Which product description phrases affect sales forecasting? An explainable AI framework by integrating WaveNet neural network models with multiple regression

Which Product Description Phrases affect Sales Forecasting? An Explainable AI Framework by Integrating WaveNet Neural Network Models with Multiple Regression

Ms. Shan Chen
Department of Management Science, School of Management,
Xiamen University, Xiamen, China
Email: chenshan@stu.xmu.edu.cn

Dr. Shengjie Ke
School of Economics and Management,
Fuzhou University of International Studies and Trade, Fuzhou, China
Email: keshengjie@fzfu.edu.cn

Dr. Shuihua Han *
Department of Management Science, School of Management,
Xiamen University, Xiamen, China
Email: hansh@xmu.edu.cn
(* : Corresponding Author)

Dr. Shivam Gupta
Department of Information Systems, Supply Chain Management & Decision Support,
NEOMA Business School,
59 Rue Pierre Taittinger, 51100 Reims, France
Email: shivam.gupta@neoma-bs.fr

Dr. Uthayasankar Sivarajah
School of Management, University of Bradford,
Richmond Road, Bradford, BD7 1DP, United Kingdom
Email: u.sivarajah@bradford.ac.uk

Which Product Description Phrases affect Sales Forecasting? An Explainable AI Framework by Integrating WaveNet Neural Network Models with Multiple Regression

Abstract

The rapid rise of many e-commerce platforms for individual consumers has generated a large amount of text-based data, and thus researchers have begun to experiment with text mining techniques to extract information from the large amount of textual data to assist in sales forecasting. The existing literature focuses textual data on product reviews; however, consumer reviews are not something that companies can directly control, here we argue that textual product descriptions are also important determinants of consumer choice. We construct an artificial intelligence (AI) framework that combines text mining, WaveNet neural networks, multiple regression, and SHAP model to explain the impact of product descriptions on sales forecasting. Using data from nearly 200,000 sales records obtained from a cross-border e-commerce firm, an empirical study showed that the product description presented to customers can influence sales forecasting, and about 44% of the key phrases greatly affect sales forecasting results, the sales forecasting models that added key product description phrases had improved forecasting accuracy. This paper provides explainable results of sales forecasting, which can provide guidance for firms to design product descriptions with reference to the market demand reflected by these phrases, and adding these phrases to product descriptions can help win more customers.

Keywords: Text Mining; Sales Forecasting; Wavenet Neural Network; Explainable AI; Cross-Border E-Commerce

1. Introduction

Sales forecasting refers to the estimation of the sales quantity and sales amount of all or specific products and services in a specific time in the future through systematic sales forecasting models based on past sales. Sales forecasting is very important in both traditional retail and e-commerce, and the accuracy of forecasting affects the planning, strategy, marketing, logistics, warehousing, and resource management decisions of retailers and other members of the supply chain [8, 23, 67]. Accurate sales forecasting is critical for companies operating in high-volume, low-margin industries. Inaccuracies in forecasts can lead to stockouts or excess inventory, resulting in an inability to meet customer demand or an increase in working capital, and for both cases, additional unnecessary costs [66].

Sales forecasting is also a key task in supply chain management. By generating more accurate sales forecasts at stock keeping unit (SKU) level, it is possible to reduce the bullwhip effect and achieve timely delivery, thereby improving the effectiveness of supply chain management [27]. In the context of the supply chain, existing research focuses on identifying and analyzing the influencing factors in sales forecasting[63], as well as constructing new sales forecasting models [62]. Additionally, the application of artificial intelligence methods to sales forecasting [13] [27] and improving the accuracy of sales forecasting[73] have always been the research priorities in sales forecasting.

Cross-border e-commerce businesses face markets with greater uncertainty, fluctuating sales demand, and intense competition, requiring companies to provide customers with competitive prices and products in a relatively short period of time, so accurate sales forecasting is critical to bridging the gap between supply and demand. Cross-border e-commerce companies improve their inventory decisions through timely and accurate sales forecasts [11, 12]. Research on improving the accuracy of forecasting models has attracted the attention of many scholars [32, 41]. Traditional forecasting models including the Winters exponential smoothing, the Box-Jenkins ARIMA model [9], and the multivariate regression method. The Winters model is a viable method for multi-step forecasting under stable economic conditions. Traditional sales forecasting methods are mainly based on historical sales data, which can lead to a certain lag. The relationship between sales volumes and their influencing factors is complex and often non-linear [74]. Therefore, in addition to traditional forecasting models, machine learning-based forecasting models combining artificial intelligence algorithms with time series are widely used. A supply chain market analysis conducted in 2018 revealed that 19% of companies use machine learning and deep learning methods to improve prediction accuracy [1]. In analyzing the impact of unstructured data such as textual data on sales, scholars have mostly used deep learning based on neural networks, and much of the literature has demonstrated that neural network models are more favorable than traditional models in terms of predictive accuracy [28, 39], and that neural network models can capture dynamic non-linear trends and seasonal patterns, as well as the interactions between trends and patterns [47]. Deep learning models exhibit exceptional performance compared to existing algorithms. However, deep learning is not as widely used in business decision making as expected due to its computational complexity, lack of big data infrastructure, and the opaqueness (black box) of deep learning [56]. How to use deep learning models to improve prediction accuracy while providing explainable decision support to firms is the key to this study? Through explainable output of results, the business decision makers can understand how decisions are made and how these data science understandings correspond to existing business knowledge to mitigate the risk of adoption [14].

In addition, the input characteristics have a direct impact on the model's forecasting effect. Identifying and capturing the factors affecting demand is the basis and basic guarantee of scientifically accurate sales forecasting results. Generally, the factors that affect the accuracy of sales forecasts can be summarized into three aspects: historical data [2], random factors such as seasons, holidays and promotions [3,4] and consumer factors [5-7]. With the development of e-commerce, more and more textual data is generated, such as product descriptions and user reviews of e-commerce platforms, online shoppers increasingly rely on "word-of-mouth" recommendations [8] and product reviews [9,10] to make purchases, thereby influencing sales forecasts for products. But from a business perspective, word-of-mouth and user reviews are external and only indirectly controllable [11]. Business owners have considerable control over their product descriptions [12]. Companies can design another type of text that is directly related to sales, specifically, they can design product descriptions that are presented to consumers. Particularly in e-commerce, products that rank high in search results tend to sell better, as customers often look for products through search engines. Companies therefore use search engine optimization (SEO) techniques to design product descriptions to improve the ranking of their products in search results. Therefore, in this study, product descriptions are used as text mining objects and important input features in sales forecasting models to investigate their impact on sales forecasts.

This study proposes an explainable framework that combines deep neural networks, text mining, multiple regression and SHAP model to analyze and explain the impact of product description on sales forecasting. We used text mining analysis to convert product descriptions into bag-of-words (BOW) vectors, which are used as explanatory parameters in a WaveNet neural network model to predict sales and construct a sales forecasting model. Then, we compared the results of a multivariate WaveNet sales forecasting model that included BOW, historical sales and price with those of a model that included only historical sales and price, and found that the accuracy of the sales forecasts improved significantly when the BOW parameters were combined. After that, we use multiple regression to identify the key product description word vectors that affect the sales prediction results as feature factors, build the WaveNet sales prediction model again, and finally interpret the deep learning model using the SHAP model to output the importance of the features to explain the impact of each word vector on the prediction results to help business decision makers find keywords that can optimize product search and product description on e-commerce platforms, and provide decision support for more timely and accurate sales forecasting. The main contributions are as follows:

Firstly, to the best of our knowledge, this paper is the first to apply the results of textual analysis of product descriptions to a sales prediction model of WaveNet neural network and to interpret the WaveNet model using SHAP. Most importantly, it is shown that the combination of textual analysis results improves the accuracy of sales forecasting. Thus, this work advances the literature on sales forecasting.

Secondly, we built an explainable framework which provides an explanation for the results of deep learning, discovering the impact of product descriptions on sales forecasts through deep neural network algorithms, multiple regression analysis was used to select the important feature factors affecting sales forecasting, to achieve dimensionality reduction of the feature variables, and finally to improve the accuracy and interpretability of sales forecasting with the combination of the reduced dimensional WaveNet model and SHAP model.

In addition, the findings presented in this paper provide a useful guide to a company's sales forecasting process. Using the explainable framework can help companies understand why product descriptions can influence sales forecasting and how the results of our analysis can be used to improve product descriptions to drive sales.

The remainder of the paper is organized as follows. Section 2 provides a literature review of text mining techniques in sales forecasting, text mining in sales forecasting and the explainability. Section 3 presents an explainable framework for sales forecasting that fuses text mining, WaveNet, multiple regression and SHAP. Section 4 describes the empirical data sources and data processing procedures. Section 5 presents and analyses the empirical results, demonstrating the significant impact of text-mining based parameters on sales forecasting, and explains the sales forecasting results and the support for business decisions. In Section 6 we discuss the implication for marketing decision support, implications from interpretable forecasts and give limitations and potential for future research. Finally, Section 7 concludes and gives summary of the paper.

2. Literature review

2.1 Sales forecasting approaches and WaveNet model

Sales forecasting is often used as an input to many operational and business decisions that affect a company's profitability and is an integral part of many core management decisions in an organization, such as pricing, ordering and inventory management [48, 58, 62]. Sales forecasting is also one of the fundamental inputs in supply chain planning decisions [55]. E-commerce companies face faster changing consumer demand [25], sales forecasting often involves generating forecasts for a large number of products in many stores within a short-term forecast horizon, and the accuracy of forecasts becomes extremely important to meet customer demand and increased expectations within shorter lead times [8].

Linear statistical methods, mainly including linear regression, moving average and other statistical time series analysis models, such as autoregressive integrated moving average (ARIMA) [13, 65], have been widely used for sales forecasting, but they are not comprehensive and flexible enough to account for dynamic changes and non-linearities in sales time series [18].To capture the dynamic changes in sales, scholars have developed new sales forecasting methods. Machine learning and deep learning algorithms have been proven to outperform traditional predictive models in terms of predictive accuracy in a number of areas [3, 4, 34, 49, 72]. Using deep learning in business analytics and operations management, deep neural networks are expected to improve forecasting performance compared to traditional machine learning models [39]. For example, deep learning methods were used to achieve more accurate sales forecasting in the fashion industry [47]. A new model based on Light GBM and LSTM was developed to obtain accurate sales predictions in the supply chain [70]. Economic indicators were employed as influencing factors, and deep learning was utilized to enhance the accuracy of retail predictions [68]. Additionally, a meta-learning framework based on a newly developed deep convolutional neural network was proposed to accurately predict the sales of different products using different ensemble models in different time periods [48].

