Big Data Research 41 (2025) 100542

Contents lists available at ScienceDirect

Big Data Research

journal homepage: www.elsevier.com/locate/bdr

E-word of mouth in sales volume forecasting: Toyota Camry case study

Domenica Fioredistella Iezzi [a,b,*], Roberto Monte [a,b]

[a] Department of Enterprise Engineering "Mario Lucertini" - Tor Vergata University, Via del Politecnico, 1, Rome, 00133, Italy
[b] Department of Civil Engineering and Computer Science Engineering - Tor Vergata University, Via del Politecnico, 1, Rome, 00133, Italy

ARTICLE INFO

Keywords:
Sentiment analysis
Social media
Machine learning
Deep learning
Multivariate autoregressive state-space

ABSTRACT

In recent years, electronic word of mouth has become a significant factor in purchasing decisions, with consumers' sentiments playing a crucial role in shaping the sales of products and services.
This paper introduces a novel approach to sales forecasting that addresses consumers' sentiments toward goods or services by combining the sales volume time series with a quantitative proxy of the unobservable true sentiment. Numerous studies have explored various methods to capture sentiment and accurately predict sales. We have integrated an estimated sentiment signal, variously built via lexicon-based, machine-learning, and deep-learning approaches, into a multivariate autoregressive state space (MARSS) model. We have tested our model on a dataset of 163,000 tweets about the Toyota Camry, covering the period from June 2009 to December 2022 and sales volumes in the US market over the same timeframe.

1. Introduction

The idea that people's feelings can influence consumers and businesses traces back to the economic psychology theorized in the 1950s and 1960s. In particular, George Katona (1964) [1] and [2] tried to discover what happens when people, such as consumers, business people, or politicians, make decisions.

This current of thought integrates the classical economic theory after Stuart Mill (1844) [3], which assesses that people analyze all information at their disposal, including prices, promotions, and opportunities, when buying a good or making an economic investment, and by sifting through all possible alternatives, choose the rationally best option to satisfy their interests ("homo oeconomicus"). George Katona, in 1946, at the University of Michigan Survey Research Center, developed the Citizen Sentiment Index through a project funded by the Federal Reserve [1]. He validated subjective variables derived from sample surveys to understand economic processes. In this way, he correctly predicted the recovery of the US economy while econometric models were still providing recessive signals [2].

When someone wants to buy a product or service, he often shares the decision-making phase among peers through word of mouth (WOM). According to Vana and Lambrecht (2021) [4], based on a McKinsey study, 67% of daily consumer decisions come from the WOM effect. The precursors of these ideas were Katz and Lagerfeld (1955) [5], who also analyzed the role of WOM in exploring counterfeiting. To deepen the understanding of this problem, the authors developed a self-assessment scale that measures consumer motivations to seek opinions and counterfeiting (see Adler, 1957 [6]). Approximately 30 years later, Brown and Reingen (1987) [7] investigated the roles that social solid bonds and homophily play in macro- and micro-WOM processes. Their results show different roles played by weak and strong social ties. At the macro level, weak links play a critical bridging function, allowing information to travel from one distinct subgroup of referral actors to another in the broader social system. At the micro-level, the flow of referral information likely activates solid and homophilic ties.

In recent years, web reviews have grown on product ratings, movie posts, restaurant recommendations, and vacations. This information can help customers make informed decisions about purchasing products and services. Moreover, it can also help companies, by becoming a virtual currency, to create or suppress products. (see Jalivand et al., 2011 [8]).

Online consumer reviews provide helpful information for making purchases by offering indirect experiences to consumers. Thus, a review has two functions: on the one hand, it informs, and on the other, it advises (see Park et al., 2007 [9]). With the development of the Internet, WOM has changed how information is transmitted. Information goes from WOM between subjects belonging to the same network of acquaintances, strong or weak, to WOM between peers who do not know each other. Gupta and Harris (2010) [10] conducted a laboratory experiment to examine the effects of electronic word of mouth (e-WOM) on consumer consideration and the choice of a product of experience.

* Corresponding author.
E-mail address: stella.iezzi@uniroma2.it (D. Iezzi).

https://doi.org/10.1016/j.bdr.2025.100542
Received 15 September 2024; Received in revised form 27 February 2025; Accepted 9 May 2025
Available online 15 May 2025

In their words: "Specifically, we manipulated the number of consumer recommendations and the optimality of the recommended product in a realistic online shopping environment. The results indicate that e-WOM will likely result in more time considering the recommended product. For consumers more motivated to process information, e-WOM recommendations lead to more time spent on the choice task overall. Further, consumers are less motivated to process information and make suboptimal decisions based on e-WOM recommendations. Consumers with a high motivation to process information are willing to accept recommendations and switch from declared attribute preferences but choose only optimal products." In decision-making, negative information appears to have a more significant impact than positive information. Moreover, the literature highlights several factors: the framing effect (see Kahneman and Tversky, 1979 [11]), impression formation (see Fiske, 1980 [12], see also Kanouse and Hanson, 1972 [13]) and persuasion (see Magdum and Megha, 2017 [14]).

Several studies focus on the role and credibility of the e-WOM source. Hussain et al. (2017) [15] explored the relationships between the credibility of e-WOM, perceived risk, and the adoption of information by consumers of food products, with the quality of argument and usefulness of the information acting as mediators. The results show that customers often rely on e-WOM during the decision-making process when purchasing food products. Ultimately, customers use e-WOM to reduce potential risks during the decision-making process, and companies can use the information derived from e-WOM for their marketing strategies.

Liu et al. (2009) [16] summarized reviews and tweets in terms of sentiments, along with additional input from the box office collection from the IMDB movies platform (see https://www.imdb.com/) using an autoregressive sentiment-aware model (ARSA) for sales prediction. Yu et al. (2012) [17] used the movie domain as a case study to predict sales performance, using online reviews as a domain-specific task to identify critical factors influencing predictions. They introduced an ARSA model for product sales forecasting to capture the impact of sentiment and past sales performance. Moreover, they also considered an autoregressive sentiment and quality-aware (ARSQA) model to address review quality. They validated the effectiveness of these models through experimental results.

Magdum and Megha (2017) [18] claimed that online reviews have become one of the most critical parts of any business. They used various algorithms to predict the sales performance of both Bollywood and Hollywood movies, leveraging sentiment data extracted from reviews and tweets. They used the probabilistic latent semantic analysis (S-PLSA) model to distill sentiment information from online reviews and social media posts. In addition, they applied an ARSA model to forecast movie sales performance by integrating sentiment insights with historical box office data.

This paper aims to use the information derived from consumers' sentiments about a product expressed by tweets in a model to predict product sales. For this purpose, inspired by the Holt-Winters decomposition (see Holt (2004) [19], Winters (1960) [20]), the derived Hyndman's error, trend, seasonality (ETS) models (see Hyndman (2002) [21], and the relatively recent availability of the R package MARSS (see Holmes (2024) [22]), we propose a space-time model in which the observed sales volume splits in trend, local trend, and seasonal components plus noise. We also observe a proxy of consumers' sentiments about the product. The individual components of the sales volume are not observable, nor is the quantification of the true sentiments about the product, which constitutes the local trend component of the sales volume. We compare various determinations of the proxy of the quantitative sentiment using different approaches. On the one hand, our model seems to improve sales volume predictions compared to other methods. On the other hand, it may constitute a way to discriminate among the several different approaches to estimating consumers' sentiments about a product and possibly to predict it.

The paper is structured as follows. In Section 2, we introduce the concepts of sentiment analysis, sentiment scoring, and opinion mining used in the paper; in Section 3, we illustrate the classification approaches used to estimate consumer mood; in Section 4, we present our data; in Section 5, we explain our model; in Section 6 we describe the main results; in Section 7, we draw conclusions and future developments.

