Soft Computing (2024) 28:4967–4986
https://doi.org/10.1007/s00500-023-09073-0

APPLICATION OF SOFT COMPUTING

Solving data-driven newsvendor problem with textual reviews through deep learning

Chuan Zhang[1] · Yu-Xin Tian[1]

Accepted: 29 July 2023 / Published online: 28 September 2023

Abstract
The production decision of a large commodity or equipment manufacturing enterprise can be modeled as a newsvendor problem. Managers must determine the optimal production volume in advance to minimize the underage cost and the overage cost. However, the traditional newsvendor problem assumes the known demand distribution, which is not the case in practice. Data-driven approaches have become the hot research topic and opened up new avenues for such issues. Recent studies have considered demand-related features but have failed to address how to optimize production and inventory using informative textual reviews, not just numerical feature data. To address this issue, we propose a data-driven newsvendor model that leverages sentiment analysis on textual reviews using a deep learning model to solve the data-driven newsvendor problem by integrating estimation and optimization. Experiments on real data show that our proposed method reduces the average cost by approximately 14.18% compared to the most advanced deep neural network method, making it the best-performing method. Furthermore, our method is more suitable for situations where unit shortage costs are greater than unit overage costs. Finally, our method is robust in terms of sample size and can still obtain good results even with insufficient historical data.

Keywords Data-driven · Newsvendor problem · Production · Online reviews · Sentiment analysis · Deep learning

1 Introduction

In practice, uncertain demand creates a dilemma for the production decision of large commodity or equipment manufacturing enterprises such as automobile manufacturing enterprises. Insufficient production of commodities would result in the *underage cost* which could potentially harm the future goodwill of the business, while over-production would lead to excess inventory, resulting in the *overage cost*, including costs of holding inventories, commodity maintenance, and repair. Additionally, the unit underage cost and the unit overage cost are asymmetric. This decision-making problem can be modeled as a newsvendor problem (Arrow et al. 1951) and is widely discussed in the field of operations research and management. However, the traditional newsvendor problem assumes that the random demand is certainly known, which is not realistic. The classical newsvendor model may not provide practical help to enterprises.

Advancements in human science and technology, along with the popularization of big data applications and breakthroughs in artificial intelligence technology, have led to data-driven approaches that offer new perspectives, theoretical paradigms, and technical means for studying production, inventory, and operations management. In the case of the typical newsvendor problem, the real-world demand is influenced by several observable indicator variables prior to the ordering date, which are known as demand-related factors. Through data analysis, the relationship between observable demand-related factors and demand can be uncovered, providing insights for production and inventory decisions. The selection of data indicators to support production and inventory management depends on the specific characteristics of the product. For example, climate temperature is an important factor to consider for perishable products and agricultural products (Keskin et al. 2022; Oroojlooyjadid et al. 2019). Calendar-related features, such

✉ Chuan Zhang
czhang@mail.neu.edu.cn

[1] School of Business Administration, Northeastern University, Shenyang 110169, China

as month of the year, day of the week, and season, are significant factors to consider for products that experience periodic demand changes, such as hospital nurse allocation (Ban and Rudin 2019; Liu et al. 2022). For high-value commodities or equipment, macroeconomic conditions can greatly affect the demand (Zhang et al. 2020a, b, c; Zhang et al. 2020a, b, c). For example, in the case of the automobile industry studied in this paper, product demand is closely related to macroeconomic indicators (such as CPI and new bank loans) and the price of crude oil. Therefore, it is recommended to utilize the additional demand-related feature data in operations management for enhanced decision-making.

In recent operations research literature, the data-driven newsvendor problem has received significant attention, and several representative methods have been proposed. The first method forecasts the demand by utilizing various additional data and uses the resulting forecast to determine the optimal order or production quantity. Although this method has achieved high predictive accuracy, it does not consider the trade-off between the underage cost and the overage cost, providing suboptimal results. Subsequently, a series of methods that integrate estimation and optimization for the data-driven newsvendor problem were proposed (Ban and Rudin 2019; Huber et al. 2019; Lin et al. 2021; Liu et al. 2022; Oroojlooyjadid et al. 2019), such as kernel optimization (KO), linear machine learning (LML), and deep neural network (DNN) models that achieve lower costs and bring a new paradigm to the study of the data-driven newsvendor problem and production management. However, the feature data utilized in existing studies are currently limited to structured and numerical formats. In fact, commercial big data can be presented in many forms, such as numbers, texts, images, videos, and so on. To realize data-driven operation further, more forms of data are worth mining. Among them, the well-known textual online review data are very important business data.

Online reviews have emerged as a prominent form of online word-of-mouth communication that significantly influences consumer decision-making and product demand in the digital age. In fact, a McKinsey study found that 67% of daily consumption decisions are driven by the word-of-mouth effects of user-generated online reviews (Vana and Lambrecht 2018). Online reviews provide customers with voluntarily published information that conveys rich emotional expressions, including joy, anger, sadness, criticism, and praise. Accordingly, such reviews serve as valuable customer-oriented information resources (Chen and Xie 2008), providing insights into the real-life experiences of product usage in various settings. Empirical studies have demonstrated that the emotional tendencies conveyed in textual online reviews have a significant impact on product demand (Chong et al. 2016; Duan et al. 2008; Ye et al. 2009), highlighting the importance of considering these factors in logistics, production, and inventory management. This paper aims to leverage textual online review data to enhance the data-driven newsvendor problem in production management by integrating sentiment features from online reviews. Specifically, the study explores two key questions: (1) Is textual online review data a useful resource for informing production decision-making? (2) How can sentiment features be extracted from textual online reviews and integrated into the newsvendor model? By answering these questions, this study offers insights into how online reviews can inform production decisions and enhance the accuracy of the newsvendor model.

To address the aforementioned two problems, this study proposes a data-driven newsvendor model that incorporates sentiment mining of textual reviews to optimize the order or product quantity for enterprises. The proposed approach comprises three main steps. Firstly, we introduce a method based on distance correlation analysis to select macroeconomic indicators that are relevant to the target product demand. By combining search indices of the product brand name and calendar-related data, we construct numerical features. Secondly, we propose a sentiment analysis technology to extract sentiment features from online textual reviews. This sentiment analysis algorithm calculates sentiment scores based on semantics. Additionally, we utilize browsing times as an auxiliary feature, and weight and aggregate sentiment scores across different time periods for each product attribute, thus obtaining sentiment values for different attributes within various time frames. Thirdly, we integrate the calculated sentiment values, along with the selected demand-related numerical features from the first step, into the newsvendor model as inputs. Subsequently, a deep neural network model is employed to map these features to the production volume that minimizes the total cost, encompassing the underage cost and the overage cost. This integrated approach facilitates the determination of the optimal order or product volume.

We conduct experiments employing real-world data and derive intriguing findings. Our proposed method demonstrates potential cost savings of approximately 55% compared to the traditional EDD method, and when compared to the most advanced DNN method, it has the potential to help enterprises achieve cost savings of approximately 14.18%. Notably, our method is particularly suitable for scenarios with limited historical data, as it ensures more balanced unit underage and unit overage costs, leading to more appropriate production and inventory decisions in the presence of unstable demand and an unknown demand distribution.

This study contributes to the recent stream of research on the data-driven newsvendor problem in two significant ways. Firstly, we extend the theory of the data-driven newsvendor problem by incorporating sentiment values derived from textual reviews into the cost function of the newsvendor model. In addition to these sentiment values, we employ a distance correlation-based feature selection method to select relevant numerical features that are included as inputs to a deep learning model. The objective is to minimize the cost function and determine the optimal production quantity. Secondly, we propose a dictionary-based sentiment analysis method specifically designed for the processing of textual data. Unlike machine learning-based sentiment analysis methods, such as recursive neural network (RNN), our approach aligns with the natural language usage of individuals. The dictionary-based sentiment analysis method offers easy interpretability, explicability, and maintainability, providing a clear and straightforward process. Overall, our research expands the understanding of the data-driven newsvendor problem and presents a new approach for future studies.

The remainder of this paper is structured as follows. Section 2 provides a concise overview of the existing literature on data-driven newsvendor problem and sentiment analysis, which highlights the difference between the current work and the literature. Section 3 describes the problem addressed in this paper. In Sect. 4, we describe the proposed method in detail. In Sect. 5, we present experiments conducted with real-world data, comparing our proposed methods with typical benchmarks and deriving managerial insights from the analysis of the experimental results. Finally, Sect. 6 outlines conclusions and explores potential avenues for future research.

2 Literature review

2.1 Data-driven newsvendor problem

In recent years, data-driven approaches have gained significant attention in the field of inventory management, particularly in relation to the newsvendor problem. The newsvendor problem serves as a fundamental starting point for data-driven research in inventory decision-making (Qi et al. 2020). The traditional newsvendor problem assumes complete knowledge of the demand distribution. However, current research has explored data-driven methods that relax this assumption. One of the early techniques involves using empirical demand distributions obtained from historical observations, named the EDD method (Besbes and Muharremoglu 2013; Levi et al. 2007; Liyanage and Shanthikumar 2005). While the EDD eliminates the need for an exact specification of the demand distribution, its reliability is questionable since demand distribution varies in practice.

The proliferation of big data has opened avenues for research aimed at addressing the newsvendor problem through the utilization of diverse demand-related feature data, as exemplified in the works of Ban and Rudin (2019) and He et al. (2012). Features encompass exogenous variables observed during the decision-making process, including factors such as raw material prices, seasonality, and economic indicators, which aid decision-makers in mitigating uncertainty and improving decision quality. Presently, three main approaches have been put forth to tackle the multi-feature data-driven newsvendor problem.

The first method is the estimate-as-solution (EAS). This method is known as the forecast as order or product quantity method, which involves forecasting demand as the optimal order or product quantity. This is done by fitting a functional relationship between demand and the related feature data via supervised learning, using observation of historical demand data. This approach is commonly found in existing literature on demand forecasting (Arunraj and Ahrens 2015; Bi et al. 2022a, b; Fan et al. 2017; Zhang et al. 2020a, b, c; Zhang et al. 2021; Zhang et al. 2020a, b, c; Zhu et al. 2021). However, the EAS approach has significant limitations, as prediction alone does not equate to decision-making. Notably, the newsvendor problem involves varying overage and underage costs, and merely predicting demand fails to account for the structure of the loss function (Qi et al. 2020).

The second method involves parameter estimation first and then optimization. This approach is known as separated estimation and optimization (SEO) by Ban and Rudin (2019) or the disjoint approach, as described by Liu et al. (2022). This approach assumes that the random demand variables obey a certain distribution with unknown parameters that are influenced by environmental variables. By utilizing machine learning methods to estimate these parameters, the distribution of demand can be obtained. Subsequently, a stochastic optimization problem is solved based on these estimated parameters. However, the objectives of the two steps are inconsistent, and a relatively large estimation error in the first stage can seriously impact the second stage's outcomes, resulting in suboptimal decisions (Liu et al. 2022). Furthermore, the method requires specifying the form of the demand distribution function in the first stage. Accurately determining the demand distribution form and its unknown parameters can be challenging, leading to inconsistencies with the actual demand distribution. To address these limitations, researchers have proposed alternative methods. For example, Elmachtoub and Grigas (2022) proposed the Smart "Predict, then Optimize" (SPO) framework, which leverages a convex substitution loss function derived from duality theory to model training.