WaveNet is a neural network model that performs prediction tasks based on time series data. The key part of the WaveNet model is causal convolution. By using causal convolution, it is ensured that the model does not violate the modelling order of the data. Stacked inflated causal convolutions allow WaveNet to have a very large receptive field with only a few layers, as well as computational efficiency, while maintaining input resolution throughout the network. As WaveNet models have obtained good predictive performance in multivariate time-series forecasting, such as electricity load forecasting [54], urban air quality forecasting [6], grid power forecasting [37] and WaveNet models are less commonly used in sales forecasting, therefore, this paper applies the WaveNet model for multivariate time series sales forecasting.

2.2 Text mining in sales forecasting

In recent years, advances in text mining techniques have made it possible to process large amounts of textual data, which is widely used in forecasting models in various fields and helps to improve forecasting errors. For example, scholars have extensively utilized textual data such as public opinion and news reports to forecast the volatility of stock trading [42], stock indices[22], crude oil market volatility [31], and visitor numbers to tourist destinations[52] by extracting text feature vectors through text mining techniques, resulting in improved forecasting results. It has also been found that misinformation/fake news (on Internet platforms) can have an impact on consumer purchase behavior [33].Companies have also found commercial value in user-created content, such as using text mining techniques to mine online reviews to obtain consumers' sentiment preferences [71] and integrating textual data into customer churn prediction (CCP) models to improve their predictive performance [17]. Textual data is also increasingly being used in sales forecasting [46, 60], such as leveraging the quantity and quality of user-generated Facebook data to enhance product predictions [15], and incorporating customer search data from Google into time series models to improve out-of-sample prediction errors[7]. However, the main way existing studies used textual data for sales forecasting was to perform sentiment analysis on online review texts, transforming the texts into sentiment factors as sales forecasting variables [20, 21, 40]. A method for sales forecasting using a multiple regression sales forecast model that utilized both numeric and textual data from customer reviews was proposed by [57]. They use the BOW method to mine textual data because this method requires minimal data pre-processing and parsing, thus making the resulting model efficient and scalable. The results showed that the forecasting performance of this approach for both tasks was strong and significantly better than those of the models that ignored the textual content of customer reviews. In this paper, textual data of product descriptions are added to a time series forecasting model to analyze the effect of product descriptors on the accuracy of sales forecasting.

2.3 Explainability of Artificial Intelligence

Explainable AI describes a process that allows one to understand how an AI system decides, predicts, and performs its operations. The weights of features in linear models, paths in decision trees can be directly examined to understand the predictions of the model, providing traceability and transparency in the decision-making process. In contrast, deep learning models such as recurrent neural networks (RNN), long short-term memory (LSTM), and convolutional neural networks (CNN) are black box. Such methods as decision trees are called intrinsically interpretable methods, and the other category is to give interpretability by using post-hoc methods after training a black box model. Combining tree-based models with deep learning models to enhance model interpretability is a common approach. For example, in a supply chain sales forecasting hybrid model, the combination of lightGBM and LSTM models has been used to improve the accuracy and efficiency of forecasts, while leveraging the intrinsic interpretability of lightGBM to explain the forecast results[70]. One example of a post-hoc method is the SHAP (SHapley Additive explanations) approach and its variants, the outputs of this type of models are usually graphs that illustrate the individual feature importance and the feature dependencies [24, 29].

While machine learning can improve the accuracy of predictions, its 'black box' nature makes it difficult to be understood and trusted by business decision makers, and fairness, accountability and transparency of AI systems are important in all areas where AI models are used [75]. The complexity of the algorithms hinders transparency and accountability of the overall system and may weaken the deployment of AI-driven decision support systems [35], making the interpretation of machine learning critical [29]. Increasingly, scholars are considering model interpretability while ensuring the accuracy of forecasting models. For example, a structured effects neural network was applied to predict remaining useful life while providing a high degree of accountability and flexibility in deep learning [38]. By using a Variable Order Bayesian Network (VOBN) model on recruitment data, a comprehensive, accurate, and interpretable analysis framework was built as a decision support tool for HR recruiters [53]. An interpretable and non-discriminatory RegTech application offered automatic assessment of financial consumer complaints[61]. Bayesian methods[50] and Local Interpretable Model-agnostic Explanations (LIME) [59] are commonly used for local post-explanations in research on model interpretability. In addition, Partial Dependence Plot (PDP) [51] and the Time Fusion Transformer (TFT) [44] have been employed to study the interpretability of time series forecasting.

Existing studies have focused on the impact of online review data on sales, but less on the relationship between product description text and sales forecasts, and rarely apply text as a direct parameter to sales forecasting models. Although the same word vectors are used to process the text, a multiple regression model is directly used in the analysis of the impact on sales, and its prediction accuracy is lower than that of neural network models, and the high-dimensional data prediction accuracy is lower than that of neural network models, and the high-dimensional data may be over-fitted, thus affecting the prediction results [57]. In this study, textual data is transformed by word vectors into a deep neural network WaveNet model to sales forecasting, using multiple regression to analyze the impact of product descriptor vectors on sales forecasting accuracy, and finally using SHAP models to interpret the deep learning forecasting results to support decision making.

3. Proposed approach for sales forecasting

A conceptual model on consumer response in an online shopping environment was proposed by[19]. The model utilizes the Stimulus-Organism-Response (S-O-R) framework as its foundation. Stimuli are defined as the total sum of visible and audible cues to online shoppers, while product descriptions, reviews, and website navigation are categorized as high task-related cues that are relevant to the shopping objective. These cue stimuli can lead online consumers to choose alternative products for purchase. Empirical analysis showed that product descriptions (online stimuli) influence consumers' emotions and attitudes towards products (organism, affective/cognitive states), which in turn affect their purchase intention (response) [36]. Product description information presented to consumers on e-commerce platforms has a significant impact on consumer product perception [64], and online shopping satisfaction [69].

Textual information of product descriptions is usually used to describe search attributes, such as product color, brand, usage, effectiveness, and marketing messages; several empirical studies have confirmed that firm-generated product descriptions influence consumer behavior [26, 30, 45]. Therefore, this study uses text mining to add product descriptions to a model for sales forecasting and uses explainable forecasting output to help decision makers obtain more accurate sales forecasting results and product description keywords that can significantly affect sales, and uses the output results of product description feature importance to improve product descriptions on e-commerce platforms.

In this section, we propose a four-step interpretable framework to sale forecasting. Figure 1 summarizes the proposed Explainable AI framework. The step 1 in our Explainable AI framework is the processing of the input data, including the cleansing of the historical sales data and the text mining of the product description, this step outputs the sales data and product description word vectors for training and testing. In the step 2, two WaveNet models are constructed, model A contains time series of historical sales data and prices, and model B adds word vector features in addition to the features in model A. Both models are trained and forecasted using the same training and testing sets, and the prediction errors of the two models are output. The step 3 uses multiple regression to analyze the key product description phrases that affect the accuracy of sales forecasting using the prediction error difference between the two models as the dependent variable and the product description phrases as the explanatory variables, while dimensionality reduction is applied to the product description features of the model. In the fourth step we constructed WaveNet model C containing feature variables of historical sales, prices, and reduced-dimensional product descriptions for sales forecasting, and used the SHAP model to output prediction results and feature importance results to further analyze the impact of product descriptions on sales forecasting to help companies make accurate sales forecasts and optimize product descriptions on e-commerce platforms using key phrases.

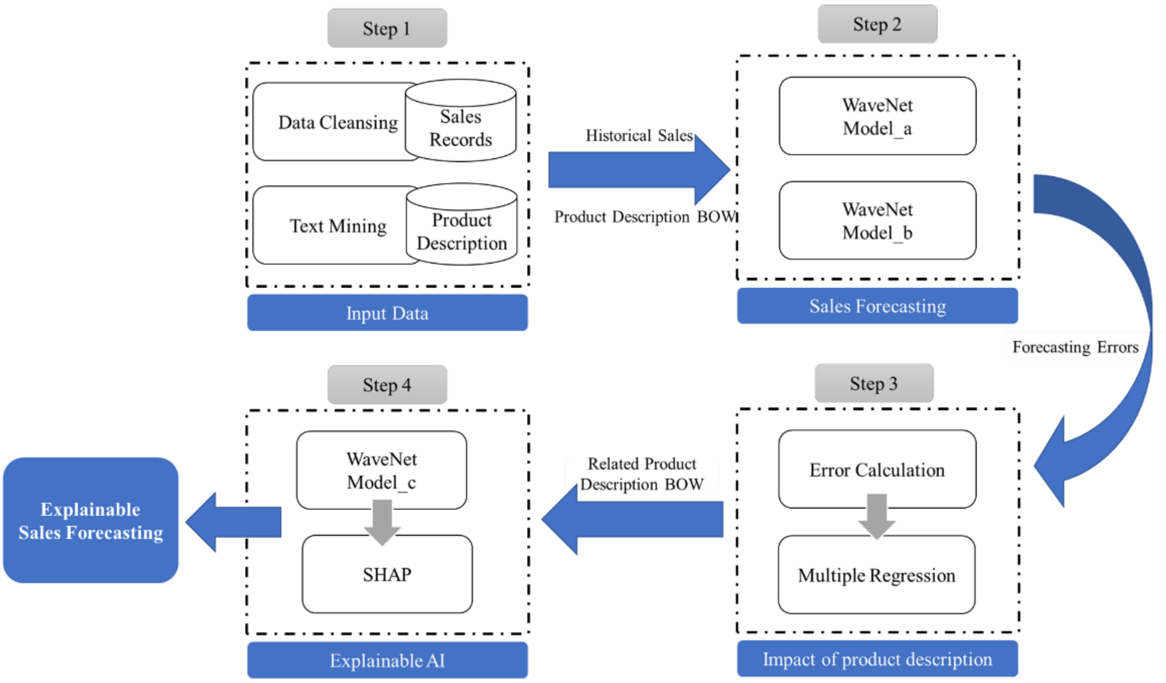

Figure 1. Explainable AI Framework for Sales Forecasting

3.1 Text mining using BOW method

The BOW vectors are constructed for the product descriptions as follows. First, a set of terms $S = \{term_i \mid i \in \{1, \dots, d\}\}$ is defined, where d denotes the total number of terms. Then, a d-dimensional vector $\mathbf{v}_j$ for product j is built as follows:

$$v_j^{(i)} = \begin{cases} 0 & term_i \notin desc_j \\ 1 & term_i \in desc_j \end{cases}, i \in \{1, \dots, d\}, \tag{1}$$

where $desc_j$ denotes the description of product j. A product description can be a word or a combination of words. As some words in product descriptions always appear at the same time, they are combined in one term. The BOW vector converted from a product description is shown in Table 1.

