Decision Analytics Journal 3 (2022) 100033

Contents lists available at ScienceDirect

Decision Analytics Journal

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

Cluster-based demand forecasting using Bayesian model averaging: An ensemble learning approach

Mahya Seyedan [a], Fereshteh Mafakheri [b,*], Chun Wang [a]

[a] Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Canada
[b] École nationale d'administration publique (ENAP), Université du Québec, Montréal, Canada

ARTICLE INFO

Keywords:
Demand forecasting
Customer segmentation
Multivariate time-series forecasting
Ensemble learning

ABSTRACT

Demand forecasting is an important aspect in supply chain management that could contribute to enhancing the profit and increasing the efficiency by aligning the supply channels with anticipated demand. In the retail industry, customers and their needs are diverse making demand forecasting a challenging task. In this regard, this study aims at developing a three-step data-driven cluster-based demand forecasting approach for the retail industry. First, customers are segmented based on their recency, frequency, and monetary (RFM) characteristics. Customers with similar buying behaviors are recognized as a segment, creating an ordered relationship between transactions made by them. In the second step, time-series analysis techniques are used to forecast demand for each customer segment. Finally, Bayesian model averaging (BMA) is adopted to ensemble the forecasting results obtained from alternative time series techniques. The applicability of the proposed approach is presented through a comparative case study analysis with presented improvement in the accuracy of daily demand prediction.

1. Introduction

Retail industry has to manage demand and supply planning processes at the operational level while dealing with demand fluctuations as well as uncertainties arising in purchase planning, distribution channels, availability of labor force, and demand for after-sales services [1–3]. Demand forecasting refers to an organization's demand estimation process [2] to support production, service, and transportation plans [4], cost-effective inventory management [5], control the safety stock [6], and consequently lowering supply chain costs. Demand forecasting has attracted a lot of attention in the retail industry. [3]. A reliable demand forecasting model can help retailers increase profit, promote products, and prevent shortages [6]. Furthermore, an accurate forecast consequently helps with developing an adaptive pricing strategy for improved revenue management. Time-series forecasting, clustering, K-nearest-neighbors (KNN), neural networks (NN), regression analysis (RA), decision tree (DT), support vector machines (SVM), and support vector regression (SVR) are common methods used to forecast demand [7–9].

In recent years, the emergence of online shopping and e-commerce has created a rich source of data on customer information and preferences, including customer demographics (postal code, date of birth, education/income status), past spending patterns, and social media activities (likes and dislikes) [10]. It is now possible for sellers to use these customer characteristics together with their purchase information to make accurate forecasts of the customers' needs and buying habits. One widely used customer information-based forecasting technique is clustering-based forecasting [11,12].

Clustering-based forecasting involves separating customers into disjoint cluster with maximum within-cluster similarity and maximum intra-cluster dissimilarity and constructing a forecasting model on top of each cluster. Due to the inter-cluster similarity, each cluster's prediction models perform better than using the complete dataset to build one prediction model. Each customer is assigned to a cluster, and the forecasting model corresponding to that cluster is used to obtain forecasting outcomes. Various factors affect the performance of the clustering-based approach, including the choice of clustering technique, the similarity measurement used, and the predictor [13]. In the literature, self-organizing map (SOM), growing hierarchical self-organizing map (GHSOM), K-means clustering approach as well as various linkage methods have been used to cluster data [13–15].