The third one is integrated estimation and optimization method. This method uses machine learning to directly determine the optimal order, or product quantity can be approached in various ways. Typical approaches are the KO methods and the empirical risk minimization (ERM) methods. The KO methods employ kernel regression to model the conditional demand distribution and employ sorting algorithms to determine the optimal order quantity. The KO models assign weights to historical data points based on similarity between current and historical features, using machine learning methods. Ban and Rudin (2019) derived performance bounds based on kernel-based weighting and validated them using a dataset on nurse staffing.

The ERM methods mainly include the LML method and deep learning methods. Ban and Rudin (2019) introduced a LML method, which formulates the data-driven newsvendor problem as a linear programming problem, directly minimizing the cost function rather than regression error. By mapping features to the order quantity through linear function parameters, LML is essentially equivalent to high-dimensional quantile regression, solvable through convex optimization. However, the limitation of the LML method is that it only describes linear relation, not nonlinear ones. In reality, the nonlinear costs in the newsvendor problem often occur (Kyparisis and Koulamas 2018). A small shortage in the company may not cause significant costs, while a significant shortage could harm the company's reputation. In recognition of this fact, studies by Oroojlooyjadid et al. (2019) have constructed neural networks to fit the nonlinear relationship and used deep learning frameworks to implement the model and optimize the solution. Cao and Shen (2019) proposed a quantile forecasting technique based on deep neural networks for newsvendor problems with non-stationary demand. Their model minimized the quantile loss function using gradient descent. Liu et al. (2022) considered the nonlinear cost of the newsvendor problem, accounting for different costs associated with shortages. Pirayesh Neghab et al. (2022) considered not only observable feature data but also unobservable features and integrated the three steps of parameter estimation, inference, and optimization into a multi-layer neural network with a hidden Markov model.

Furthermore, researchers have explored data-driven newsvendor problems with constraints and multi-period inventory problems. For instance, Lin et al. (2021) studied a risk-averse newsvendor problem and proposed a data-driven approximate solution using machine learning methods and historical demand information. Their solution included a risk constraint, was computationally efficient, and asymptotically optimal. Qi et al. (2023) employed multi-layer perceptron (MLP) and deep neural network (DNN) models for end-to-end multi-period inventory management. Their experiments indicated that DNN generally exhibited superior prediction ability and robustness compared to MLP, across different datasets and scenarios.

Table 1 summarizes recent research on using machine learning methods to solve the newsvendor problem, comparing our approach to existing methods. In contrast with the previous research, we incorporate both numerical feature data and textual reviews data. The current literature primarily focuses on numerical data, but we contend that the non-numerical data, such as online reviews, can provide more information (Archak et al. 2011). Online reviews have been commonly utilized for product selection and ranking (J. W. Bi et al. 2022a, b; Yang Liu et al. 2017b; Y. Liu et al. 2017a, b, c; Zhang et al. 2020a, b, c), product or service enhancement (Liu et al. 2018), and demand forecasting (Fan et al. 2017; Zhang et al. 2021; Zhang et al. 2020a, b, c). However, their application in the field of order or product decision-making is still underexplored.

2.2 Sentiment analysis

Sentiment analysis is a specialized area within text mining that aims to extract positive and negative sentiments from text based on their expressions (Akhtar et al. 2020; Kang et al. 2020; Peng et al. 2022). It can be conducted at the document level or the sentence/phrase level. Document-level sentiment analysis involves building classifiers to distinguish positive from negative reviews (Alam et al. 2016), while sentence-/phrase-level analysis focuses on tasks such as multi-perspective questions, summaries, opinion extraction, and customer review mining (Kotelnikova et al. 2022; Najmi et al. 2015). In this study, we focus on sentence-/phrase-level sentiment analysis since online reviews tend to be written in short sentences or phrases.

Existing research has developed various sentiment analysis technologies, mainly falling into two categories: (1) lexicon-based methods (Mohamad Beigi and Moattar 2021; Weichselbraun et al. 2011; Zhang et al. 2020a, b, c; Zhang et al. 2021; Zhang et al. 2020a, b, c), which leverage sentiment lexicons, grammatical analysis, and syntactic patterns and (2) machine learning-based models, including support vector machines (SVM) (Khorsand et al. 2020; Martinez-Torres and Toral 2019), Naïve Bayes (Fan et al. 2017; Zhang et al. 2022), convolutional neural network (CNN) (Chang et al. 2020), and long short-term memory (LSTM) (Cao et al. 2020; Chang et al. 2020). Machine learning-based sentiment analysis techniques can be further classified into supervised, unsupervised, or semi-supervised approaches. Supervised machine learning techniques involve the association of texts with emotional polarity through the utilization of intelligent algorithms that are trained on labeled data (Bai 2011; Yang Liu et al. 2017a; Moreo et al. 2012; Wu et al. 2016). Unsupervised machine

Table 1 Major studies on data-driven newsvendor problem using machine learning

Method	Examples	Driven by feature data	Newsvendor cost function integration	Nonlinear model	Driven by textual data
Linear machine learning	Ban and Rudin (2019)	√	√		
Kernel optimization models	Ban and Rudin (2019)	√	√		
Deep learning models	Oroojlooyjadid et al. (2019), Cao and Shen (2019), Liu et al. (2022), Pirayesh Neghab et al. (2022)	√	√	√	
Our study		√	√	√	√

learning techniques involve learning algorithms that develop without labeled training samples, measuring text similarities based on keyword lists and clustering texts accordingly (Fernández-Gavilanes et al. 2016). Semi-supervised machine learning utilizes partially labeled data in the training process (Moreno-Ortiz and Fernández-Cruz 2015). In contrast, lexicon-based sentiment analysis techniques can be divided into dictionary-based and corpus-based approaches. Dictionary-based methods involve constructing sentiment dictionaries, starting with a small set of manually determined sentiment words and expanding the list by searching for synonyms and antonyms in corpora using resources such as *WordNet* or *HowNet* (Akter and Aziz 2016; Dey et al. 2018). Corpus-based techniques address context-specific orientation of opinion words by identifying syntactic patterns and using seed lists of opinion words.

The previous research indicates that machine learning-based techniques are more effective in addressing document-level sentiment analysis, while lexicon-based techniques are better suited for sentence-level analysis (Liu et al. 2017b). Therefore, in this study, we adopt a lexicon-based approach to calculate the sentiment score for each online review. To accomplish this, we introduce sentiment degree lexicons and propose an algorithm that incorporates sentiment degree to derive sentiment scores. This algorithm is particularly well-suited for our investigation into Chinese online reviews.

2.3 Research gap

In contrast with the existing literature, our work differs in the following aspects. Firstly, the existing literature lacks exploration of multi-source data. In the previous data-driven inventory management studies, researchers primarily focused on using single-source data, such as historical demand data (Liyanage and Shanthikumar 2005) or pre-specified numerical feature data (Ban and Rudin 2019; Pirayesh Neghab et al. 2022), with limited attention given to the selection of demand-relevant features. However, the choice of data is critical for effective inventory decision-making. In our study, we place significant emphasis on the selection of demand-related features and propose a distance correlation-based feature selection method. Most importantly, we explore the value of incorporating textual online reviews into the decision-making process. Secondly, there is a scarcity of research on sentiment analysis of Chinese reviews, particularly with regard to fine-grained text analysis. In our study, we provide an alternative solution by developing a sentiment analysis algorithm specifically designed for the characteristics of Chinese expressions. Our method utilizes sentiment word lexicons and degree adverb lexicons, offering a transparent, interpretable, and easily maintainable approach compared to methods like recursive neural networks (RNNs). Thirdly, the previous studies have overlooked the investigation of the impact of sample size on order/production decision outcomes (Liu et al. 2022; Pirayesh Neghab et al. 2022). In contrast, our research explicitly examines the performance of different models under varying sample sizes, highlighting the robustness of our proposed method in the face of limited data availability.

Overall, our work significantly diverges from the existing literature by exploring multi-source data, addressing sentiment analysis of Chinese reviews, and investigating the influence of sample size on decision-making outcomes. These distinctions underscore the novelty and unique contributions of our research.

3 Problem description

The cost function of the original newsvendor problem can be represented as follows:

$$C(q,d) = c_h(q-d)^{+} + c_b(d-q)^{+}, \tag{1}$$

where q represents the order quantity, d represents the demand volume, c_h represents the unit overage cost, and c_b represents the unit underage cost, and $(x)^{+} = \max(0,x)$.

In the problem of data-driven newsvendor, the objective function is constructed based on the principle of empirical risk minimization (ERM), which is a rule commonly utilized in machine learning. We denote historical demand data as $\mathbf{d} = [d_1, d_2, \ldots, d_n]$, and the demand-related feature data as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n]$. The data-driven newsvendor problem is as follows:

$$\min_{\boldsymbol{\beta}} \frac{1}{n-1} \sum_{i=2}^{n} C(\theta(\mathbf{x}_{i-1}; \boldsymbol{\beta}), d_i) + \lambda R(\boldsymbol{\beta}), \tag{2}$$

where $\theta(\mathbf{x}_{i-1}; \boldsymbol{\beta})$ is the optimal order or product quantity obtained by observing the environmental variables $\mathbf{x}_{i-1}$, and $\boldsymbol{\beta}$ represents the fitted parameters of the neural network θ. $\lambda R(\boldsymbol{\beta})$ represents the regularization term aimed at preventing over-fitting, λ is the regularization coefficient, and $R(\boldsymbol{\beta})$ represents the norm of the vector $\boldsymbol{\beta}$.