Table 1. BOW vector converted from the product description (ex. Product CZ1213)

Term	Miss Rose	Lip gloss	...	Waterproof	Sexy red	New	...
Component of BOW vector	$v_1^{(1)}$	$v_1^{(2)}$	…	$v_1^{(9)}$	$v_1^{(10)}$	$v_1^{(11)}$	…
Value	1	1	…	1	0	0	…

3.2 WaveNet neural network model

WaveNet [2] is a neural network model consisting of a multilayer CNN (Convolutional Neural Network) that can predict based on time series data. The core of WaveNet is the Dilated Causal Convolutions layer, which allows it to correctly handle temporal order and deal with long-term dependencies without leading to an explosion of model complexity. The WaveNet model is adopted to obtain the discrete probability distribution of the logarithmic sales at time step ($t+1$), and it is expressed as follows:

$$\Pr_{t+1}\left(S = s_{t+1}\right), s_{t+1} \in \{0,\cdots,255\}, \tag{2}$$

First, the logarithmic sale $\log\left(sales_i + 1\right)$ is quantified to 256 possible values using data cleansing, as explained in detail in Section 4.2. Then, the value with the highest probability is taken as a forecast value. Let the forecast value at time step i be s_i, then it holds that:

$$s_{t+1} = \arg\max_{s_{t+1}} \Pr_{t+1}\left(S = s_{t+1}\right). \tag{3}$$

It is necessary to input all sales of previous time steps, that are $\{sales_1,\cdots,sales_{t-1}\}$ to the WaveNet model. The model then generates a discrete probability distribution function of logarithmic sales from time step 2 to time step t:

$$\Pr_i\left(S = s_i\right) = d\left(s_i \mid s_1,\cdots,s_{i-1};\theta\right), i \in \{2,\cdots,t\}, \tag{4}$$

where $d(\cdot)$ denotes the discrete probability distribution function, and θ represents the set of parameters; function $d(\cdot)$ uses a Softmax mapping. A Softmax function maps a set of values to a set of probability distributions whose sum is one, which is expressed as:

$$d\left(s_i \mid s_1,\cdots,s_{i-1};\theta\right)=\frac{\exp\left(a\left(s_i \mid s_1,\cdots,s_{i-1};\theta\right)\right)}{\sum_{S=0}^{255}\exp\left(a\left(S \mid s_1,\cdots,s_{i-1};\theta\right)\right)}, \tag{5}$$

where $\sum_{S=0}^{255} d\left(S \mid s_1,\cdots,s_{i-1};\theta\right)=1$, and $a(.)$ is an activation function of the WaveNet model, which is expressed as:

$$\begin{gathered} a(s_i|s_1,\cdots,s_{i-1};\theta) = Linear(ReLU(Linear(ReLU(\textstyle\sum_{k=1}^{n} Skipped_k)))) \\ Skipped_k = tanh(W_{g_1,k} * s) \odot \sigma(W_{g_2,k} * s) \end{gathered} \tag{6}$$

where $*$ denotes a convolution operator, $\odot$ denotes an element-wise multiplication operator, $\sigma(\cdot)$ is a sigmoid function, and k is the layer index; g_1 and g_2 denote filter and gate, respectively; $\boldsymbol{W}$ is a learnable convolution filter that includes a parameter set θ; $\boldsymbol{s}$ is the vector of dilated causal convolution operated sales; n is the number of layers, ReLU denotes a rectifier, and Linear denotes a linear operation with the same number of input and output dimensions.

Since the actual values of $sales_2,\cdots,sales_t$ are known, the parameter set θ can be optimized so that the actual values correspond to the highest probabilities. Specifically, the joint probability of historical sales data is factorized as a product of conditional probabilities; let $p_i\left(s_i\right)=\Pr_i\left(S=s_i\right)$, $p(\cdot)$ denotes the joint probability function. The goal is to obtain a parameter set θ that will increase the value of $p(\cdot)$ as much as possible, which is expressed as:

$$\theta=\arg\max_{\theta} p\left(s_2,\cdots,s_t\right). \tag{7}$$

Then, the ideal θ can be iteratively obtained by:

$$\theta:=\theta+\alpha\cdot\frac{\partial p}{\partial\theta}, \tag{8}$$

In this work, a logarithmic operation is performed on $p(\cdot)$ without changing its monotonicity; this, Eq. (8) becomes:

$$\begin{aligned} &\theta=\arg\max_{\theta}\ln(p(s_2,\cdots,s_t)) \\ &\ln(p(s_2,\cdots,s_t))=\sum_{i=2}^{t}\ln d(s_i \mid s_1,\cdots,s_{i-1};\theta) \\ &\qquad=\sum_{i=2}^{t}(a(s_i \mid s_1,\cdots,s_{i-1};\theta)-\ln\sum_{S=0}^{255}\exp(a(S \mid s_1,\cdots,s_{i-1};\theta))) \end{aligned}. \tag{9}$$

By training the model on other input variables, the WaveNet can be optimized to forecast sales with the required characteristics. For instance, a product can be chosen by adding the product price to the model as an extra input parameter. Similarly, for forecasting using the product description text mining results, the BOW vectors converted from the product description are added as an extra input parameter.

The model is trained on the other inputs in the way of global training, which is characterized by a single latent representation $\boldsymbol{h}$ that influences the output distribution across all time steps. Then, the activation function given by Eq. (6) becomes:

$$a(s_i|s_1,\cdots,s_{i-1};\theta) = Linear(ReLU(Linear(ReLU(\sum_{k=1}^{n} Skipped_k))))$$
$$Skipped_k = tanh(W_{g_1,k} * s + V_{g_1,k}^{T} h) \odot \sigma\left(W_{g_2,k} * s + V_{g_2,k}^{T} h\right) \quad (10)$$

where $\boldsymbol{V}$ denotes a learnable linear projection, and vector $\boldsymbol{V}^{\boldsymbol{\cdot}}\boldsymbol{h}$ represents broadcast a time dimension.

3.3 Explainable framework for sales forecasting

The developed model is specified to explain the variation in $\log\left(sales_{jt}+1\right)$, where $sales_{jt}$ represents the sales of product *j* in time step *t*, for products $j \in \{1,\cdots,J\}$ and over time steps $t \in \{1,\cdots,T\}$.

Model A considered only variables in traditional sales forecasts, such as historical sales and prices (*p*), regardless of the impact of product descriptions on sales. The variables were fed to the WaveNet neural network input for sales forecasting as follows:

$$Model\ A: log\left(\widehat{sales_{j(t+1)}}+1\right) = WaveNet\left(log\left(sales_{j1}+1\right),\ldots,log\left(sales_{jt}+1\right),ln\left(p_j\right)\right)+\varepsilon_{j(t+1)}. \quad (11)$$

Model B considered the influence of the product description on sales, based on Model A:

$$Model\ B: log\left(\widehat{sales_{j(t+1)}}+1\right) = WaveNet\left(log\left(sales_{j1}+1\right),\ldots,log\left(sales_{jt}+1\right),ln\left(p_j\right),\boldsymbol{v}_j\right)+\varepsilon_{j(t+1)}^{'}. \quad (12)$$

A variable $\boldsymbol{v}_j$ that was a d-dimensional vector that represented information related to the product description was introduced using the BOW method.

The difference between the forecasting results of Models A and B could be regarded as a difference in the impacts of the product descriptions on sales. It was examined which terms have a more significant impact through a multiple regression analysis:

$$\varepsilon_{jt} - \varepsilon'_{jt} = \beta_0 + \boldsymbol{\beta}^{\bullet}_{BOW}\boldsymbol{v}_j . \tag{13}$$

where $\boldsymbol{\beta}^{\bullet}_{BOW}\boldsymbol{v}_j$ expressed the effects of the description of product j, and $\boldsymbol{\beta}_{BOW}$ denoted a d-dimensional parameter vector that represented the effects of product descriptions.

Model C adds product description characteristics with significance based on the results of multiple regression, based on Model A:

$$Model\ C: log\left(\widehat{sales_{j(t+1)}} + 1\right) = WaveNet\left(log\left(sales_{j1} + 1\right), \ldots, log\left(sales_{jt} + 1\right), ln\left(p_j\right), \boldsymbol{v}_k\right) + \varepsilon'_{j(t+1)}. \tag{14}$$

Finally, SHAP values were calculated to explain model C. We generated explanatory visualizations using the public python package. The SHAP values give accurate information about which features contribute to the predicted values and an estimate of their overall impact on the predictions.

4. Data preparation

4.1 Data overview

We conducted this study in collaboration with a cosmetic sales firm operating on the AliExpress cross-border e-commerce platform which is the China's largest cross-border export B2C platform and the world's third largest English online shopping site, is Alibaba's cross-border e-commerce platform for the international market, through the Alipay international account for guaranteed transactions, and the use of international logistics channels transport delivery for overseas buyers and customers. The historical sales data is obtained from partner cosmetic sales firm's sales records on the AliExpress platform (Figure 2). The sales record data consisted of 524,253 collected from July 26, 2016, to September 7, 2019, during a total of 1,139 days. The parameters of one record example of the data are shown in Table 2.