In this regard, depending on the nature of the dataset, various machine learning and statistical tools can be employed for forecasting. In other words, there is no "one fits all" solution. Time-series forecasting is mostly used when the data has an ordered relationship. For example, customer transactions are considered time-series data. Therefore, the forecasts depend on customers' previous purchase patterns [5,16]. Machine learning and deep learning methods are mostly generating more accurate forecasts for large time-series data. However,

* Corresponding author.
E-mail addresses: mahya.seyedan@concordia.ca (M. Seyedan), Fereshteh.Mafakheri@enap.ca (F. Mafakheri), chun.wang@concordia.ca (C. Wang).

https://doi.org/10.1016/j.dajour.2022.100033
Received 24 November 2021; Received in revised form 9 February 2022; Accepted 2 March 2022
Available online 7 March 2022

Nomenclature

ANFIS	Adaptive neuro-fuzzy inference system
ARCH	Autoregressive conditional heteroscedastic
ARIMA	Auto-regressive integrated moving average
ANN	Artificial neural network
BMA	Bayesian model averaging
DT	Decision tree
GARCH	General autoregressive conditional heteroscedastic
GHSOM	Growing hierarchical self-organizing map
KNN	K-nearest neighbor
LOO	Leave-One-Out cross-validation
LSTM	Long short-term memory
LR	Logistic regression
MAE	Mean absolute error
MAPE	Mean absolute percentage error
NN	Neural networks
RFM	Recency, frequency and monetary
RMSE	Root mean square error
RNNs	Recurrent neural networks
SOM	Self-organizing map
SVM	Support vector machine
SVR	Support vector regression
WAIC	Widely applicable information criterion

in some forecasting problems, classical methods (such as seasonal autoregressive integrated moving average (ARIMA) and exponential smoothing) would outperform especially in case of one-step forecasting problems with univariate datasets [16,17]. Therefore, it is important to understand how traditional time-series forecasting methods work and evaluate them before exploring more data-intensive techniques.

The literature suggests using ensemble learning to combine forecasting results as a means of achieving higher accuracy [18–21]. In the ensemble learning approach, different predictors (each with different performance) are combined to arrive at predictions. The predictors are combined such that each model covers the weakness(es) of another approach and improves the overall accuracy. A random forest algorithm is an example of ensemble learning where a collection of trees is used instead of a single tree predictor [13]. Majority voting is the most commonly used ensemble learning method, where no parameter tuning is required once each predictor is trained [13]. In addition to increased accuracy, superiorities in robustness, stability, confidence of estimation, parallelization, and scalability are other benefits of ensemble learning [22]. Bayesian model averaging (BMA) is also a well-known aggregation tool for combining results to improve forecast accuracy [23]. BMA uses the entire dataset for the inference-making process to avoid individual model dependence. In BMA aggregation process, combinational weights are assigned to each individual model. Models with higher accuracy gain a higher weight than the lower-performing ones [24].

In this paper, we propose a multi-stage demand forecasting approach that combines clustering and ensemble learning techniques to improve the accuracy of the customer behavior forecasts for retail sector. The main novelty rests in the fact that by means of clustering, the dataset will be segmented to ensure generation of more accurate forecasts for each cluster. Then, using ensemble learning, the clustered forecasts will be combined to map the whole dataset again. A combinatorial forecasting model is then used to generate the combined forecasts. To show the applicability and usefulness of the proposed approach, it is implemented in a real-world case where demand data for sports products is used to provide forecasts for the next demand cycle.

The rest of this paper is structured as follows: Section 2 presents a review of the literature on demand forecasting as well as the applications of various machine learning algorithms, highlighting the contributions of this paper. Section 3 presents the proposed three-step methodology for demand forecasting. The results and related discussions will be presented in Section 4. We conclude the paper in Section 5, providing an overview of the approach and directions for future research.

2. Background and literature review

Demand forecasting is a highly needed and challenging issue in the retail industry. Demand forecasts are used to support inventory control [25], supply chain management [26], and replenishment [27] decisions. Lack of demand planning could lead to shortages and stock reduction or over storage resulting in high storage costs [28], delays, the bullwhip effect [29], need to reordering, and missed customers. Customer segmentation in demand forecasting is an approach that identifies customers with similar demand behavior and used as a means of improving prediction accuracy [30]. According to McDonald [31], market segmentation is "the process of splitting customers, or potential customers, within a market into different groups, or segments, within which customers have the same or similar requirements satisfied by a distinct marketing mix".