4 Methodology

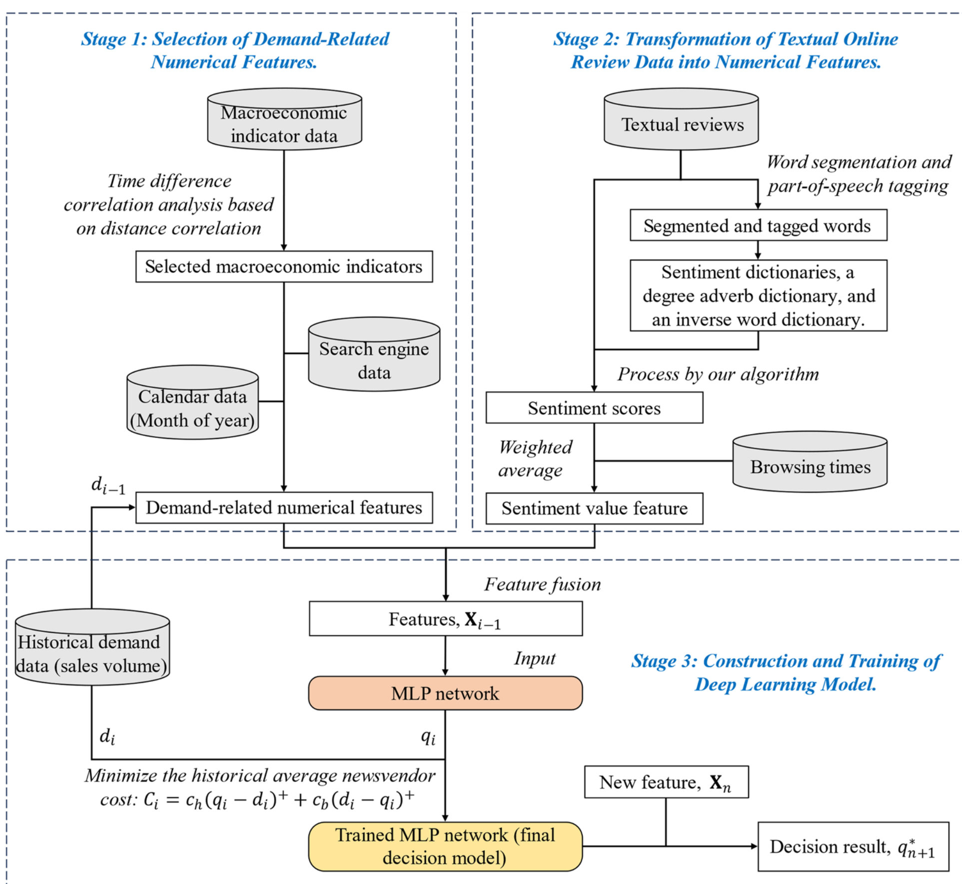
4.1 Framework and steps

The method proposed in this paper, named the Sentiment Analysis-Deep Neural Network (SA-DNN), is depicted in Fig. 1 and consists of three main stages.

Stage 1: Selection of Demand-Related Numerical Features. Firstly, we employ a time difference correlation analysis method based on distance correlation to select macroeconomic indicators (Zhang et al. 2022) that are relevant to the target product demand. There are numerous macroeconomic indicators, and using all of them would result in two consequences: First, considering all indicators would lead to low training efficiency due to the large computational load, and second, it could cause over-fitting and inaccurate results due to the introduction of information with relatively weak correlations. Additionally, we include search engine data (i.e., search index of the product brand name) and calendar data (i.e., month of the year) as features to provide richer information and clues about periodical changes in demand.

Stage 2: Transformation of Textual Online Review Data into Numerical Features. Firstly, we collect the online reviews of the product for each period along with their corresponding browsing times information. Then, we preprocess the text review data, including word segmentation and part-of-speech tagging, and establish sentiment dictionary of each attribute, a degree adverb dictionary, and an inverse word dictionary. Next, we apply our proposed sentence-grained sentiment analysis algorithm to calculate the sentiment scores for each review. Finally, we aggregate the sentiment scores of all reviews with the same attribute

Fig. 1 Framework of the proposed method



within the same time period to generate a numerical sentiment value, which is utilized as a feature.

Stage 3: Construction and Training of Deep Learning Model. In this stage, we construct a MLP network, with the features generated in Stage 1 and Stage 2 as inputs and order or product quantity as the output. The MLP network is trained by minimizing the historical average newsvendor cost, and then, a final decision model is developed.

In the following sections, we provide a detailed description of the specific implementation process for each stage.

4.2 Selection of demand-related numerical features

We employ the time difference correlation analysis based on distance correlation to select the macroeconomic indicator data. The core idea of this method is to eliminate the indices whose distance correlation with demand is below a certain threshold within the specified lag period. To measure both linear and nonlinear correlations, we utilize the distance correlation metric (Székely et al. 2007). Let $\mathbf{x}$ and $\mathbf{y}$ represent paired continuous variables of length N, and the calculation procedures of distance correlation are as follows:

(1) Calculate the distance matrix between the respective elements of $\mathbf{x}$ and $\mathbf{y}$:

$$\alpha_{i,j} = \left\| x_i - x_j \right\|, \quad i,j = 1,2,\ldots,N, \tag{3}$$

$$\beta_{i,j} = \left\| y_i - y_j \right\|, \quad i,j = 1,2,\ldots,N, \tag{4}$$

(2) Perform centralization on all pairwise distances:

$$A_{i,j} = \alpha_{i,j} - \overline{\alpha_{i\cdot}} - \overline{\alpha_{\cdot j}} + \overline{\alpha_{\cdot\cdot}}, \tag{5}$$

$$B_{i,j} = \beta_{i,j} - \overline{\beta_{i\cdot}} - \overline{\beta_{\cdot j}} + \overline{\beta_{\cdot\cdot}}, \tag{6}$$

where $\overline{\alpha_{i\cdot}}$ $(\overline{\beta_{i\cdot}})$ is the average of row i, $\overline{\alpha_{\cdot j}}$ $(\overline{\beta_{\cdot j}})$ is the average of column j, and $\overline{\alpha_{\cdot\cdot}}$ $(\overline{\beta_{\cdot\cdot}})$is the average of all elements.

(3) Calculate the arithmetic mean of the squared covariance of the sample distances:

$$\mathrm{dCov}^2(\mathbf{x},\mathbf{y}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{i,j}B_{i,j}, \tag{7}$$

$$dVar^2(\mathbf{x}) = \frac{1}{N^2}\sum_{i,j=1}^{N} A_{i,j}^2, \tag{8}$$

$$dVar^2(\mathbf{y}) = \frac{1}{N^2}\sum_{i,j=1}^{N} B_{i,j}^2, \tag{9}$$

(4) Calculate the distance correlation:

$$\mathrm{dCor}(\mathbf{x},\mathbf{y}) = \frac{\mathrm{dCov}(\mathbf{x},\mathbf{y})}{\sqrt{dVar(\mathbf{x}) \cdot dVar(\mathbf{y})}}. \tag{10}$$

In this paper, we denote M as the total number of alternative features, $[\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_M]$ as the vector of alternative features, and $\mathbf{f}_m^L$ as the m - th alternative feature with L $(L > 0)$ period lagging behind the dependent variable $\mathbf{y}$. Then, the maximum absolute distance correlation between $\mathbf{f}_m^L$ and $\mathbf{y}$ at different lag orders is

$$\delta\left(\mathbf{f}_m^L\right) = \max_{0 < L \le L_{\max} \cap L \in \mathbb{N}} \left| \mathrm{dCor}\left(\mathbf{f}_m^L, \mathbf{y}\right) \right|, \tag{11}$$

where $L_{\max}$ is the given maximum lag order (default value is 12).

Finally, the selection requirement for the m - th alternative feature is $\delta\left(\mathbf{f}_m^L\right) \ge \delta_0$, where δ_0 is the threshold, which means the absolute value of the minimum distance correlation coefficient that should be met between the adopted features with a certain lag order and the demand.

In general, for the correlation coefficient, if the absolute value falls within the range [0, 0.2), there is no correlation; if the absolute value falls within the range [0.2, 0.4), there is a weak correlation; if the absolute value falls within the range [0.4, 0.6), there is a moderate correlation; and if the absolute value falls within the range [0.6, 1.0), there is a strong correlation (Zhang et al. 2020a, b, c). To reduce the introduction of noise, the absolute distance correlation threshold δ_0 should be set larger than 0.4 (Zhang et al. 2020a, b, c; Zhang et al. 2022). Let the macroeconomic indicator data after selection be denoted as $\left[\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_p, \ldots, \tilde{\mathbf{f}}_{Q-2}\right]$, and combined with the brand search index $\tilde{\mathbf{f}}_{Q-1}$ and the calendar feature $\tilde{\mathbf{f}}_Q$ (i.e., month of the year), the final representation of numerical features is as follows:

$$\left[\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_p, \ldots, \tilde{\mathbf{f}}_{Q-2}, \tilde{\mathbf{f}}_{Q-1}, \tilde{\mathbf{f}}_Q\right] = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^Q \\ x_2^1 & x_2^2 & \cdots & x_2^Q \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^Q \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}. \tag{12}$$

For each period, there are Q observable numerical type feature data.

4.3 Transformation of textual online review data into numerical features

In this paper, we propose a sentiment analysis method for textual reviews based on dictionary and grammatical analysis, which is similar to the one presented in Liu et al. (2017b) and Zhang et al. (2021). This method aims to

	Date	Browsing Times	Space	Power	Controllability	…	Cost performance
			m attributes				
s_1	2015-01-01	2080	非常满意… (Very satisfied ...)	动力充沛… (Full of power ...)	方向盘省力… (Steering wheel saves effort ...)	…	没有啥性价比… (There is no cost performance ...)
	2015-01-01	865	没有压力… (No pressure ...)	勉强够用… (Barely enough…)	转向精准… (Accurate steering ...)	…	一般… (Just so so …)
	…	…	…	…	…	…	…
s_i	2015-01-06	437	只能说够用… (I can only say that it is enough ...)	够用了… (Enough…)	刹车灵敏… (Sensitive braking …)	…	还是可以的… (It's ok ...)
	2015-01-06	1010	值得赞扬… (Commendable ...)	还可以… (Not bad ...)	很好… (Very good…)	…	性价比很高… (Very cost -effective …)
	2015-01-06	1060	前排空间很足… (There's plenty of room in the front row …)	油门不够猛… (The throttle is not strong enough ...)	感觉还好… (I feel fine ...)	…	性价比很低… (The cost performance is very low …)
	…	…	…	…	…	…	…
	2021-12-31	67	还可以… (Not bad ...)	挺给力的… (It's awesome ...)	挺灵活的… (Very flexible ...)	…	很好的… (Very good ...)

Fig. 2 The structure of the online review data

identify the sentiment polarity and calculate the sentiment degree.

The structure of the online review data used in this study is illustrated in Fig. 2. In this scenario, the product has m attributes, and there are s_i groups of online reviews related to different attributes in the period i. The online reviews for each attribute in period i can be represented as follows:

$$\mathbf{r}_i = \left\{ \left\{ r_i^{1,1}, \ldots, r_i^{1,k}, \ldots, r_i^{1,s_i} \right\}, \ldots, \left\{ r_i^{j,1}, \ldots, r_i^{j,k}, \ldots, r_i^{j,s_i} \right\}, \ldots, \left\{ r_i^{m,1}, \ldots, r_i^{m,k}, \ldots, r_i^{m,s_i} \right\} \right\}, \tag{13}$$

where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, and $k = 1, 2, \ldots, s_i$. The element $r_i^{j,k}$ in $\mathbf{r}_i$ denotes the k - th textual online review concerning the j - th attribute of the product in the period i. Then, the steps to obtain the sentiment values for each attribute in period i are as follows:

Step 1: Word segmentation and part-of-speech tagging.

In this study, we focus on sentiment analysis of online reviews in the Chinese context. Unlike English, Chinese reviews do not have spaces to separate words. Therefore, the first step is to segment the reviews into individual words and identify their respective part-of-speech tags. We utilize the *Jieba* word segmentation tool for this purpose, enabling us to construct dictionaries and calculate sentiment scores in the subsequent steps.

Step 2: Construction of attribute-specific sentiment word dictionaries.