2. From words to sentiment metrics

Sentiment analysis, sentiment score, and opinion analysis are three expressions used to quantify the opinions of customers who buy products and services. Sentiment measurements allow you to estimate the propensity to buy based on the mood of customers. The data collected to evaluate customer sentiment is mainly based on reviews, where the customer can detail aspects that concern all the characteristics of the product/service. Social media corpora[1] display a distinctive language abundant in slang expressions, linguistic blends, emoticons, emojis, videos, and photos. This information is part of the broader application domain that deals with natural language processing (NLP), and its purpose is to extract from unstructured input (e.g., text) a structured knowledge that can be used as a support for decision-making (see Devika et al., 2016 [23]). In particular, the category of sentiment analysis includes various natural language processing tasks, e.g., sarcasm detection, topic extraction, and subjectivity detection (see Carvalho et al., 2014 [24]). In other words, sentiment analysis is a process that automates the extraction of attitudes from a text or speech (see Vishal et al., 2016 [25]). More formally, according to Liu (2015) [26], an opinion can be seen as a quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, in which:

- e_i is the name of the ith entity to which the opinion (or sentiment) refers;
- a_{ij} is the specific jth aspect of the ith entity that generates the opinion (and to which it refers);
- s_{ijkl} is the sentiment of the kth opinion holder reffered to a_{ij}th, expressed at lth time. The polarity of the latter can refer to the direct judgment of the writer (opinion) or his emotional state during the writing of the text (sentiment);
- h_k is the kth opinion holder;
- t_l is the lth time in which the opinion is expressed.

Millions of people use social networks daily to stay informed, express their moods, share their opinions, and discuss various topics. This makes social media one of the most promising tools for "measuring" people's moods. Since 2016, the Italian National Institute of Statistics (ISTAT) has developed the Social Mood on Economy Index (SMEI). This experimental index provides daily measures of the Italian sentiment on the economy, derived from samples of public posts in the Italian language, captured in real-time. The index production procedure selects and processes only posts whose text contains at least one word from a filter set prepared by domain experts. The index accounts for about 26,000 daily tweets (see Righi et al., 2022 [27]). In our case study, we also use a set of tweets to measure consumers' sentiments (see Iezzi and Monte (2024) [28]).

The process of quantifying consumers' sentiments, which allows one to pass from text to numerical data, is expressed in several types of classification.

- **binary**: The sentiment of the corpus is categorized as positive and negative. Our analysis uses a dataset called "Sentiment140".[2] It is one of the most extensive datasets for the analysis of sentiments using Twitter data (see Go et al., 2009 [29], Pandya et al., 2021 [30]). We can use tweets to detect sentiments based on their encoding of vocabulary, e.g. *"obama", "mandanicole", "How can you*

[1] A "corpus" is a collection of texts of various types, such as articles, books, conversations, and messages from social media.

[2] See https://www.kaggle.com/datasets/kazanova/sentiment140.

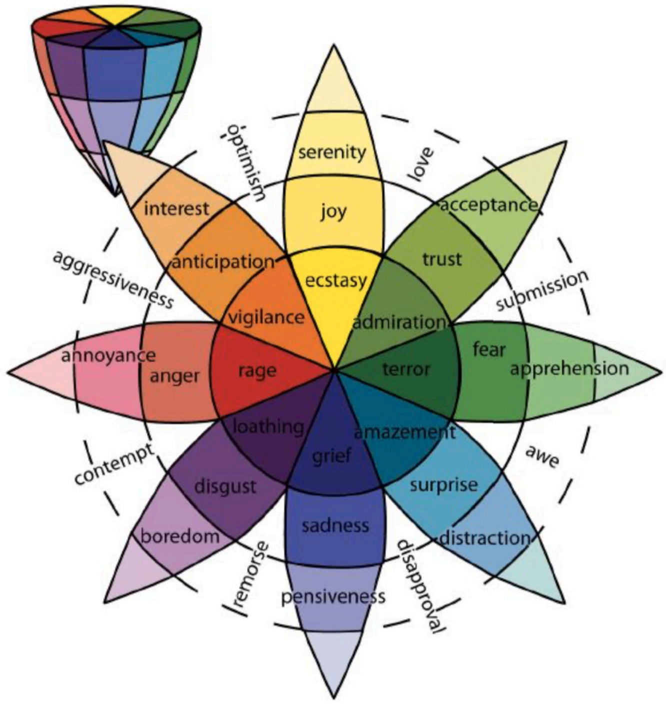

Fig. 1. Multidimensional model of Plutchik - Source: Maupomé and Isyutina, 2013 [35].

not love Obama? he makes jokes about himself." is coded positive. In contrast, the tweet *"Twitter API", "raykolbe", "@morind45 Because the Twitter API is slow and most clients aren't good."* is coded as negative. Due to the importance that the Sentiment140 dataset plays in our study, we will give a more detailed presentation of this dataset in Section 3.

- **categorical**: The sentiment is classified into emotional groups. Researchers have studied emotions for a long time. The first studies date back to Darwin, who studied the expression of emotions on the face and through body gestures. In an attempt to identify a universal evolutionary model, his study of the expression of emotions has not only focused on humans but has also considered animals' emotional expressions and their similarity to humans (see Niedenthal and Brauer, 2012 [31]). In this approach, a numeric polarity score is not the primary determinant of sentiment. Instead, with emotional classification, a document is determined to exhibit an emotional quality, such as anger, disgust, fear, joy, sadness, and surprise. In the 80s, American psychologist Robert Plutchik [32] dedicated himself to the study of emotions, starting from the theories of Paul Eckman [33], who had identified six basic emotions, considered specific to each individual, declensed by age, culture, or other demographic and social factors. Russel's circumplex model (see [34]) represents emotions in a two-dimensional circular space, in which valence is the horizontal axis and excitement (arousal) is the vertical one. Plutchik's multidimensional model (see [32]) extends the complex model by placing emotions on a wheel analogous to the color wheel, in which similar emotions are placed close together. In contrast, opposite emotions are placed 180° apart. In addition to the dimensions identified by Russell, the multidimensional model has a third dimension that represents the intensity of the emotion, giving the model a cone-shaped structure (see Fig. 1). In his model [32], Plutchik states that there are eight basic or primary emotions, each determining behavior, whose combination generates more complex emotions, called secondary emotions. The basic idea is that primary emotions are primitive and innate in the individual at a biological level. In contrast, the so-called secondary emotions are considered an evolution of these. According to Plutchik, primary emotions are defined by four couples: joy - sadness; trust - disgust; anger - fear; surprise - anticipation.
- **sentiment scoring**: It assigns a score to a text. The procedures for quantifying sentiment can rely on continuous or discrete scales between finite endpoints. In the first case, when we choose the endpoints to be zero and one so that the score can be interpreted as a probability. In the second case, we usually measure the score between zero and ten or two opposite endpoints. According to Iacus and Porro (2021) [36], scoring consists of ordering texts along a spectrum of continuous opinions rather than classifying them according to a discrete and finite set of categories. This theory assumes the existence of a latent dimension, which is the fictitious axes on which the text lies. We can order texts of a corpus along the left-right axis.

Scoring techniques can be implemented in both supervised and unsupervised approaches.

Table 1
Sentiment Classification.

Measure	Unsupervised	Supervised
SENTIMENT	A) Dictionary (Bing)	C) Machine Learning (LR, NB, RF, SVM)
SCORING	B) VADER	D) Deep Learning (Word2Vec, LSTM, BERT)

3. Approaches to sentiment classification

Sentiment classification is based on linguistic rules or machine-learning algorithms and is widely used in marketing, online reputation management, and customer feedback analysis. According to Priyavrat (2017) [37], there are three approaches to sentiment classification: 1. Lexicon-based, 2. Machine learning, 3. Hybrid approaches. Fig. 2 summarizes these approaches. The algorithms used in this manuscript are included in parentheses. In lexicon-based approaches, sentiment is quantified based on the semantic orientation of lexicons using a dictionary or corpus (see Tyagi and Sharma, 2017 [38]). A lexicon-based approach utilizes a predefined lexicon with associated specific polarity scores. A lexicon sentiment comprises a list of lexical items generally labeled as positive or negative according to their semantic orientation.

