



The derived demand for advertising expenses and implications on sustainability: a comparative study using deep learning and traditional machine learning methods

Sule Birim¹ · Ipek Kazancoglu² · Sachin Kumar Mangla³ · Aysun Kahraman⁴ · Yigit Kazancoglu⁵

Accepted: 10 November 2021 / Published online: 7 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

In recent years, machine learning models based on big data have been introduced into marketing in order to transform customer data into meaningful insights and to make strategic decisions by making more accurate predictions. Although there is a large amount of literature on demand forecasting, there is a lack of research about how marketing strategies such as advertising and other promotional activities affect demand. Therefore, an accurate demand-forecasting model can make significant academic and practical contributions for business sustainability. The purpose of this article is to evaluate machine learning methods to provide accuracy in forecasting demand based on advertising expenses. The study focuses on a prediction mechanism based on several Machine Learning techniques—Support Vector Regression (SVR), Random Forest Regression (RFR) and Decision Tree Regressor (DTR) and deep learning techniques—Artificial Neural Network (ANN), Long Short Term Memory

✉ Yigit Kazancoglu
yigit.kazancoglu@yasar.edu.tr

Sule Birim
sule.ozturk@cbu.edu.tr

Ipek Kazancoglu
ipek.savasci@ege.edu.tr

Sachin Kumar Mangla
smangla@jgu.edu.in; sachinmangl@gmail.com; sachin.kumar@plymouth.ac.uk

Aysun Kahraman
aysun.kahraman@cbu.edu.tr

- ¹ Department of Business Administration, Salihli Faculty of Economics and Administrative Sciences, Manisa Celal Bayar University, Manisa, Turkey
- ² Department of Business Administration, Faculty of Economics and Administrative Sciences, Ege University, İzmir, Turkey
- ³ OP Jindal Global University, Jindal Global Business School, Operations Management, Haryana, India
- ⁴ Department of Business Administration, Salihli Faculty of Economics and Administrative Sciences, Manisa Celal Bayar University, Manisa, Turkey
- ⁵ Yasar Universitesi, Izmir, Turkey

(LSTM),—to deal with demand forecasting based on advertising expenses. Deep learning is a powerful technique that can solve marketing problems based on both classification and regression algorithms. Accordingly, a television manufacturer's real market dataset consisting of advertising expenditures, sales and demand forecasting via chosen machine learning methods was analyzed and compared in terms of the accuracy of demand forecasting. As a result, Long Short Term Memory has been found to be superior to other models in providing highly accurate prediction results for demand forecasting based on advertising expenses.

Keywords Advertisement · Demand forecasting · Machine learning · Marketing intelligence

1 Introduction

Over the past several decades, advances in statistics, computing and communication science have been responsible for the creation of new sources (e.g. point of sales data, in store path data, user generated content such as comments, blogs, tweets, likes, videos, conversations etc.) for data collecting (Boone et al., 2019). This abundance of data has given rise to a concept called “big data” leading to the creation of large-scale datasets; this motivates businesses using data to improve their decision-making processes (Bag et al., 2020; Bajari et al., 2015; Kraus et al., 2020; Mustak et al., 2021; Rahman et al., 2016). As big data has become increasingly important, various types of research in different fields related to big data have been conducted, such as agriculture, accounting and finance, education, marketing, medicine and healthcare, manufacturing and operations, supply chain management, tourism, transportation, weather forecasting etc. (Abkenar et al., 2021; Bag et al., 2021; Daniel, 2017; Fathi et al., 2021; Galetsi & Katsaliaki, 2020; Hasan et al., 2020; Khanra et al., 2020a; Khanra et al., 2020b; Kumar et al., 2021; Kamilaris et al., 2017; Li et al., 2016; Lian et al., 2020; Talwar et al., 2021; Qayyum et al., 2021). Bag et al. (2021) discussed the impact of big data empowered by artificial intelligence on marketing performance in B2B markets by focusing on creation of knowledge, user knowledge, and external market knowledge. Kumar et al. (2021) and Qayyum et al. (2021) presented studies on forecasting COVID-19 pandemic dynamics and detection of the areas affected by COVID-19. Another study about big data usage in healthcare, which was conducted by Khanra et al. (2020b), has reviewed the prior literature and discovered five main themes which are health awareness, stakeholders of the healthcare ecosystem, hospital management practices, specific medical conditions, and healthcare service delivery through technology use. Khanra et al. (2020a) conducted a bibliometric analysis of applications of big data analytics in enterprises and four main themes were discovered which are strategic decision making, concept development for big data analytics, trends in applications of big data analytics, efficient supply chain management. Talwar et al. (2021) have reviewed the literature about big data in operations and supply chain management and offered actionable implications for managers. In studies on the use of big data in supply chain management, the issue of sustainability has also been addressed. Kittichotsawat et al. (2021) and Mageto (2021) discussed using big data analytics in sustainable supply chain towards sustainable growth. Choi and Chen (2021) conducted a study on the usage of big data in order to make effective decisions in circular supply chain applications from a sustainability perspective. Kwon et al. (2021), Mariani and Borghi (2020), Woo and Mishra (2020), studied the use of big data from a marketing point of view and conducted studies on online reviews and ratings. Li et al. (in press) and Wang et al. (in press) presented studies on the impact of big data analytics on intellectual manufacturing. Hasan et al. (2020) identified impacts of big data

on financial products and services in the framework of exploratory literature review. Daniel (2017) discussed the opportunities afforded by big data to prepare educational researchers and applications.

The increase in the volume, velocity, variety and volatility of data has made it extremely difficult to make meaningful analyses; classical methods of analysis have been found to be insufficient (Suma, 2020). Thus, different data analysis tools have begun to gain importance as new methods, such as artificial intelligence (AI), have been developed (Agrawal et al., 2019; Keitzman et al., 2018; Mishra et al., 2016; Rahman et al., 2016). AI is systems that could improve and renew themselves with data by imitating human intelligence (Wirth, 2018). AI can independently interpret data and think like humans to solve complex problems by learning patterns and making flexible adaptations (Almaghrebi et al., 2020; Kaplan & Haenlein, 2019). AI can operate in many different areas through its tools, one of them being machine learning (ML). ML, unlike other artificial intelligence applications, is a set of algorithms that imitates human intelligence without needing rules to be interpreted and entered manually. It can solve problems based on classification and regression algorithms, density estimation or clustering algorithms (Bennett & Parrado-Hernandez, 2006; Rahman et al., 2016; Ribeiro, 2016). ML is expected to completely change the business environment in the near future as using ML has a positive impact on strategic performance of businesses (Reis et al., 2020). ML techniques are used for many functions of a company; among these functions, marketing stands out because it aims to create value and satisfaction by focusing on phenomena such as need, want, demand, target customers, relationship management, customer attraction, sales and competition (Hwang et al., 2020; Khokhar & Chitsimran, 2019; Mariani & Wamba, 2020; Wirtz, 2020). ML techniques are helpful to design customized services and advertisement content for shared-home rental industry (Sengupta et al., 2021). Therefore, ML helps marketing by studying large amounts of data and their interactions to solve many inherent problems such as understanding consumer sentiments, service excellence, customer satisfaction, loyalty, market segmentation, customer communication, brand management, dynamic pricing, business sustainability, customized sales and promotion strategies, general sales as well as demand forecasting (Davenport et al., 2020; Leminen et al., 2018; Luo et al., 2021; Ma, Fildes and Huang, 2016; Mariani & Wamba, 2020; Mustak et al., 2021; Oztekin, 2017; Shahid & Li, 2019; Wirtz, 2020). The Marketing Management Support System (MMSS) allows managers to make better decisions, analyze data and present information with the help of AI.

Given the predictive ability of ML, one of the most prominent problems of marketing is to improve demand forecasting as this is vital for customer satisfaction, product planning and supply chain management (Boone et al., 2019; Carbonneau et al., 2008; Kharfan et al., 2020; Li et al., 2016; Zhu et al., 2021) and predict the stock market movement (Eachempati et al., 2021). ML uses several forecasting methods such as Neural Network, Fuzzy Neural Network, Artificial Neural Network, Recurrent Neural Networks, Genetic Algorithm, Support Vector Regression, Random Forest Regressor, Decision Tree Regressor etc. to tackle demand forecasting problems (Abbasimehr et al., 2020; Bünning et al., 2020; Carbonneau et al., 2008; Chang et al., 2009; Chen & Lu, 2017; Claveria et al., 2016; Feng & Wang, 2017; Qiu et al., 2017). Solving demand forecasting problems is not easy because there are many factors that affect demand; one of them is advertisements (Ali et al., 2009; Okrent & MacEwan, 2014; Zheng & Henneberry, 2010). Advertising plays an important role in attracting consumer attention to products, creating a positive attitude and increasing demand (Mukherjee & Banerjee, 2017). In ML applications, the use of very large datasets may reduce performance of the prediction as many noisy variables impact on the algorithm (Nguyen, Nguyen, et al., 2020; Van Nguyen, Zhou, et al., 2020). Moreover, while there is a vast array of literature on

sales and demand forecasting, it has been found that the effects of promotions, especially advertising, has been neglected (Di Pillo et al., 2013; Kumar et al., 2020; Rai et al., 2019). Besides, current studies generally examine the promotion or only the advertising effect on demand/sales using traditional methods (Divakar et al., 2005; Trapero et al., 2015; Van Don-selaar et al., 2016). A few studies have used ML methods to estimate demand based on promotions (Ferreira et al., 2016; Abolghasemi et al., 2020; Ali & Gürlek, 2020). The number of studies using ML methods and estimating demand by focusing on just advertisements is even smaller (Bollapragada et al., 2008; Güler, 2019; Rai et al., 2019). Therefore, to the best of our knowledge, there is a gap in current literature about predicting demand based on advertising data using ML methods. It is hoped that this study will help to fill this gap in academic literature and provide guidelines for businesses in successful demand forecasting that will assist them to make their marketing decisions more effectively.

It is noted that although there are several papers in literature on demand forecasting, there is a lack of studies about how marketing strategies such as advertising and other promotional activities affect demand (Di Pillo et al., 2013; Kumar et al., 2020). Therefore, an accurate demand forecasting model has significant academic and practical implications for business sustainability. Thus, the research motivation of this study is to explain how marketers can use ML to deliver more effective advertising decisions through demand forecasting. With this motivation, the aim of this study is formulating and evaluating ML methods to provide accuracy in demand forecasting based on advertising expenses. In this context, the following research questions have been structured:

RQ1: How can advertising expenses be used to make more accurate demand forecasting with ML?

RQ2: Depending on the advertising expenses, how can the accuracy of ML based forecasting methods be evaluated?

In order to answer research question 1, the demand forecasting performance of ML techniques will be investigated for situations in which advertising expenses are low and high. To answer research question 2, the accuracy of ML techniques will be compared for advertising expenditure levels and the best technique will be determined for each level.

The main contribution of this study is to propose a ML model in predicting demand based on advertising expenses for business sustainability. In addition, it is also analyzed that how advertising changes the performance of the prediction; in order to investigate this effect, the amounts of advertising expenses are considered. Thus, the performance of the proposed ML model is evaluated when advertising expenses are low and high. Therefore, it is examined that how the performance of ML models in demand forecasting changes when the data on advertising expenses also changes.