Online retailers can use customer segment profiles to predict the expected demand from each segment for their products and thus be able to adjust product processes accordingly [32,33]. Customer segmentation is often done based on individuals' purchasing power [15]. It can improve the accuracy of demand predictions because the data in each segment belongs to a separate cluster of (similar) customers, making it easier for prediction models to extract the patterns in data [34–37]. For example, forecasting for a cluster of customers who make seasonal purchases differs from customer segments that follow monotonically increasing and decreasing purchase patterns [38]. Several methods exist for customer segmentation [34]. A basic segmentation can be done based on customers' geographic location. However, some demand characteristics of customers in the same location could still be different [15]. Partitioning, hierarchical, density-based, grid-based, and model-based methods are some of more advanced segmentation methods [39]. Many studies in the marketing area have focused on data-mining techniques for segmenting customers. The data mining is conducted through analysis of recency, frequency and monetary (RFM) value, partitioned clustering, logistic regression (LR), DT, and neural networks (NN) [35,40,41]. Partitioned clustering has low computational costs and is applicable when the datasets are large. K-means is one of the common partitioning methods [15].

After segmentation of customers, various algorithms can be used to forecast demand for each cluster of customers. These algorithms are generally categorized into statistical and artificial intelligence methods [42]. ARIMA is a statistical method used for time-series forecasting. The method requires complete datasets (without missing information) and aims at identifying linear relationships considering fixed temporal dependence and univariate data. It can only provide one-step forecasts [16]. The autoregressive conditional heteroscedastic (ARCH) and general autoregressive conditional heteroscedastic (GARCH) models are two other statistical methods that can be used for time-series data with nonlinear patterns [42]. Prophet is another time-series forecasting method that is attracting growing attention. It is an algorithm to forecast time-series data based on an additive model (adding a seasonal trend to forecast) where nonlinear trends are fitted with daily, weekly, yearly, and holiday seasonality effects. In addition, Prophet can handle datasets with missing data and outliers [43].

Artificial intelligence techniques such as machine learning and deep learning are also effective in solving time-series forecasting problems [16]. Artificial neural networks (ANN), SVM, KNN, and adaptive neuro-fuzzy inference systems (ANFIS) are examples of methods for predicting time-series data [44]. Among these methods, ANNs have several advantages, including universal approximation, being data-driven, and better capturing of nonlinear patterns in data [45]. Although different types of ANNs can capture nonlinear patterns in time-series data, researchers have indicated that ANNs with shallow architectures cannot accurately model time series with a high degree of nonlinearity, longer range, and heterogeneous characteristics [46]. A specific class of ANNs is recurrent neural networks (RNNs) which is capable of learning temporal representation of data [47]. Unlike feed forward ANNs, the connections between nodes in an RNN establish a (feedback) cycle, allowing signals to move both forward and backward, thus, providing a structure that supports a short-term memory suitable for processing sequential data. However, RNNs have a vanishing and exploding gradient problem for longer term predictions, making them sometimes hard to train. The prevalent solution to this problem is the addition of gated architectures such as long short-term memory (LSTM) that can exploit a long-range timing information [44,48].