Table 2. Data overview

Field	Example
Order time	2017/11/12 0:00:43
Customer	L…
Country	Israel
City	Nazareth-illit
Product description	Miss Rose Brand lips Matte Moisturizing Lipstick Makeup Lipsticks Lip Stick Waterproof Lipgloss Mate Lipsticks Make up [38]
SKU	CZ1213-38
Order quantity	1

Amount (USD)	1.12

The product descriptions are obtained from the e-commerce platform of the partner companies, and contain the product descriptions of similar best-selling products, and we obtained the product descriptions of other companies for similar products through the search of keywords. The product description represents the content presented to the customers on AliExpress cross-border e-commerce platform, as indicated by the dashed box in Figure 2, but the product descriptions that are shown in the Table 2 and the figure are different. The product-description content represents a description of the same product (main SKU, stock keeping unit), but without any punctuation; also, the letter case is not standardized, and there are even misspellings.

SKU is a coding and classification method used for products after they are stocked, and it is also the smallest unit of inventory control. It can be measured in terms of pieces, boxes, pallets, etc. Each product corresponds to a unique SKU number, which includes attributes such as brand, model, configuration, grade, packaging capacity, unit, production date, shelf life, purpose, price, place of origin, etc. When the attributes of a product are different from those of other products, it is defined as a unique item corresponding to a specific SKU number. In our data, a product's SKU is generally composed of the main SKU and a sub-SKU. For instance, for the product "CZ1213-38," the main SKU was "CZ1213," and the sub-SKU was "38." Sub-SKUs were only used to distinguish the details of the product, such as different colors of the same type of lipstick. Therefore, the products with the same main SKU were regarded as the same products, and their sales were summarized. The sales referred to the order quantity, not the amount. Then, each six-hour period was divided into a time step, starting from July 26, 2016, at 6:00. The product sales were summarized in order to form historical sales data.

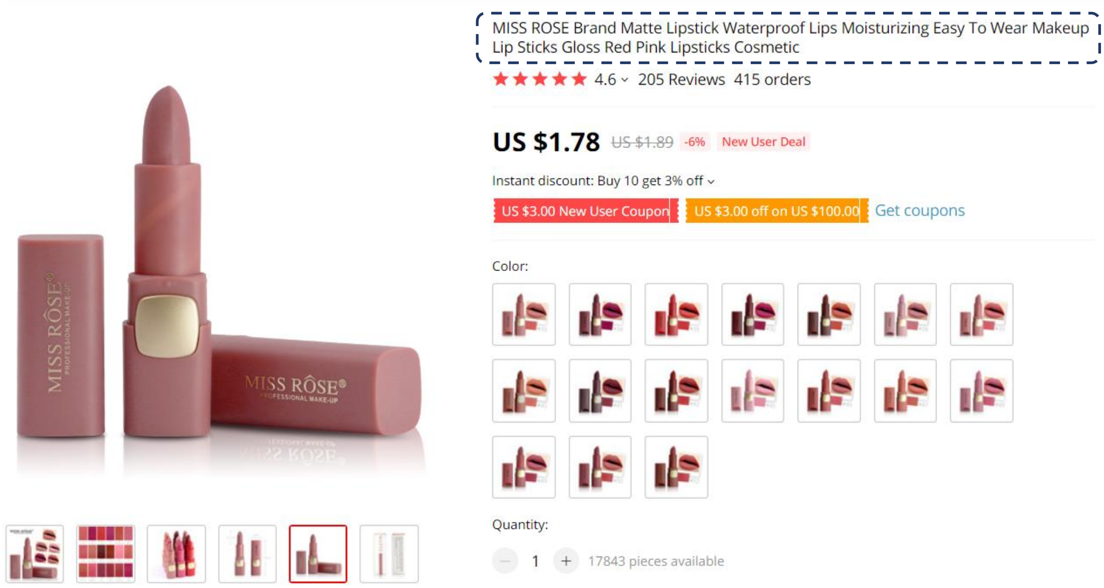

Figure 2. A screenshot of AliExpress cross-border e-commerce platform (website)

First, the life cycle of the product, which was defined as a time interval between the first and last sales records of the product, was analyzed. The data included the sales of 1,083 products, but only the sum of sales of 37 products with the highest sales accounting for more than 80% of the total was considered. The statistical description of the life cycles is shown in Table 3.

Table 3. The statistical description of life cycles of 37 representative products*

Size	37	25% Quantile	431
Mean	585.97	Median	598
Std	197.62	75% Quantile	750
Min	192	Max	991

* The unit of life cycle is day.

As shown in Table 3, both the mean and median of the life cycles of 37 representative products were approximately 590 days. Thus, six products with life cycles of approximately 590 days were selected to examine specifically. The historical sales data of these six products are shown in Figure 3, where the outlier values that deviate from the mean of historical sales by more than three standard deviations are replaced with the mean of non-outlier values.

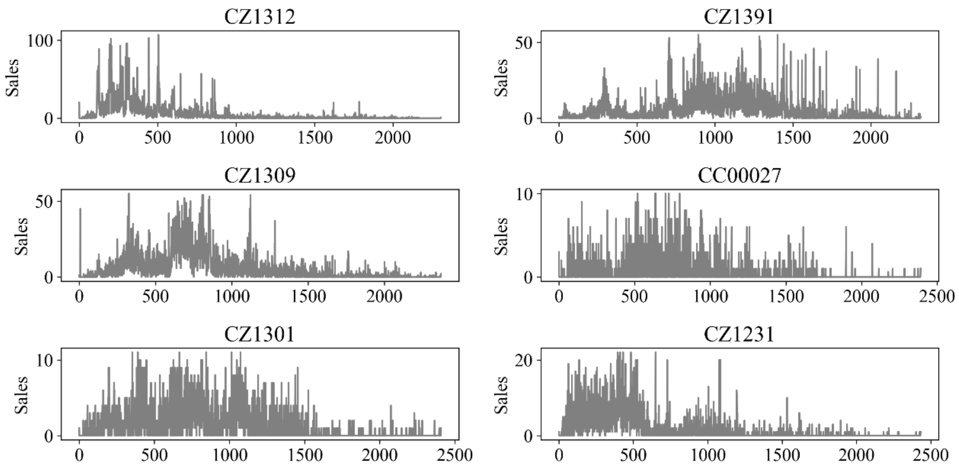

Figure 3. The historical sales data of six products with the life cycles of approximately 590 days* (*: In the figure, each unit on the horizontal axis represents six hours; every four units represent a period of one day)

As shown in Figure 3, the main sales of products were generally concentrated between half of the life cycle and 2/3 of the life cycle, which was approximately 300-400 days. Considering the full seasonal cycle, one-year sales data, from November 12, 2017 to November 10, 2018, were used to obtain data samples for further analysis. The data on November 11 were avoided because that day is the busiest sales day of the year on the analyzed shopping platform, so the sales of most products could be much higher on that day than on any of the other days of the year. The total number of extracted sales records was 192,876. The products were first sorted by sales in descending order, and then the cumulative sales were calculated, and finally, the top 80% of the products were selected to be used in the forecasting-model development process. The sum of the sales of these products accounted for 80% of the total sales of all the products in the collected data.

4.2 Sales data cleansing

The sales data was cleaned through several steps. First, the values that deviated from the mean of historical sales by more than three standard deviations were replaced with the mean to mitigate the impact of outliners. Next, the historical cumulative sales of each product were calculated, and the cumulative sales were converted into the cumulative sales percentage (cumulative sales / total sales), and then the main part, which was from 10% to 90% of the cumulative sales, was taken as a sample. Then, the logarithm of sales was calculated to mitigate the impact of scale problems. Namely, $\log(sales+1)$ was calculated, where the sales value was increased by one to avoid the log parameter being zero. Finally, the logarithm of sales was normalized and mapped to 256 integers in the range of [0, 255]. This was done because, in the product sales forecasting, the SoftMax function was used to output the occurrence probability of all possible sales values. Thus, it was necessary to convert continuous sales into a limited number of discrete values. The z-transformation was applied to the logarithm of sales, and it was quantized to 256 possible values. The data cleansing results are shown in Figure 4.

4.3 Product description data cleansing

The product description data was subjected to several steps to obtain text mining data. First, product descriptions were converted into lowercase, and non-letter characters were removed. The spelling mistakes were corrected, and the expressions of synonyms were unified. Next, the description of each product was divided into three categories: brand, phrases, and words. Then, it was ensured that the resulting form of a word was a known word in a dictionary, which has been known as the lemmatization process, and the duplicate words were removed. The lemmatization process removes affixes only if the resulting word is in its dictionary. The product descriptions before and after data cleansing are shown in Table 4.