The study is structured as follows. Section 2, the literature review, provides an overview of related work with statements about research gaps that will be addressed. Section 3 covers the methodological framework and the analysis of the study. In this section, there is a description of how the framework can be applied to our case. Section 4 includes experimental results and analysis. Finally, in Sect. 5 research results are discussed; in Sect. 6, theoretical and managerial implications, and in Sect. 7, conclusion is outlined in terms of academic and managerial perspective.

2 Literature review

Demand forecasting is concerned with predicting demand for future periods based on real-time and historical data by using various factors related to demand signals usage (Ahmad & Chen, 2020; Ravikumar et al., 2005). Demand forecasting is one of the biggest challenges for retailers, wholesalers and manufacturers in any industry and it is a vital part of business intelligence (Kang et al., 2020; Khan et al., 2020; Pavlyshenko, 2019; Ren et al., 2019). Demand forecasting plays a very important role in marketing activities, especially sales & operations planning and supply chain management due to its demand volatility reducing effect (Ali et al., 2009; Bandara et al., 2019; Gilliland, 2010; Narayanan et al., 2019; Silveira-Netto and Brei, 2017; Fildes, et al., 2008; Khakpour, 2020). Accurate demand forecasting creates value in terms of predicting next purchases, avoiding under/over-stocking, sale opportunities, robust replenishment systems, new product planning, operational efficiency, customer satisfaction, competitive pricing and cost reductions (Chintagunta & Nair, 2011; Hu et al., 2019; Liu et al., 2013; Lu et al., 2012; Martinez et al., 2020; Trapero et al., 2015; Kumar et al., 2020; Yue et al., 2007). In order for businesses to reach their sustainability goals, they need to analyze and manage their data and information (Garg et al., 2019). ML techniques provide an effective method for sustainable businesses to identify and reduce waste, reduce cost, and increase time efficiency (Nishant, Kennedy and Corbett, 2020; Savarimuthu et al., 2020; Singhal et al., 2020; De Las Heras et al., 2020). In addition, ML can improve the decision processes in terms of both sustainable production and forecasting. Decision-making problems such as time, cost and resource constraints in the sustainable supply chain are tried to be solved by using machine learning techniques. In this way, ML is utilized while optimizing consumption and costing models to reduce the waste and unavailability of resources in supply chain management (Wang & Zhang, 2020).

The demand forecasting process uses data analytics to support the selection, training and transformation of unstructured and structured data types via different methods. Some of these methods are included under the heading of traditional demand forecasting methods. Traditional demand forecasting methods predict future values based on time series that analyze limited historical demand data. Some of them include Delphi method, grass roots forecasting, trend analysis, multi-regression model, moving averages, Croston's method, time series model, exponential correction and ARIMA (Alon et al., 2001; Bollapragada et al., 2008; Chen et al., 2000; Kharfan et al., 2020). These methods have been found to be effective only when factors such as nature trends and seasonal factors are not considered. In addition, since traditional estimation methods are based on assumptions such as sample size, linearity, stationary and distribution being normal, unsuccessful demand forecasting can occur when working with complex nonlinear large datasets. However, when the relationship between past time series data and demand/sales is complex and not linear (Jaipuria & Mahapatra, 2014; Nguyen, Nguyen, et al., 2020; Van Nguyen, Zhou, et al., 2020), forecasting becomes more difficult. Also, traditional methods lack the ability to dynamically update demand change as new data becomes available (Lawrence et al., 2006). Today, the business environment is very dynamic due to intense competition with many different factors affecting demand; a robust demand forecasting system may not be straightforward (Abbasimehr et al., 2020; Kraus et al., 2020; Schmidhuber, 2015). For these reasons, artificial intelligence-based methods such as ML are now being used for demand forecasting instead of classical methods to make more accurate predictions. Scientific studies on this subject are also continuing (Ali & Gürlek, 2020; Wolters & Huchzermeier, 2021).

Demand forecasts based on ML provide stronger estimations to businesses due to an ability to make better computation of a large number of data, high modeling capacity and nonlinear models as ML modeler can be tailored based on problem formulation (Bennett & Parrado-Hernandez, 2006; Carboneau et al., 2008; Huber & Stuckenschmidt, 2020; Kaneko, 2019). In this way, ML provides a model for unexplained and complex events (Di Pillo et al., 2013). Besides, using ML allows companies to focus on more value-added activities by halving the demand planner workload. It decreases demand estimation errors by 20% by increasing the service level to 98.6% (Bodenstab, 2017). Another situation that forces businesses towards demand forecasting is that demand is relative to many factors such as historic sales data, price, seasonal factors, promotion strategies, product category or demographics of the customer (Dave & Saffer, 2012; Ma et al., 2016; Okrent & MacEwan, 2014; Trapero et al., 2015; Yue et al., 2007). Among these variables, promotion highly affects demand as it includes price changes and shifts the demand curve by affecting consumer perceptions (Divakar et al., 2005; Yue et al., 2007). An advertising shift in the demand curve changes consumers' evaluations, price elasticity and influences willingness to pay (Bagwell, 2005; Dong and Kaiser, 2007; Okrent & MacEwan, 2014; Rosenthal et al., 2003; Zheng & Henneberry, 2010). Advertising achieves this through its persuasive, informative and complementary functions (Bagwell, 2007; Dave & Saffer, 2012).

When looking at studies where promotional activities—especially advertisements—are used in forecasting demand, it is seen that in general, traditional forecasting methods have been used (Divakar et al., 2005; Ma et al., 2016; Ramanathan & Muiyldermans, 2010, 2011; Van Donselaar et al., 2016; Van Heerde et al., 2002). However, a few studies using ML-based demand prediction methods have been used involving promotions as a variable (Ali et al., 2009; Di Pillo et al., 2013; Ferreira et al., 2015; Trapero et al., 2015; Makridakis et al., 2018; Abolghasemi et al., 2020; Aguilar-Palacios et al., 2019; Ali & Gürlek, 2020; Wolters & Huchzemeier, 2021). Other research has been conducted that aims to forecast demand based on advertising. Many of these studies have been conducted by using traditional forecasting methods (Brooker, et al., 1994; Dawes et al., 2018; Jedidi et al., 1999; Kronenberg et al., 2016; Zheng & Henneberry, 2010). However, sales and demand data generated by the effect of advertisements are irregular and nonlinear; in addition, advertising creates volatility in the whole demand series (Divakar et al., 2005; Huber & Stuckenschmidt, 2020). Therefore, as Bodenstab (2017) and Ali et al. (2009) have stated, ML is suitable for accurate forecasting demand arising from advertising and promotional activity. On the other hand, only a few researchers have used ML techniques by focusing on demand forecasting and advertising (Adnane et al., 2019; Bollapragada et al., 2008; Di Pillo et al., 2013; Güler, 2019; Khakpour, 2020; Rai et al., 2019; Shilpi and Sharma, 2016; Tugay and Oğuducu, 2017). Furthermore, Rai et al. (2019) have revealed that advertisements are effective in determining product demand by using Artificial Neural Networks, a tool of ML.

In this study ML methods are chosen as demand forecasting tools due to their high ability in understanding the nature of the data. ML methods can automatically analyses the patterns and dependencies in the input data and use this knowledge for prediction (Moroff et al., 2021). ML methods do not require previously accepted assumptions like parametric models. Therefore, they bridge the gap between parametric methods and non-parametric models (Bajari et al., 2015).

When traditional forecasting methods as opposed to ML models are considered, they are based on time-series dependencies. Traditional time-series methods assume that the future demand depend on the past demand. However this may not always be the case and demand depends on exogenous factors which are not characterized by past values (Cadavid et al., 2018). For example demand for driving services like UBER or Lyft cannot be represented by

past values since demand also depends on weather, time or day (Ke et al., 2017). Traditional methods provides simplicity and accurate results when the data is completely proper for the specific model (Yu et al., 2011). However, they are not successful in handling exogenous variables. From this point of view, ML can bring multi-perspective and it can handle complex data including time series, categorical variables, text, images, fuzzy elements and other variables (Cadavid et al., 2018). Especially LSTM networks better understand spatio-temporal features and correlations between input variables (Ke et al., 2017).

Several studies compared forecasting performance of ML models with traditional techniques and showed ML models bring superior performance over traditional methods. Cadavid et al (2018) found ML models outperforms various types of regression. Bajari et al. (2015) showed ML models provided better predictive accuracy when compared with linear and logistic regression.

Moroff et al. (2021) compared forecasting quality of six different models including statistical and ML methods. They found ML models bring better results in most of the products. Calkoen et al., (2021) also compared traditional and ML models and showed that ML methods were successful than Ordinary Least Squares (OLS) for most of the analysis. other studies also showed superior performance of ML methods over traditional forecasting models (Slimani et al., 2016; Yu et al., 2011).

Previous studies show that ML models can better capture exogenous variables and dependencies between input variables. This capability makes ML models perform better when compared to traditional forecasting methods as shown in previous studies. Superior capabilities and performance of ML techniques are the reasons for choosing them in demand forecasting in this study.

As can be seen from the literature review above, demand forecasting is crucial for marketing decisions. Besides, advertising is an important factor in creating demand and ML techniques are useful for more accurate demand prediction. However, there is a lack of studies on demand forecasting based on advertisements by using ML techniques. Accordingly, the research described in the next section was conducted to fill this gap.

3 Proposed methodology

In order to answer the research questions, a study has been designed based on the real-market dataset of a television manufacturing company. An attempt has been made to find an accurate model in predicting demand. Additionally, the study aimed to explore how advertising can change the performance of the prediction. To investigate this effect, the levels of advertising expenses will be examined. The performance of the proposed ML models will be analyzed when advertising expenses are low and high. It will then be possible to observe how the performance of ML models in demand forecasting change when the advertising expenses change. The steps of the proposed methodology are shown in Fig. 1 with steps of the methodology explained in more detail in the following sections.

3.1 Data preparation

Time series data should be pre-processed at the outset before being used in a ML analysis. Missing value analysis and data normalization are two crucial tasks that are used to prepare data for the forecasting calculations. It is also useful to analyze whether the data is stationary or not is also a useful task, since stationary data has a time dependent component which can

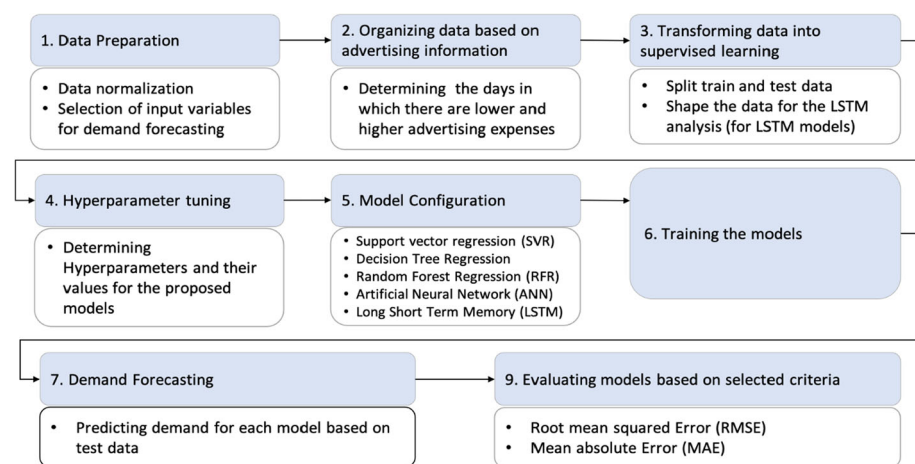


Fig. 1 Steps of the proposed methodology

affect the accuracy of forecasts. After conducting missing value and stationary analysis, input variables to be used for demand forecasting will be selected. Next, a data normalization task will be applied for further analysis.