In machine-learning approaches, sentiment is quantified based on a computational method that enables computers to detect patterns in the data and make predictions or classifications without explicit programming, using algorithms to reveal relationships in data sets. Continuously improves accuracy by training on labeled or unlabeled data.

The hybrid approach mixes lexicon-based and machine-learning techniques (see Fang and Chen, 2011 [39]).

Table 1 summarizes the approaches used in this paper to classify consumers' sentiments. We aim to analyze and evaluate how these methods capture sentiment nuances across diverse datasets to improve the forecast.

The advantages of unsupervised classification methods lie in their ability to operate without labeled data; instead, they rely on predefined lexicons or rule-based models to infer sentiment.

Supervised methods rely on labeled datasets to train models that learn from examples, enabling a more accurate sentiment classification. These methods can be broadly categorized into traditional machine-learning and deep-learning models, each providing distinct advantages and limitations. In our study, we used Sentiment140, a widely used dataset for sentiment analysis that consists of tweets labeled positive, negative, or neutral based on emoticons. However, this dataset, which is based on general Twitter content, may not perform well in our domain-specific sentiment analysis.

In A), the lexicon-based classification approach uses Bing's Sentiment Dictionary (see Bing, 2015 [26]). In this predefined dichotomous sentiment dictionary, each word in the text is assigned a positive or negative score. This method is simple and interpretable and does not require

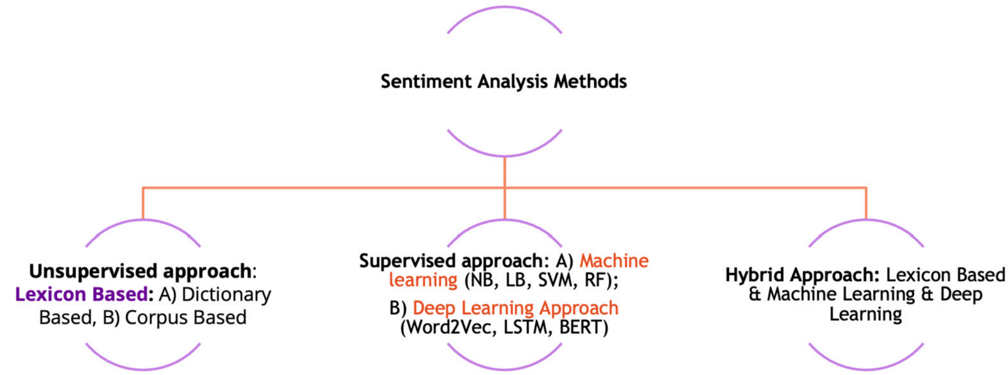

Fig. 2. Sentiment analysis approaches.

training data, making it easy to implement. However, it has limitations, as it cannot handle contextual meaning or detect sarcasm, which impacts its accuracy in sentiment classification.

In B), the rule-based model Valence Aware Dictionary for Sentiment Reasoning (VADER) (see Hutto and Gilbert, 2014 [40]) adopts a model-based approach to sentiment analysis. Unlike simple lexicons, VADER assigns sentiment intensity scores instead of merely binary labels, making it more effective at capturing the strength of sentiment. It is specifically optimized for social networks and short texts, where informal language, emojis, and punctuation are crucial to express sentiment. Although VADER surpasses traditional lexicons by incorporating contextual sentiment intensity, it still faces limitations with nuanced expressions and domain-specific terminology, which can lead to misinterpretations, particularly relevant in the car market. The current literature (see Zhu and Goldberg, 2009 [41]) indicates that semi-supervised methods may be as accurate as a fully supervised method with fewer labeled data to use as input. The technique depends on the type of textual data under analysis: full documents of text clusters are classified differently from social media data, such as Twitter or Facebook posts (or text messages).

In C), we apply four traditional algorithms that use statistical and probabilistic methods for sentiment classification.

- **Logistic Regression (LR)**: A statistical model that predicts sentiment based on weighted input features, making it a widely used baseline for text classification (see Lakshmi et al., 2018 [42]).
- **Naive Bayes (NB)**: A probabilistic model that assumes word independence and computes likelihoods for sentiment classification, often proving effective for short texts such as tweets (see Siswanto et al., 2022 [43]).
- **Random Forest (RF)**: An ensemble method that constructs multiple decision trees and combines their outputs to enhance classification robustness (see Breiman, 2001 [44]).
- **Support Vector Machine (SVM)**: A model that finds the optimal hyperplane to separate sentiment classes, ensuring a well-defined decision boundary (see James et al., 2023 [45]).

These machine-learning approaches offer advantages such as generalizability, interpretability, and efficiency with medium-sized datasets, but they require feature engineering for optimal performance. However, they may struggle to capture deep contextual meanings, limiting their effectiveness in complex sentiment analysis tasks.

In D), deep-learning models offer more advanced sentiment analysis by capturing semantic relationships and context within text data. We apply three sophisticated neural network-based models that employ word embeddings but differ in their generation and usage.

- **Word2Vec**: Uses static embeddings, assigning each word a fixed vector based on its context within a training corpus, with similar words having close numerical representations (see Mikolov et al., 2013 [46]).
- **Long Short-Term Memory (LSTM):** Can utilize pre-trained word embeddings like Word2Vec or GloVe, which can be dynamically updated during training because LSTMs process text sequentially, effectively capturing temporal dependencies, enhancing understanding in longer texts (see Sari et al. 2019 [47]).
- **Bidirectional Encoder Representations from Transformers (BERT):** Utilizes contextual embeddings where a word's vector changes based on surrounding words, leveraging a bidirectional attention mechanism, making it adept at understanding nuanced meanings meaning and semantic disambiguation (see Devil et al., 2019 [48]).

3.1. Preprocessing for text classification

Preprocessing is crucial for improving the performance of text classification. This step cleans and transforms raw text into a format that machine-learning and deep-learning models can easily interpret. In a classification pipeline, effective preprocessing reduces noise and ensures that models can focus on the relevant aspects of the text, improving both accuracy and efficiency.

We used a cleaning procedure composed of six phases for the LR, NB, RF, SVM, Word2Vec, and LSTM algorithms. The lexicon-based approach and BERT do not require these cleaning operations. Fig. 3 shows the sequential phases of cleaning: first, a simple normalization with the extraction of HTML texts, the conversion of words to lowercase, the removal of words with fewer than two characters and more than 15 characters, the removal of stop words and theme words (e.g., Toyota, Camry), the removal of punctuation, and finally the morphological modification with stemming.[3] In sentiment analysis, stemming is useful for improving sentiment classification. We applied simple and morphological normalization during the pre-processing phase to reduce matrix sparsity and limit the number of vocabulary features.

4. Our data

We considered the Toyota Camry US monthly sales volume time series from June 1, 2009, to December 31, 2022.[4] We scraped 163,000 English tweets about the Toyota Camry[5] in the US, selecting a sample of 1,000 tweets for each month from June 1, 2009, to December 31, 2022.

[3] Stemming is a natural language processing (NLP) technique used to reduce words to their base or root form, called the "stem".

[4] Courtesy of GOODCARBADCAR - Automotive Sales Data & Statistics https://www.goodcarbadcar.net/.

[5] Courtesy of Twitter Academic Research Product Track https://developer.twitter.com/en/blog/product-news/2021/enabling-the-future-of-academic-research-with-the-twitter-api.

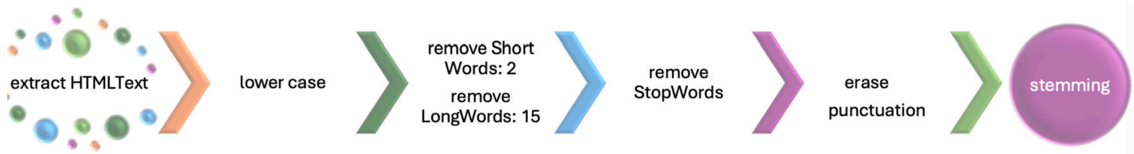

Fig. 3. Pre-processing pipeline.

Fig. 4. Word cloud with the 50 most frequent words.