Table 4. Comparison of product descriptions before and after data cleansing

		Product description (ex. Product CZ1213)
Before cleansing		Miss Rose Brand lips Matte Moisturizing Lipstick Makeup Lipsticks Lip Stick Waterproof Lipgloss Mate Lipsticks Make up
After cleansing	Brand	Miss Rose
	Phrases	Lipgloss
	Words	Brand, lip, matte, moisturizing, lipstick, makeup, waterproof

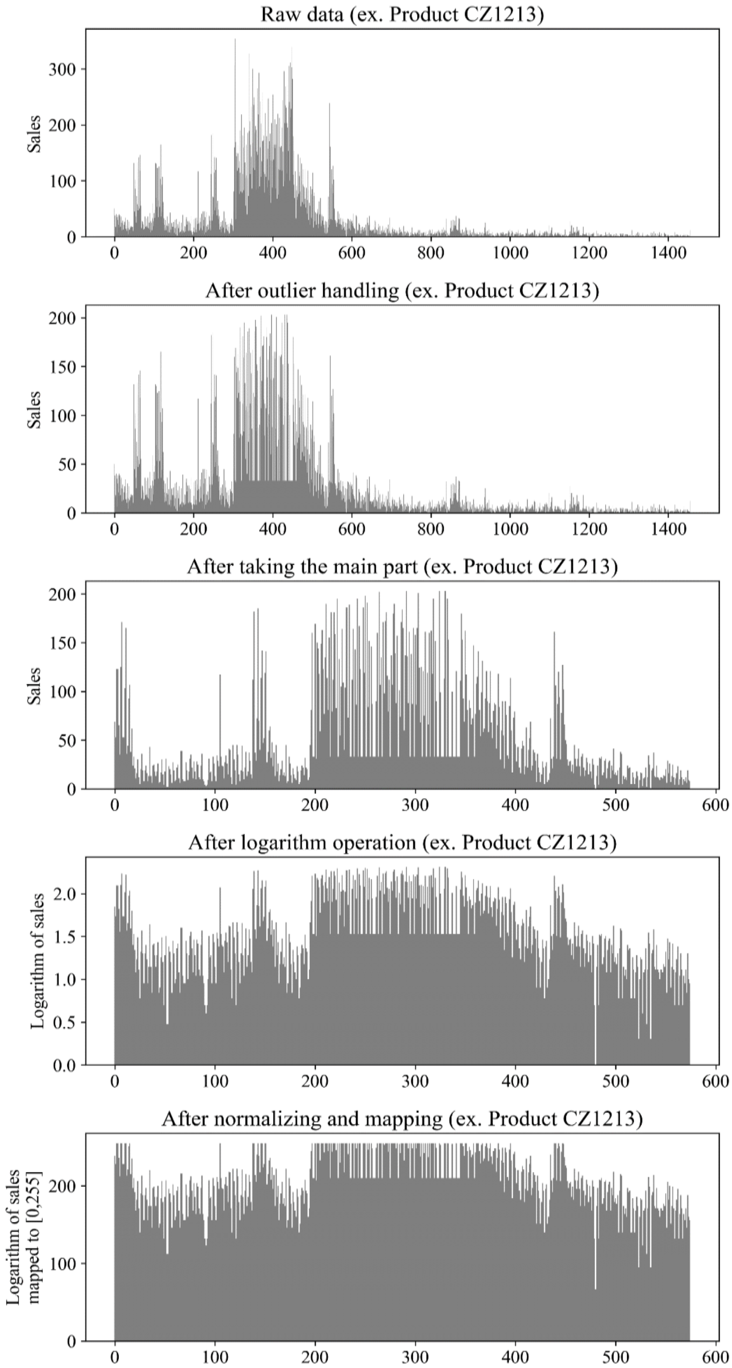

Figure 4. Sales data cleansing results

5. Empirical results

5.1 Model evaluation and cross-validations

WaveNet neural network is capable of extracting features of sales variations across different time spans, if the product's sales exhibit seasonality, the model will expect the neural network to learn these patterns and incorporate them into the forecasted results. Since deep learning models can automatically capture features from high-dimensional data, we use WaveNet model A as a benchmark model to compare the sales prediction performance with CNN and LSTM model based on the metrics of the regression task, i.e., mean absolute error, root mean square error, and mean square error. Table 5 summarizes the sales forecast results of the three models, and it can be seen that the WaveNet model has better forecasting results. Therefore, in our article we chose the WaveNet deep learning model as the sales forecasting model.

Table 5. Comparison of model forecasting performance

Forecasting model	Mean Absolute Error	Mean Square Error	Root Mean Square Error
WaveNet	23.07	14.3	27.95
LSTM	23.12	15.12	30.56
CNN	38.39	24.42	47.2

Then, we divide the time series dataset into 5 contiguous blocks for cross-validation. Each block is first used to test the model and then retrain it. The first time, the first data block was used as the training level and the second as the test set, the second time, the first and second data blocks were used as the training set and the third as the test set, the third time, the first, second, and third data blocks were used as the training set and the fourth as the test set, and the fourth time, the first, second, third, and fourth data blocks were used as the training set and the fifth as the test set. The mean absolute error, root mean square error and mean square error are used to evaluate the prediction results of the model. Table 6 summarizes the prediction results of four different datasets, and the differences in the prediction performance of the four datasets are small, and the model predicts better as the training set data increases, from which reflects the robustness of our WaveNet model.

Table 6. Cross-validation results

Forecasting model	Mean Absolute Error	Mean Square Error	Root Mean Square Error
DataSet1	24.24	13.04	32.56
DataSet2	23.71	30.62	30.62
DataSet3	22.96	24.49	24.65
DataSet4	21.18	24.19	26.35

5.2 WaveNet Models A and B forecasting results

The one-year sales data, as well as the price of the products and the BOW vectors converted from the product descriptions, were used to train Model A and Model B. The training objective was that the mean cross-entropy of 10 training epochs had to be less than 0.02. Then, each trained model was used to forecast sales for the next seven days. As mentioned above, each day was divided into six-hour time intervals: 0:00 - 6:00, 6:00 - 12:00, 12:00 - 18:00, and 18:00 - 24:00; thus, each day included four time steps. Thus, there were 28 (seven days times four time steps) forecasted values for each product. The comparison between the actual and forecasted sales values of Model A and Model B are shown in Figure 5 and Figure 6, respectively.

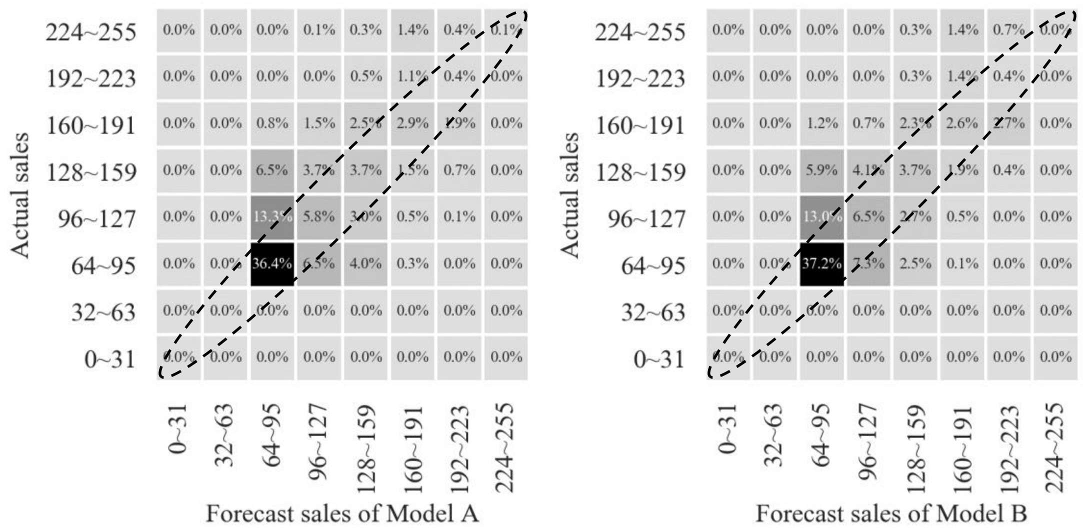

Figure 5. Confusion matrices of the forecasting results of Model A and Model B

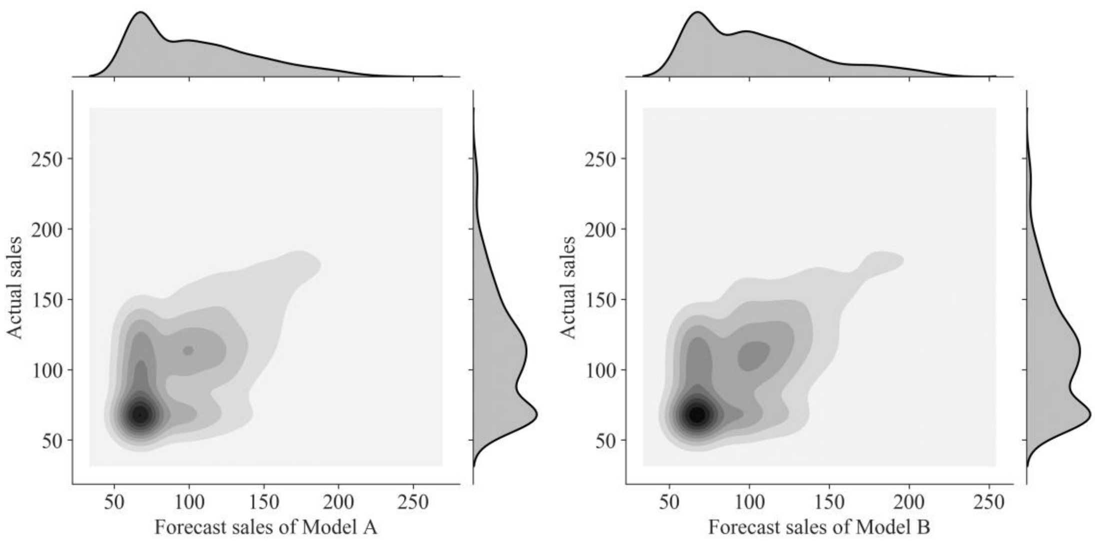

Figure 6. The kernel density estimation (KDE) curves of the actual and forecasted sales values of Model A and Model B

Figure 5 shows the confusion matrices of forecasting results of Model A and Model B. As shown in Figure 5, most of the forecasting results were within the diagonal squares marked by the dashed box, indicating that they were close to the actual values.

From the confusion matrix, both Model A and Model B perform best when the actual value is in the 64-95 range, with the predicted value also tending to fall within this range. Additionally, compared to Model A, Model B, which incorporates BOW vector parameters, shows a slight improvement in prediction accuracy in this range. Furthermore, the KDE curves of actual and forecasted sales values of Model A and Model B are shown in Figure 6, where the same trend as that of the confusion matrices can be observed.

In order to validate the forecasting results of the WaveNet neural network model, it was investigated whether the forecast errors of Model A and Model B obeyed the mean normal distribution. The frequency distributions of the forecasting errors of Model A and Model B are shown in Figure 7. In addition, the statistical descriptions and normality test results of the forecasting errors of Model A and Model B are given in Table 7.