Selecting the independent variables to predict a dependent variable is an important task in forecasting models. This task can be difficult due to various reasons—the number of available variables is too large, there is a relationship between the input variables or some of the variables can have little prediction power (May et al., 2011). Choosing the right input variables not only increases the predictive power of the model, but also decreases the necessary computing duration since computations will be conducted with a smaller number of variables.

Neural Network (NN) models deal with big data and include massive amounts of variables. The rise in popularity of NN models has increased the necessity of efficient input variable selecting models (May et al., 2011). Therefore, variable selection has become a critical step in ML applications (Nguyen, Nguyen et al., 2020; Van Nguyen, Zhou, et al., 2020).

In this study, a stepwise regression method was used to select the best input variable combination. Stepwise regression is a regular technique that aims to find input variables which do not have a statistically significant explanatory effect on output variables (Jiao & Li, 2010). Among the selection types, the one which best combines forward and backward selection properties of regression is chosen. The following paragraphs describe the types of selection methods.

Forward Selection: In forward selection, the method starts with no input variables and one variable which has the best selected criterion value added in each step of the regression (Yamashita et al., 2007). Statistical criteria may have significance value of F statistic showing that the input variable has an influence on the predictive power of the model (Massart et al., 2003).

Backward Selection: Backward selection as opposed to forward selection is a top-down technique that starts with all the input variables and in each step of the regression, this method eliminates the one variable that brings the worst criterion value (Meyer-Baese & Schmid, 2014).

Stepwise Selection: In this method, a combination of forward and backward procedures is used. In each step of the regression after a new input variable is added, the model is checked to see whether some variables having non-significant effect on the prediction can be removed. The stepping process ends when there is no variable left to add or remove (Jiao & Li, 2010). Stepwise selection has been previously used by studies to select variables to predict a model with ANN technique (Jiao et al., 2013). Since stepwise selection applies both forward and backward properties, in this study it is chosen as the variable selection technique.

Before the analysis, each variable in the time series data should be scaled to be used in ML algorithms. Data is scaled based on the minimum and the maximum value in each variable and computed as follows:

$$v_n(t_i) = \frac{v(t_i) - \min(v)}{\max(v) - \min(v)} \quad (1)$$

In this study, it is wanted to explore the effect of advertising on the sales prediction performance with ML techniques. In order to elaborate this effect, selection of the parts from the dataset as having low and high advertising expenses is necessary. The selection process is explained in Sect. 4.2.

Time series data should transform in a way that is proper for supervised learning. This transformation is necessary to accurately apply forecasting methodology. The first step in transforming data into supervised learning is splitting data into train and test sets. The second step is to shape the data to represent the necessary time lag in LSTM analysis.

Datasets should be split into train and test sets to be analyzed with ML methods. The train set is used for the ML algorithm to learn the data and conduct a model. The test set is used to make predictions based on the conducted model with the train set.

For LSTM analysis, data should be appropriately shaped first. Lag size, which shows the time step in LSTM, is used to convert the data into a set of input and output shapes. The variable to be predicted is reshaped in a way that is used as output and as one of the input variables including observations from the previous time step. This form of restructuring is necessary to run a LSTM model. In this study lag size is used as 1. Based on this restructuring, the input variable and output variable to be predicted is depicted as follows:

$$v_{input} = [s(t_1), s(t_2), \dots, s(t_{n-1})] \quad (2)$$

$$v_{output} = [s(t_2), s(t_3), \dots, s(t_n)] \quad (3)$$

3.2 Hyperparameter tuning

Hyperparameters are used in the design of ML algorithms and have an important effect on the model performance. A model that runs with a tuned method can learn tasks better than a model that runs with a non-tuned method (Amirabadi et al., 2020).

LSTM networks are recognized as having a promising performance in times series prediction; the training quality of these networks heavily depends on the selection and tuning of many hyperparameters (Reimers & Gurevych, 2017). In this study, the number of hidden layers, the number of neurons in each layer, the epoch size and batch size are chosen to be tuned to improve the performance of NN methods.

Number of hidden layers: For NN, some studies recommend a low value for better performance (Hewamalage et al., 2021). The reason for this is the over fitting effect due to the increased number of parameters and increased computational complexity when there are high numbers of hidden layers (Hewamalage et al., 2021).

Number of neurons in each layer: When the number of neurons is large, the method is expected to give precise results. However, when the number of neurons is too large, the model is faced with the over fitting risk along with a decrease in practicality (Ismailov, 2014).

Epoch size: An epoch is defined as one full forward and backward run throughout the whole dataset. The number of epochs resembles how many times such runs occur in the analysis (Hewamalage et al., 2021). If the number of epochs is too small, the model will not learn enough from the training data. Additionally, if the epoch size is too large, the model will overfit the training data resulting in poor performance for the testing set (Abbasimehr et al., 2020).

Batch size: As stated above, one epoch occurs when the whole data set is passed through the network. Since the entire dataset is too large, it should be divided into several smaller batches. Batch size is the total number of training observations in a single batch (Amirabadi et al., 2020).

Other ML methods utilized in this study also have hyperparameters to be tuned. In the next section, hyperparameters of each ML method, other than NN models, are explained.

3.3 Model configuration

Applied models in this study are configured based on the proposed hyperparameter combinations. With each parameter combination, each applied model is trained and the combination that yields the best performance is selected. In this section, the utilized models in this study are described.

3.3.1 Support Vector Regression (SVR)

SVR is a regression version of support vector machines which indirectly map inputs into a higher dimensional space with kernel functions (Ali et al., 2009). SVR aims to find a linear function in the stated higher dimensional space and compute distances between the linear function (error terms) and the mapped output values (Ali et al., 2009). The algorithm penalizes deviations from the linear function when they are more than a predetermined value (Ribeiro et al., 2011).

The hyperparameters mostly used in SVR are the kernel function and the C parameter which controls the tradeoff between the generalization performance and the sum of deviations value (Ali et al., 2009). The type of kernel function and C parameter to be tuned for SVR is selected.

3.3.2 Decision tree regressor (DTR)

A decision tree constructs predictions by partitioning the input space recursively into subspaces; these establish a basis for prediction (Rokach, 2016). A tree has three basic components—a root node for the entire dataset, internal nodes and terminal nodes (Xu, et al., 2005). Internal nodes are also called splits. These have outgoing edges causing the tree to spread; terminal nodes, so called leaves, do not have continuing edges (Dev & Eden, 2019). Each node of the decision tree makes the decision of separating a class from the remaining ones (Rokach, 2016).

DTR is a type of decision tree classifier used for predicting numerical values. The splitting continues until the tree reaches a predetermined value of minimum node size and the recent nodes become terminal (Xu et al., 2005).

In this study, three parameters are chosen to be tuned for DTR; these are maximum depth of decision tree showing maximum number of levels in the tree, the minimum number of samples necessary to split an internal node and the minimum number of observations required to be at a terminal node.

3.3.3 Random forest regressor (RFR)

RFR is a variant of the decision tree method. This method aims to overcome any weaknesses of DT such as sensitivity to small variations in the data and to improve the prediction accuracy of DT (Loureiro et al., 2018). In random forest, many trees are formed by using a subset of randomly selected features; the mean of predicted values at the terminal nodes of each tree is calculated (Loureiro et al., 2018).

In this study, the number of trees to construct, maximum depth of each tree and the minimum number of observations required to be at a terminal node are chosen as hyperparameters for tuning.

3.3.4 Artificial neural network (ANN)

ANNs are developed to imitate human brain in the learning activity. ANN is fault tolerant systems which can learn from the given information, have ability to handle complex problems such as non-linear data and they are able to perform predictions with high speeds (Kalogirou, 2000). The smallest unit of an ANN is a perceptron, having a single input and an output layer (Güven & Şimşir, 2020). ANN is popularly used in many applications such as robotics, manufacturing, pattern recognition and forecasting (Kalogirou, 2000). Deep learning fields today use ANNs that have multilayers, using the perceptron as the building block (Güven & Şimşir, 2020). Figures 2 represent the structure of a perceptron.

A traditional Multilayer Perceptron (MLP) network includes an input layer which receives external inputs, at least one hidden layer and an output layer which creates predictions (Yan et al., 2006). Data enters the system at the input layer, calculations are performed in the consecutive layers and stops in the output layer in which the predicted value is produced (Yan et al., 2006).

Fig. 2 Traditional Feed Forward ANN model

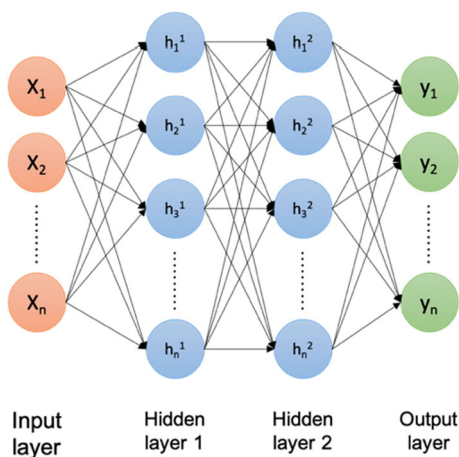
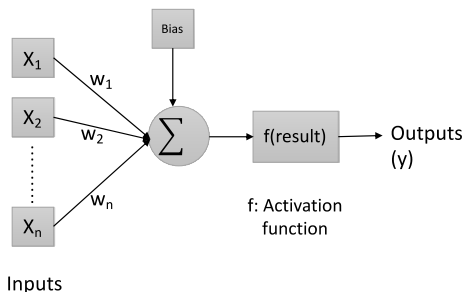


Fig. 3 A perceptron structure

Inputs ($x_1, x_2, x_3, \dots, x_n$) are information that enter the network from external environments. Weights ($w_1, w_2, w_3, \dots, w_n$) are values that express the effect of inputs or another processing element in the previous layer on the current layer (Yakak, et al., 2018). Each input is multiplied by its own weight and then these products are summated to be processed into a predetermined activation or transfer function.

The most popular NN is the feed forward architecture using a back propagation learning algorithm. In feed-forward architecture, data moves from the input layer to the output layer. In the meantime, weights at each layer are forward ANN model respectively, determined (Øyen, 2018). The difference between the target values and the actual values in the output layer is calculated. Computed differences between the target and actual values are sent back, weights assigned to each input are updated and the model runs again. This iterative process continues until the error reaches a pre-determined threshold (Reese, 2021). In the back propagation learning algorithm, the weights determined in the layers are updated backwards until the calculated differences are minimized by the gradient descent or similar optimization function (Azzouni & Pujolle, 2017). Figures 2 and 3 represent the structure of a perceptron and a traditional feed.

3.3.5 Long short-term memory networks

Recurrent neural networks (RNN) is a special type of NN that has strong capability in time series forecasting (Shi et al., 2019). The basic difference between a traditional RNN and a LSTM network lies in the LSTM's ability to store time-dependent information for long sequences (Nakisa et al., 2018). This ability allows LSTM to better map between input and output data (Abbasimehr et al., 2020). Therefore, among the evolving NN methods, LSTM is found to be remarkably successful to model time series data in which each observation depends on a specific time (Adam et al., 2018).

