019.112896

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., & Seaman, B. (2019). Sales demand forecast in E-commerce using a long short-term memory neural network methodology. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11955 LNCS*, 462–474. https://doi.org/10.1007/978-3-030-36718-3_39

Blattberg, R. C., & Briesch, R. A. (2012). Sales Promotions. In *The Oxford Handbook of Pricing Management* (Issue April 2018). https://doi.org/10.1093/oxfordhb/9780199543175.013.0024

Brau, R., Aloysius, J., & Siemsen, E. (2019). When Models Meet Managers: Optimal Integration of Statistical Model-Based and Judgmental Forecasting. *APICS 2019*.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chybalski, F. (2017). Forecast value added (FVA) analysis as a means to improve the efficiency of a forecasting process. *Operations Research and Decisions*, *27*(1), 5–19. https://doi.org/10.5277/ord170101

Cohen, M. C., Kalas, J. J., & Perakis, G. (2020). Promotion Optimization for Multiple Items in Supermarkets. *Management Science*, *June*. https://doi.org/10.1287/mnsc.2020.3641

Cooper, L. G., Baron, P., Levy, W., Swisher, M., & Gogos, P. (1999). PromoCast™: A New Forecasting Method for Promotion Planning. *Marketing Science*, *18*(3), 301–316. https://doi.org/10.1287/mksc.18.3.301

De Baets, S., & Harvey, N. (2016). *Enhanced anchoring effects produced by the presence of statistical forecasts: Effects on judgmental forecasting*. http://hdl.handle.net/20.500.12127/5696

De Baets, S., & Harvey, N. (2018). Forecasting from time series subject to sporadic perturbations: Effectiveness of different types of forecasting support. *International Journal of Forecasting*, *34*(2), 163–180. https://doi.org/10.1016/j.ijforecast.2017.09.007

DelVecchio, D., Henard, D. H., & Freling, T. H. (2006). The effect of sales promotion on post-promotion brand preference: A meta-analysis. *Journal of Retailing*, *82*(3), 203–213. https://doi.org/10.1016/j.jretai.2005.10.001

Divakar, S., Ratchford, B. T., & Shankar, V. (2005). Practice Prize Article—*CHAN4CAST*: A Multichannel, Multiregion Sales Forecasting Model and

Decision Support System for Consumer Packaged Goods. *Marketing Science*, *24*(3), 334–350. https://doi.org/10.1287/mksc.1050.0135

Feiler, D. C., Tong, J. D., & Larrick, R. P. (2013). Biased Judgment in Censored Environments. *Management Science, 59*(3), 573–591. https://doi.org/10.1287/mnsc.1120.1612

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570–576. https://doi.org/10.1287/inte.1070.0309

Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, *42*(1), 351–361. https://doi.org/10.1016/j.dss.2005.01.003

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23. https://doi.org/10.1016/j.ijforecast.2008.11.010

Fildes, R., Goodwin, P., & Önkal, D. (2018). Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*, *35*(1), 144–156. https://doi.org/10.1016/j.ijforecast.2017.12.006

Fildes, R., Ma, S., & Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting*, *xxxx*. https://doi.org/10.1016/j.ijforecast.2019.06.004

Fildes, R., & Petropoulos, F. (2015). Improving Forecast Quality in Practice. *Foresight: The International Journal of Applied Forecasting*, *36*, 5–12. https://econpapers.repec.org/RePEc:for:ijafaa:y:2015:i:36:p:5-12

Franses, P. H., & Legerstee, R. (2013). Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*, *29*(1), 80–87. https://doi.org/10.1016/j.ijforecast.2012.05.008

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

Goodwin, P. (2000). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, *16*(1), 85–99. https://doi.org/10.1016/S0169-2070(99)00026-6

Gür Ali, Ö. (2013). Driver moderator method for retail sales prediction. *International Journal of Information Technology and Decision Making*, *12*(6), 1261–1286. https://doi.org/10.1142/S0219622013500363

Gür Ali, Ö., & Gürlek, R. (2020). Automatic Interpretable Retail forecasting with promotional scenarios. *International Journal of Forecasting*, *36*(4), 1389–1406. https://doi.org/10.1016/j.ijforecast.2020.02.003

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

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

Hewage, H. C., Perera, H. N., & De Baets, S. (2021). Forecast adjustments during post-promotional periods. *European Journal of Operational Research*, *300*(2), 461–472. https://doi.org/10.1016/J.EJOR.2021.07.057

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *International Journal of Forecasting*, *37*(1), 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008

Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, *237*(2), 738–748. https://doi.org/10.1016/j.ejor.2014.02.022

Huang, T., Fildes, R., & Soopramanien, D. (2019). Forecasting retailer product sales in the presence of structural change. *European Journal of Operational Research*, *279*(2), 459–470. https://doi.org/10.1016/j.ejor.2019.06.011

Huber, J., & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, *36*(4), 1420–1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

Hübner, A., Amorim, P., Kuhn, H., Minner, S., & Van Woensel, T. (2018). Retail operations. *OR Spectrum*, *40*(4), 831–835. https://doi.org/10.1007/s00291-018-0535-1

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd editio). OTexts: Melbourne, Australia. OTexts.com/fpp3

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.

https://doi.org/10.1016/j.ijforecast.2006.03.001

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493–518. https://doi.org/10.1016/j.ijforecast.2006.03.007

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, *23*(3), 377–390. http://dx.doi.org/10.1016/j.ijforecast.2007.02.006