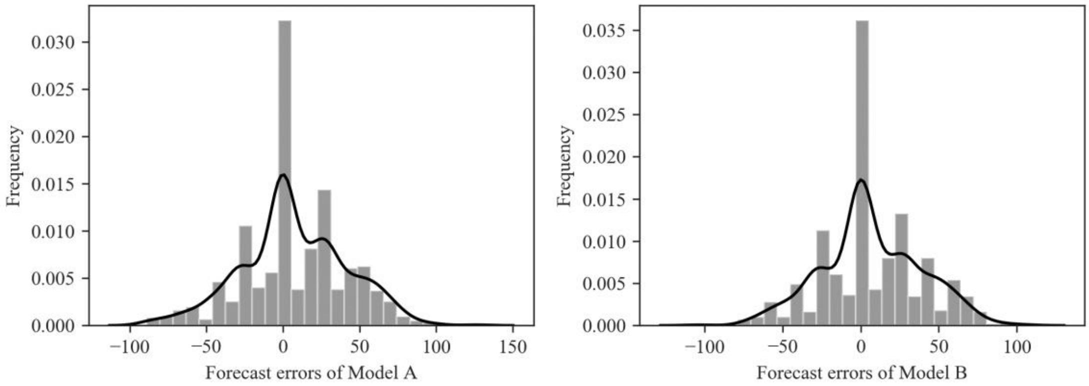

Figure 7. Frequency distributions of forecasting errors of Model A and Model B

Table 7. Statistical descriptions and normality test results of forecasting errors of Model A and Model B

		Forecasting errors	
		Model A	Model B
Statistic	Size	728	728
	Min	-89	-104
	Max	126	106
	Mean	7.3530	6.7102
	Std	34.3798	31.9775
Skew test	Skew of data set	-0.0275	0.0409
	Skew test p-value	0.7596	0.6493
Kurtosis test	Kurt of data set	0.0477	-0.0711
	Kurt test p-value	0.6941	0.7809
Normal test	Norm test p-value	0.8832	0.8675

It was tested from three aspects, whether the distribution of the dataset differed from the normal distribution. The null hypothesis (H0) was that the data set obeyed the normal distribution, while the opposite hypothesis (H1) was that the dataset did not obey the normal distribution. The two-sided p-value was calculated for the hypothesis test. The results showed that the p-value was sufficiently large not to reject H0, i.e., the forecasting errors of Models A and B obeyed the normal distribution. Thus, the forecasting result of the WaveNet neural network model could be considered to be valid.

Furthermore, as shown in Table 7, Model B that included the product description BOW variables achieved more accurate forecasting. The error mean of Model B was closer to zero than that of Model A, and the error standard deviation was smaller. Therefore, adding product descriptions to the forecasting model positively affected forecasting results accuracy.

5.3 Multiple regression analysis of product description features

In order to examine the impact of each specific product sales parameter, according to Eq. (13), the results that Model B forecasted more accurately than Model A were selected and used as a dependent variable that expressed the difference between the forecasting errors of the two models. Also, the BOW vectors were converted from the product descriptions as explanatory variables to conduct a multiple regression analysis. The results are given in Table 8.

Table 8. Multivariate regression analysis results of product sales and product description BOW vectors

		Coef	Std err	P-value
	Constant	3.206	1.997	0.109
	Terms of BOW vectors			
1	Lipgloss	-26.066	6.838	0.000**
2	Moisturizing	13.565	2.975	0.000**
3	Liquid	-10.864	3.000	0.000**
4	Glitter	-4.816	1.268	0.000**
5	Beauty	13.384	3.336	0.000**
6	Powder	7.293	1.811	0.000**
7	Moisturizer	13.423	3.719	0.000**
8	New	-10.372	3.131	0.001**
9	Makeup	4.975	1.730	0.004**
10	Velvet, glossy, balm	6.178	2.206	0.005**
11	Gold	-3.531	1.252	0.005**
12	Brighten	-4.992	1.785	0.005**
13	Bronzer	4.819	1.705	0.005**
14	Concealer	5.563	1.957	0.005**
15	Lip	-9.241	3.406	0.007**
16	Red	-4.500	1.668	0.007**
17	Stamp, eye	4.878	1.839	0.008**
18	Eyeliner	5.008	1.928	0.010**

19	Silver	-3.661	1.452	0.012*
20	Waterproof	7.245	2.935	0.014*
21	Illuminator	-5.026	2.038	0.014*
22	Black	4.564	1.915	0.018*
23	Metal color	6.654	2.960	0.025*
24	Natural	-3.905	1.743	0.026*
25	Snail (cream), whitening, nourish, skin tone, acne treatment, day	-4.952	2.241	0.028*
26	Anti-wrinkle, deep facial, hyaluronic acid serum, intensive lifting, firming	5.195	2.480	0.037*
27	Double	-3.318	1.594	0.038*
28	Pudaier (brand)	5.339	2.611	0.041*
29	Miss Rose (brand)	-4.822	2.504	0.055
…	…	…	…	…
62	Glow kit, highlighter	-0.207	1.428	0.885
63	Anti-aging, skin care essence	0.243	1.707	0.887
64	White	0.130	1.349	0.923

∗ Denotes significance at the 5%-level.

∗∗ Denotes significance at the 1%-level.

As shown in Table 8, although the BOW vectors with high correlation were combined, there were still 64 BOW vectors converted from the product descriptions. Also, the coefficient estimates of 28 of 64 BOW vectors were statistically significant, having a p-value of less than 0.05 and accounting for 43.75% of the total BOW vectors. Thus, the information reflected by these terms could be considered to affect the sales of the related products significantly and positively. Consequently, adding BOW vectors from product description can improve sales. Hence, the text mining analysis on product descriptions can be helpful to sales forecasting.

5.4 Explanation of sales forecasting results

We obtained 28 BOW vectors with significant effects on sales forecasting from the multiple regression results of the error difference between Model A and Model B and the product description BOW, so we added these 28 BOW vectors to Model A to form the new WaveNet Model C. Similarly, we used the same training and test sets as Model A and Model B for sales forecasting, and the performance of the forecasting results is shown in Table 9. From Table 9, we can find that the forecasting accuracy of the model is improved after the dimensionality reduction of the product description feature vectors.

Table 9. Comparison of WaveNet Model forecasting performance

Forecasting model	Mean Absolute Error	Mean Square Error	Root Mean Square Error
Model_A	23.07	14.30	27.95
Model_B	18.14	10.72	22.68
Model_C	17.07	10.30	20.93

Finally, we use the SHAP model to interpret the output results of WaveNet Model C. By SHAP value, we can get the influence of each feature variable on the output results of the model. Figure 8 and Figure 9 show the importance of ranking of each feature variable. Using our proposed explainable AI framework for sales forecasting can help firms improve the accuracy of forecasting, and firm decision makers can make better operational management strategies for purchasing and inventory based on the model's forecasting results.

Meanwhile, our product descriptions are not only from the firms' own descriptions of their products, but also from the descriptions of other firms on the e-commerce platform for products in the same category, especially the product descriptions of those best-selling products. Therefore, the keyword vectors output from the model and their impact on sales forecasting showed in Figure 8 and Figure 9, can help decision makers to better formulate marketing strategies, optimize product descriptions on product pages on e-commerce platforms, and add product description keywords to the titles of relevant products to gain more attention from consumers and thus increase product sales.

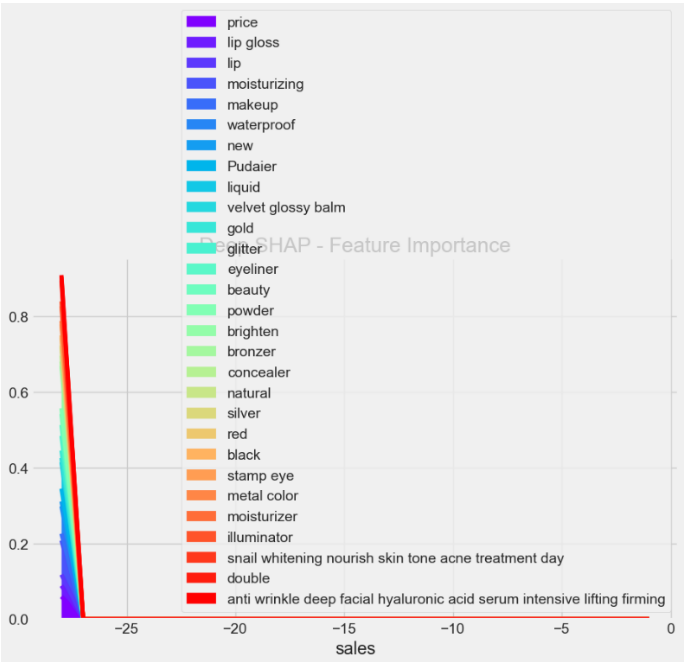

Figure 8. Feature Importance of Model C

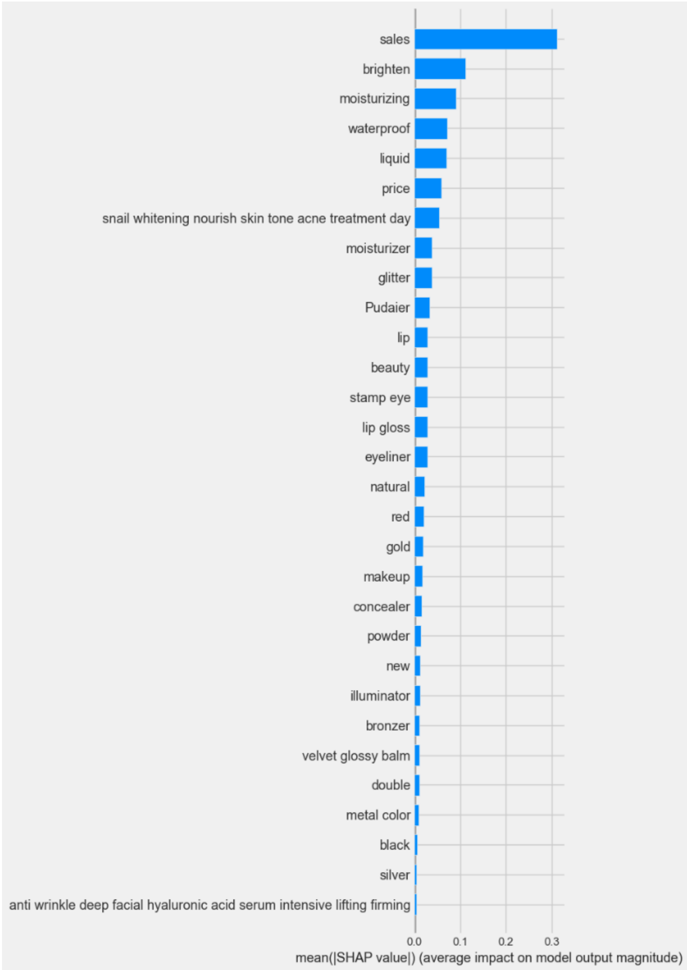

Figure 9. Mean SHAP value (average impact on model output magnitude) of Model C

6. Discussion

6.1 Implication for marketing decision support

Previous research has shown that information disseminated by sellers exhibits strong persuasive power, and it has as much impact on purchase likelihood as user-generated information such as word-of-mouth reviews [5]. Product descriptions, including textual descriptions, have a significant impact on perceived convenience [30] and credibility [43], making them one of the factors influencing consumer purchasing behavior [26]. The existing research literature mainly analyzed the influence of a part of types of web information, such as online reviews and service guarantees, on consumers' purchase decisions [20, 21, 57, 71, 74]. This study focuses on product descriptions to obtain product description keywords that can impact sales forecasting and consumer buying behavior, and proposes an explainable approach using text mining and deep learning to accurately identify product selling points to support enterprise operation and marketing decisions