The online reviews collected for a specific attribute are merged and segmented into individual words. All the sentiment words are extracted and sorted in descending order based on their frequency of occurrence. Through manual identification, these words are added to the dictionaries D_j^+ or D_j^-, depending on their sentiment orientation. Here, D_j^+ (D_j^-) denotes the set (or dictionary) of positive (negative) sentiment words corresponding to the j - th attribute of the product. The reason for constructing attribute-specific sentiment word dictionaries is that the same sentiment word may have opposite sentiment orientations when describing different product attributes. For instance, "高 (high)" is considered positive for the cost performance attribute but negative for the fuel consumption attribute. Figure 3 illustrates the program interface developed for the manual construction of attribute-specific sentiment word dictionaries, while Table 2 presents the construction results and the statistics of the vocabulary size.

Step 3: Construction of degree adverb dictionaries and the inverse word dictionary.

Based on *HowNet* (Ku and Chen 2007) and incorporating human intuition, we assign values to degree adverbs and divide degree adverb dictionaries into six intensity levels. The different degree adverbs and their corresponding degree values are presented in Table 3. Additionally, to identify inverse words within sentences, we construct an inverse word dictionary using *HowNet*. This dictionary includes words such as "不是 (not)" and "非 (no)."

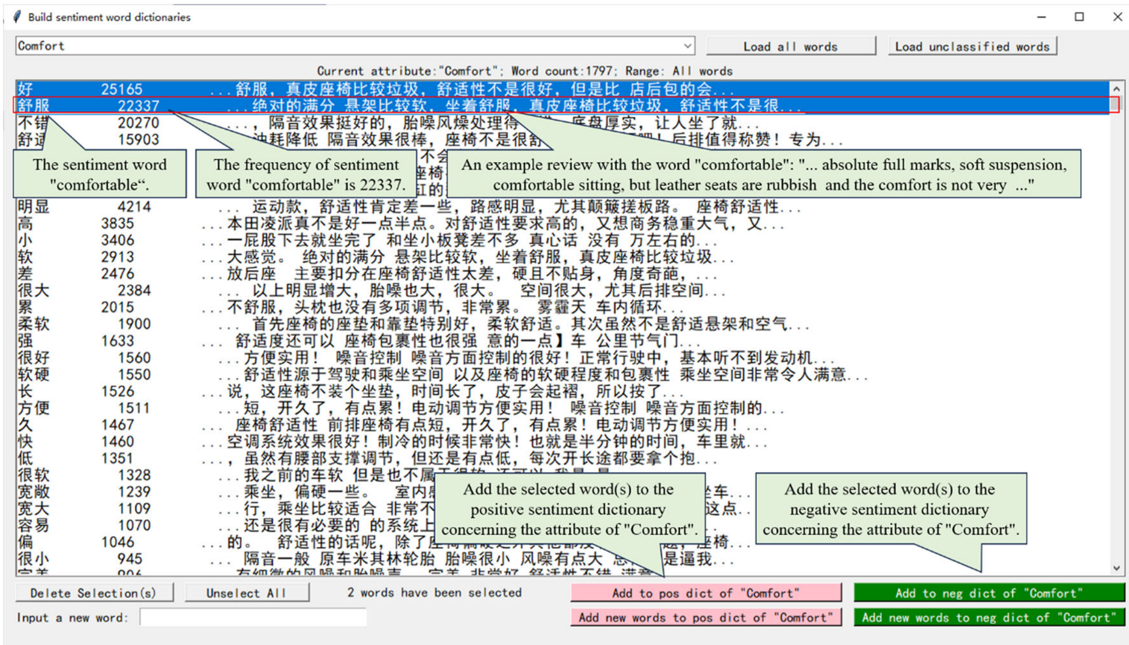


Fig. 3 Software for manually building sentiment word dictionaries concerning attributes

Table 2 Results of building sentiment word dictionaries concerning attributes

Attribute	Examples	Count
Space	宽敞(capacious), 脏(dirty), 小(narrow)	$D_1^+=276$; $D_1^-=143$
Power	强(strong), 低(weak), 稳(steady), 充沛(energic)	$D_2^+=341$; $D_2^-=174$
Controllability	容易(easy), 灵活(flexible), 复杂(complex)	$D_3^+=334$; $D_3^-=149$
Fuel consumption	低(low), 平稳(stable), 高(high), 一般(common)	$D_4^+=114$; $D_4^-=75$
Comfort	舒服(comfortable), 好(good), 差(bad)	$D_5^+=225$; $D_5^-=138$
Appearance	豪华(luxury), 漂亮(beautiful),优雅(elegant)	$D_6^+=370$; $D_6^-=193$
Car interior	细腻(exquisite), 精致(delicate), 完美(perfect)	$D_7^+=355$; $D_7^-=189$
Cost performance	不错(nice), 昂贵(expensive), 高(high), 低(low)	$D_8^+=160$; $D_8^-=108$

Step 4: Calculation of sentiment scores for each online review.

We develop a dictionary-based sentiment score calculation algorithm to convert textual reviews into corresponding numerical sentiment scores. The pseudo-code of this algorithm is presented in Algorithm 1. The algorithm generates a sentiment score, where a positive value indicates positive emotion, a negative value indicates negative emotion, and an absolute value represents a sentiment intensity. For ease of reference in subsequent discussions, we denote the sentiment analysis algorithm as the function $\varphi(\cdot)$, and the corresponding sentiment score $v_i^{j,k}$ of the review sentence $r_i^{j,k}$ is calculated as $v_i^{j,k}=\varphi\left(r_i^{j,k}\right)$.

Table 3 Degree adverb dictionaries and corresponding degree scores based on *HowNet*

Dictionaries	Categories	Examples	Degree scores
S_1	Extreme/Most	无比 (incomparably), 至极 (extremely), 最为 (most)	$\rho_1=2.0$
S_2	Over	过于 (over), 何止 (beyond), 超级 (super)	$\rho_2=1.8$
S_3	Very	非常 (very), 格外 (especially), 很是 (much)	$\rho_3=1.5$
S_4	More	那么 (so), 越加 (more)	$\rho_4=1.2$
S_5	Slightly/A bit	稍微 (slightly), 有点 (a bit), 较为 (relatively)	$\rho_5=0.5$
S_6	Insufficiently	轻度 (mild), 微 (insufficiently)	$\rho_6=0.2$

Algorithm 1. Calculating the sentiment score $\varphi(\cdot)$.

Input: The online review sentence $r_i^{j,k}$.

Output: The sentiment score $v_i^{j,k}$.

function $\varphi\left(r_i^{j,k}\right)$

$\Gamma_i^{j,k}=\left\{\varpi_{i,1}^{j,k},\ldots,\varpi_{i,g}^{j,k},\ldots\varpi_{i,G}^{j,k}\right\}$ $\leftarrow$ Segment sentence $r_i^{j,k}$ into G separate words by the segmentation tool *jieba*.

$v\leftarrow 0$

for $g=1\rightarrow G$ **do**

Calculate the sentiment score of each word by $O_{i,g}^{j,k}=\begin{cases}1 & ,\varpi_{i,g}^{j,k}\cap D_j^{+}\neq\varnothing\\ -1 & ,\varpi_{i,g}^{j,k}\cap D_j^{-}\neq\varnothing\\ 0 & ,\text{else}\end{cases}$.

if here is neither degree adverb nor inverse word in front of $\varpi_{i,g}^{j,k}$ **then**

$\eta_g\leftarrow 1$

else if there is only degree adverb $\overline{s}$ and no inverse word before $\varpi_{i,g}^{j,k}$ **then**

$\eta_g\leftarrow\rho_f, f=\left\{F\mid\overline{s}\in\mathcal{S}_F\right\}$

else if there are only inverse words and no degree adverb before $\varpi_{i,g}^{j,k}$ **then**

$\eta_g\leftarrow(-1)^{Ng}$

else if there are both inverse words and degree adverb $\overline{s}$ before $\varpi_{i,g}^{j,k}$ **then**

$\eta_g\leftarrow\begin{cases}(-1)^{Na}\cdot\rho_f & ,\text{if } (-1)^{Nb}=1\\ -1/\left[(-1)^{Na}\cdot\rho_f\right] & ,\text{if } (-1)^{Nb}=-1\end{cases}$ and $f=\left\{F\mid\overline{s}\in\mathcal{S}_F\right\}$

end if

$v\leftarrow v+\eta_g\cdot O_{i,g}^{j,k}$

end for

$v_i^{j,k}\leftarrow v/G$

return $v_i^{j,k}$

end function

Algorithm 1 can be symbolically described as follows: $\varpi_{i,g}^{j,k}$ represents the g-th word in the review sentence $r_i^{j,k}$; $O_{i,g}^{j,k}$ represents the sentiment score of the word $\varpi_{i,g}^{j,k}$; and η_g is the degree coefficient of the word $\varpi_{i,g}^{j,k}$ and is a temporary variable. The degree adverb dictionary with level F is denoted as $S_F(F=1,2,3,4,5,6)$, and the corresponding degree score is $\rho_F(F=1,2,3,4,5,6)$ (see Table 3). Na denotes the number of negation words following a degree adverb; Nb represents the number of negation words preceding a degree adverb, and Ng represents the number of inverse words preceding a sentiment word. All of these are temporary variables.

Step 5: Calculation of sentiment value using weighted average based on browsing times.

During a given period, there are multiple online reviews, and each review has a different impact on the demand. Generally, the browsing times reflect the extent of influence of a particular review, where a higher number of browsing times indicate a greater impact of the sentiment score on the demand. We calculate the sentiment value e_i^j for the j - th attribute in period i by taking the weighted average of the sentiment scores of all reviews based on browsing times. The calculation formula is as follows:

$$e_i^j=\frac{\sum_{k=1}^{s_i}\zeta_i^k v_i^{j,k}}{\sum_{k=1}^{s_i}\zeta_i^k}=\frac{\sum_{k=1}^{s_i}\zeta_i^{j,k}\varphi\left(r_i^{j,k}\right)}{\sum_{k=1}^{s_i}\zeta_i^k}, \tag{14}$$

where ζ_i^k represents the browsing times of the k-th review in the period i.

After the aforementioned steps, we obtain a set of features related to the online reviews of the product. The number of features for each period is equal to the number of product attributes, denoted as $\left[e_i^1,\ldots,e_i^m\right]$, $(i=1,2,\ldots,n)$.

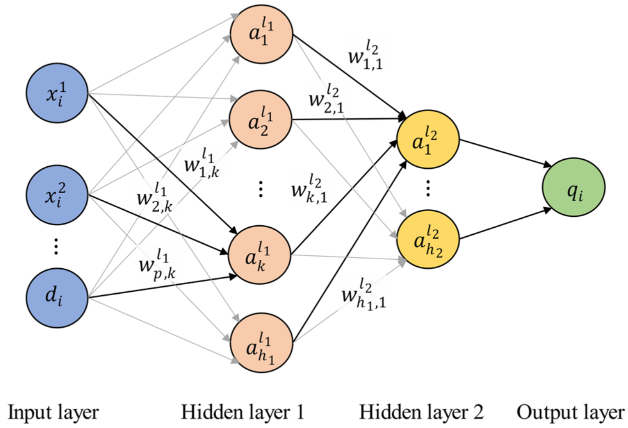


Fig. 4 A simple example of DNN

4.4 Construction and training of deep learning model

By combining the features selected and constructed in Sect. 4.2 with the online review features and historical demand, we obtain the input dataset:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^Q & e_1^1 & \cdots & e_1^m & d_1 \\ x_2^1 & x_2^2 & \cdots & x_2^Q & e_2^1 & \cdots & e_2^m & d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^Q & e_n^1 & \cdots & e_n^m & d_n \end{bmatrix}, \tag{15}$$

where d_i represents the demand volume in the period i.