The basic unit of a LSTM network is different from a traditional perceptron since it has a cell and some gates to manage information flow. In detail, A LSTM unit contains a cell state, an input gate, a forget gate and an output gate (Greff et al., 2017). Figure 4 represents a basic LSTM unit.

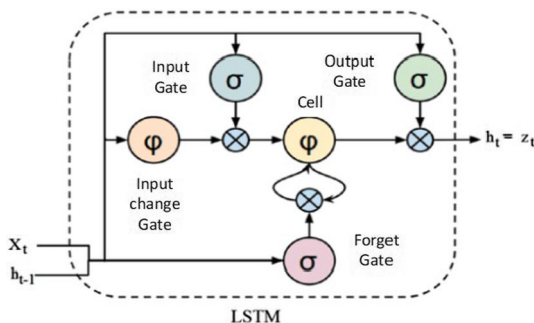
Computations for LSTM elements are as follows (Bui, et al., 2018; Kang, 2017):

$$f_t = \sigma * (w_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma * (w_i[h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{c}_t = \tanh * (w_c[h_{t-1}, x_t] + c) \quad (6)$$

Fig. 4 LSTM Cell—Source: (Kang, 2017)



$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (7)$$

$$o_t = \sigma * (w_0[h_{t-1}, x_t] + b_0) \quad (8)$$

$$h_t = o_t * \tanh(c_t) \quad (9)$$

where f_t is the forget gate, i_t is the input gate, \tilde{c}_t is the input change gate and calculated for cell state equation, c_t is the cell state, o_t is the output gate, h_t is the hidden layer and b is the bias: σ and \tanh represent the activation functions of sigmoid and hyperbolic tangent respectively.

3.4 Training, demand forecasting and the evaluation of the models

With the selected hyperparameter combinations, each model is trained with the training set. Then, information learned in the training phase is used to conduct demand forecasts for the testing set. The best configuration for each model is chosen based on Root Mean Squared Error (RMSE). The best configurations of ML methods are compared with each other based on RMSE and Mean Absolute Error (MAE). RMSE and MAE are calculated respectively as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (10)$$

$$MAE = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{n} \quad (11)$$

where t represents the observation number depending on time, y_t is the real value of the output variable on time t and \hat{y}_t is the predicted value of the output variable on time t .

4 Experimental results and analysis

Five ML models including SVR, DTR, RFR, ANN and LSTM are utilized in this study. The last two of these models, ANN and LSTM, are NN models which may have deep learning architecture depending on the number of hidden layers used. The effect of NN on predicting time series data can be observed when the structure of the data changes depending on the selected feature information. More specifically, it observes how the demand forecasting performance of ML models changes depending on the increased amount of advertising expenses.

It is therefore valuable to test whether NN models reveal a notable performance when the high advertising expense information is embedded into the data.

All of the proposed models and the stepwise regression analysis were performed with Python 3.6. SVR, DTR and RFR were implemented with the Scikit-learn package in Python. ANN and LSTM were executed using Keras in Python. Stepwise regression was conducted using the Statsmodels package in Python.

The following part of the paper will describe the dataset. The next section will explore how the dataset is used to understand the effect of advertising on the demand forecasting performance. The results of each method will then be given.

4.1 Dataset description

Products that require technological advancements have high variation in their demand due to rapid changes in specifications and price; this makes demand forecasting a difficult task (Lu et al., 2012). A TV manufacturing company was chosen to conduct demand forecasting since a company such as this plays an important role in the technological production industry. Besides, TV manufacturing is a technology-based industry and as San Kim and Sohn (2020) emphasized technology-based industries has fluctuating and competitive environments.

Proposed models on a time series data that was recently used by Kumar et al. (2020) have been tested. This dataset has been previously used by James Rawlins for the TV manufacturing company and can be downloaded from his Github repository (www.github.com). The dataset has 17 features; these are daily information of demand, sales in dollars, POS, unit price, consumer price index (CPI), consumer confidence index (CCI), producer price index (PPI), advertising expenses for SMS, newspaper, radio, TV, internet and gross rating points (GRP) for the stated advertising types. The dataset has 2614 daily observations from January 1st 2010 to February 25th 2017. The aim of GRP is to measure the effectiveness of each advertising channel; this is calculated as a weighted sum of the number of ads aired for a TV company or brand in a specific week (Kumar et al., 2020). Weights are the rating points of a specific rating agency on radio or TV during the time when the ads are aired (Kumar et al., 2020).

Among the 17 variables, the aim is to forecast demand. Therefore, demand is used as the target variable for all the utilized models. Before starting the analysis, it is necessary to prepare the time series data to be used in the methods.

4.1.1 Preparing data

Missing value analysis is not needed since the time series data used in this study has no missing values for all 17 variables. The next step to be taken for a time series data preparation is to determine the degree of stationary. An Augmented Dickey Fuller (ADF) unit root test showed that the target variable is stationary as shown in Table 1.

The next step for data preparation is variable selection for demand forecasting. Stepwise regression is used to select which variables have the highest impact on the target variable. Among the 16 independent variables, 8 of them are selected in stepwise regression to be used in demand forecasting. The steps and the results of stepwise regression are shown in Table 2. These 8 variables are Sales (\$), Consumer Price Index (CPI), POS/ Supply Data, Advertising Expenses (Internet), Unit Price (\$), Advertising Expenses (TV), Advertising Expenses (SMS) and Consumer Confidence Index (CCI).

Descriptive statistics for the selected variables are shown in Table 3.

Table 1 Stationary analysis results

	Value
ADF Statistic	– 3.433
<i>p</i> value	0.034205
Critical Value 1%	– 2.863
Critical Value 5%	– 2.567
Critical Value 10%	– 2.567

Table 2 The steps and the results of stepwise regression

Steps	Included Variable	Excluded Variable	<i>P</i> value
1	Point of Sales(POS) Data		0.0000 < 0.05
2	Consumer Price Index (CPI)		0.0000 < 0.05
3	Sales (\$)		0.0000 < 0.05
4	Advertising Expenses(Internet)		0.0000 < 0.05
5	Unit Price (\$)		0.0003 < 0.05
6	Advertising Expenses(Radio)		0.0069 < 0.05
7	Advertising Expenses(TV)		0.0045 < 0.05
8	Advertising Expenses (SMS)		0.0000 < 0.05
		Advertising Expenses(Radio)	0.1589 > 0.05
9	Consumer Confidence Index(CCI)		0.0359 < 0.05
10	Loops have ended since there are no variables with <i>p</i> values < 0.05		
Selected Variables based on stepwise regression	['POS/ Supply Data', 'Consumer Price Index (CPI)', 'SALES (\$)', 'Advertising Expenses(Internet)', 'Unit Price (\$)', 'Advertising Expenses(TV)', 'Advertising Expenses (SMS)', 'Consumer Confidence Index(CCI)']		

After selection of the input variables, data is normalized to be used in ML methods. The data was normalized using min–max scaler as shown in Eq. 1.

4.2 The effect of advertising on the demand prediction

In this study is to establish the effect of advertising on the demand prediction performance with ML techniques. To elaborate this effect, the dataset needs to be classified based on total advertising expenses as low and high. To construct this division, firstly total advertising expenses is calculated by summing expenses of all advertising types for each day. Then the data is sorted in ascending order based on total advertising expenses. To separate the data into

Table 3 Descriptive statistics for the variables in the models

	Demand	SALES (\$)	Consumer Price Index (CPI)	POS Data	Advertising Expenses (Internet)	Unit Price (\$)	Advertising Expenses (TV)	Advertising Expenses (SMS)	Consumer Confidence Index(CCI)
Number of observations	2613	2613	2613,00	2613	2613	2613	2613	2613	2613
Mean	5021	1,641,507	102,61	4523	3079	363	1,325	60	103
Std. Deviation	2681	941,667.3	1,38	2604	1521	26	124	14	3
Minimum	1610	462,709.6	101,30	1510	0	282	1067	38	96
Maximum	18,565	5,960,221.0	106,50	16,482	6355	400	1479	90	108

Table 4 F-Test: Two-Sample for Variances

	Low Ad expenses Group (Q1)	High Ad expenses Group (Q3)
Mean	2425,400,795	6274,31,072
Variance	388,516,802	291,278,2142
Number of Observations	652	653
F	1,333,834,057	
P value one-tail	0,000,121,901	
F Critical one-tail	1,137,661,401	

Table 5 t-Test: Two-Sample Assuming Unequal Variances

t Statistic	– 119,2,381,117
P value (one-tail)	< 0.01
t Critical one-tail	1,646,048,676
P value (two-tail)	< 0.01
t Critical two-tail	1,961,824,866

low and high advertising expenses, a similar method applied in the study of Wattanakamolchai (2016) was used.

The data was divided into four quartile groups based on values of the total advertising expenses. The highest and lowest quartile groups were used to represent the days that have high ($n = 652$) and low ($n = 653$) expenses respectively. In order to see whether low and high expenses groups significantly differed from each other, a t-Test has been applied. Before the t-Test, a F-test was applied to see if the two groups have equal or unequal variances. F-test results are shown in Table 4, indicating that compared groups have unequal variances. Now two sample t-Tests assuming unequal variances can be executed. The t-Tests indicate significant differences between the low advertising and high advertising groups. This is shown in Table 5.

After deciding the low and high advertising groups, corresponding days in each group are selected with each group transformed into time-series data. Demand forecasting has been concluded with the proposed ML algorithms for each part to see the effect of advertising.

4.3 Transforming data into supervised learning

In this study, a ML method will be applied to two groups of data including low and high advertising expenses. Instances in each group should be split into train and test sets. For both groups, the first 80% of data is used as the train set, while the remaining 20% is used as the test set.

For LSTM models, it is necessary to shape the data to represent time lag in LSTM analysis. Lag size is selected as 1. A subset of the formed instances is shown in Table 6.

Table 6 A part of the formed instances

Time	Demand($t-1$)	Demand(t)
1	4384	4366
2	4366	4006
3	4006	4076
4	4076	4834

4.4 Identifying the hyperparameters and training the model

For all the utilized methods, a list of the values of the chosen hyperparameters is initially determined. The possible values of the hyperparameters for the ML models can be seen in Table 7. For every parameter combination, a ML model is trained.

In ANN and LSTM models, Adam is used as the optimization algorithm. The reason for that is because the Adam optimizer is expected to decrease the error faster than the stochastic gradient algorithm (Kingma and Ba, 2015). Additionally, for all designed ANN and LSTM models, MAE was used as the loss function.

The designed models for every hyperparameter combination are evaluated on the test dataset using RMSE parameter. Based on the RMSE results, the best model for each ML technique is selected. The selected hyperparameters that reveal the best performance can be seen in Table 8.