2.1. Statement of novelty

The literature review revealed a gap in investigating the use of ensemble clustering approach in demand prediction and its potential benefits. Table 1 presents the recent studies in demand forecasting domain along with the techniques adopted. It is shown that the literature in ensemble learning domain is still emerging. To the best of authors' knowledge, applications of ensemble learning in integrating demand clustering and forecasting have not yet been presented in the literature. This study proposes a data-driven framework for ensemble cluster-based demand forecasting. First, customers are segmented using the RFM method. Then, a K-means algorithm along with four hierarchical clustering algorithms (single linkage, complete linkage, centroid linkage, and Ward's linkage) is used to cluster the segmented customers. An ensemble learning (based on a majority voting scheme) is then utilized to combine the results obtained from the above-mentioned clustering methods. Afterward, "LSTM" and "Prophet" machine learning techniques are compared in establishing demand forecasts corresponding to each customer segment. Finally, to map the forecasting for the whole dataset, the results obtained from each cluster are combined using averaging methods. The proposed approach is implemented in a real-world case with demand data of sports products. The aim is to arrive at forecasting of the demand for next cycle. We then conduct a scenario/sensitivity analysis to gain insights on the performance of the proposed demand forecasting approach. The scenarios include with or without customer segmentation and with or without ensemble learning in the clustering and forecasting sections. This analysis shows that the proposed ensemble learning approach provides much more accurate predictions compared to using conventional models.

3. Methodology

As mentioned earlier, the main objective of this study is to propose an effective demand forecasting approach with improved accuracy. The proposed approach has two main steps: customer segmentation and multivariate time-series demand forecasting. An overview of the proposed approach is shown in Fig. 1. First, data is preprocessed, and the outliers interpolated (details provided in Section 3.1). Then, target and predictor variables are determined. The target variable is considered as sales quantity (demand). However, it could also be considered as sales in dollars, discounts in dollars, or demand in quantity. The training phase includes customer segmentation and demand forecasting. The segmentation deals with clustering of customers. First, data is clustered using different algorithms (K-means, single linkage, complete linkage, centroid linkage, and Ward's linkage). Then, the resulting clusters are combined to form three segments of low, mid, and high representing customer affordability (more details are provided in Section 3.2.1).

After customer segmentation, LSTM and Prophet algorithms are used to establish the forecasts of demand corresponding to each customer segment. These algorithms are selected due to their superior performance in time-series forecasting problems [44,57]. LSTM is a recurrent neural network with long and short-term memory units suitable for forecasting order-dependent data. It is designed to direct the neural network in extracting longer-term trends and learning about longer-range dependencies in data [47]. As purchasing habits usually follow a cyclic pattern, investigating the application of LSTM in our dataset is appealing. Also, Prophet is suitable for data containing seasonality and holiday effects [58]. The details on implementation of LSTM and Prophet methods are discussed in Section 3.2.2.

The testing phase consists of a procedure for evaluating the model performance. The training data is fed into the clustering and segmentation algorithms. Then, trained prediction models are applied to each segment. The outputs of individual predictors are then ensembled using the BMA to generate forecast outputs (Section 3.3 provides more details on testing phase).

3.1. Data preprocessing

To ensure accuracy of the proposed demand clustering and forecasting, outlier detection and interpolation for collected test data are conducted using an unsupervised SVM. The data is scaled, and then, *OneClassSVM* is used to map the data into labels of 0 (for normal) or 1 (for abnormal) (see [59]). The outliers can then be replaced by values obtained using linear interpolation. The selected dataset have 52 features in total. However, not all these features are valuable for (and contribute to) demand forecasting. Each record of dataset corresponds to a unique transaction ID and sale quantity (i.e., target variable). The selected set of predictor variables are date, time, sale price, holidays, and demand in the previous period.

3.2. Training phase

The training phase includes two steps of (1) clustering and ensemble segmentation using majority voting, and (2) training of prediction models of each segment described as follows:

3.2.1. Clustering

The choice of clustering methodology and feature selection depends on the business goals and needs. For example, if the goal is to increase the retention rate, segmentation can be done using churn probability [14]. Among various existing segmentation methods, a RFM method is chosen due to its ability to segment customers who are highly likely to respond to a marketing campaign [40]. This clustering method classifies customers as low, mid, and high value-added customers. The customer segmentation is done using five clustering algorithms (K-means, single linkage, complete linkage, centroid linkage, and Ward's linkage), and then an ensemble learning will be done using majority voting method [13]. In doing so, three segments of customers are identified:

- Low-value: Those who are less active than others and less frequent buyers/visitors. They generate very low to zero or even negative revenue.
- Mid-value: These customers are in the middle. They often buy/visit (but not as much as the high-value customers) frequently and generate moderate revenue.
- High-value: These customers are most frequent buyers, and businesses do not want to lose them. They correspond to a high share of revenue and a high frequency of purchase.