For each given data input $\mathbf{X}_i = \left[x_i^1, \ldots, x_i^Q, e_i^1, \ldots, e_i^m, d_i\right]$, we construct a multi-layer deep neural network model, which is denoted as the function $\theta(\cdot)$. Figure 4 presents the network structure of $\theta(\cdot)$.

MLP imitates the biological structure of human brain (Lecun et al. 2015). In Fig. 4, each circle represents a node, and each line represents a weight. Taking the first hidden layer, l_1, as an example, the output value of the node k is $a_k^{l_1} = f_a\left(\sum_{i=1}^p x_i w_{i,k}^{l_1} + b^{l_1}\right)$, in which b^{l_1} is the threshold, and $f_a(\cdot)$ is the activation function. Commonly used activation functions are *relu*, *sigmoid,* and *tanh*, and the formulas are $\mathrm{relu}(x) = \max(0, x)$, $\mathrm{sigmoid}(x) = 1/(1 + e^{-x})$, and $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$, respectively. In this paper, we use *relu* as the activation function.

In contrast with a regression problem, the newsvendor problem training set lacks the label of order or product quantity, including only feature data and corresponding demand data. In the training stage, we use the sentiment value yielded by our sentiment analysis method in Sect. 4.3 and selected numerical demand-related feature data in Sect. 4.2 as the inputs of the MLP model, and then determine the optimal order or product quantity through minimization of the newsvendor problem's objective cost function. Thereafter, in the execution stage, numerical feature inputs and sentiment values (obtained from new online reviews) are applied in real time to the trained model, which then produces an approximate optimal order or product quantity of the next period.

Denote the order or product quantity obtained for each period as q_i, and the objective function is constructed by.

$$q_i = \theta(\mathbf{X}_{i-1}; \mathbf{W}), \tag{16}$$

where $\mathbf{W}$ represents all parameters in the network to be estimated. The loss function of the newsvendor problem is

$$C_i = c_h(q_i - d_i)^+ + c_b(d_i - q_i)^+. \tag{17}$$

The data-driven newsvendor problem, i.e., the optimize objective of this MLP, can be expressed as follows:

$$\begin{aligned} \min_{\mathbf{W}} C &= \frac{1}{n-1}\sum_{i=2}^n C_i + \lambda\|\mathbf{W}\|_2^2 \\ &= \frac{1}{n-1}\sum_{i=2}^n \left[c_h(q_i - d_i)^+ + c_b(d_i - q_i)^+\right] + \lambda\|\mathbf{W}\|_2^2 \\ &= \frac{1}{n-1}\sum_{i=2}^n \begin{bmatrix} c_h(\theta(\mathbf{X}_{i-1}; \mathbf{W}) - d_i)^+ + \\ c_b(d_i - \theta(\mathbf{X}_{i-1}; \mathbf{W}))^+ \end{bmatrix} + \lambda\|\mathbf{W}\|_2^2. \end{aligned} \tag{18}$$

In the problem (18), $\lambda\|\mathbf{W}\|_2^2$ is the ℓ_2 regularization term, which is used for alleviating over-fitting, and λ is the regularization coefficient. After estimating the parameters $\mathbf{W}^*$, the optimal order or product quantity for the next period can be obtained by $q^* = q_{n+1} = \theta(\mathbf{X}_n; \mathbf{W}^*)$.

The neural network training process involves adjusting the weight parameters and thresholds in the neural network through optimizer. This process can obtain asymptotic optimal solutions of these trainable parameters. In this paper, we use the *Adam* algorithm to optimize the MLP network because it has faster convergence and better learning performance compared with other learning rate adaptive algorithms. The *Adam* optimizer also addresses the limitations of other optimization techniques, such as loss of learning rates, slow convergence, or high-variable parameter updates, which can cause large fluctuations in the loss function. For a detailed description of the *Adam*, see Kingma and Ba (2015).

5 Experiments with real-world data

5.1 Data description

The performances of our method and benchmarks were evaluated using real-world data of *Lavida*, a popular Volkswagen car model in China (Zhang et al. 2020a, b, c). The data of automobile products are representative of other products because the data of different products have similar feature-demand relationships. It is important for automobile manufacturers to accurately determine the monthly production quantity in advance.

To facilitate optimal production decision-making for the automobile manufacturer, we gathered feature data that are closely associated with demand. Firstly, we retrieved 92 macroeconomic indicators relevant to the automotive industry from the wind database (Zhang et al. 2020a, b, c). Utilizing the approach outlined in Sect. 4.2, we selected three indicators that exhibit the strongest correlation with the *Lavida* car: CPI of transportation facility, crude oil price data, and new RMB loans. Among them, the optimal lag order for the CPI of transportation facility feature is 5, with an absolute distance correlation coefficient of 0.48. Both the crude oil price and new RMB loans features share an optimal lag order of 11, with absolute distance correlation coefficients of 0.40 each. Secondly, we collected the search index of the query "Lavida" from the Baidu index platform.[1] This index reflects the monthly search frequency of the term "Lavida" on China's largest search engine, Baidu. Thirdly, we derived the month of the year for each data point based on the corresponding date, thereby incorporating the periodic feature pertinent to the automotive sales industry. Fourthly, employing web scraping software, we gathered textual online reviews and their associated browsing times for various attributes of the *Lavida* car from the *Autohome Koubei* platform.[2] Lastly, we obtained the historical monthly sales volume of the *Lavida* car product from the *Chezhuzhijia* website,[3] which serves as the demand data for model training. A detailed description of the aforementioned data is presented in Table 4.

The data collection frequency is monthly, and the time range is from January 2014 to January 2022. Figure 5 presents the time series of the historical demand. Additionally, Table 5 provides descriptive statistics for these data. It can be observed that the demand fluctuates greatly and is susceptible to external environmental interference. For example, in February 2020, its demand was affected by the COVID-19 epidemic and fell sharply in the short term.

5.2 Experiment setting

The problem of automobile production can be represented by the newsvendor model. The goal of this model is to determine the product quantity that minimizes the total cost by considering the uncertainty of demand, the underage cost, and the overage cost. We employed Python 3.10 as the programming language and TensorFlow 2.0 as the deep learning development framework to implement our method. The operating environment was Windows 11 X64 with an Intel Core i5-10500 T CPU @ 2.30 GHz and 16-GB RAM.

To train the deep learning model, we assumed that the historical sales volume is regarded as the demand. We used the data from February 2021 to January 2022 (a total of 12 samples) as the test set and conducted experiments in the following two aspects.

1. Comparison of models under different sample sizes.

   We selected four different time ranges as the training set for our experiment: (a) February 2020–January 2021 (12 samples in total), (b) February 2019–January 2021 (24 samples in total), (c) February 2018–January 2021 (36 samples in total), and (d) February 2017–January 2021 (48 samples in total). The purpose of this experiment is to investigate whether the number of available sample size has an influence on the output results. This exploration aims to ascertain whether the proposed method is applicable to both large- and small-scale datasets. Additionally, this experiment demonstrates the robustness of our model to varying sample sizes, thereby showcasing its ability to yield favorable outcomes even with insufficient historical data.

2. Comparison of models under different c_b/c_h *values.*

   In many enterprises, the loss caused by shortage is often greater than the inventory cost, that is, $c_b \geq c_h$ is almost always true (Oroojlooyjadid et al. 2019). However, to comprehensively compare and fully explain the issue, we also compared cases where $c_b < c_h$. We assumed that $c_b + c_h = 2$ and selected seven different values of c_b/c_h, i.e., $c_b/c_h = \{0.1, 0.5, 1, 1.5, 2, 5, 10\}$, to conduct the experiments, respectively. The purpose of this experiment is to further explain that the integrated estimation and optimization model is generally better than the *EAS* model or the *SEO* model as it offers an optimal balance between the underage and the overage costs. Furthermore, this experiment illustrates the superiority and robustness of our proposed model.

   The evaluation metric used for model comparison is total cost. For each point in the test set, the newsvendor cost is $C = c_h(q-d)^+ + c_b(d-q)^+$. To evaluate the overall performance of a model in our experiment, we calculated the average cost of the newsvendor problem, i.e.,

   $\text{Average Cost} = \sum_{t=1}^{T}\left[c_h(q_t-d_t)^+ + c_b(d_t-q_t)^+\right]$, where T is the number of samples in the test set.

The deep neural network encompasses several hyperparameters, such as the learning rate, number of neurons, and neural network layers, which require setting in advance. The selection of these hyperparameters has a certain influence

[1] URL of Baidu index: https://index.baidu.com/.

[2] URL: https://k.autohome.com.cn/614#pvareaid=102519.

[3] URL: https://xl.16888.com/s/57415/.

Table 4 The selected data

Data	Explanation	Type	Example
Crude oil price	Unit: dollars per barrel	Number	39.4
CPI-Trans	Concerning transportation facility, month-on-month ratio	Number	−10%
New RMB loans	Current month value calculated by financial institutions (Unit: CNY 100 million)	Number	10,497
Search index of "Lavida"	Baidu search index of the product name	Number	7548
Month of year	The month in the next period	Number	8
Online reviews	Textual reviews for each attribute of the automobile	String	"非常省油…" (Very fuel efficient …)
Browsing times	The number of times an online review is viewed	Number	2080
Sales volume	We assume that demand equals sales (Unit: ten thousand vehicles)	Number	5.2428

Fig. 5 Time series of the monthly demand for *Lavida* from January 2015 to January 2022

Table 5 Descriptive statistic of the selected data

Data	Count	Maximum	Minimum	Mean	Variance
Spot prices for crude oil	84	81.48	16.55	53.12	12.09
Search index	84	360,717	111,985	237,573.88	67,963.79
CPI-Trans	84	0.5	−0.5	−0.117294	0.180986
New RMB loans	84	35,800	4636	13,149.607	6341.13
Month of year	84	12	1	6.5	3.472786
Online reviews	9206	–	–	–	–
Browsing times	9206	1,474,443	2	12,564.65	55,452.06
Sales volume	84	6.2644	0.3303	3.7019	0.97325

on the performance of the model (Kuhn and Johnson 2013; Zhang et al. 2021). In this study, we employed the cross-validation and grid search methods to determine the proper hyperparameters for the network model. Specifically, for a given hyperparameter combination, we divided the training set further and conducted an internal cross-validation (Pirayesh Neghab et al. 2022) by utilizing 70% of the sample data in the training set for training and the remaining 30% for validation, thus evaluating the model's performance under the hyperparameter combination. Subsequently, each group of hyperparameters was tested in turn, and the model with the best hyperparameter combination was obtained through a comparison of the various models tested. Finally, the model with the best hyperparameter combination underwent testing on the unused test set data to evaluate its performance.