Table 7 Candidate values for the hyperparameters

Model	Hyperparameters	Values
SVR	Kernel function	Radial Basis Function (RBF), Linear, Polynomial, Sigmoid
	C Parameter	0.1, 1, 10, 100, 1000
DTR	Maximum depth	[10, 20, 30, 40, 50, 60, 70, 80]
	The minimum number of samples in an internal node	[0.5, 2, 4, 6]
	The minimum number of observations at a terminal node	[1, 2, 4, 6]
RFR	Number of trees	[400, 600, 800, 1000, 1200, 1400, 1600]
	Maximum depth	[10, 20, 30, 40, 50, 60, 70, 80]
	The minimum number of observations at a terminal node	[1, 2, 4, 6]
ANN, LSTM	Number of Hidden layers	[1,2,3]
	Number of Neurons	[4, 8, 16, 32, 64]
	Epoch size	[50, 100, 250, 500, 750]
	Batch Size	[1, 5, 10, 25, 50, 75, 100]

Table 8 The best hyperparameters for the utilized models

Model	Hyperparameters	Values	
		Low Ad Expenses	High Ad expenses
SVR	Kernel function	RBF	RBF
	C Parameter	10	10
DTR	Maximum depth	10	10
	The minimum number of samples in an internal node	0.5	2
	The minimum number of observations at a terminal node	1	6
RFR	Number of trees	1000	1600
	Maximum depth	30	60
	The minimum number of observations at a terminal node	6	6
ANN	Number of Hidden layers	1	1
	Number of Neurons	4	32
	Epoch size	750	50
	Batch Size	5	1
LSTM	Number of Hidden layers	1	1
	Number of Neurons	8	8
	Epoch size	50	250
	Batch Size	25	25

4.5 Findings

The accuracy of the predictions is assessed with metrics of RMSE and MAE. RMSE and MAE values are estimated by using the best hyperparameter combination for each model. Additionally, to determine the impact of advertising on the demand prediction performance of the models, RMSE and MAE values for the low and high advertising expenses groups were reported separately. RMSE and MAE values are shown in Table 9.

4.5.1 Comparing the utilized models

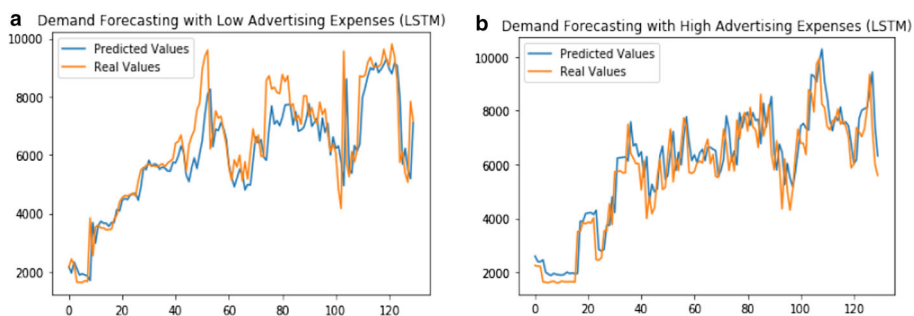
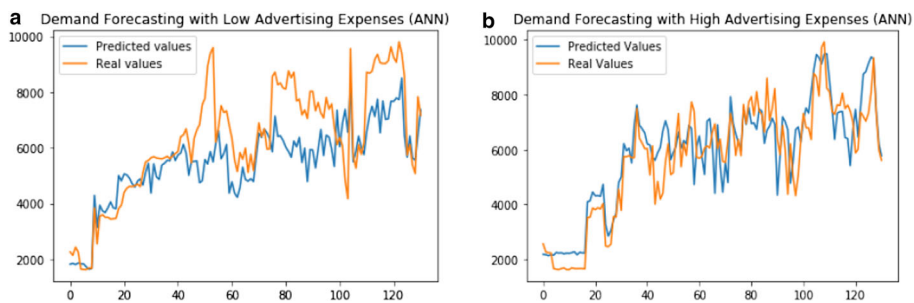
According to Table 9, LSTM network gives the lowest RMSE and MAE, meaning the highest accuracy in prediction when compared with the other four models, SVR, DTR, RFR and ANN. LSTM showed best accuracy for both the low and high advertising groups. Second best model for the two advertising groups is ANN, the other NN method in this study. The two NN models outperformed SVR, DTR and RFR with regard to both RMSE and MAE. Figures 5a and b show how actual values and predicted values of LSTM for the testing dataset changes in the low and high advertising group respectively. Similarly, forecasting results of ANN jointly with the real values is also shown in Figs. 6a and b for both the low and high advertising groups. These figures show that the lines of predicted values are closest to the real values when demand is predicted with LSTM for the high advertising expenses group.

The results demonstrate the power of NN in forecasting with time-series data when compared with other ML models. In particular, the LSTM model showed an outstanding

Table 9 RMSE and MAE values for the utilized models

Method	Low Ads		High Ads		Change Between Low and High Ads	
	RMSE	MAE	RMSE	MAE	Improvement in RMSE (%)	Improvement in MAE (%)
SVR	1473.96	1166.37	1056.29	879.76	28	25
DTR	1581.91	1196.18	956.71	716.94	40	40
RFR	1678.37	1308.17	953.96	723.81	43	45
ANN	1365.38	1035.64	918.61	686.54	33	34
LSTM	971.74	686.54	791.93	604.56	19	12

*The lowest RMSE and MAE values mean the highest accuracy

**Fig. 5** a LSTM results for Low Ad Group. b. LSTM results for High Ad Group**Fig. 6** a ANN results for Low Ad Group. b. ANN results for High Ad Group

performance in demand forecasting. The reason for this lies in LSTM's capability in times series forecasting while storing time dependent information for long sequences (Nakisa et al., 2018).

4.5.2 The effect of advertising expenses on demand forecasting

The aim of this study is to evaluate how prediction performance is affected by advertising expenses. This effect is examined by comparing RMSE and MAE results between low and high expenses groups. All ML techniques showed that the prediction performance increased when advertising expenses are high. Improvement in prediction performance for all ML models can be observed in Table 8. Improvements range between 19 and 43% for the RMSE metric. LSTM revealed 19% improvement in RMSE when high advertising expenses are spent, while RTM showed a 43% increase. This high improvement in RFR when there is more information in the data about advertising expenses can be explained by the multi tree structure of RFR. RFR conducts modeling with multiple trees interconnected to each other with internal nodes. This multiple tree structure of RFR enables the model to handle and analyze the complex data better.

Although RFR showed the highest improvement in predicting demand when high advertising expenses are spent, when the overall prediction accuracy is observed, LSTM outperformed all the other ML models in the high group. Second best model is ANN while RFR comes third as shown in Table 8. When the prediction accuracy in the low advertising expenses group is examined, RFR is the worst ML model while LSTM is the best. This result shows that LSTM can be used to predict demand when advertising expenses are either high or low, while RFR works better when advertising expenses are high.

5 Discussion

In this study, demand was predicted using ML techniques with a dataset containing advertising expenses and sales; it was determined which model gave the best demand forecast. Additionally, in this study the effect of advertising expenses on the accuracy of demand forecasting is explored. To obtain forecasts of demand with different advertising expenses, five ML techniques (SVM, DTR, RFR, ANN and LSTM) have been used. ML techniques have been chosen for demand forecasting since previous studies have shown that ML techniques are superior to traditional statistical techniques in time series demand forecasting (Bajari et al., 2015; Carbonneau et al., 2008; Di Pillo et al., 2013; Rai, et al., 2019).

The overall findings showed that ML techniques can be successfully applied in demand forecasting. NN methods yield best prediction performance, especially LSTM; this shows superior performance when compared with other ML techniques in time-series forecasting. This finding is consistent with previous studies using LSTM for time series forecasting. Kong et al. (2019) used LSTM in short-term load forecasting and found that it outperformed other algorithms such as ANN and KNN. Muzaffar and Afshari (2019) also showed that LSTM predicts electrical load time series with less error than traditional statistical methods.

In this study, LSTM's highest accuracy results in demand forecasting are also consistent with other studies using LSTM for demand prediction. Xu et al. (2018) used LSTM to predict bike sharing demand over time and found that LSTM yielded best prediction results when compared with both traditional statistical methods and several other ML techniques. Bandara et al. (2019) used LSTM to predict e-commerce demand for Walmart. They found that LSTM showed a superior performance over traditional demand forecasting techniques. Abbasimehr et al. (2020) found that forecasts made with LSTM for the demands in a furniture company outperformed several techniques including statistical and ML algorithms. The research of

Husna et al. (2021) which conducted on predicting online grocery sales forecasting, shows that LSTM performs better than CNN and ANN.

The findings of this study also show that changing the structure of the data can improve prediction accuracy for ML techniques. Analysis showed that using more detailed input data increases prediction accuracy in demand forecasting. When advertising expenditure is high, demand accuracy improves dramatically for all ML models. This finding suggests that when advertising expenses increase, ML techniques model the data better and make more accurate predictions. This finding is also consistent with some previous studies. Ali et al. (2009) found that ML models predict demand more accurately than statistical models in the weeks with promotional activities. Similarly, Ma et al. (2016) showed that in non-promotion weeks the statistical models performed better when compared to promotion weeks. The authors did not test the performance of ML methods in their study.

6 Implications

6.1 Theoretical implications

This study contributes to the literature of marketing to determine how to measure the effect of advertising and media expenses on demand, which have a significant share in marketing expenditures, using ML techniques. In this way, with the help of ML techniques used in the study, data interpretation in marketing will be facilitated, the understanding of increasing market opportunities will be enriched and improvement in marketing decision-making processes will be achieved. It also shows that more accurate predictions can be made which will enhance business sustainability. The potential impacts on sustainability in terms of cost savings, reduction in CO2 emissions, energy and time savings can be achieved with accurate demand prediction by ML.

6.2 Managerial implications

Findings of the proposed model contribute to the advertising sector in understanding under which advertising environment ML demand forecasting methods work best. Results of the study demonstrated that when advertising expenses are high, prediction accuracy improves considerably for all ML models. This finding indicates that when advertising expenses increase, ML techniques understand the nature of the data better and make more accurate predictions. Our study compared the performance of the ML methods in two different advertising settings which are high and low advertising expenses. Results showed that for all the ML techniques, the prediction performance increased when advertising expenses are high.

The results of this research will contribute to policy makers and managers. The advertising sector can benefit from this study when making the right decision on the amount of advertising expenses. Since advertising is a promotional activity with a high cost, companies want to reduce advertising costs and maximize the effect of advertising on sales and demand. The results of the study will also support most marketers in making the right decision in terms of measuring the return on advertising investments and determining the effect of different advertising expenditures on demand. For this reason, the results of the study are important in terms of evaluating the return of advertising expenses and the effect on demand.

According to the results of this research, if the firm estimates the demand correctly and thinks that there are exogenous variables affecting the demand, it may decide to spend more

on advertising. Spending more on advertising not only increases sales, but with higher advertising expenses, ML techniques lead to accurate results and aid for future planning. If the firm has the available resources and large amount of data to be analysed with ML techniques, the firm may decide to spend more on advertising. On the other hand, if the firm does not have sufficient financial resources and technical infrastructure, statistical methods can be used for demand forecasting, since the amount spent on advertising will be low.

The demand forecasting is also crucial for business sustainability. The lack of demand forecasting can be explained in two folds. First, if the forecast is less than demand then there will be shortage causing a need for backordering which will cause inevitable excess transportation and handling operations causing additional carbon emission, traffic congestion and cost. Therefore, the consequences of poor demand forecast will embrace all dimensions of Triple Bottom Line, environmental, social and economic. On the other side of the continuum, if the forecast is greater than the demand then the excess inventory will accumulate causing an inefficient use of resources such as energy, workforce and cost. Similar to shortage of inventory, in the case of excess inventory the aforementioned consequences are also covering the all three pillars of sustainability.