Leeflang, P. S. H., van Heerde, H. J., & Wittink, D. R. (2005). How Promotions Work: SCAN*PRO-Based Evolutionary Model Building. *SSRN Electronic Journal*, *September 2018*. https://doi.org/10.2139/ssrn.321003

Ma, S., & Fildes, R. (2017). A retail store SKU promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, *260*(2), 680–692. https://doi.org/10.1016/j.ejor.2016.12.032

Ma, S., & Fildes, R. (2021). Retail sales forecasting with meta-learning. *European Journal of Operational Research*, *288*(1), 111–128. https://doi.org/10.1016/j.ejor.2020.05.038

Ma, S., Fildes, R., & Huang, T. (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*(1), 245–257. https://doi.org/10.1016/j.ejor.2015.08.029

Macé, S., & Neslin, S. A. (2004). The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research*, *41*(3), 339–350. https://doi.org/10.1509/jmkr.41.3.339.35992

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, *13*(3), 1–26. https://doi.org/10.1371/journal.pone.0194889

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

Mou, S., Robb, D. J., & DeHoratius, N. (2018). Retail store operations: Literature review and research directions. *European Journal of Operational Research*, *265*(2), 399–422. https://doi.org/10.1016/j.ejor.2017.07.003

Pankratz, A. (1989). Time series forecasts and extra-model information. *Journal of Forecasting*, *8*(2), 75–83. https://doi.org/10.1002/for.3980080202

Perera, H. N., Fahimnia, B., & Tokar, T. (2020). Inventory and ordering decisions: a systematic review on research driven through behavioral experiments. *International Journal of Operations & Production Management*, *Volume 40*(7/8). https://doi.org/10.1108/IJOPM-05-2019-0339

Perera, H. N., Hurley, J., Fahimnia, B., & Reisi, M. (2019). The human factor in supply chain forecasting: A systematic review. *European Journal of Operational Research*, *274*(2), 574–600. https://doi.org/10.1016/j.ejor.2018.10.028

Perera, H. N., & Perera, H. Y. R. (2022). *Applications of Pixel Oriented Mobility Modelling in Transport & Logistics BT - Dynamics in Logistics* (M. Freitag, A. Kinra, H. Kotzab, & N. Megow (eds.); pp. 337–348). Springer International Publishing.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., Browell, J., Carnevale, C., Castle, J. L., Cirillo, P., Clements, M. P., Cordeiro, C., Cyrino Oliveira, F. L., De Baets, S., Dokumentov, A., … Ziel, F. (2022). Forecasting: theory and practice. *International Journal of Forecasting*. https://doi.org/https://doi.org/10.1016/j.ijforecast.2021.11.001

Petropoulos, F., & Svetunkov, I. (2020). A simple combination of univariate models. *International Journal of Forecasting*, *36*(1), 110–115. https://doi.org/10.1016/j.ijforecast.2019.01.006

Ramanathan, U., & Muyldermans, L. (2010). Identifying demand factors for promotional planning and forecasting: A case of a soft drink company in the UK. *International Journal of Production Economics*, *128*(2), 538–545. https://doi.org/10.1016/j.ijpe.2010.07.007

Ramanathan, U., & Muyldermans, L. (2011). Identifying the underlying structure of demand during promotions: A structural equation modelling approach. *Expert Systems with Applications*, *38*(5), 5544–5552. https://doi.org/10.1016/j.eswa.2010.10.082

Rob J. Hyndman, & Yeasmin Khandakar. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3), 22. http://www.jstatsoft.org/%0Ahttp://www.jstatsoft.org/v27/i03/paper

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(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., Makridakis, S., Semenoglou, A. A., & Assimakopoulos, V. (2020). Comparison of statistical and machine learning methods for daily SKU demand forecasting. *Operational Research*, *0123456789*. https://doi.org/10.1007/s12351-020-00605-2

Sprangers, O., Schelter, S., & de Rijke, M. (2022). Parameter-efficient deep probabilistic forecasting. *International Journal of Forecasting*, *xxxx*. https://doi.org/10.1016/j.ijforecast.2021.11.011

Tong, J., Feiler, D., & Larrick, R. (2018). A Behavioral Remedy for the Censorship Bias. *Production and Operations Management*, *27*(4), 624–643. https://doi.org/10.1111/poms.12823

Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, *66*(2), 299–307. https://doi.org/10.1057/jors.2013.174

Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, *29*(2), 234–243. https://doi.org/10.1016/j.ijforecast.2012.10.002

Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications*, 116993. https://doi.org/10.1016/j.eswa.2022.116993

Van den Broeke, M., De Baets, S., Vereecke, A., Baecke, P., & Vanderheyden, K. (2019). Judgmental forecast adjustments over different time horizons. *Omega*, *87*, 34–45. https://doi.org/10.1016/j.omega.2018.09.008

van Steenbergen, R. M., & Mes, M. R. K. (2020). Forecasting demand profiles of new products. *Decision Support Systems*, *139*(January), 113401. https://doi.org/10.1016/j.dss.2020.113401

Wang, X. (Shane), Ryoo, J. H. (Joseph), Bendle, N., & Kopalle, P. K. (2020). The role of machine learning analytics and metrics in retailing research. *Journal of Retailing*. https://doi.org/10.1016/j.jretai.2020.12.001

Wilcoxon, F., & Wilcox, R. A. (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