5.3 Benchmarks

To demonstrate the effectiveness of our proposed method, we conducted experiments using the same data with recent representative benchmarks to draw interesting conclusions. These benchmarks include:

- EDD: The empirical demand distribution approach. This approach uses only the demand observations and derives the optimal solution based on an empirical distribution. In this paper, the optimal production quantity is calculated following Bertsimas and Thiele (2005), i.e., assuming that all demand observations in the sample are assigned an equal probability of $1/n$.
- EAS-LR: The demand estimated by the linear regression model is regarded as the optimal production of the enterprise. The historical demand data and numerical demand-related feature data are inputs of the linear regression model.
- EAS-SVR: An approach similar to EAS-LR. The difference is that the vector support regression (SVR) model is used to forecast the future demand.
- EAS-ANN: An approach similar to EAS-LR and EAS-SVR. The difference is that the artificial neural network (ANN) model is used to forecast the future demand.
- SEO: Suppose the demand follows a normal distribution. In the first step, the expectation and variance of the demand distribution are estimated by linear regression model using available historical demand data and numerical demand-related feature data. In the second step, the estimated demand distribution is substituted into the newsvendor model for optimal solution.
- KO: The kernel optimization method proposed by Ban and Rudin (2019), which integrates estimation and optimization. The method uses machine learning method as kernel to approximate the conditional distribution of $D|\mathbf{x}$, given $\mathbf{x}$. $D|\mathbf{x}$ can be expressed as the local weighted average of historical data, i.e.,

$$\min \frac{\sum_{i=1}^{n} K_w(\mathbf{x}_{n+1} - \mathbf{x}_i) C(q, d_i)}{\sum_{i=1}^{n} K_w(\mathbf{x}_{n+1} - \mathbf{x}_i)}, \tag{19}$$

where K_w is a kernel function with bandwidth w. $C(q, d_i)$ is the loss function of the newsvendor problem. In this experiment, the kernel function is the Gaussian function $K_w(\mathbf{u}) = \frac{1}{2\pi}\exp\left(-\frac{\|\mathbf{u}\|_2^2}{2w}\right)$. Intuitively speaking, the KO method selects points whose distance from $\mathbf{x}_{n+1}$ is less than w, and then calculates their losses and takes the arithmetic average.

LML: The linear machine learning method proposed by Ban and Rudin (2019). This method also integrates estimation and optimization. The optimal production quantity is obtained through a linear programming problem of the form $\theta(\mathbf{x}_i; \boldsymbol{\beta}) = \beta_0 + x_i^1\beta_1 + \ldots + x_i^p\beta_p = \left[1, \mathbf{x}_i^{\mathrm{T}}\right]\boldsymbol{\beta}$, which transforms the problem into a linear programming problem represented by:

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n}(c_h o_i + c_b u_i) \\ &\text{s.t.} \quad o_i \geq d_i - \left[1, \mathbf{x}_i^{\mathrm{T}}\right]\boldsymbol{\beta} \\ &\quad u_i \geq \left[1, \mathbf{x}_i^{\mathrm{T}}\right]\boldsymbol{\beta} - d_i \\ &\quad u_i, o_i \geq 0, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{20}$$

- DNN: The nonlinear extension of the LML model by applying the deep neural network to solve the data-driven newsvendor problem (Liu et al. 2022; Oroojlooyjadid et al. 2019). Similarly, it is a model integrating estimation and optimization.
- SA-DNN: Data-driven method using textual reviews proposed by this paper, i.e., the Sentiment Analysis-Deep Neural Network model.

5.4 Comparison and analysis

We conducted 12 experiments for each of the above benchmark methods according to different sample sizes and different c_b/c_h values. The total number of experiments was 12×9=108. Each experiment produced 12 test results, i.e., 12 newsvendor costs (*underage cost*+*overage cost*), totaling 108×12=1296 results.

Table 6 presents the average costs of production decisions obtained by each experiment with different models under different sample sizes n when $c_b/c_h = 2$. Table 7 shows average costs of production decisions obtained by each experiment with different c_b/c_h values when the sample size $n = 48$. The bolded values in the table represent the best result for each case. To provide a visual comparison of the performance of different methods under different situations, we have plotted Figs. 6 and 7, which, respectively, depict the average costs of order/production decisions obtained by each model against the change of the sample size and the c_b/c_h value.

As shown in Table 6 and Fig. 6, the average cost of production decisions obtained by all methods except for the EDD method, and the proposed SA-DNN method decreases with an increase in the sample size. This indicates that the performance of existing methods is greatly influenced by

Table 6 The average costs of production decision by each method with respect to different sample sizes n

Benchmarks	n=12	n=24	n=36	n=48	n=60	n=72
EDD	0.8848	1.0520	1.0702	1.0709	1.0861	1.0414
EAS-LR	1.0923	2.7810	1.2816	0.9313	0.9479	0.8788
EAS-SVR	1.1621	2.0139	1.4148	0.9384	0.5524	0.7287
EAS-ANN	3.4939	1.1651	1.4078	0.9863	**0.2235**	0.2911
SEO	0.8775	0.9740	0.7730	0.6754	0.6275	0.5879
KO	0.5584	0.6953	0.3729	0.3445	0.3445	0.3445
LML	0.6854	1.1947	0.8771	0.2778	0.2280	0.2326
DNN	0.4380	0.5138	0.2784	0.2662	0.2597	0.2381
SA-DNN	**0.2002**	**0.2176**	**0.2169**	**0.2411**	0.2276	**0.2150**

Table 7 The average costs of benchmarks with respect to different c_b/c_h values

Benchmarks	$\frac{c_b}{c_h} = 0.1$	$\frac{c_b}{c_h} = 0.5$	$\frac{c_b}{c_h} = 1$	$\frac{c_b}{c_h} = 1.5$	$\frac{c_b}{c_h} = 2$	$\frac{c_b}{c_h} = 5$	$\frac{c_b}{c_h} = 10$
EDD	0.5198	1.1616	1.0158	0.7667	0.8024	1.1957	1.0709
EAS-LR	0.6073	0.7033	0.7693	0.8089	0.8353	0.9013	0.9313
EAS-SVR	0.5212	**0.6448**	0.7298	0.7808	0.8148	0.8998	0.9384
EAS-ANN	0.8273	0.7350	**0.6839**	0.7756	0.8501	0.8212	0.9863
SEO	0.4822	0.8198	0.9099	1.0144	1.0389	0.9460	0.6754
KO	0.8080	1.0328	0.9744	0.9030	0.9846	0.7999	0.3445
LML	0.4407	0.6850	0.7424	0.7834	0.8106	0.4797	0.2778
DNN	0.4946	0.7549	0.7469	0.7480	0.6805	0.4251	0.2662
SA-DNN	**0.3395**	0.8616	0.7979	**0.6756**	**0.6589**	**0.3743**	**0.2411**

the sample size, and an inadequate sample size can result in poor decision-making outcomes. On the other hand, our proposed method consistently achieves the lowest average cost of order/production decisions, particularly in instances of small sample sizes, except when $n = 60$. This demonstrates that leveraging textual reviews enable enterprises to make optimal production decisions even with limited data samples, and our proposed method is not vulnerable to fluctuations in sample size. In summary, our method for calculating the optimal production quantity using textual reviews is both robust in terms of the number of training samples and well-suited to cases where sample size is insufficient.

In Table 7 and Fig. 7, it can be observed that the EAS methods, including EAS-LR, EAS-SVR, and EAS-ANN, perform more poorly as the c_b/c_h value increases. This outcome can be attributed to the fact that forecasting aims to minimize the difference between predicted and actual values, whereas the newsvendor problem considers the trade-off between the unit underage cost and the unit overage cost. Additionally, our proposed method demonstrates the best overall performance when $c_b > c_h$, which is more reflective of actual scenarios. In summary, our method is more suitable for solving newsvendor problems in practice, especially when the unit underage cost outweighs the unit overage cost.

We plot all experimental results (newsvendor costs of every test point) for each method under all conditions into boxplots, as shown in Fig. 8. In the figure, the solid line in the body of each box represents the median, and the dotted line represents the mean value.

As shown in Fig. 8: (1) The performances of the decision-making methods based on EAS are generally comparable to those of EDD overall, with EAS outperforming EDD in the majority of cases. In this experiment, EAS-LR, EAS-SVR, and EAS-ANN outperform EDD in 50.75%, 57.58%, and 56.0% of the cases, respectively. However, it is evident that some results from EAS are poor, indicating unstable performance. (2) The SEO method outperforms EDD and EAS methods in 32.58% of cases, with an average cost of production decisions that are 14%, 21.62%, 12.74%, and 23.65% lower than those of EDD, EAS-LR, EAS-SVR, and EAS-ANN, respectively. However, the SEO method separates the estimation and optimization steps, and the discrepancy between the objective functions of the two steps may degrade its overall performance. In contrast, integrated estimation and optimization methods, such as KO, LML, DNN, and SA-DNN, exhibit superior performance. (3) The DNN method is generally more effective than the KO and LML methods because it considers the nonlinear relationship between features and demand. Statistics show that the DNN method outperforms the KO method in 51.51% of cases and the LML method in 51.51% of cases as well. (4) Our proposed SA-DNN method demonstrates the best overall performance among all methods, because it incorporates unstructured feature data, i.e., textual online reviews, while considering the nonlinear relationship between demand and structured features. Through sentiment analysis of review sentences, SA-DNN can effectively extract useful information to produce the lowest-cost solution compared to other methods. Specifically, the average cost of the SA-DNN method is approximately 54.72% lower than EDD and approximately 14.18% lower than the most advanced DNN method. Furthermore, the statistics demonstrate that our SA-DNN method is superior to EDD in 85.6% of cases, DNN in 54.55% of cases, and all other benchmark methods in 19.7% of cases.

In conclusion, the method of integrating estimation and optimization performs better than the EDD method and the SEO method, while the data-driven method using textual reviews outperforms the nonlinear deep learning method that exclusively considers structured data.

To demonstrate the superiority of each method over the others, we conducted *t*-tests on cost values between pairs of methods. The resulting *p*-values are presented in Table 8, which indicate whether there is a meaningful and statistically significant difference between the cost values of any two methods. Specifically, if the cost calculated by the method in row i is lower than that calculated by the method in column j, the corresponding cell ij in the table displays their *p*-value obtained from the *t*-test; otherwise, the cell is left blank. A *p*-value below 5% indicates that the average cost of one method is lower than that of another at a 5% significance level, marked with "*" in the table. Similarly, a *p*-value lower than 1% expresses that the average cost of a method is lower than another at a 1% significance level, marked with "**" in the table. It can be seen that our proposed SA-DNN method has the lowest cost compared to other methods, with *p*-values close to zero confirming this finding. Specifically, SA-DNN is superior to all other methods except DNN at the 1% level and superior to the DNN method at the 5% level.