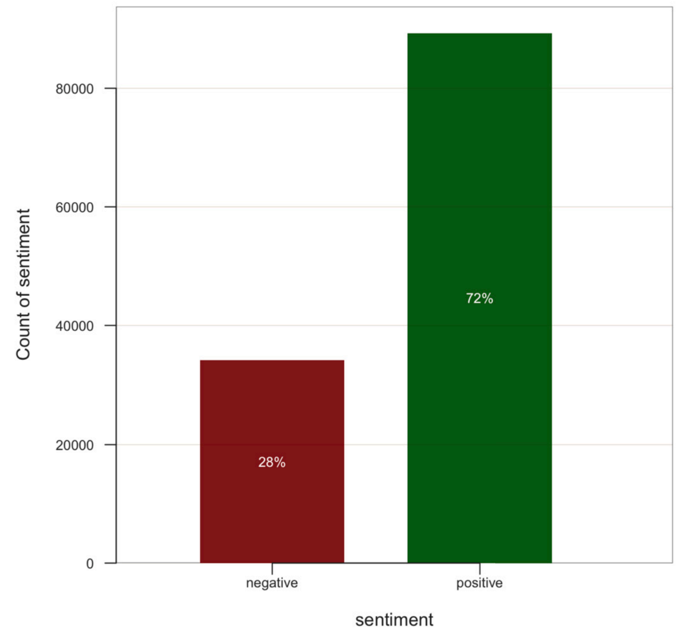

Fig. 5. Sentiment Analysis of corpus using the Bing Dictionary.

The corpus consists of 3,521,852 tokens and 185,345 types. The vocabulary density[6] is 0.053, indicating low lexical variety. Approximately 5.3% of the text's words are unique, while the rest are repetitions. A text with a higher vocabulary density is generally more complex and varied. The readability index[7] (see Wang et al., 2013 [50]) is 16.636. Thus, the tweets have low readability. The score of this index ranges from 1 to 100, where a higher score means greater ease of reading. Scores below 30 indicate that the text is challenging to read and likely requires an academic degree to be understood. Since these are tweets from a specific domain, many technical terms are related to cars and the specific market.

The word average per sentence is 30.9. The most frequent words in the corpus, excluding the theme-specific terms "Toyota" and "Camry", include "car" (23,604 tokens), "used" (22,559), "my" (18,109), "new" (18,003), and "price" (17,520). Fig. 4 presents the fifty most frequent words, which, as shown, mainly focus on descriptions rather than sentiment.

Using the Bing Dictionary, the lexical approach allows the classification of positive and negative words within a list of 6,848 sentiment words. It is aimed at product reviews and natural language for consumers, specifically product reviews, opinions, blog articles, and other subjective texts. Therefore, it is well-suited for analyzing messages from social media. The Bing Dictionary has classified 123,492 sentiment expressions, 34,179 negative words, and 89,313 positive words (see Fig. 5). The words classified as positive account for 72%, compared to 28% negative.

The results from the model-based approach using VADER closely resemble those from the Bing Dictionary, showing a ratio of 1 negative value for every three positive ones.

To implement supervised machine-learning models, we used, for the training, the Sentiment140 dataset comprising 1.6 million positively and negatively labeled tweets (see Go et al., 2009 [29]). We quantified the sentiment of the tweets using R and Matlab software. Among the packages we used, we mention *syuzhet* (see Jockers, 2015 [51]) and *VADER* (see Roehrick, 2020 [52]), regarding to the lexical-based approach, and *quanteda* (see Benoit et al., 2018 [53]), *caret* (see Kuhn, 2008 [54]), e1071 (see Meyer et al. 2023 [55]), *xgboost* (see Davison and Hinkley, 1997 [56]; Canty and Ripley, 2024 [24]), and *rpart* (see Therneau et al., 2023 [57]), for machine learning. We also used the Toolboxes *Text Analytics for Statistics*, *Machine Learning Tools*, and *Deep Learning* (see The MathWorks, 2023 [58], see also Beale et al., 2023 [59]).

5. Our model

Our research aims to explore whether we can extract valuable information from the quantification of sentiment contained in tweets to improve the prediction of Toyota Camry US sales volume time series. Additionally, we intend to compare the effectiveness of several popular methods for sentiment quantification presented in previous sections. To achieve this, we propose applying a state-space model that facilitates a synergistic treatment of the quantitative proxy of consumers' sentiment alongside the sales volume signal. However, we will assert the alleged improvement in predictions compared to benchmark models, which do not incorporate sentiment quantification. Our univariate benchmark models are two ETS models and a MARSS model with univariate observation. ETS models are a well-known broad family of time series models. In particular, we refer to the ETS-ANA and ETS-AAA models, where ANA [resp. AAA] stands for Additive error, No trend, and Additive seasonality [resp. Additive error, Additive trend, and Additive seasonality]. The ETS-AAA model differs from the ETS-ANA model due to an additive slope component in the local trend. More specifically,

[6] Vocabulary density refers to the ratio between the number of unique words and the total number of words in a text.

[7] In this work, we used the Flesch–Kincaid (F-K) formula (see Kincaid et al., 1975 [49]): F = 206.835 - (84.6 * S) - (1.015 * P), where F is readability, S is the average number of syllables per word, and P is the average number of words per sentence.

ETS-ANA

$$\begin{cases} y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t = \ell_{t-1} + \alpha\varepsilon_t, \\ s_t = s_{t-m} + \gamma\varepsilon_t; \end{cases}$$

ETS-AAA

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \\ b_t = b_{t-1} + \beta\varepsilon_t, \\ s_t = s_{t-m} + \gamma\varepsilon_t. \end{cases}$$

Here y_t is the univariate time series value at time t, the hidden variable ℓ_t [resp. b_t, resp. s_t] is the *local level* [resp. *slope*, resp. *seasonality*] of the Holt-Winters decomposition of y_t, and the variable ε_t represents the innovation term with variance σ^2 (see Goodwin, 2012 [60]). The estimation procedure determines the parameters α, β, γ, along with σ^2, and the initial states $y_0, \ell_0, b_0, s_{1-m}, \ldots, s_0$ of the model, where m is the seasonal period. The reason for choosing our ETS benchmarks is twofold. First, the exponential smoothing feature of the considered ETS makes them well suited for handling time series in which the trend and the seasonality evolve, underweighting older data compared to more recent data. This characteristic is particularly relevant for consumer goods (see Hyndman and Billah, 2003 [61], Hyndman and Athanasopoulos 2014 [62], Makridakis et al., 2018 [63]). Second, our idea is to generalize the considered ETS-AAA model in the context of state-space models by attributing the role of quantitative representation of consumers' sentiments to the slope variable b_t. On the other hand, we cannot expect to observe this variable, which we treat as a hidden variable. Instead, we assume that we observe a proxy differently estimated by each lexicon, machine-learning, and deep-learning procedures for sentiment quantification. Consequently, we have applied a multivariate state-space model in which we link the unobserved states and the observations by the respective equations MARSS bivariate

$$\begin{cases} \ell_t = \beta_{\ell,\ell}\ell_{t-1} + \beta_{\ell,b}b_{t-1} + \sigma_{\ell,\ell}w_t^{(l)}, \\ b_t = \beta_{b,b}b_{t-1} + \sigma_{b,b}w_t^{(b)}, \\ s_t = s_{t-m} + \sigma_{s,s}w_t^{(s)}. \end{cases}$$

$$\begin{cases} y_t = \beta_{y,\ell}\ell_t + \beta_{y,b}b_t + \beta_{y,s}s_t + \sigma_{y,y}w_t^{(y)}, \\ z_t = \beta_{z,b}b_t + \sigma_{z,z}w_t^{(z)}. \end{cases}$$