The proposed methodology can be used by policy making authorities, such as governments and municipalities, to increase the efficiency of their promotional activities on social issues and social marketing activities. In this way, they can use their limited budgets more efficiently and determine their optimal promotion mix as Zhou et al. (2018) emphasized in their research. As Aaamer et al. (2021) emphasized, from a managerial point of view, ML-based demand forecasting methods can create value in the field of marketing by giving companies the ability to achieve better, faster and more accurate forecasts of their future sales and remove uncertainty in demand forecasting. In this way, companies will be able to increase their profitability through revenue growth, cost reduction and productivity increase. In addition, these methods will be able to support companies in optimizing their supply chains, improving sales and marketing operations and maintaining balanced stock levels through accurate demand forecasting.

Within the framework of the findings obtained as a result of the study, marketing experts and advertising agencies can plan more effective advertising strategies by using their marketing communication budgets more rationally via ML techniques. ML can also be suggested to analyze and compare companies that use different advertising channels such as Internet, radio, TV, SMS and digital advertising campaigns.

7 Conclusion

ML methods are popularly used in marketing data in recent years to understand customer data better and to obtain meaningful insights. In this study the effect of advertising on demand forecasting performance is analyzed with using ML techniques. In other words, demand forecasting performance of ML methods are compared based on the amount of advertising expenses. Five ML techniques are used for demand forecasting which are SVR, NN, LSTM, RFR and DTR. Proposed techniques are used for time series demand forecasting according to the levels of advertising expenditure based on the actual market dataset of a single TV manufacturer. The proposed methods are compared with each other in terms of both RMSE and MAE. Overall results showed that LSTM outperformed ANN, SVM, DTR and RFR for both low and high advertising groups. ANN showed second best performance for the two advertising groups. The other notable finding of the study is that the accuracy of the prediction

increases as advertising expenses increase. For all ML methods, prediction accuracy is best when advertising expenses are high. This result demonstrates that as the detail in the input data increases, the performance of the ML techniques improves substantially.

Nowadays, the use of ML in marketing has become important for complex big data, transforming information into meaningful customer insights and making strategic marketing decisions for business sustainability. These models are used to make realistic forecasting using multiple data such as pricing, sales promotions, advertising, store inventory and in-store sales and demand (Mishra et al., 2016; Shah et al., 2020). Advertising is a promotional activity that draws the attention of current and potential customers to a company's products or services, gives information, spreads messages and convinces the target audience to purchase their products/services (Karimova & Shirkhanbeik, 2019). However, advertising is a promotional activity with a high cost and needs a large investment. Therefore, companies aim to reduce the cost of advertising while maximizing the effect of advertising on both sales and demand (Wei et al., 2021). In addition, it is difficult to measure the return on investment of advertising and to determine whether the advertising investment has an effect on demand (Kim et al., 2005). Hence, the aim of this study is to provide accuracy in demand forecasting based on advertising expenses with ML techniques. Therefore, in this study, demand forecasting was conducted by using ML techniques over a dataset containing advertising expenses and sales; a best model was thus determined.

There are some limitations and future research suggestions. One of the limitations of the study is the real market dataset of a single TV. Therefore, first future study, different datasets from different sectors can be used and the same ML demand forecasting methods can be applied to these datasets. Additionally, by using ML methods, in order to achieve business sustainability, it is possible to examine whether there is a difference in results by forecasting demand in different sectors such as service sector on transportation, health industry, fashion and retailing, where the demand structure may be volatile, dynamic and uncertain. In addition to industrial differences, different ML techniques can be used by making differences in the models used in the method; even the accuracy of the prediction can be tested with hybrid models such as fuzzy neural networks. The other limitation of this study has focused on a forecasting mechanism based on ML techniques—SVR, NN, LSTM, RFR, DTR—to deal with demand forecasting based on advertising expenses. Apart from these ML techniques, second further research is that the performance of models can be compared by making demand forecasting studies based on neural networks. According to a second future research proposal, the performance of the models can be compared by performing demand forecasting studies based on neural networks as well as ML techniques.

Third limitation is that demand was forecast using ML techniques with a dataset containing advertising expenses and sales. The dataset was limited, therefore the effect of other variables could not be considered. In third future study, other marketing indicators such as customer profile data, other marketing mix elements or other corporate marketing communication expenditure data can be included; deep learning algorithms that perform better with larger datasets can be introduced. Additionally, various ML techniques can also be used to forecast different areas such as customer future behavior. For instance, the researchers can investigate how the financial attitude of retail investors are affected during economic disruptions or pandemics such as Covid-19 by using ANN and reveal the behavioural biases. The results of the future study may be better to decide which technique is better with which indicator.

References

- Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2020). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, 14(1), 1–13.
- Aamer, A., Eka Yani, L., & Alan Priyatna, I. (2021). Data analytics in the supply chain management: Review of machine learning applications in demand forecasting. *Operations and Supply Chain Management: An International Journal*, 14(1), 1–13.
- Abkenar, S. B., Haghi Kashani, M., Mahdipour, E., & Mahdi Jameii, S. (2020). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics*, 101517.
- Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers and Industrial Engineering*, 143, 106435.
- 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, 106380.
- Adam, K., Smagulova, K., & James, A. P. (2018). Memristive LSTM network hardware architecture for time-series predictive modeling problems. Preprint Arxiv ArXiv:1809.03119v1.
- Adnane, M., El, M., El Fkihi, S., & Thami, R.O.H. (2019, March). Prediction demand for classified ads using machine learning: an experiment study. In *Proceedings of the 2nd, International Conference on Networking, Information Systems and Security* (pp.1–6).
- Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2), 31–50.
- Aguilar-Palacios, C., Muñoz-Romero, S., & Rojo-Álvarez, J. L. (2019). Forecasting promotional sales within the neighbourhood. *IEEE Access*, 7, 74759–74775.
- Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society*, 54, 102010.
- Ali, Ö. G., & Gürlek, R. (2020). Automatic interpretable retail forecasting with promotional scenarios. *International Journal of Forecasting*, 36(4), 1389–1406.
- 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.
- Almoghrebi, A., Aljuheshi, F., Rafaie, M., James, K., & Alahmad, M. (2020). Data-driven charging demand prediction at public charging stations using supervised Machine Learning regression methods. *Energies*, 13(16), 4231.
- Alon, I., Qi, M., & Sadowski, R. J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, 8(3), 147–156.
- Amirabadi, M. A., Kahaei, M. H., & Nezamalhosseini, S. A. (2020). Novel suboptimal approaches for hyperparameter tuning of deep neural network [under the shelf of optical communication]. *Physical Communication*, 41, 101057.
- Azzouni, A., & Pujolle, G. (2017). A long short-term memory recurrent neural network framework for network traffic matrix prediction. *Arxiv Preprint Arxiv: 1705.05690v2*. <https://arxiv.org/abs/1705.05690>
- Bag, S., Gupta, S., & Wood, L. (2020). Big data analytics in sustainable humanitarian supply chain: Barriers and their interactions. *Annals of Operations Research*, 1–40.
- Bag, S., Gupta, S., Kumar, A., & Sivarajah, U. (2021). An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance. *Industrial Marketing Management*, 92, 178–189.
- Bagwell, K. (2007). The economic analysis of advertising. *Handbook of Industrial Organization*, 3, 1701–1844.
- Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, 105(5), 481–485.
- 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. In *International Conference on Neural Information Processing* Springer, Cham, 462–474.
- Bazzaz Abkenar, S., Haghi Kashani, M., Mahdipour, E., & Mahdi Jameii, S. (2020). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics*, 101517.
- Bennett, K. P., & Parrado-Hernandez, E. (2006). The interplay of optimization and machine learning research. *Journal of Machine Learning Research*, 7, 1265–1281.
- Bodenstab (2017). Traditional statistics versus machine learning. What's the difference?, <https://www.toolsgroup.com/blog/traditional-statistics-versus-machine-learning-whats-the-difference/>
- Bollapragada, S., Gupta, S., Hurwitz, B., Miles, P., & Tyagi, R. (2008). NBC-Universal uses a novel qualitative forecasting technique to predict advertising demand. *Interfaces*, 38(2), 103–111.

- Boone, L. E., Kurtz, D. L., & Berston, S. (2019). *Contemporary business*. John Wiley & Sons.
- Brooker, J. R., Eastwood, D. B., & Gray, M. D. (1994). The Impact of advertising on consumer demand for beef: An application of scan data. *Journal of Food Products Marketing*, 2(1), 17–35.
- Bui, T. C., Le, V. D., & Cha, S. K. (2018). A Deep learning approach for forecasting air pollution in South Korea using LSTM. *Arxiv Preprint Arxiv*: 1804.07891. <https://arxiv.org/abs/1804.07891>
- Bünning, F., Heer, P., Smith, R. S., & Lygeros, J. (2020). Improved day ahead heating demand forecasting by online correction methods. *Energy and Buildings*, 211, 6109821.
- Cadavid, J. P. U., Lamouri, S., & Grabot, B. (2018). Trends in machine learning applied to demand & sales forecasting: A review. *ILS 2018—Information Systems, Logistics and Supply Chain, Proceedings*, July 2018, Lyon, France, 307–316.
- Calkoen, F., Luijendijk, A., Rivero, C. R., Kras, E., & Baart, F. (2021). Traditional vs. Machine-learning methods for forecasting sandy shoreline evolution using historic satellite-derived shorelines. *Remote Sensing*, 13(5), 1–21. <https://doi.org/10.3390/rs13050934>
- Carboneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140–1154.
- Carreón, E. C. A., Nonaka, H., Hentona, A., & Yamashiro, H. (2019). Measuring the influence of mere exposure effect of TV commercial adverts on purchase behavior based on machine learning prediction models. *Information Processing and Management*, 56(4), 1339–1355.
- Chang, P. C., Fan, C. Y., & Hsieh, J. C. (2009). A weighted evolving fuzzy neural network for electricity demand forecasting. 1st Asian Conference on Intelligent Information and Database Systems.
- Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, 46(3), 436–443.
- Chen, I. F., & Lu, C. J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing. *Neural Computing and Applications*, 28(9), 2633–2647.
- Chintagunta, P. K., & Nair, H. S. (2011). Structural workshop paper—discrete-choice models of consumer demand in marketing. *Marketing Science*, 30(6), 977–996.
- Choi, T.-M., & Chen, Y. (2021). Circular supply chain management with large scale group decision making in the big data era: The macro-micro model. *Technological Forecasting and Social Change*, 169, 120791.
- Christopher, M., & Holweg, M. (2017). Supply chain 20 revisited: A framework for managing volatility-induced risk in the supply chain. *International Journal of Physical Distribution and Logistics Management*, 47(1), 2–17.
- Claveria, O., Monte, E., & Torra, S. (2016). Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters*, 23(6), 428–431.
- Daniel, B. K. (2017). Big data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, 50(1), 101–113.
- Dave, D., & Saffer, H. (2012). Impact of direct-to-consumer advertising on pharmaceutical prices and demand. *Southern Economic Journal*, 79(1), 97–126.
- Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42.
- Dawes, J., Kennedy, R., Green, K., & Sharp, B. (2018). Forecasting advertising and media effects on sales: Econometrics and alternatives. *International Journal of Market Research*, 60(6), 611–620.
- De Las Heras, A., Luque-Sendra, A., & Zamora-Polo, F. (2020). Machine learning technologies for sustainability in smart cities in the post-covid era. *Sustainability*, 12(22), 9320.
- Dev, V. A., & Eden, M. R. (2019). Gradient boosted decision trees for lithology classification. *Computer Aided Chemical Engineering*, 47, 113–118.
- Divakar, S., Ratchford, B. T., & Shankar, V. (2005). Practice prize article—chan4cast: A multichannel, multi-region sales forecasting model and decision support system for consumer-packaged goods. *Marketing Science*, 24(3), 334–350.
- Dong, D., Kaiser, H. M., & Myrland, Ø. (2007). Quantity and quality effects of advertising: A demand system approach. *Agricultural Economics*, 36(3), 313–324.
- Eachempati, P., Srivastava, P. R., Kumar, A., Tan, K. H., & Gupta, S. (2021). Validating the impact of accounting disclosures on stock market: A deep neural network approach. *Technological Forecasting and Social Change*, 170, 120903.
- Fathi, M., Haghi Kashani, M., Jameii, S. M., & Mahdipour, E. (2021). Big data analytics in weather forecasting: A systematic review. *Archives of Computational Methods in Engineering* (Published online).
- Feng, Y., & Wang, S. (2017). A forecast for bicycle rental demand based on random forests and multiple linear regression. *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*.
- Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing and Service Operations Management*, 18(1), 69–88.