The BOW method and WaveNet neural network model are used to demonstrate that product descriptions presented to customers can influence sales forecasting. Forecasting models with a vector of product description words are more accurate than models without texture data. Furthermore, using multiple regression analysis, we identified word groups in product descriptions that can significantly influence sales (p-value < 0.05), and the forecasting accuracy was again improved by adding the significantly influential word groups back into the baseline deep learning model, and we used the SHAP values to output the importance of each product description feature that is, the degree of influence on sales forecasting results. Our results show that approximately 44% of the terms used in product descriptions can significantly influence the level of sales forecasts, and these product descriptions can be considered as selling points for the product. Therefore, companies can add certain types of terms to product descriptions to promote product sales. In addition, companies and marketers should refer to the customer needs reflected in these terms when designing product descriptions. The method proposed in this study has broad applicability. As long as the descriptive text of the product displayed to the consumer at the time of sale is clarified and historical product sales data over a certain time span is available, it is possible to find out which phrases have a significant impact on the sales of the product. In addition, textual data, such as typically unstructured data, companies are faced with the urgent problem of how to eliminate invalid distracting content from the data during analysis. This study considers textual data in the form of product descriptions and our method can also be considered as a text mining technique that can be used to remove noise, i.e. invalid interfering content, from the data.

6.2 Implications from inter-pretable forecasts

There is a very rich research literature on time series forecasting [6, 10, 13, 31, 72]. Traditional statistical models, as well as machine learning and deep learning models, have been widely used in time series forecasting, and the use of deep learning models [4, 18, 25, 47] has increased rapidly especially with the increase in data volume and the high dimensionality and nonlinearity of features. Sales forecasting is typical of time series forecasting, and scholars have continued to improve the accuracy of sales forecasting in terms of research methods and feature engineering [16, 27]. However, using complex deep learning models for critical decision making faces the challenge that these models are often used as "black boxes". This means that we need to explain the output of the models to the decision makers and increase their trust in the models and the results.

Explainable AI is essential for users to understand, trust and manage these AI 'partners'. The range of explainability includes algorithmic transparency, global explainability and local explainability. Using the inherent interpretability of the tree model to explain the model output is a common research approach [29, 70]. In addition to the explainable algorithm TFT on time series forecasting [44], existing locally explainable algorithms such as LIME (locally explainable model agnostic explanation) and SHAP (SHapley additive explanation) are mostly used for machine learning of classification problems. And, we found that the existing literature lacks explainable studies of the WaveNet model. Therefore, we provide new ideas for decision-supported interpretable machine learning, providing a framework for applying explainability to deep learning, especially for time series prediction based on convolutional neural network model. Our model combines WaveNet with multiple regression based on textual data as well as SHAP model to provide improved forecasting performance compared to traditional machine learning, while also being highly explainable. Examining each product description phrase through multiple regression analysis of the error difference between two WaveNet deep learning models, merging product description phrases with significant impact into the forecasting model indicates further improvement in forecasting performance. Using the SHAP model to output the importance of features and their impact on output results allows business decision makers to use our explainable AI model for more accurate sales forecasting, providing decision support for business operations management, while helping them find key phrases that impact sales forecasting to optimize product description information on e-commerce platforms to help businesses gain more consumer attention.

6.3 Limitations and potential for future research

The method proposed in this study can be further extended to other areas. For example, the method could be used to examine and determine which keywords in media reports have had a significant impact on stock market prices over a certain period of time. This would involve replacing product descriptions with the text of media reports and replacing product sales data with historical stock market price data. It is also possible to check which topics in the search engine have a significant impact on the GDP of a particular country. This can be found by replacing product descriptions with search engine search history and replacing product sales with the country's GDP. The results of the study found that the BOW vector obtained by text mining of product descriptions had a smaller impact on product sales compared to the historical product sales and price variables. Whether the reasons for this are that product descriptions naturally have less impact on product sales than historical sales and prices, or that some information is lost when converting product descriptions to word vectors, is unclear and therefore future research should be extended to explore these reasons. At the same time, the current forecasts are still based on intra-SKU forecasts, and taking into account the influence of inter-SKU factors in the predictor variables is also one of the future research directions, as well as on interpretability future research could further expand on the transparency and global interpretability studies of the WaveNet model's time series forecasting algorithm.

7. Conclusion

This study presents an explainable framework for sales forecasting based on deep learning. We transform product descriptions, which are unstructured textual data, into word vector variables and add them to a WaveNet sales forecasting model, which has higher forecasting performance than traditional time series forecasting methods and forecasting models that do not consider product descriptions. Multiple regression and SHAP model were also used to explain the forecasting results, helping companies to better understand and trust the model and the forecasting results, and to improve the interpretability of the forecasting model and results. The main conclusions based on the results obtained are as follows: (1) Adding product descriptions to the forecasting model helps to improve forecasting accuracy. (2) If historical sales data is available for a certain time span, then the company can identify the terms that have the greatest impact on product sales and change the product descriptions accordingly. (3) Companies can increase sales by adding key phrases to describe their products to customers in a more appealing way at the point of sale.

References

[1] C. A, How Artificial Intelligence is Transforming Business' Supply Chains, in, (2021).

[2] S.D. Aaron van den Oord, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, in, (2016), pp. arXiv:1609.03499.

[3] M.Z. Abedin, M.H. Moon, M.K. Hassan, P. Hajek, Deep learning-based exchange rate prediction during the COVID-19 pandemic, Ann Oper Res, (2021).

[4] S. Ahmed, R.K. Chakrabortty, D.L. Essam, W.P. Ding, Poly-linear regression with augmented long short term memory neural network: Predicting time series data, Inform Sciences, 606(2022) 573-600.

[5] Barbara, Bickart, Robert, M, Schindler, Internet forums as influential sources of consumer information, Journal of Interactive Marketing, (2001).

[6] M. Benhaddi, J. Ouarzazi, Multivariate Time Series Forecasting with Dilated Residual Convolutional Neural Networks for Urban Air Quality Prediction, Arab J Sci Eng, 46(4) (2021) 3423-3442.

[7] T. Boone, R. Ganeshan, R.L. Hicks, N.R. Sanders, Can Google Trends Improve Your Sales Forecast?, Prod Oper Manag, 27(10) (2018) 1770-1774.

[8] T. Boone, R. Ganeshan, A. Jain, N.R. Sanders, Forecasting sales in the supply chain: Consumer analytics in the big data era (vol 35, pg 170, 2019), Int J Forecasting, 37(3) (2021) 1311-1311.

[9] G.E.P. Box, An Unexpected Route to Time-Series - a Citation Classic Commentary on Time-Series Analysis - Forecasting and Control by Box,G.E.P., and Jenkins,G.M., Cc/Phys Chem Earth, (44) (1989) 22-22.

[10] T.M. Choi, C.L. Hui, N. Liu, S.F. Ng, Y. Yu, Fast fashion sales forecasting with limited data and time, Decis Support Syst, 59(2014) 84-92.

[11] T.M. Choi, D. Li, H. Yan, Optimal two-stage ordering policy with Bayesian information updating, J Oper Res Soc, 54(8) (2003) 846-859.

[12] T.M. Choi, D. Li, H.M. Yan, Quick response policy with Bayesian information updates, Eur J Oper Res, 170(3) (2006) 788-808.

[13] T.M. Choi, Y. Yu, K.F. Au, A hybrid SARIMA wavelet transform method for sales forecasting, Decis Support Syst, 51(1) (2011) 130-140.

[14] K. Coussement, D.F. Benoit, Interpretable data science for decision making, Decis Support Syst, 150(2021).

[15] R.M. Cui, S. Gallino, A. Moreno, D.J. Zhang, The Operational Value of Social Media Information, Prod Oper Manag, 27(10) (2018) 1749-1769.

[16] D.F. Davis, J.T. Mentzer, Organizational factors in sales forecasting management, Int J Forecasting, 23(3) (2007) 475-495.