5.5 Managerial insights

Analysis of Table 6 and Fig. 6 reveals the impact of sample size on the performance of existing decision-making methods, particularly for methods that do not incorporate textual online reviews, such as EAS and SEO. In contrast, our method demonstrates robustness and is less affected by changes in sample size. This robustness is attributed to enriched sentiment analysis from textual data, which reduces variability in demand. The insight derived from this analysis is that for machine learning models, having an adequate amount of data samples can lead to better decision outcomes. However, even in cases of limited data samples, incorporating sentiment analysis of textual online reviews can still yield good decision results. Therefore, managers should focus on the collection of historical data related to demand, particularly online reviews on e-commerce platforms, as these data contain valuable features that influence demand. For businesses with a shorter operating history and a smaller sample period, it is especially important to expand the variety of selected feature data, gather online reviews, and increase the dimensionality of demand-related features.

Analysis of Table 7, Figs. 7 and 8 demonstrates the superiority of multi-source data-driven methods over single-source data-driven methods, deep learning methods over linear machine learning methods, integrated estimation and optimization methods over separate estimation and optimization methods, as well as prediction-based decision-making methods. The managerial insight gained from these findings is that in ordering/production decision-making, it is not advisable to separate the forecasting and optimization departments. Additionally, incorporating textual online reviews into production and inventory decision-making processes can significantly improve cost optimization. Managers can implement our proposed method and develop corresponding systems or tools that enable effective analysis and utilization of textual reviews from large volumes of unstructured textual data. In practice, leading companies have recognized the importance of online review data and utilize it to understand customer preferences and sentiments, thereby making more informed production and inventory decisions. For example, Amazon leverages sentiment analysis to optimize its inventory management and product recommendations based on customer reviews and preferences.[4]

Synthesizing the results of all experiments, our proposed method demonstrates numerous advantages. The insight derived from this study is that investing in data analysis capabilities and fostering a data-driven culture can provide organizations with a competitive advantage. By combining deep learning techniques with carefully selected multi-source feature data, organizations can effectively capture the complex nonlinear relationships between features and demand. Managers should embrace the use of data and advanced analytical techniques to inform their decision-making processes.

[4] https://www.wonderflow.ai/blog/sentiment-analysis-examples.

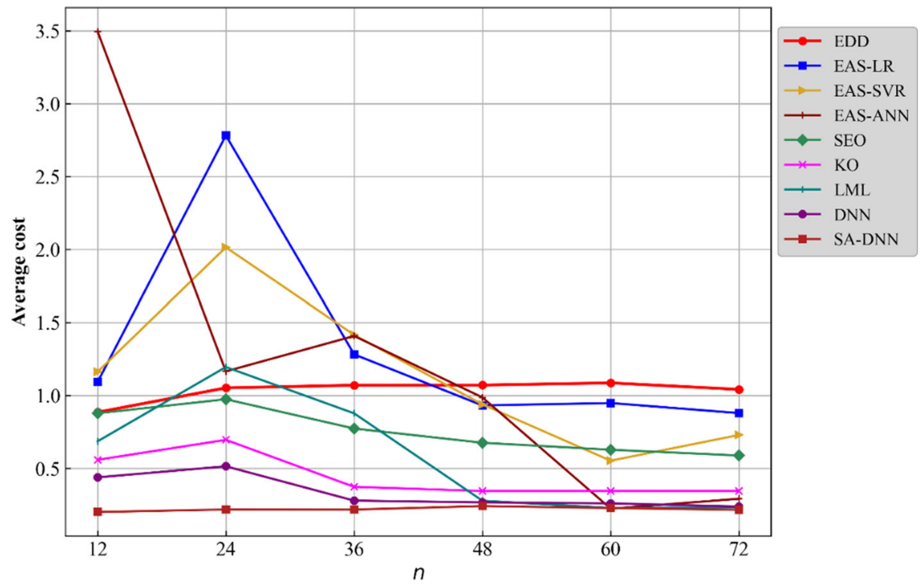


Fig. 6 Comparison of models with respect to different sample sizes n

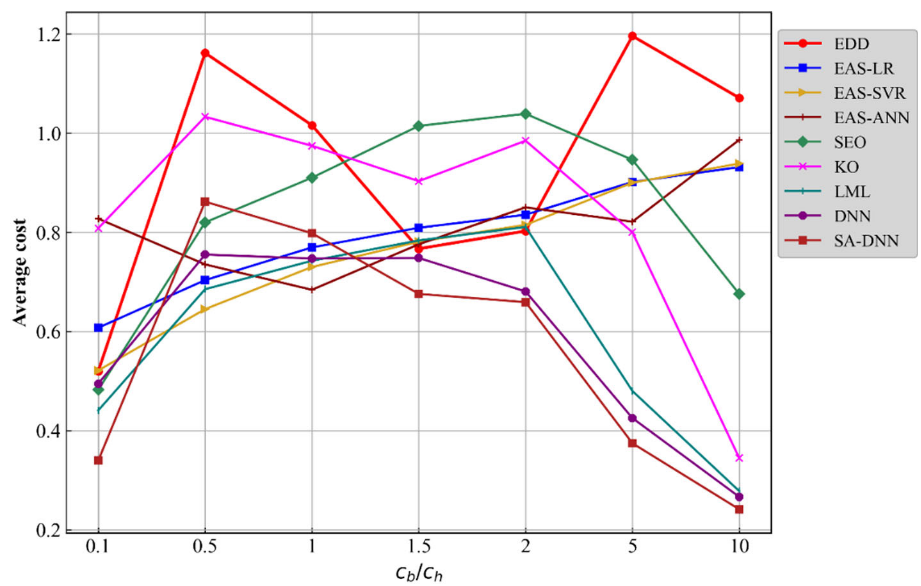


Fig. 7 Comparison of models with respect to different c_b/c_h ratio values

6 Conclusions

The purpose of this study is to propose a new integrated learning and optimization method for ordering or production decisions. The proposed method utilizes numerical feature data in combination with textual online review data to address the newsvendor problem. The approach leverages sentiment analysis and deep learning techniques to extract pertinent information from large volumes of unstructured textual data. This information assists in the production and inventory decision-making processes, by considering the vast amounts of reviews provided by shoppers regarding products and services, to optimize production and inventory cost, thereby reducing overhead costs while enhancing profitability mightily. The experimental results from real-world data demonstrate the superiority and robustness of the proposed method. The proposed method offers significant cost savings for enterprises, with reductions of nearly 55%

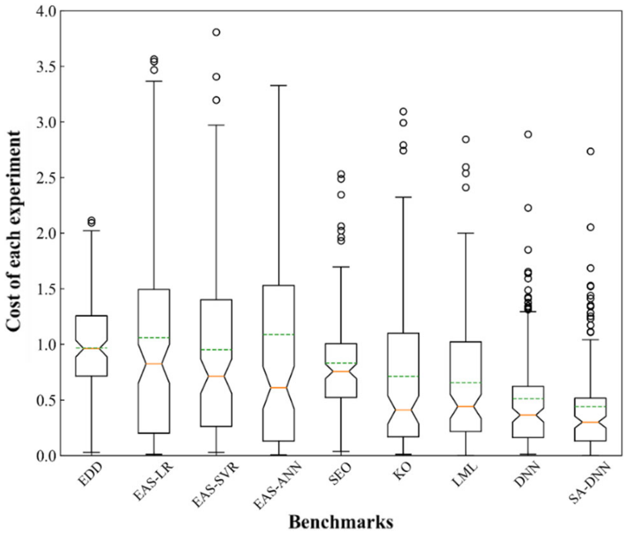


Fig. 8 The statistical results of all tested methods across all cases

compared to the EDD method and 14.18% compared to the most advanced DNN method.

The primary managerial insights derived from this study can be summarized as follows: (1) It is unwise to separate the forecasting and optimization departments in ordering/production decision-making. Integrating these two departments can result in improved decision outcomes. (2) Managers should leverage sentiment analysis and deep learning techniques and implement systems or tools that enable effective analysis and utilization of textual reviews from large volumes of unstructured textual data to obtain decision support. (3) Investing in data analysis capabilities and cultivating a data-driven culture can provide organizations with a competitive advantage. Managers should embrace the use of data and advanced analytical techniques to inform their decision-making processes. In conclusion, incorporating textual reviews and adopting a data-driven approach are essential managerial insights that can assist managers in enhancing their product management strategies. By leveraging these insights, organizations can improve cost optimization, customer satisfaction, and overall operational efficiency.

The proposed method offers a more comprehensive approach to data-driven production and inventory problems, with theoretical value and practical applications. However, it is important to acknowledge its limitations and identify areas for future research. Firstly, there is a need to deepen the sentiment analysis technology employed and explore more advanced sentiment analysis technologies in practical applications. Secondly, the proposed method only considers the production decision of a single product in a single cycle, and future research can extend this approach to address multi-stage and multi-product production and inventory problems. Furthermore, in this study, we assume that demand equals sales volume, which is not always the case when product supply fails to meet demand. Future research can address this issue by considering the demand censoring problem, which would highlight the decision support role of textual reviews in a more realistic scenario.

Table 8 *p*-values of the *t*-test between the cost of benchmarks

p-value	EDD	EAS-LR	EAS-SVR	EAS-ANN	SEO	KO	LML	DNN	SA-DNN
EDD		1.833e−01		1.802e−01					
EAS-LR				4.061e−01					
EAS-SVR	4.372e−01	4.839e−03**		9.293e−02					
EAS-ANN									
SEO	1.069e−03**	1.585e−02*	9.116e−02	2.562e−02*					
KO	9.642e−04**	8.579e−04**	6.892e−03**	1.473e−03**	4.809e−02*				
LML	8.450e−06**	2.777e−08**	7.151e−08**	4.297e−05**	4.952e−03**	2.202e−01			
DNN	2.457e−12**	8.998e−08**	4.284e−07**	3.669e−06**	2.787e−11**	1.142e−03**	8.413e−03**		
SA-DNN	2.582e−18**	2.314e−08**	6.374e−08**	1.199e−06**	9.980e−23**	1.159e−05**	4.632e−04**	1.851e−02*	

Author contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by CZ and Y-XT. The first draft of the manuscript was written by Y-XT, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National Social Science Fund of China [Grant Number, 19BGL229].