For us, y_t is the observed sales volume time series or, possibly, a Box-Cox transformation of the sales volume, and z_t is the estimated proxy of the hidden representative variable of consumers' sentiments b_t. To add more flexibility to our model, we have introduced the additional parameters $\beta_{\ell,\ell}, \ldots, \beta_{z,b}$. Moreover, we have introduced independent innovations $w_t^{(l)}, \ldots, w_t^{(z)}$, with variances $\sigma^2_{\ell,\ell} \ldots, \sigma^2_{z,z}$, respectively. Our univariate benchmark MARSS model differs from the above-presented multivariate model for suppressing the observed variable z_t. We implemented it to contribute to assessing whether consumers' sentiments can effectively improve the accuracy of sales forecasts. The estimation procedure allows us to determine all parameters $\beta_{\ell,\ell}, \ldots, \beta_{z,b}$ and $\sigma^2_{\ell,\ell} \ldots, \sigma^2_{z,z}$. However, we encountered difficulties when trying to estimate the initial state $y_0, \ell_0, b_0, s_{1-m}, \ldots, s_0$. Given our observed time series, it turns out that the *MARSS() R* function has limited capabilities for estimating the initial state, and the time series themselves are likely not long enough for an effective application of the *diffuse* option. On the other hand, a good choice of the initial state is crucial for an accurate estimation. So far in our studies, we adopted two techniques to circumvent the "curse" of the initial state: a "brute force" approach with a random choice of the initial state and a "perturbative" approach, which assumes as initial state a random perturbation of the initial state estimated by the benchmark ETS-AAA model. In particular, the results presented in this paper come from the brute force approach. To avoid overfitting, we obtained the estimates of the model by an in-sample set of nearly the 90% of the observed time series, and we tested the accuracy of our estimates by the remaining out-of-sample set, which covers 17 months. We acknowledge that the "brute-force" method employed for initial state estimation in our MARSS model is computationally intensive and may lead to the risk of converging to local minima, resulting in systematic errors in the model's parameter estimates. However, it is important to emphasize that the impact of initial state estimation is primarily significant in the short run. The Kalman filter's updating mechanism used in state-space models allows for the continuous refinement of predictions as new data become available. Ultimately, this will lead to accurate estimates in the long term. This characteristic enhances the robustness of state-space models, enabling them to adapt to exogenous shocks and changing conditions, which is particularly relevant in forecasting applications such as the automotive market. For this reason, we believe that state-space models have greater capabilities than other methodologies to handle future scenarios that are not necessarily "surprise-free." Eventually, since the early 1980s (see, e.g., McGee and Schmidt, 1985 [64]), the Kalman filter has been widely employed across numerous industries, including automotive, ship navigation, aerospace, and robotics, for its resilience in guidance and control systems, effectively addressing challenges posed by exogenous sudden shocks. The real challenge comes from the continuous availability of new data. While a brute-force approach may not be suitable for all forecasting situations, especially those requiring real-time analysis or scalability, it can still be a helpful starting point when available data are insufficient to estimate the initial state or apply a diffuse approach. Afterward, the Kalman filter mechanism allows for ongoing adjustments of estimates as new data becomes available. As a final note, the reason for considering the ETS-ANA benchmark model is that it has been selected as the best ETS model to fit the Toyota Camry US time series based on information criteria. Not surprisingly, the estimation of the ETS-ANA benchmark model, which shows a good fit, suggests a small size for the slope signal b_t, as it is estimated. This is why the significant difference in size between the sales volume signal and the proxy of consumers' sentiments led us to interpret the latter as the slope in our MARSS model.

6. Results and discussion

Table 2 summarizes the values of key information criteria and accuracy measures for Toyota Camry US monthly sales volumes, (y), without and with the estimated proxy of consumers' sentiments (z).

The information criteria AIC, BIC, and AICc[8] lead us to prefer the MARSS univariate benchmark over the ETS-ANA and ETS-AAA models. This preference is particularly interesting due to the larger number of parameters in the former model, which penalizes it in terms of the information criteria. These information criteria are somewhat redundant when comparing bivariate MARSS models with each other, as these models share the same structure and parameters. Consequently, all the information criteria used reflect the model's log-likelihood. Furthermore, we should be cautious when comparing univariate benchmarks with bivariate MARSS models because of significant differences in their structures and the number of parameters involved. Nevertheless, Regarding information criteria, it is interesting to note that the MARSS model utilizing the SVM estimated proxy of consumers' sentiments outperforms its bivariate competitors and all univariate benchmarks.

The accuracy measures[9] we present here constitute a selected subset of the many currently used in the literature, aiming to provide a comprehensive assessment of model performance while avoiding redundancy. MAE is a fundamental metric that quantifies the average magnitude of errors in the model predictions, providing a straightforward interpretation. RMSE retains the simplicity of MAE while overweighting larger errors and underweighting smaller ones, which is particularly useful in contexts where larger deviations are of greater concern, especially in economically significant forecasts. SMAPE offers valuable information on the forecast accuracy expressed as a percentage, which is easy to interpret in different contexts. Considering the actual and forecast values

[8] AIC = Akaike Information Criterion, BIC = Bayes Information Criterion, AICc = Akaike Information Criterion corrected.

[9] MAE = Mean Absolute Error, RMSE = Root Mean Square Error, MAPE = Mean Absolute Percentage Error, SMAPE = Symmetric Mean Absolute Percentage Error, MASE = Mean Absolute Scaled Error, RMSSE = Root Mean Square Error.

Table 2
Information parameters and accuracy measures for Toyota Camry US monthly sales volumes: ETS-ANA model, ETS-AAA model, MARSS univariate, and MARSS bi-variate with VADER, LR, NB, RF, SVM, Word2Vec, LSTM, and BERT estimated proxy of consumers' sentiments.

Models	AIC	BIC	AICc	MAE	RMSE	SMAPE%	MASE	RMSSE
ETS-ANA (y)	3212.869	3257.623	3216.561	8610.602	9440.055	15.902	1.669	1.389
ETS-AAA (y)	3218.252	3268.973	3223.033	10241.120	10809.330	1.985	18.321	1.590
MARSS (y) - univ.	3154.555	3184.319	3156.185	6087.463	7291.802	11.831	1.180	1.073
MARSS (y)	3273.601	3309.405	3275.947	2539.489	2975.810	5.586	0.492	0.438
MARSS (z) - BING est.	3273.601	3273.601	3275.947	0.070	0.089	27.590	0.614	0.588
MARSS (y)	3148.469	3184.272	3150.814	3608.202	4599.050	7.811	0.699	0.676
MARSS (z) - VADER est.	3148.469	3184.272	3150.814	0.018	0.026	24.169	1.156	0.971
MARSS (y)	3621.910	3653.713	3624.255	3119.797	3947.773	6.925	0.605	0.580
MARSS (z) - LR est.	3621.910	3653.713	3624.255	0.062	0.118	5.319	0.818	0.990
MARSS (y)	3687.089	3722.892	3689.435	3028.798	3824.414	6.659	0.587	0.563
MARSS (z) - NB est.	3687.089	3722.892	3689.435	0.087	0.112	9.360	0.875	0.877
MARSS (y)	3706.703	3742.506	3709.049	3361.706	4061.289	7.359	0.605	0.580
MARSS (z) - RF est.	3706.703	3742.506	3709.049	0.218	0.231	30.879	1.198	1.596
MARSS (y)	3032.004	3067.807	3034.350	2334.179	2887.931	5.211	0.452	0.425
MARSS (z) - SVM est.	3032.004	3067.807	3034.350	0.104	0.123	18.152	0.947	0.851
MARSS (y)	3629.910	3665.713	3631.028	3066.996	3956.932	6.795	0.594	0.582
MARSS (z) - Word2Vec est.	3629.910	3665.713	3631.020	0.037	0.076	3.291	0.995	0.991
MARSS (y)	3486.898	3522.701	3489.244	2987.291	3826.338	6.614	0.579	0.563
MARSS (z) - LSTM est.	3486.898	3522.701	3489.244	0.072	0.096	5.685	0.797	0.822
MARSS (y)	3646.188	3681.991	3648.534	3002.287	3662.178	6.633	0.579	0.582
MARSS (z) - BERT est.	3646.188	3681.991	3648.534	0.103	0.146	23.029	0.909	0.977

in its calculation, SMAPE is a balanced measure. MASE allows for relatively evaluating forecasting performance by comparing the model's predictions against the random-walk naive benchmark. Lower MASE values indicate better predictive performance compared to this naive benchmark, which is crucial when assessing the efficacy of our model against simple predictive strategies. Lastly, RMSSE stands in relation to MASE similarly to RMSE's to MAE, overweighting larger prediction errors and underweighting smaller ones against the naive benchmark.