- Fildes, R., Nikolopoulos, K., Crone, S. F., & Syntetos, A. A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society*, 59(9), 1150–1172.
- Galetsis, P., & Katsaliaki, K. (2020). A review of the literature on big data analytics in healthcare. *Journal of the Operational Research Society*, 71(10), 1511–1529.
- Garg, R., Aggarwal, H., Centobelli, P., & Cerchione, R. (2019). Extracting knowledge from big data for sustainability: A comparison of machine learning techniques. *Sustainability*, 11(23), 6669.
- Gilliland, M. (2010). Defining "demand" for demand forecasting. *Foresight: The International Journal of Applied Forecasting*, (18), 4–8.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space Odyssey. *IEEE-Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232.
- Güler, M. G. (2019). Advertising and forecasting investments of a newsvendor. *4OR*, 17(1), 45–73.
- Güven, İ., & Şimşir, F. (2020). Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods. *Computers & Industrial Engineering*, 147(July).
- Hasan, M. M., Popp, J., & Oláh, J. (2020). Current landscape and influence of big data on finance. *Journal of Big Data*, 7(1).
- 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.
- Hu, Y. C., Jiang, P., & Lee, P. C. (2019). Forecasting tourism demand by incorporating neural networks into Grey–Markov models. *Journal of the Operational Research Society*, 70(1), 12–20.
- 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.
- Husna, A., Amin, S. H., & Shah, B. (2021). Demand forecasting in supply chain management using different Deep Learning methods. In demand forecasting and order planning in supply chains and humanitarian logistics, 140–170. IGI Global.
- Hwang, S., Kim, J., Park, E., & Kwon, S. J. (2020). Who will be your next customer: A machine learning approach to customer return visits in airline services. *Journal of Business Research*, 121, 121–126.
- Ismailov, V. E. (2014). On the approximation by neural networks with bounded number of neurons in hidden layers. *Journal of Mathematical Analysis and Applications*, 417(2), 963–969.
- Jaipuria, S., & Mahapatra, S. S. (2014). An improved demand forecasting method to reduce bullwhip effect in supply chains. *Expert Systems with Applications*, 41(5), 2395–2408.
- Jedidi, K., Mela, C. F., & Gupta, S. (1999). Managing advertising and promotion for long-run profitability. *Marketing Science*, 18(1), 1–22.
- Jiao, L., & Li, H. (2010). Chemometrics and Intelligent Laboratory Systems QSPR studies on the aqueous solubility of PCDD/Fs by using artificial neural network combined with stepwise regression. *Chemometrics and Intelligent Laboratory Systems*, 103(2), 90–95.
- Kalogirou, S. A. (2000). Applications of artificial neural-networks for energy systems. *Applied Energy*, 67(1–2), 17–35. [https://doi.org/10.1016/S0306-2619\(00\)00005-2](https://doi.org/10.1016/S0306-2619(00)00005-2)
- Kamilaris, A., Kartakoulis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture*, 143, 23–37.
- Kaneko, Y. (2019). Customer-Base sequential data analysis: An application of attentive Neural Networks to sales forecasting. 2019 International Conference on Data Mining Workshops (ICDMW).
- Kang, E. (2017). Long Short-Term Memory (LSTM): Concept. Retrieved from <https://medium.com/@kangeugine/long-short-term-memory-lstm-concept-cb3283934359>
- Kang, Y., Spiliotis, S., Petropoulos, F., Athinotis, N., Li, F., & Assimakopoulos, V. (2020). Déjà vu: A data-centric forecasting approach through time series cross-similarity. *Journal of Business Research*. In press.
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.
- Karimova, G. Z., & Shirkhanbeik, A. (2019). Marketing artificial intelligence: Creating the AI archetype for evoking the personality trust. *Academy of Marketing Studies Journal*, 23(4), 1–13.
- Ke, J., Zheng, H., Yang, H., & Chen, X. (Michael). (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 85, 591–608. <https://doi.org/10.1016/j.trc.2017.10.016>
- Khakpour, A. (2020). Data science for decision support: Using machine learning and big data in sales forecasting for production and retail, Master's thesis, Østfold University College.
- Khan, M. A., Saqib, S., Alyas, T., Rehman, A. U., Saeed, Y., Zeb, A., Zareei, M., & Mohamed, E. M. (2020). Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE Access*, 8, 116013–116023.

- Khanra, S., Dhir, A., Islam, N., & Mäntymäki, M. (2020b). Big data analytics in healthcare: A systematic literature review. *Enterprise Information Systems*, 14(7), 878–912.
- Khanra, S., Dhir, A., & Mäntymäki, M. (2020a). Big data analytics and enterprises: A bibliometric synthesis of the literature. *Enterprise Information Systems*, 14(6), 737–768.
- Kharfan, M., Chan, V. W. K., & Firdolas Efendigil, T. (2020). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Annals of Operations Research*, 1–16.
- Khokhar, P. & Chitsimran, (2019). Evolution of Artificial Intelligence in marketing, Comparison with Traditional Marketing. *Our Heritage*, 67(5), 375–389.
- Kialashaki, A., & Reisel, J. R. (2013). Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks. *Applied Energy*, 108, 271–280.
- Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial Intelligence in advertising. *Journal of Advertising Research*, 58(3), 263–267.
- Kim, D. Y., Hwang, Y. H., & Fesenmaier, D. R. (2005). Modeling tourism advertising effectiveness. *Journal of Travel Research*, 44(1), 42–49.
- Kittichotsawat, Y., Jangkrajarn, V., & Tippayawong, K. Y. (2021). Enhancing coffee supply chain towards sustainable growth with big data and modern agricultural technologies. *Sustainability*, 13(8), 4593.
- Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-Term residential load forecasting based on LSTM recurrent Neural Network. *IEEE Transactions on Smart Grid*, 10(1), 841–851.
- Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*, 281, 628–641.
- Kremer, M., Siemsen, E., & Thomas, D. J. (2016). The sum and its parts: Judgmental hierarchical forecasting. *Management Science*, 62(9), 2745–2764.
- Kronenberg, K., Fuchs, M., Salman, K., Lexhagen, M., & Höpken, W. (2016). Economic effects of advertising expenditures—a Swedish destination study of international tourists. *Scandinavian Journal of Hospitality and Tourism*, 16(4), 352–374.
- Kumar, A., Choi, T. M., Wamba, S. F., Gupta, S., & Tan, K. H. (2021). Infection vulnerability stratification risk modelling of COVID-19 data: a deterministic SEIR epidemic model analysis. *Annals of Operations Research*, 1–27.
- Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial Marketing Management*, 90, 493–507.
- Kwon, W., Lee, M., Back, K. J., & Lee, K. Y. (2021). Assessing restaurant review helpfulness through big data: Dual-process and social influence theory. *Journal of Hospitality and Tourism Technology*, 12(2), 177–195.
- Lawrence, M., Goodwin, P., O'Connor, M., & Önköl, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493–518.
- Leminen, S., Rajahonka, M., Westerlund, M., & Wendelin, R. (2018). The future of the internet of things: Toward heterarchical ecosystems and service business models. *Journal of Business and Industrial Marketing*, 33(6), 749–767.
- Li, L., Chi, T., Hao, T., & Yu, T. (2016). Customer demand analysis of the electronic commerce supply chain using Big Data. *Annals of Operations Research*, 268(1–2), 113–128.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2016). Big data in tourism research: A literature review. *Tourism Management*, 68, 301–323.
- Lian, Y., Zhang, G., Lee, J., & Huang, H. (2020). Review on big data applications in safety research of intelligent transportation systems and connected/automated vehicles. *Accident Analysis and Prevention*, 146, 105711.
- Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: A review. *Mathematical Problems in Engineering*, 2013, 1–9.
- Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114, 81–93.
- Lu, C., Lee, T., & Lian, C. (2012). Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. *Decision Support Systems*, 54(1), 584–596.
- Luo, J., Klein, T., Ji, Q., & Hou, C. (2021). Forecasting realized volatility of agricultural commodity futures with infinite Hidden Markov HMM models. *International Journal of Forecasting*, in press, <https://doi.org/10.1016/j.ijforecast.2019.08.007>.
- 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), 245–257.2.

- Mageto, J. (2021). Big data analytics in sustainable supply chain management: A focus on manufacturing supply chains. *Sustainability*, 13(13), 7101.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one*, 13(3), e0194889.
- Mariani, M., & Borghi, M. (2021). Are environmental-related online reviews more helpful? A big data analytics approach. *International Journal of Contemporary Hospitality Management*, 33(6), 2065–2090.
- Mariani, M. M., & Wamba, S. F. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies. *Journal of Business Research*, 121, 338–352.
- Martinez, E. Z., Aragon, D. C., & Nunes, A. A. (2020). Short-term forecasting of daily COVID-19 cases in Brazil by using the Holt's model. *Revista da Sociedade Brasileira de Medicina Tropical*, 53.
- Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., & Kaufman, L. (2003). *Supervised Pattern Recognition. Data Handling in Science and Technology*, Amsterdam: Elsevier.
- May, R., Dandy, G., & Maier, H. (2011). Review of input variable selection methods for artificial neural networks. In Suzuki, K. (ed.) *Artificial neural networks-methodological advances and biomedical applications* (pp.19–44).
- Meyer-Baese, A., & Schmid, V. (2014). Feature Selection and Extraction. In: V. S. Anke Meyer-Baese (ed.), *Pattern Recognition and Signal Analysis in Medical Imaging* (2nd ed. pp. 21–69). Academic Press.
- Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, S. J. (2016). Big Data and supply chain management: A review and bibliometric analysis. *Annals of Operations Research*, 270, 313–336.
- Moroff, N. U., Kurt, E., & Kamphues, J. (2021). Machine Learning and statistics: A Study for assessing innovative demand forecasting models. *Procedia Computer Science*, 180, 40–49. <https://doi.org/10.1016/j.procs.2021.01.127>
- Mukherjee, K., & Banerjee, N. (2017). Effect of social networking advertisements on shaping consumers' attitude. *Global Business Review*, 18(5), 1291–1306.
- Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2021). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. *Journal of Business Research*, 124, 389–404.
- Muzaffar, S., & Afshari, A. (2019). Short-term load forecasts using LSTM networks. *Energy Procedia*, 158, 2922–2927.
- Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F., & Chandran, V. (2018). Long short-term memory hyperparameter optimization for a neural network-based emotion recognition framework. *IEEE Access*, 6, 49325–49338.
- Narayanan, A., Sahin, F., & Robinson, E. P. (2019). Demand and order-fulfillment planning: The impact of point-of-sale data, retailer orders and distribution center orders on forecast accuracy. *Journal of Operations Management*, 65(5), 468–486.
- Nguyen, K. T., Nguyen, Q. D., Le, T. A., Shin, J., & Lee, K. (2020). Analyzing the compressive strength of green fly ash based geopolymer concrete using experiment and machine learning approaches. *Construction and Building Materials*, 247, 118581.
- Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, 53, 102104.
- Okrent, A. M., & MacEwan, J. P. (2014). The effects of prices, advertising, expenditures, and demographics on demand for nonalcoholic beverages. *Agricultural and Resource Economics Review*, 43(1), 31–52.
- Øyen, S. (2018). *Forecasting multivariate time series data using neural networks*. Master Thesis, Norwegian University of Science and Technology Department of Engineering Cybernetics.
- Oztekin, A. (2017). Creating a marketing strategy in healthcare industry: A holistic data analytic approach. *Annals of Operations Research*, 270, 361–382.
- Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data*, 4(1), 15.
- Di Pillo, G., Latorre, V., Lucidi, S., & Procacci, E. (2013). An application of learning machines to sales forecasting under promotions, DIAG Technical Reports 2013–04, Department of Computer, Control and Management Engineering, Università degli Studi di Roma "La Sapienza".
- Qayyum, A., Razzak, I., Tanveer, M., & Kumar, A. (2021). Depth-wise dense neural network for automatic COVID19 infection detection and diagnosis. *Annals of Operations Research*, 1–21.
- Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. (2017). Empirical mode decomposition-based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, 54, 246–255.
- Rahman, M. Z., Sajib, M. N., Rifat, M. S. H., Hossam-E-Haider, M., & Khan, M. A. A. (2016). Forecasting the long-term energy demand of Bangladesh using SPSS from 2011–2040. 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT).
- Rai, S., Gupta, A., Anand, A., Trivedi, A., & Bhadauria, S. (2019). Demand prediction for e-commerce advertisements: A comparative study using state-of-the-art machine learning methods. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