[17] A. De Caigny, K. Coussement, K.W. De Bock, S. Lessmann, Incorporating textual information in customer churn prediction models based on a convolutional neural network (vol 36, pg 1563, 2019), Int

J Forecasting, 37(3) (2021) 1306-1306.
[18] M.I.A. Efat, P. Hajek, M.Z. Abedin, R.U. Azad, M. Al Jaber, S. Aditya, M.K. Hassan, Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales, Ann Oper Res, (2022).
[19] S.A. Eroglu, K.A. Machleit, L.M. Davis, Atmospheric qualities of online retailing - A conceptual model and implications, J Bus Res, 54(2) (2001) 177-184.
[20] Z.P. Fan, Y.J. Che, Z.Y. Chen, Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis, J Bus Res, 74(2017) 90-100.
[21] E. Fernandes, S. Moro, P. Cortez, F. Batista, R. Ribeiro, A data-driven approach to measure restaurant performance by combining online reviews with historical sales data, Int J Hosp Manag, 94(2021).
[22] S. Feuerriegel, J. Gordon, Long-term stock index forecasting based on text mining of regulatory disclosures, Decis Support Syst, 112(2018) 88-97.
[23] R. Fildes, S. Kolassa, S.H. Ma, Post-script-Retail forecasting: Research and practice, Int J Forecasting, 38(4) (2022) 1319-1324.
[24] B.K. Finale Doshi-Velez, Towards a rigorous science of interpretable machine learning, in, (2017), pp. arXiv:1702.08608.
[25] Q.Q. He, C. Wu, Y.W. Si, LSTM with particle Swam optimization for sales forecasting, Electron Commer R A, 51(2022).
[26] C.F. Ho, W.H. Wu, Antecedents of Customer Satisfaction on the Internet: An Empirical Study of Online Shopping, in, (2013).
[27] T. Huang, R. Fildes, D. Soopramanien, Forecasting retailer product sales in the presence of structural change, Eur J Oper Res, 279(2) (2019) 459-470.
[28] J. Huber, H. Stuckenschmidt, Daily retail demand forecasting using machine learning with emphasis on calendric special days, Int J Forecasting, 36(4) (2020) 1420-1438.
[29] M.E. Irarrazaval, S. Maldonado, J. Perez, C. Vairetti, Telecom traffic pumping analytics via explainable data science, Decis Support Syst, 150(2021).
[30] L. Jiang, Z.L. Yang, M. Jun, Measuring consumer perceptions of online shopping convenience, J Serv Manage, 24(2) (2013) 191-214.
[31] X.R. Jiao, Y.P. Song, Y. Kong, X.L. Tang, Volatility forecasting for crude oil based on text information and deep learning PSO-LSTM model, J Forecasting, 41(5) (2022) 933-944.
[32] Y.F. Kang, E. Spiliotis, F. Petropoulos, N. Athiniotis, F. Li, V. Assimakopoulos, D & eacute;j & agrave; vu: A data-centric forecasting approach through time series cross, J Bus Res, 132(2021) 719-731.
[33] A.K. Kar, S.N. Tripathi, N. Malik, S. Gupta, U. Sivarajah, How Does Misinformation and Capricious Opinions Impact the Supply Chain-A Study on the Impacts During the Pandemic, Ann Oper Res, (2022).
[34] J.P. Karmy, S. Maldonado, Hierarchical time series forecasting via Support Vector Regression in the European Travel Retail Industry, Expert Syst Appl, 137(2019) 59-73.
[35] B. Kim, J. Park, J. Suh, Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information, Decis Support Syst, 134(2020).
[36] J.H. Kim, M. Kim, S.J. Lennon, Effects of web site atmospherics on consumer responses: music and product presentation, Direct Marketing An International Journal, 3(1) (2009) p.4-19.
[37] M.S. Ko, K. Lee, K. Hur, Feedforward Error Learning Deep Neural Networks for Multivariate Deterministic Power Forecasting, Ieee T Ind Inform, 18(9) (2022) 6214-6223.
[38] M. Kraus, S. Feuerriegel, Forecasting remaining useful life: Interpretable deep learning approach

via variational Bayesian inferences, Decis Support Syst, 125(2019).
[39] M. Kraus, S. Feuerriegel, A. Oztekin, Deep learning in business analytics and operations research: Models, applications and managerial implications, Eur J Oper Res, 281(3) (2020) 628-641.
[40] R.Y.K. Lau, W.P. Zhang, W. Xu, Parallel Aspect-Oriented Sentiment Analysis for Sales Forecasting with Big Data, Prod Oper Manag, 27(10) (2018) 1775-1794.
[41] K.B. Lee, S. Joo, H. Baik, S. Han, J. In, Unbalanced data, type II error, and nonlinearity in predicting M&A failure, J Bus Res, 109(2020) 271-287.
[42] B.L. Lei, Z.D. Liu, Y.P. Song, On stock volatility forecasting based on text mining and deep learning under high-frequency data, J Forecasting, 40(8) (2021) 1596-1610.
[43] Z.Q. Liao, M.T. Cheung, Internet-based e-shopping and consumer attitudes: an empirical study, Inform Manage, 38(5) (2001) 299-306.
[44] B. Lim, S.O. Arik, N. Loeff, T. Pfister, Temporal Fusion Transformers for interpretable multi-horizon time series forecasting, Int J Forecasting, 37(4) (2021) 1748-1764.
[45] M. Limayem, M. Khalifa, A. Frini, What makes consumers buy from Internet? A longitudinal study of online shopping, Ieee T Syst Man Cy A, 30(4) (2000) 421-432.
[46] X. Liu, P.V. Singh, K. Srinivasan, A Structured Analysis of Unstructured Big Data by Leveraging Cloud Computing, Market Sci, 35(3) (2016) 363-388.
[47] A.L.D. Loureiro, V.L. Migueis, L.F.M. da Silva, Exploring the use of deep neural networks for sales forecasting in fashion retail, Decis Support Syst, 114(2018) 81-93.
[48] S.H. Ma, R. Fildes, Retail sales forecasting with meta-learning, Eur J Oper Res, 288(1) (2021) 111-128.
[49] S. Maldonado, A. Gonzalez, S. Crone, Automatic time series analysis for electric load forecasting via support vector regression, Appl Soft Comput, 83(2019).
[50] C. Moreira, Y.L. Chou, M. Velmurugan, C. Ouyang, R. Sindhgatta, P. Bruza, LINDA-BN: An interpretable probabilistic approach for demystifying black-box predictive models, Decis Support Syst, 150(2021).
[51] A. Mouakher, W. Inoubli, C. Ounoughi, A.D. Ko, Expect: EXplainable Prediction Model for Energy ConsumpTion, Mathematics-Basel, 10(2) (2022).
[52] E. Park, J. Park, M.M. Hu, Tourism demand forecasting with online news data mining, Ann Tourism Res, 90(2021).
[53] D. Pessach, G. Singer, D. Avrahami, H.C. Ben-Gal, E. Shmueli, I. Ben-Gal, Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming, Decis Support Syst, 134(2020).
[54] G.T. Ribeiro, V.C. Mariani, L.D. Coelho, Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting, Eng Appl Artif Intel, 82(2019) 272-281.
[55] Y.R. Sagaert, E.H. Aghezzaf, N. Kourentzes, B. Desmet, Tactical sales forecasting using a very large set of macroeconomic indicators, Eur J Oper Res, 264(2) (2018) 558-569.
[56] M.A. Schmitt, Deep Learning in Business Analytics: A Clash of Expectations
and Reality, International Journal of Information Management Data Insights, 3(1), 100146.(2023).
[57] M.J. Schneider, S. Gupta, Forecasting sales of new and existing products using consumer reviews: A random projections approach, Int J Forecasting, 32(2) (2016) 243-256.
[58] B. Seaman, Considerations of a retail forecasting practitioner (vol 34, pg 822, 2019), Int J Forecasting, 37(3) (2021) 1299-1299.
[59] R. Sharma, A. Kumar, C. Chuah, Turning the blackbox into a glassbox: An explainable machine

learning approach for understanding hospitality customer, International Journal of Information Management Data Insights, 1(2) (2021).
[60] H. Shi, Z.M. Ma, D.Z. Chong, W. He, The impact of Facebook on real estate sales, J Manag Anal, 8(1) (2021) 101-112.
[61] M. Siering, Explainability and fairness of RegTech for regulatory enforcement: Automated monitoring of consumer complaints, Decis Support Syst, 158(2022).
[62] V. Sohrabpour, P. Oghazi, R. Toorajipour, A. Nazarpour, Export sales forecasting using artificial intelligence, Technol Forecast Soc, 163(2021).
[63] Z.L. Sun, T.M. Choi, K.F. Au, Y. Yu, Sales forecasting using extreme learning machine with applications in fashion retailing, Decis Support Syst, 46(1) (2008) 411-419.
[64] D.M. Szymanski, R.T. Hise, e-satisfaction: An initial examination, J Retailing, 76(3) (2000) 309-322.
[65] T. Van Calster, B. Baesens, W. Lemahieu, ProfARIMA: A profit-driven order identification algorithm for ARIMA models in sales forecasting, Appl Soft Comput, 60(2017) 775-785.
[66] G. Verstraete, E.H. Aghezzaf, B. Desmet, A data-driven framework for predicting weather impact on high-volume low-margin retail products, J Retail Consum Serv, 48(2019) 169-177.
[67] G. Verstraete, E.H. Aghezzaf, B. Desmet, A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts, Comput Ind Eng, 139(2020).
[68] C.H. Wang, Considering economic indicators and dynamic channel interactions to conduct sales forecasting for retail sectors, Comput Ind Eng, 165(2022).
[69] Q.Z. Wang, X.L. Cui, L.Q. Huang, Y.L. Dai, Seller reputation or product presentation? An empirical investigation from cue utilization perspective, Int J Inform Manage, 36(3) (2016) 271-283.
[70] T.Y. Weng, W.Y. Liu, J. Xiao, Supply chain sales forecasting based on lightGBM and LSTM combination model, Ind Manage Data Syst, 120(2) (2020) 265-279.
[71] X. Xu, Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants, J Bus Res, 106(2020) 196-210.
[72] S.G. Yang, A novel study on deep learning framework to predict and analyze the financial time series information, Future Gener Comp Sy, 125(2021) 812-819.
[73] B. Zhang, M.L. Tseng, L.L. Qi, Y.H. Guo, C.H. Wang, A comparative online sales forecasting analysis: Data mining techniques, Comput Ind Eng, 176(2023).
[74] C.A. Zhang, Y.X. Tian, Z.P. Fan, Forecasting sales using online review and search engine data: A method based on PCA-DSFOA-BPNN, Int J Forecasting, 38(3) (2022) 1005-1024.
[75] D. Zhdanov, S. Bhattacharjee, M.A. Bragin, Incorporating FAT and privacy aware AI modeling approaches into business decision making frameworks, Decis Support Syst, 155(2022).