Three features – recency (R), frequency (F), and monetary (M) – were calculated for each row of dataset as the basis for clustering. R implies the most recent purchase date of a customer. It is defined as the number of days since the last purchase of a customer. F is the total number of

Table 1
Literature review on ensemble learning approach.

Authors	Clustering method	Ensemble learning in clustering	Forecasting method	Ensemble learning in forecasting	Ensemble learning in clustering & forecasting
Lemke & Gabrys (2010) [49]	–	–	ARIMA Moving average Single exponential smoothing NN	Simple average Trimming average Variance-based model Outperformance method Variance-based pooling Regression combination	–
Li et al. (2011) [50]	–	–	Adaptive linear element network Backpropagation network Radial basis function network	BMA	–
Lu & Kao (2016) [13]	Single linkage Complete linkage Centroid linkage Median linkage Ward's linkage	Majority voting	Extreme learning machine	–	–
Nilashi et al. (2017) [21]	SOM Expectation Maximization	Hypergraph-Partitioning Algorithm	ANFIS SVR	–	–
Raza et al (2017) [23]	–	–	Backpropagation neural network Elman neural network ARIMA Feed forward neural network Radial basis function Wavelet transform	BMA	–
Adhikari et al. (2018) [51]	–	–	Time-Series Model Regression Model	Weights Generation	–
Papageorgiou et al. (2019) [52]	–	–	Fuzzy cognitive maps ANN	Bootstrapping method	–
Bandara et al (2020) [53]	K-Means PAM DBSCAN Snob	Boosting approach	LSTM	–	–
Abbasimehr & Shabani (2021) [54]	Agglomerative hierarchical clustering	–	ARIMA Simple moving average KNN	Combined method (as described in their paper)	–
Gastinger et al. (2021) [19]	–	–	ARIMA NN Naive	Stacking ensemble-based	–
Massaoudi et al. (2021) [55]	–	–	Light Gradient Boosting Machine eXtreme Gradient Boosting machine Multi-Layer Perceptron	Stacking ensemble-based	–
Zhang et al. (2021) [56]	–	–	Multi-layer perceptron neural network Convolutional neural network LSTM	Stacking ensemble-based	–
This Study	K-means Single linkage Complete linkage Average linkage Ward's linkage	Majority voting	LSTM Prophet	BMA Simple average	✔

purchase orders for a customer in each period of time. The higher the frequency, the more active the customer. Similar to recency, a high-frequency customer is more value-added. M indicator shows how much money a customer has spent on his/her purchases to date.

Five clustering algorithms were then used, considering the above features. These are K-means, single linkage, complete linkage, centroid linkage, and Ward's linkage. We chose these algorithms to cover a broad range of selection criteria and to comparatively observe/evaluate their performance. As clustering methods could have different distance criteria, each could assign a record to a different cluster. We then use majority voting to combine the results of these clustering algorithms [13]. We chose the cluster with the highest vote and assigned it to the customer ID. In case of equal votes, a random selection is made. In this process, the Elbow method is used to find the optimal number of clusters [60].