In light of the accuracy measures applied, the best performance of the MARSS univariate benchmark is confirmed against the other benchmarks. However, it is overshadowed by the MASE and RMSSE values above unity. In contrast, all bivariate MARSS models outperform univariate benchmarks. Notably, the LSTM-estimated proxy of consumers' sentiments demonstrates good performance, whereas the VADER and RF estimated proxies show somewhat unsatisfactory results.

Fig. 6 highlights how the sentiment signal generated by the LSTM appears to be effective in improving sales forecasting. We have achieved the best results using the LSTM, but the BING and SVM models have also performed well.

Fig. 7 shows the monthly time series of Sales Volumes and negative tweets analyzed using BING, SVM, and LSTM. Visual examination reveals that significant deviations from the average in the three algorithms are predominantly negative and occur during 09/2019, 05/2021, and 07/2022. In particular, the signals produced by the SVM and LSTM models exhibit some similarity. The Bing dictionary-based approach provides us with a pretty accurate sentiment proxy because a simple strategy that classifies tweets unsupervised but relies on a list of sentiment words specifically used in a social media context may be valid. In this case, it was not necessary to use a training set. In fact, due to the investigation domain, we would have needed a training set of tweets focused on car sales. The results of the corpus analysis revealed a complex and highly specialized text. SVM has demonstrated strong performance among machine-learning algorithms, mainly due to its capability to handle high-dimensional data, such as text. As a result, SVM is both efficient and versatile, excelling especially in contexts with complex and high-dimensional datasets. LSTM models have effectively captured long-term dependencies and managed text sequences, making them well-suited for text classification and providing the most accurate representation of consumers' sentiments. The superior performance of LSTM is likely due to its design, which explicitly addresses long-term dependencies in data sequences. This capability is essential in textual analysis, where the meaning of a word or phrase can depend on distant contextual information. In contrast, traditional machine-learning models rely on a vocabulary derived from the training set and use vector representations, which do not account for the sequential nature of text. Consequently, they produce very sparse term-document matrices. Although techniques like morphological normalization and the removal of low-frequency words can reduce sparsity, they still leave behind background noise that can affect the quality of the analysis. An idea to further improve the quality of the estimated proxy of consumers' sentiments could be to experiment with other deep-learning models like BERT and manually label a small training set from a corpus belonging to the same social media and domain.

7. Conclusions and future steps

This paper presents a case study on Toyota Camry sales volume. Although this study has the standard limitations of the case study, we are optimistic regarding the possibility of its generalization, as the adopted methodology sounds effective. For instance, the proposed approach is easily adaptable to other automotive brands and industries, allowing for a broader range of market analyses. Indeed, we have also partially applied this methodology to FIAT data in previous experiments, illustrating its flexibility and potential for a broader implementation (see Iezzi and Monte, 2024 [65]). Our research suggests that the bivariate MARSS model for Toyota Camry monthly sales volume, which incorporates an estimated proxy of consumer sentiment, significantly outperforms the univariate benchmark models that do not account for this proxy. Furthermore, the univariate benchmark MARSS model performs better than other ETS benchmarks. We present various approaches for estimating the consumer sentiment proxy due to the lack of consensus on the best method for extracting a quantitative sentiment signal from textual data. Our exploration and previous studies (see Iezzi and Monte, 2024 [65]) indicate that specific methods, such as VADER and Random Forest, may be less effective. In contrast, methods such as LSTM, SVM, LR, and,

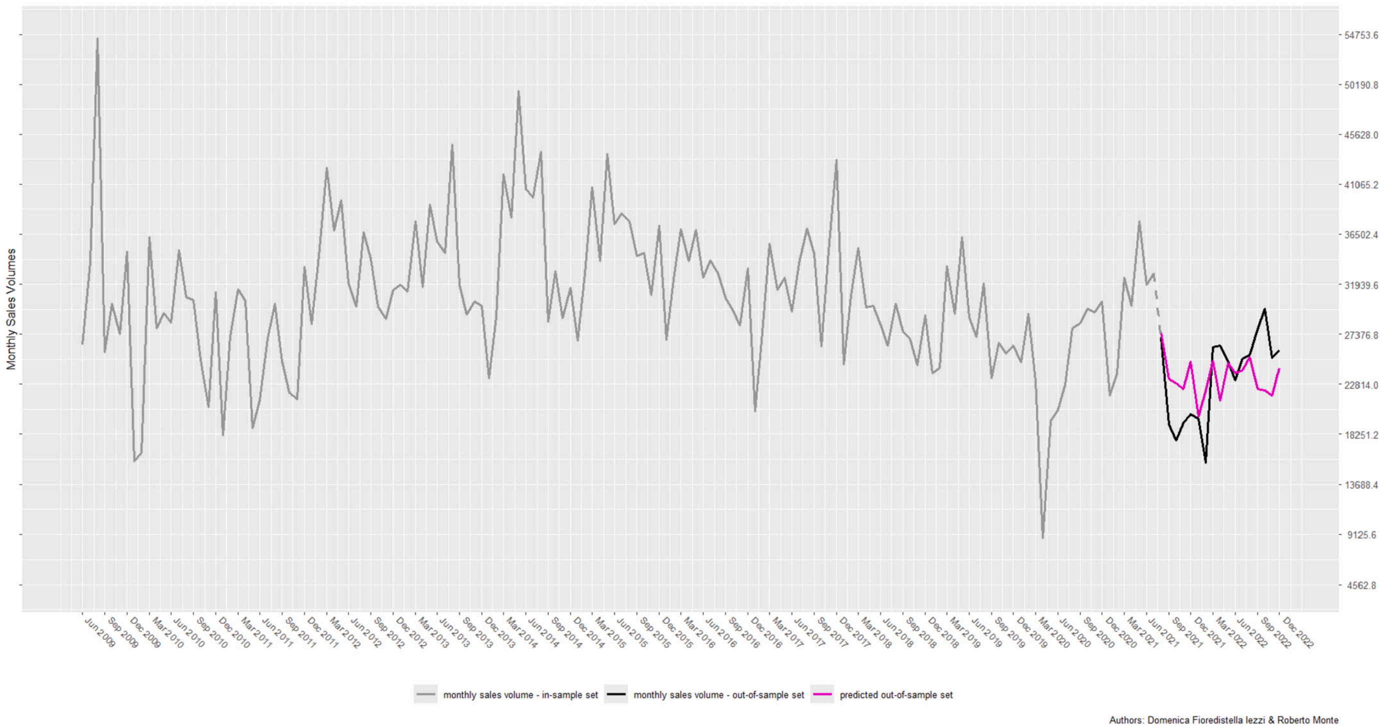

Fig. 6. MARSS bivariate model - Toyota Camry US monthly sales volume and negative sentiment percentage LSTM estimated proxy - predicted sales volume vs out-of-sample set.