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

References

Akhtar MS, Ekbal A, Cambria E (2020) How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble [application notes]. IEEE Comput Intell Mag 15(1):64–75. https://doi.org/10.1109/MCI.2019.2954667

Akter S, Aziz MT (2016) Sentiment analysis on facebook group using lexicon based approach. Paper presented at the 2016 3rd international conference on electrical engineering and information communication technology (ICEEICT), 22–24 Sept. 2016

Alam MH, Ryu W-J, Lee S (2016) Joint multi-grain topic sentiment: modeling semantic aspects for online reviews. Inf Sci 339:206–223. https://doi.org/10.1016/j.ins.2016.01.013

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. Manag Sci 57 (8):1485–1509. https://doi.org/10.1287/mnsc.1110.1370

Arrow KJ, Harris T, Marschak J (1951) Optimal inventory policy. Econometrica 19:250–272

Arunraj NS, Ahrens D (2015) A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. Int J Prod Econ 170:321–335. https://doi.org/10.1016/j.ijpe.2015.09.039

Bai X (2011) Predicting consumer sentiments from online text. Decis Support Syst 50(4):732–742. https://doi.org/10.1016/j.dss.2010.08.024

Ban GY, Rudin C (2019) The big Data newsvendor: practical insights from machine learning. Oper Res 67(1):90–108. https://doi.org/10.1287/opre.2018.1757

Bertsimas D, Thiele A (2005) A data-driven approach to newsvendor problems. Working paper, Massachusetts Institute of Technology, 51

Besbes O, Muharremoglu A (2013) On implications of demand censoring in the newsvendor problem. Manag Sci 59(6):1407–1424. https://doi.org/10.1287/mnsc.1120.1654

Bi JW, Han TY, Yao Y, Li H (2022a) Ranking hotels through multi-dimensional hotel information: a method considering travelers' preferences and expectations. Inf Technol Tour 24(1):127–155. https://doi.org/10.1007/s40558-022-00223-y

Bi X, Adomavicius G, Li W, Qu A (2022b) Improving sales forecasting accuracy: a tensor factorization approach with demand awareness. Informs J Comput 34(3):1644–1660. https://doi.org/10.1287/ijoc.2021.1147

Cao Y, Shen Z-JM (2019) Quantile forecasting and data-driven inventory management under nonstationary demand. Oper Res Lett 47(6):465–472. https://doi.org/10.1016/j.orl.2019.08.008

Cao N, Ji S, Chiu DKW, He M, Sun X (2020) A deceptive review detection framework: combination of coarse and fine-grained features. Expert Syst Appl 156:113465. https://doi.org/10.1016/j.eswa.2020.113465

Chang Y-C, Ku C-H, Chen C-H (2020) Using deep learning and visual analytics to explore hotel reviews and responses. Tour Manag 80:104129. https://doi.org/10.1016/j.tourman.2020.104129

Chen Y, Xie J (2008) Online consumer review: Word-of-mouth as a new element of marketing communication mix. Manag Sci 54 (3):477–491. https://doi.org/10.1287/mnsc.1070.0810

Chong AYL, Li B, Ngai EWT, Ch'ng E, Lee F (2016) Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. Int J Oper Prod Manag 36(4):358–383. https://doi.org/10.1108/IJOPM-03-2015-0151

Dey A, Jenamani M, Thakkar JJ (2018) Senti-N-Gram: an n-gram lexicon for sentiment analysis. Expert Syst Appl 103:92–105. https://doi.org/10.1016/j.eswa.2018.03.004

Duan W, Gu B, Whinston AB (2008) Do online reviews matter?—an empirical investigation of panel data. Decis Support Syst 45 (4):1007–1016. https://doi.org/10.1016/j.dss.2008.04.001

Elmachtoub AN, Grigas P (2022) Smart "predict, then optimize." Manag Sci 68(1):9–26. https://doi.org/10.1287/mnsc.2020.3922

Fan Z-P, Che Y-J, Chen Z-Y (2017) Product sales forecasting using online reviews and historical sales data: a method combining the Bass model and sentiment analysis. J Bus Res 74:90–100. https://doi.org/10.1016/j.jbusres.2017.01.010

Fernández-Gavilanes M, Álvarez-López T, Juncal-Martínez J, Costa-Montenegro E, Javier González-Castaño F (2016) Unsupervised method for sentiment analysis in online texts. Expert Syst Appl 58:57–75. https://doi.org/10.1016/j.eswa.2016.03.031

He B, Dexter F, Macario A, Zenios S (2012) The timing of staffing decisions in hospital operating rooms: incorporating workload heterogeneity into the newsvendor problem. Manuf Serv Oper Manag 14(1):99–114. https://doi.org/10.1287/msom.1110.0350

Huber J, Müller S, Fleischmann M, Stuckenschmidt H (2019) A data-driven newsvendor problem: from data to decision. Eur J Oper Res 278(3):904–915. https://doi.org/10.1016/j.ejor.2019.04.043

Kang Y, Cai Z, Tan C-W, Huang Q, Liu H (2020) Natural language processing (NLP) in management research: a literature review. Journal of Management Analytics 7(2):139–172. https://doi.org/10.1080/23270012.2020.1756939

Keskin NB, Li Y, Song J-S (2022) Data-driven dynamic pricing and ordering with perishable inventory in a changing environment. Manag Sci 68(3):1938–1958. https://doi.org/10.1287/mnsc.2021.4011

Khorsand R, Rafiee M, Kayvanfar V (2020) Insights into TripAdvisor's online reviews: the case of Tehran's hotels. Tour Manag Perspect 34:100673. https://doi.org/10.1016/j.tmp.2020.100673

Kingma DP, Ba JL (2015) Adam: a method for stochastic optimization. Paper presented at the 3rd international conference on learning representations, ICLR 2015

Kotelnikova A, Paschenko D, Bochenina K, Kotelnikov E (2022) Lexicon-based methods vs. BERT for text sentiment analysis. Paper presented at the analysis of images, social networks and texts, Cham, 2022

Ku L-W, Chen H-H (2007) Mining opinions from the Web: Beyond relevance retrieval. J Am Soc Inform Sci Technol 58(12):1838–1850. https://doi.org/10.1002/asi.20630

Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York

Kyparisis GJ, Koulamas C (2018) The price-setting newsvendor with nonlinear salvage revenue and shortage cost. Oper Res Lett 46 (1):64–68. https://doi.org/10.1016/j.orl.2017.11.001

Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521 (7553):436–444. https://doi.org/10.1038/nature14539

Levi R, Roundy RO, Shmoys DB (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. Math Oper Res 32(4):821–839. https://doi.org/10.1287/moor.1070.0272

Lin S, Chen Y, Li Y, Shen ZJM (2021) Data-driven newsvendor problems regularized by a profit risk constraint. Prod Oper Manag. https://doi.org/10.1111/poms.13635

Liu Y, Bi J-W, Fan Z-P (2017a) Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms. Expert Syst Appl 80:323–339. https://doi.org/10.1016/j.eswa.2017.03.042

Liu Y, Bi J-W, Fan Z-P (2017b) Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. Inf Fus 36:149–161. https://doi.org/10.1016/j.inffus.2016.11.012

Liu Y, Bi JW, Fan ZP (2017c) A method for ranking products through online reviews based on sentiment classification and interval-valued intuitionistic fuzzy topsis. Int J Inf Technol Decis Mak 16 (6):1497–1522. https://doi.org/10.1142/S021962201750033X

Liu Y, Jiang C, Zhao H (2018) Using contextual features and multi-view ensemble learning in product defect identification from online discussion forums. Decis Support Syst 105:1–12. https://doi.org/10.1016/j.dss.2017.10.009

Liu C, Letchford AN, Svetunkov I (2022) Newsvendor problems: an integrated method for estimation and optimisation. Eur J Oper Res 300(2):590–601. https://doi.org/10.1016/j.ejor.2021.08.013

Liyanage LH, Shanthikumar JG (2005) A practical inventory control policy using operational statistics. Oper Res Lett 33(4):341–348. https://doi.org/10.1016/j.orl.2004.08.003

Martinez-Torres MR, Toral SL (2019) A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. Tour Manag 75:393–403. https://doi.org/10.1016/j.tourman.2019.06.003

Mohamad Beigi O, Moattar MH (2021) Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. Knowl Based Syst 213:106423. https://doi.org/10.1016/j.knosys.2020.106423

Moreno-Ortiz A, Fernández-Cruz J (2015) Identifying polarity in financial texts for sentiment analysis: a corpus-based approach. Procedia Soc Behav Sci 198:330–338. https://doi.org/10.1016/j.sbspro.2015.07.451

Moreo A, Romero M, Castro JL, Zurita JM (2012) Lexicon-based comments-oriented news sentiment analyzer system. Expert Syst Appl 39(10):9166–9180. https://doi.org/10.1016/j.eswa.2012.02.057

Najmi E, Hashmi K, Malik Z, Rezgui A, Khan HU (2015) CAPRA: a comprehensive approach to product ranking using customer reviews. Computing 97(8):843–867. https://doi.org/10.1007/s00607-015-0439-8

Oroojlooyjadid A, Snyder LV, Takáč M (2019) Applying deep learning to the newsvendor problem. IISE Trans 52(4):444–463. https://doi.org/10.1080/24725854.2019.1632502

Peng S, Cao L, Zhou Y, Ouyang Z, Yang A, Li X, Jia W, Yu S (2022) A survey on deep learning for textual emotion analysis in social networks. Digital Commun Netw 8(5):745–762. https://doi.org/10.1016/j.dcan.2021.10.003

Pirayesh Neghab D, Khayyati S, Karaesmen F (2022) An integrated data-driven method using deep learning for a newsvendor problem with unobservable features. Eur J Oper Res. https://doi.org/10.1016/j.ejor.2021.12.047

Qi M, Mak HY, Shen ZJM (2020) Data-driven research in retail operations—a review. Nav Res Logist 67(8):595–616. https://doi.org/10.1002/nav.21949

Qi M, Shi Y, Qi Y, Ma C, Yuan R, Wu D, Shen Z-JM (2023) A practical end-to-end inventory management model with deep learning. Manage Sci 69(2):759–773. https://doi.org/10.1287/mnsc.2022.4564

Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35(6):2769–2794. https://doi.org/10.1214/009053607000000505

Vana PM, Lambrecht AM (2018) Online reviews: star ratings, position effects and purchase likelihood. Social Science Electronic Publishing, Rochester

Weichselbraun A, Gindl S, Scharl A (2011) Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons. Paper presented at the proceedings of the 20th ACM international conference on Information and knowledge management, Glasgow, Scotland, UK. https://doi.org/10.1145/2063576.2063729

Wu H, Gu Y, Sun S, Gu X (2016) Aspect-based opinion summarization with convolutional neural networks. Paper presented at the 2016 international joint conference on neural networks (IJCNN), 24–29 July 2016

Ye Q, Law R, Gu B (2009) The impact of online user reviews on hotel room sales. Int J Hosp Manag 28(1):180–182. https://doi.org/10.1016/j.ijhm.2008.06.011

Zhang C, Tian YX, Fan LW (2020a) Improving the Bass model's predictive power through online reviews, search traffic and macroeconomic data. Ann Oper Res 295(2):881–922. https://doi.org/10.1007/s10479-020-03716-3

Zhang C, Tian YX, Fan LW, Li YH (2020b) Customized ranking for products through online reviews: a method incorporating prospect theory with an improved VIKOR. Appl Intell 50(6):1725–1744. https://doi.org/10.1007/s10489-019-01577-3

Zhang C, Tian YX, Fan ZP, Liu Y, Fan LW (2020c) Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis. Soft Comput 24(9):6213–6226. https://doi.org/10.1007/s00500-018-03742-1

Zhang C, Tian YX, Fan ZP (2021) Forecasting sales using online review and search engine data: a method based on PCA–DSFOA–BPNN. Int J Forecast. https://doi.org/10.1016/j.ijforecast.2021.07.010

Zhang C, Tian YX, Fan ZP (2022) Forecasting the box offices of movies coming soon using social media analysis: a method based on improved Bass models. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2021.116241

Zhu XD, Ninh A, Zhao H, Liu ZM (2021) Demand forecasting with supply-chain information and machine learning: evidence in the pharmaceutical industry. Prod Oper Manag 30(9):3231–3252. https://doi.org/10.1111/poms.13426