At this point, the cluster numbers (please refer to Section 4 for more details) are summed up for all features (using RFM method). The overall scores of segmentations were 0–2 (for low-value), 3–4 (for mid-value), and 5+ (for high-value). The distance criterion for each clustering algorithm will be discussed below. In this regard, C.-J. Lu and Kao [13] could serve as a reference for more detailed information about the distance criteria and the corresponding clustering methods discussed as follows:

K-means method: K-means algorithm separates the data points into k cluster of equal variances. More precisely, the algorithm splits a set of n records into $j = 1, 2, \ldots k$ disjoint clusters (c), each characterized by a mean of μ_j of samples x_i (values) in the cluster. The mean is called centroid and is representative of the cluster. The algorithm works based on minimizing a within-cluster sum of square [61] as follows:

$$\sum_{i=0}^{n} \min_{\mu_j} \left\| x_i - \mu_j \right\| \tag{1}$$

Single linkage method: In this method, the designated cluster for a sample is the one with minimum Euclidean distance. Assuming that $x_{kj}^{(d)}$ is the kth value of jth predictor in the cluster d identified in training phase, the Euclidean distance between the predictor x_i of testing data y_i^* and the predictors $x_{kj}^{(d)}$ of training data $y^{(d)}$ in cluster d can be computed