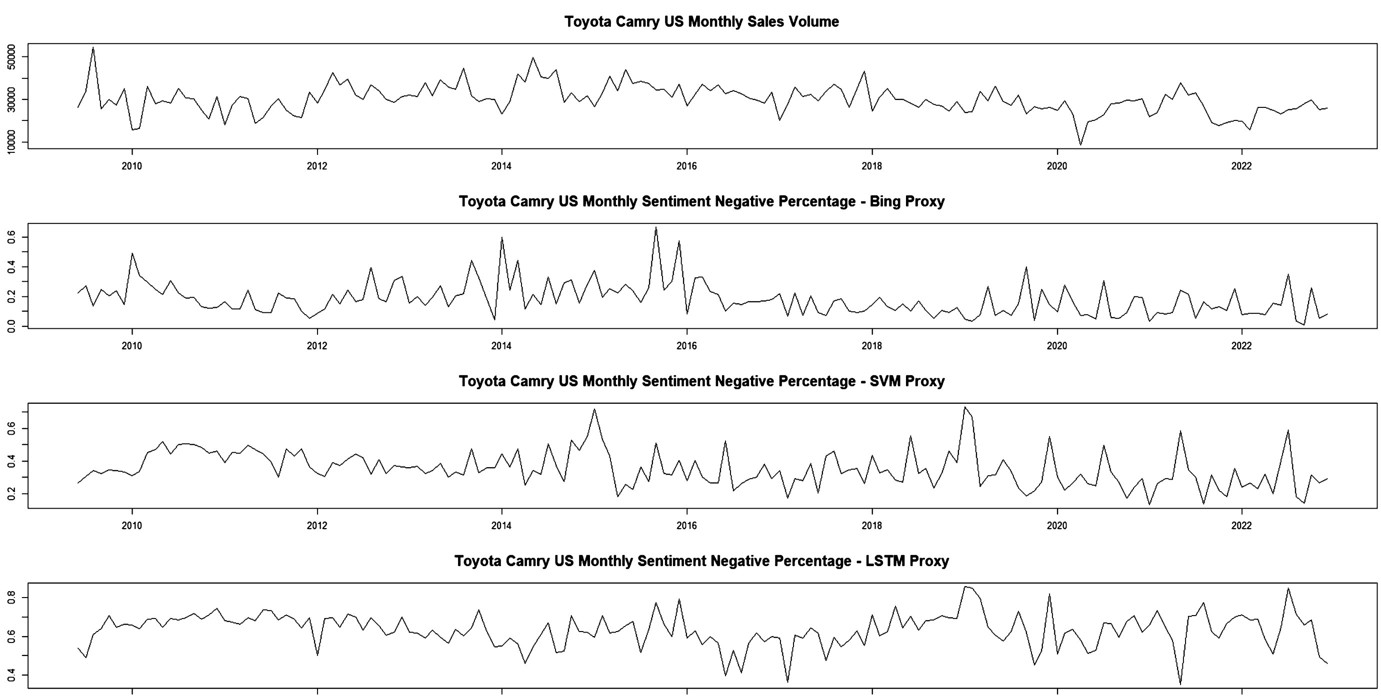

Fig. 7. Monthly time series for Sales Volume and the percentage of negative tweets are analyzed using BING, SVM, and LSTM.

somewhat unexpectedly, Bing appear to perform better. However, further investigations are necessary to draw definitive conclusions.

Another crucial issue is the optimal synchronization between the consumer sentiment proxy and sales volume dynamics. Specifically, identifying the optimal delay by which consumer sentiment influences sales is of utmost importance.

Finally, we should also consider the robustness of the model in light of the nontrivial choice of the initial state combined with the arrival of sudden large shocks shortly, such as the COVID-19 pandemic that emerged in early 2020, after ten years and seven months of monthly data collection. We believe that the estimation update mechanism of the Kalman filter used in the state-space model should be robust enough to handle such shocks. However, our belief needs to be supported by empirical data.

We are working on these issues and aim to address them in a forthcoming paper.

CRediT authorship contribution statement

Domenica Fioredistella Iezzi: Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Roberto Monte:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] G. Katona, Psychological analysis of business decisions and expectations, American Economic Review 36 (1) (1946) 44–62.
[2] G. Katona, The Mass Consumption Society, McGraw-Hill, New York, 1964.
[3] J.S. Mill, On the Definition of Political Economy, and of the Method of Investigation Proper to It: Essays on Some Unsettled Questions of Political Economy, Routledge/Thoemmes, London, 1997.
[4] P. Vana, A. Lambrecht, Online reviews: star ratings, position effects and purchase likelihood, Marketing Science 55 (6) (2021) 708–730, https://doi.org/10.2139/ssrn.3108086.
[5] E. Katz, P.F. Lazarsfeld, Personal Influence: the Part Played by People in the Flow of Mass Communications, Free Press, London, 1955.
[6] K.P. Adler, Elihu Katz, Paul F. Lazarsfeld, Personal influence: the part played by people in the flow of mass communication, The Annals of the American Academy of Political and Social Science 309 (1957) 176–177, https://doi.org/10.2139/ssrn.3108086.
[7] J.J. Brown, P.H. Reingen, Social ties and word-of-mouth referral behavior, Journal of Consumer Research 14 (3) (1987) 350–362, https://doi.org/10.1086/209118.
[8] M.R. Jalilvand, S.S. Esfahani, N. Samiei, Electronic word-of-mouth: challenges and opportunities, Procedia Computer Science 3 (2011) 42–46, https://doi.org/10.1016/j.procs.2010.12.008, https://www.sciencedirect.com/science/article/pii/S1877050910003832.
[9] D.H. Park, J. Lee, I. Han, The effect of on-line consumer reviews on consumer purchasing intention: the moderating role of involvement, International Journal of Electronic Commerce 11 (4) (2007) 125–148, https://doi.org/10.2753/JEC1086-4415110405.
[10] P. Gupta, J. Harris, How e-wom recommendations influence product consideration and quality of choice: a motivation to process information perspective, in: Advances in Internet Consumer Behavior & Marketing Strategy, Journal of Business Research 63 (9) (2010) 1041–1049, https://doi.org/10.1016/j.jbusres.2009.01.015, https://www.sciencedirect.com/science/article/pii/S0148296309001969.
[11] D. Kahneman, A. Tversky, Prospect theory: an analysis of decision under risk, Econometrica 47 (2) (1979) 263–291, https://EconPapers.repec.org/RePEc:ecm:emetrp:v:47:y:1979:i:2:p:263-91.
[12] S. Fiske, Attention and weight in person perception: the impact of negative and extreme behavior, Journal of Personality and Social Psychology 38 (6) (1980) 889–906, https://doi.org/10.1037/0022-3514.38.6.889.
[13] D.E. Kanouse, L.R. Hanson, Negativity in Evaluations, General Learning Press, Morristown, NJ, 1972, https://books.google.it/books?id=sTjFwgEACAAJ.
[14] B.E. Meyerowitz, S. Chaiken, The effect of message framing on breast self-examination attitudes, intentions, and behavior, Journal of Personality and Social Psychology 52 (3) (1987) 500–510.
[15] S. Hussain, W. Ahmed, R. Jafar, A. Rabnawaz, Y. Jianzhou, ewom source credibility, perceived risk and food product customer's information adoption, Computers in Human Behavior 66 (2017) 96–102, https://doi.org/10.1016/j.chb.2016.09.034.
[16] Y. Liu, X. Yu, X. Huang, A. An, Blog Data Mining: The Predictive Power of Sentiments, Springer US, Boston, MA, 2009, pp. 183–195.
[17] X. Yu, Y. Liu, X. Huang, A. An, Mining online reviews for predicting sales performance: a case study in the movie domain, IEEE Transactions on Knowledge and Data Engineering 24 (4) (2012) 720–734, https://doi.org/10.1109/TKDE.2010.269.
[18] S. Magdum, J. Megha, Mining online reviews and tweets for predicting sales performance and success of movies, in: International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 334–339.
[19] C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting 20 (1) (2004) 5–10, https://doi.org/10.1016/j.ijforecast.2003.09.015, https://www.sciencedirect.com/science/article/pii/S0169207003001134.
[20] P. Winters, Forecasting sales by exponentially weighted moving averages, Management Science 6 (3) (1960) 324–342.
[21] R. Hyndman, A. Koehler, R. Snyder, S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting 18 (2002) 439–454, https://doi.org/10.1016/S0169-2070(01)00110-8.
[22] E.E. Holmes, E.J. Ward, M.D. Scheuerell, K. Wills, MARSS: multivariate autoregressive state-space modeling, R package version 3.11.9, https://CRAN.R-project.org/package=MARSS, 2024.
[23] M.D. Devika, C. Sunitha, A. Ganesh, Sentiment analysis: a comparative study on different approaches, in: Fourth International Conference on Recent Trends in Computer Science & Engineering (ICRTCSE 2016), Procedia Computer Science 87 (2016) 44–49, https://doi.org/10.1016/j.procs.2016.05.124, https://www.sciencedirect.com/science/article/pii/S187705091630463X.
[24] J. Carvalho, A. Prado, A. Plastino, A statistical and evolutionary approach to sentiment analysis, in: 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), vol. 2, 2014, pp. 110–117.
[25] A. Vishal, S.S. Sonawane, Sentiment analysis of Twitter data: a survey of techniques, International Journal of Computer Applications 139 (11) (2016) 5–15, https://doi.org/10.5120/ijca2016908625.
[26] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, Cambridge, 2015.
[27] A. Righi, E. Catanese, L. Valentino, D. Zardetto, The Italian social mood on economy index during the covid-19 crisis, in: S.M.N. Salvati, C. Perna, R. Chambers (Eds.), Studies in Theoretical and Applied Statistics, Springer International Publishing, Cham, 2022, pp. 475–485.
[28] D.F. Iezzi, R. Monte, Building Deep-Learning Consumers' Sentiment Signals for Sales Forecasting: A Comparative Study, Presses Universitaires de Louvain, 2024, pp. 419–428.
[29] A. Go, A.R. Bhayani, L. Huang, Twitter sentiment classification using distant supervision, Processing 150 (2009) 1–12.
[30] V. Pandya, A. Somthankar, S.S. Shrivastava, M. Patil, Twitter sentiment analysis using machine learning and deep learning techniques, in: 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4), 2021, pp. 1–5.
[31] P. Niedenthal, M. Brauer, Social functionality of human emotion, Annual Review of Psychology 63 (2012) 259–285, https://doi.org/10.1146/annurev.psych.121208.131605.
[32] R. Plutchik, A psychoevolutionary theory of emotions, Social Science Information 21 (4–5) (1982) 529–553, https://doi.org/10.1177/053901882021004003.
[33] P. Ekman, Universals and cultural differences in facial expressions of emotion, Nebraska Symposium on Motivation 19 (1971) 207–283.
[34] J. Russell, A circumplex model of affect, Journal of Personality and Social Psychology 39 (1980) 1161–1178, https://doi.org/10.1037/h0077714.
[35] G. Maupomé, O. Isyutina, Dental students' and faculty members' concepts and emotions associated with a caries risk assessment program, Journal of Dental Education 77 (11) (2013) 1477–1487, https://doi.org/10.1002/j.0022-0337.2013.77.11.tb05624.x.
[36] S. Iacus, G. Porro, Subjective Well-Being and Social Media: Reconciling Big Data and Statistics, Taylor & Francis, 2021.
[37] P. Chauhan, Sentiment analysis: a comparative study of supervised machine learning algorithms using rapidminer, International Journal for Research in Applied Science and Engineering Technology V (2017) 80–89, https://doi.org/10.22214/ijraset.2017.11011.
[38] E. Tyagi, D. Sharma, Sentiment analysis of product reviews using support vector machine learning algorithm, Indian Journal of Information Science and Technology 10 (2017) 1–9, https://doi.org/10.17485/ijst/2017/v10i35/118965.
[39] J. Fang, B. Chen, Incorporating lexicon knowledge into svm learning to improve sentiment classification, in: Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP 2011), 2011, pp. 94–100.
[40] C. Hutto, E. Gilbert, Vader: a parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the International AAAI Conference on Web and Social Media, 2014, pp. 94–100.
[41] X. Zhu, A.B. Goldberg, Introduction to semi-supervised learning, Synthesis Lectures on Artificial Intelligence and Machine Learning 3 (1) (2009) 1–130.
[42] R. Lakshmi, R.B. Divya, S. Valarmathi, Analysis of sentiment in Twitter using logistic regression, International Journal of Engineering and Technology 7 (2018) 619–662, https://doi.org/10.30812/varian.v6i1.2205.
[43] S. Siswanto, Z. Mariah, A. Sabir, T. Hidayat, F. Adhel, W. Amni, The sentiment analysis using Naïve Bayes with lexicon-based feature on tiktok application, Jurnal Varian 1 (2022) 89–96, https://doi.org/10.30812/varian.v6i1.2205.
[44] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[45] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in Python, Springer, 2023.
[46] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in Neural Information Processing Systems 26 (2013).
[47] W.K. Sari, D.P. Rini, R.F. Malik, Text classification using long short-term memory, in: 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), 2019, pp. 150–155, https://api.semanticscholar.org/CorpusID:211057709.