- Ramanathan, U., & Muiyldermans, 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.
- Ramanathan, U., & Muiyldermans, L. (2011). Identifying the underlying structure of demand during promotions: A structural equation modelling approach. *Expert Systems with Applications*, 38(5), 5544–5552.
- Ramya, B. S. S., & Vedavathi, K. (2020). An advanced sales forecasting using machine learning algorithm. *International Journal of Innovative Science and Research Technology*, 5(5), 342–345.
- Ravikumar, K., Saroop, A., Narahari, H. K., & Dayama, P. (2005). Demand sensing in e-business. *Sadhana*, 30(2–3), 311–345.
- Rawlins, J. (2017). DDF-MMM—Data. https://github.com/jamesrawlins1000/DDF-MMM--Data/blob/959a87a888f481876a8ca13b6615f17aaed5270a/MMM_data.xlsx
- Reese, K. M. (2021). Deep learning artificial neural networks for non-destructive archaeological site dating. *Journal of Archaeological Science*, 132(June 2020), 105413. <https://doi.org/10.1016/j.jas.2021.105413>
- Reimers, N., & Gurevych, I. (2017). Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. *arXiv: Computation and Language*, Volume abs/1707.06799, 2017.
- Reis, C., Ruivo, P., Oliveira, T., & Faroleiro, P. (2020). Assessing the drivers of machine learning business value. *Journal of Business Research*, 117, 232–243.
- Ren, S., Chan, H.-L., & Siqin, T. (2019). Demand forecasting in retail operations for fashionable products: Methods, practices, and real case study. *Annals of Operations Research*, 291, 761–777.
- Ribeiro, G. H. T., De Neto, P. S. G. M., Cavalcanti, G. D. C., & Tsang, I. R. (2011). Lag selection for time series forecasting using Particle Swarm Optimization. *The 2011 International Joint Conference on Neural Networks*.
- Ribeiro, F. M. (2016). Interdisciplinarity in ferment: The role of knowledge networks and department affiliation. *Technological Forecasting and Social Change*, 113, 240–247.
- Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, 27, 111–125.
- Rosenthal, M. B., Berndt, E. R., Donohue, J. M., Epstein, A. M., & Frank, R. G. (2003). Demand effects of recent changes in prescription drug promotion. In David M. Cutler & Alan M. Garber (ed). 1–26, *Frontiers in health policy research*, 6, Cambridge, MA: MIT Press.
- San Kim, T., & Sohn, S. Y. (2020). Machine-learning-based deep semantic analysis approach for forecasting new technology convergence. *Technological Forecasting and Social Change*, 157, 120095.
- Savarimuthu, B. T. R., Corbett, J., Yasir, M., & Lakshmi, V. (2020). Using Machine Learning to Improve the Sustainability of the Online Review Market. In *ICIS*.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.
- Sengupta, P., Biswas, B., Kumar, A., Shankar, R., & Gupta, S. (2021). Examining the predictors of successful Airbnb bookings with Hurdle models: Evidence from Europe, Australia, USA and Asia-Pacific cities. *Journal of Business Research*, 137, 538–554.
- Shah, N., Engineer, S., Bhagat, N., Chauhan, H., & Shah, M. (2020). Research trends on the usage of machine learning and artificial intelligence in advertising. *Augmented Human Research*, 5(1), 1–15.
- Shahid, M. Z., & Li, G. (2019). Impact of artificial intelligence in marketing: A perspective of marketing professionals of Pakistan. *Global Journal of Management and Business Research*, 19(2), 27–33.
- Shi, H., Hu, S., & Zhang, J. (2019). LSTM based prediction algorithm and abnormal change detection for temperature in aerospace gyroscope shell. *International Journal of Intelligent Computing and Cybernetics*, 12(2), 274–291.
- Shilpi & Sharma, B. K. (2016). Prediction of preferred advertising formats in e-advertisements by using data mining techniques. *International Journal of Applied Information Systems*, 10(6), 27–32.
- Silveira Netto C. F. & Brei V. A. (2017). Demand forecasting in marketing: Methods, types of data, and future research. https://forecasters.org/wp-content/uploads/gravity_forms/7-c6dd08fee7f0065037affb5b74fec20a/2017/07/SilveiraNetto_Carla_ISF2017.pdf.
- Singhal, S., Ahuja, L., & Monga, H. (2020, December). State of the art of machine learning for product sustainability. In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 197–202). IEEE.
- Slimani, I., Farissi, I. El, & Al-Qualsadi, S. A. (2016). Configuration of daily demand predicting system based on neural networks. *Proceedings of the 3rd IEEE International Conference on Logistics Operations Management, GOL 2016*. <https://doi.org/10.1109/GOL.2016.7731709>
- Suma, V. (2020). Data mining based prediction of demand in Indian market for refurbished Electronics. *Journal of Soft Computing Paradigm (JSCP)*, 2(02), 101–110.
- Talwar, S., Kaur, P., Wamba, S. F., & Dhir, A. (2021). Big Data in operations and supply chain management: A systematic literature review and future research agenda. *International Journal of Production Research*, 59(11), 3509–3534.

- Tan, T. F., & Ko, C. H. (2016). Application of artificial intelligence to cross-screen marketing: a case study of AI technology company, 2nd International Conference on Artificial Intelligence and Industrial Engineering, Atlantis Press.
- Tiautrakul, J., & Jindakul, J. (2019). The Artificial Intelligence (AI) with the future of digital marketing. SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3405184.
- 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.
- Tugay, R. & Ögüdücü Ş.G. (2017) Demand prediction using Machine Learning methods and stacked generalization. DATA 2017. 6th International Conference on Data Science, Technology and Applications, 216–222.
- Van Donselaar, K. H., Peters, J., de Jong, A., & Broekmeulen, R. A. (2016). Analysis and forecasting of demand during promotions for perishable items. *International Journal of Production Economics*, 172, 65–75.
- Van Heerde, H. J., Leeflang, P. S., & Wittink, D. R. (2002). How promotions work: SCAN* PRO-based evolutionary model building. *Schmalenbach Business Review*, 54(3), 198–220.
- Van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020b). Predicting customer demand for remanufactured products: A data-mining approach. *European Journal of Operational Research*, 281(3), 543–558.
- Wang, D., & Zhang, Y. (2020). Implications for sustainability in supply chain management and the circular economy using machine learning model. *Information Systems and e-Business Management*, 1–13.
- Wang, J., Xu, C., Zhang, J., & Zhong, R. (2021). Big data analytics for intelligent manufacturing systems: A review. *Journal of Manufacturing Systems*. (in press)
- Wattanakamolchai, S. (2016). Socially responsible customers and the evaluation of service quality. *Journal of Hospitality and Tourism Research*, 40(6), 715–738.
- Wei, Z., Dou, W., Jiang, Q., & Gu, C. (2021). Influence of incentive frames on offline-to-online interaction of outdoor advertising. *Journal of Retailing and Consumer Services*, 58, 102282.
- Wirth, N. (2018). Hello marketing, What can artificial intelligence help you with? *International Journal of Market Research*, 60(5), 435–438.
- Wirtz, B. W. (2020). *Business Model Management*. Springer.
- Wolters, J., & Huchzermeier, A. (2021). Joint in-season and out-of-season promotion demand forecasting in a retail environment. *Journal of Retailing*. In press.
- Woo, J., & Mishra, M. (2020). Predicting the ratings of Amazon products using Big Data. *Wires Data Mining and Knowledge Discovery*, 11(3), 1–10.
- Xu, C., Ji, J., & Liu, P. (2018). The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies*, 95(July), 47–60.
- Xu, M., Watanachaturaporn, P., Varshney, P. K., & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment*, 97(13), 322–336.
- Yakak, E., Ugurlu, U., & Tas, O. (2018). Using artificial neural network and a statistical method for the estimation of Euro/Turkish Lira exchange rate. *Pressacademia*, 7(1), 414–417.
- Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A Stepwise AIC method for variable selection in linear regression. *Communications in Statistics-Theory and Methods*, 36(13), 2395–2403.
- Yan, H., Jiang, Y., Zheng, J., Peng, C., & Li, Q. (2006). A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Systems with Applications*, 30(2), 272–281. <https://doi.org/10.1016/j.eswa.2005.07.022>
- Yu, Y., Choi, T. M., & Hui, C. L. (2011). An intelligent fast sales forecasting model for fashion products. *Expert Systems with Applications*, 38(6), 7373–7379. <https://doi.org/10.1016/j.eswa.2010.12.089>
- Yue, L., Yafeng, Y., Junjun, G., & Chongli, T. (2007). Demand forecasting by using support vector machine. *3rd International Conference on Natural Computation (ICNC 2007)*.
- Zheng, Z., & Henneberry, S. R. (2010). The impact of changes in income distribution on current and future food demand in urban China. *Journal of Agricultural and Resource Economics*, 51–71.
- Zhou, Y.-W., Chen, C., Zhong, Y., & Cao, B. (2018). The allocation optimization of promotion budget and traffic volume for an online flash-sales platform. *Annals of Operations Research*, 291, 1183–1207.
- Zhu, X., Ninh, A., Zhao, H., & Liu, Z. (2021). *Demand forecasting with supply-chain information and machine learning: Evidence in the pharmaceutical industry*. Early Review (in press).