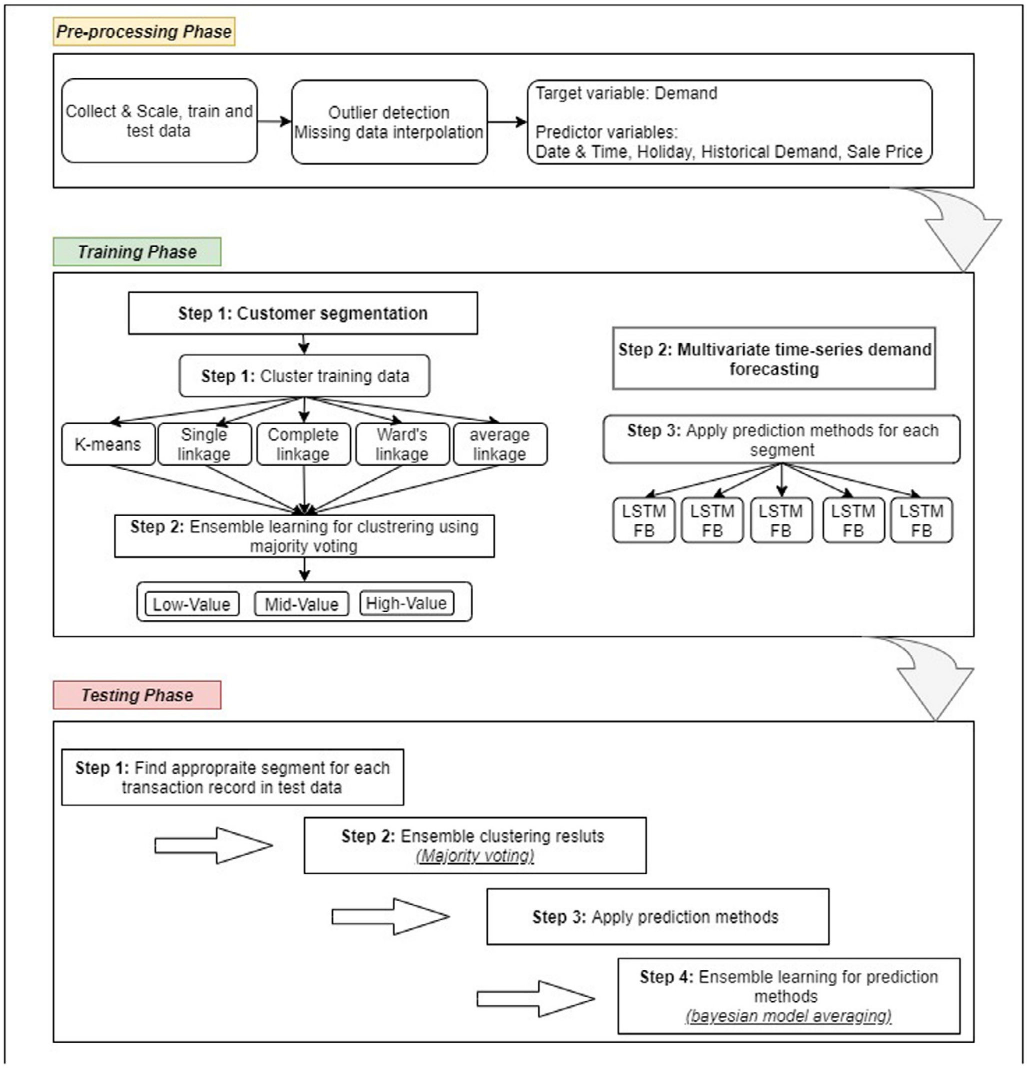

Fig. 1. The proposed ensemble modeling and clustering-based multivariate time-series demand forecasting.

as:

$$\lambda_{ik}^{(d)} = \sum_{j=1}^{p} \left(\sqrt{(x_{ij} - x_{kj}^{(d)})^2} \right) \quad for\ i = 1, 2, \ldots, n, d = 1, 2, \ldots, g \tag{2}$$

.

In single linkage method, distance $S_i^{(d)}$ between the test data y_i^* and the cluster d is the minimum value of $\lambda_{ik}^{(d)}$, i.e. $S_i^{(d)} = min(\lambda_{ik}^{(d)})$. In this sense, the best representative cluster for test data y_i^* is cluster SG_i that corresponds to smallest $S_i^{(d)}$, i.e. $SG_i = argmin(S_i^{(d)})$ [13,62,63].

Complete linkage method: The difference between a single linkage method and a complete linkage method is that the distance between test data y_i^* and cluster d is the maximum value of $\lambda_{ik}^{(d)}$, i.e. $C_i^{(d)} = max(\lambda_{ik}^{(d)})$. Thus, the best representative cluster for test data y_i^* will be a cluster CG_i that corresponds to the smallest $C_i^{(d)}$, i.e. $CG_i = argmin(C_i^{(d)})$ [13,63].

Average linkage method: In contrary to the methods that use farthest or nearest points to calculate a similarity value, an average linkage method measures the distance between clusters using their averages, advocating the fact that a mean is an indicator of data centricity. In doing so, the mean of a jth predictor in cluster d is defined as:

$$\alpha_j^{(d)} = \frac{1}{n_d} \sum_{i=1}^{n_d} x_{ij}^{(d)} = mean(\{x_{ij}^{(d)}, \forall i = 1, 2, \ldots, n_d\}) \tag{3}$$

Thus, the best representative cluster for test data y_i^* will be a cluster AG_i that corresponds to the smallest $\alpha_j^{(d)}$, i.e. $AG_i = argmin(\alpha_j^{(d)})$ [13].

Ward's linkage method: Ward's linkage method defines the distance between two clusters by calculating and minimizing a within-cluster variance. First, test data y_i^* is included in cluster d with a centroid value of $o_j^{(d)}$ calculated as follows:

$$o_j^{(d)} = \frac{x_{ij} + \sum_{k=1}^{n_d} x_{kj}^{(d)}}{1 + n_d} \quad \forall j = 1, 2, \ldots, p. \tag{4}$$

Then, Ward's linkage distance $W_i^{(d)}$ between test data y_i^* and cluster d is calculated as:

$$W_i^{(d)} = \sum_{j=1}^{p} (\sqrt{(x_{ij} - o_j^{(d)})^2} + n_d \times \sqrt{(a_j^{(d)} - o_j^{(d)})^2}) \tag{5}$$

where n_d is number of data points in cluster d.

In this sense, the best representative cluster WG_i for test data y_i^* is determined as $WG_i = argmin(W_i^{(d)})$ [13,63].

3.2.2. Demand forecasting

In previous section, we reviewed several clustering methods that can provide an analysis of transaction records and conduct a segmentation of customers based on RFM features. The output of clustering step (customer clusters) can be used as an input for demand forecasting.