[48] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding, cite, arXiv:1810.04805, 2018, http://arxiv.org/abs/1810.04805.
[49] J.P. Kincaid, R.P. Fishburne, R.L. Rogers, B.S. Chissom, Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel, Tech. Rep., Chief of Naval Technical Training, 1975.
[50] L. Wang, M.J. Miller, M.R. Schmitt, F.K. Wen, Assessing readability formula differences with written health information materials: application, results, and recommendations, Research in Social and Administrative Pharmacy 9 (5) (2013) 503–516, https://doi.org/10.1016/j.saph.
[51] M.L. Jockers, Syuzhet: extract sentiment and plot arcs from text, https://github.com/mjockers/syuzhet, 2015.
[52] K. Roehrick, vader: valence aware dictionary and sEntiment reasoner (VADER), R package version 0.2.1, https://CRAN.R-project.org/package=vader, 2020.
[53] K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, A. Matsuo, quanteda: an R package for the quantitative analysis of textual data, Journal of Open Source Software 3 (30) (2018) 774, https://doi.org/10.21105/joss.00774.
[54] M. Kuhn, Building predictive models in R using the caret package, Journal of Statistical Software 28 (5) (2008) 1–26, https://doi.org/10.18637/jss.v028.i05, https://www.jstatsoft.org/index.php/jss/article/view/v028i05.
[55] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien, R package version 1.7-14, https://CRAN.R-project.org/package=e1071, 2023.
[56] A.C. Davison, D.V. Hinkley, Bootstrap Methods and Their Applications, Cambridge University Press, Cambridge, 1997.
[57] T. Therneau, B. Atkinson, B. Ripley, rpart: recursive partitioning and regression trees, R package version 4.1-19, https://CRAN.R-project.org/package=rpart, 2023.
[58] The MathWorks, Inc., Text Analytics Toolbox: For Use with MATLAB, The MathWorks, Inc., Natick, Massachusetts, 2023, https://www.mathworks.com/products/text-analytics.html.
[59] M.H. Beale, M.T. Hagan, H.B. Demuth, Deep Learning Toolbox™: For Use with MATLAB, The MathWorks, Inc., Natick, Massachusetts, 2023, https://www.mathworks.com/products/deep-learning.html.
[60] P. Goodwin, The Holt-winters approach to exponential smoothing: 50 years old and going strong, Foresight: The International Journal of Applied Forecasting 19 (2010) 30–33, https://ideas.repec.org/a/for/ijafaa/y2010i19p30-33.html.
[61] R.J. Hyndman, B. Billah, Unmasking the theta method, International Journal of Forecasting 19 (2) (2003) 287–290.
[62] R.J. Hyndman, G. Athanasopoulos, Optimally reconciling forecasts in a hierarchy, Foresight: The International Journal of Applied Forecasting 35 (2014).
[63] S. Makridakis, E. Spiliotis, V. Assimakopoulos, The m4 competition: results, findings, conclusion and way forward, International Journal of Forecasting 34 (4) (2018) 802–808.
[64] L.A. McGee, S.F. Schmidt, Discovery of the Kalman filter as a practical tool for aerospace and industry, Tech. Rep., NASA Technical Memorandum 86847, retrieved March 26, 2016, November 1985.
[65] D.F. Iezzi, R. Monte, Social media effects on sales: consumer sentiment in a state-space model, in: G. Giordano, M. Misuraca (Eds.), New Frontiers in Textual Data Analysis, Springer Nature, Switzerland, Cham, 2024, pp. 77–90.