To predict daily demand for each segment, the following procedure is proposed in this paper. First, a summation of daily transaction records is established. Then, other features such as sales price, holiday and other dates effect, as well as demand in the previous periods, are used as predictor variables to establish forecasts for the target variable, i.e., daily demand through alternative forecasting algorithms of LSTM and Prophet.

3.2.2.1. LSTM. LSTM method uses recurrent NNs with memory units and loops that allow the information to persist. LSTM has an input gate, a forget gate, an internal state, and an output gate. The terms used in LSTM formulation are:

$x(t_i)$: input value

$h\left(t_{i-1}\right), h\left(t_i\right)$: output values

$c\left(t_{i-1}\right), c\left(t_i\right)$: cell states

$b = \{b_a, b_f, b_c, b_o\}$ biases of input gate, forget gate, internal state and output gate

$W_1 = \{w_a, w_f, w_c, w_o\}$: weights of input gate, forget gate, internal state, and output gate

$W_2 = \{w_{ha}, w_{hf}, w_{hc}, w_{ho}\}$: recurrent weights

$a = \{a(t_i), f(t_i), c(t_i), o(t_i)\}$: output results for input gate, forget gate, internal state, and output gate

The following equations present the integration of the above elements in forming an LSTM cell:

$$a(t_i) = \sigma(w_a x(t_i) + w_{ha} h(t_{i-1}) + b_a) \tag{6}$$

$$f(t_i) = \sigma(w_f x(t_i) + w_{hf} h(t_{i-1}) + b_f) \tag{7}$$

$$c(t_i) = f_t \times c(t_{i-1}) + a_t \times tanh(w_c x(t_i)) + w_{hc}(h(t_{i-1}) + b_c) \tag{8}$$

$$o(t_i) = \sigma(w_o x(t_i) + w_{ho} h(t_{i-1}) + b_o) \tag{9}$$

$$h(t_i) = o(t_i) \times tanh(c(t_i)) \tag{10}$$

After identifying $c\left(t_i\right)$ and $h\left(t_i\right)$, error (variance) between the predicted data and input data is calculated. This error is propagated back to input gate, cell gate, and forget gate, as a feedback, to update the weights following an optimization algorithm (such as gradient descent approach or similar) [44].

3.2.2.2. Prophet. Turning to Prophet, this forecasting tool consists of autoregression models to predict time-series data capable of handling seasonality, holidays effect, strong shifts in trends, and outlier existence in data [43,57]. This forecasting tool is automated in tuning time-series [58]. Prophet fits several linear and nonlinear time functions according to the following equation:

$$y(t) = g(t) + s(t) + h(t) + e(t) \tag{11}$$

where $g(t)$ is a trend representing non-periodic changes (such as growth or decay over time), $s(t)$ is a seasonality term representing periodic changes (such as daily, weekly, or monthly), $h(t)$ represents the holiday effects, and $e(t)$ represents noise in data. The $g(t)$ follows either a saturating growth model or a piecewise growth model. As such, the saturating trend is given by:

$$g(t) = \frac{\mathrm{C}}{1 + \exp(-\mathrm{k}(\mathrm{t} - \mathrm{m}))} \tag{12}$$

where C is carrying capacity, k is the growth rate, and m is an offset parameter.

In the case of a piecewise growth model (i.e. the default method in Prophet), the trend is defined as:

$$g(t) = \left(k + a(t)^T \delta\right) t + (m + a(t)^T \gamma) \tag{13}$$

where k is the growth rate, and δ is a vector of rare adjustments, where δ_j is the change in rate at time t. The rate at time t is represented by a base rate k plus all adjustments up to that point. This is represented by the following function $a(t)$

$$a_j(t) = \begin{cases} 1, & if\, t \geq s_j, \\ 0, & otherwise. \end{cases}$$

As for the seasonality trend $s(t)$, the aim is to provide adaptability in the model to periodic changes following a sub-daily, daily, weekly, and yearly seasonality. Use of a standard Fourier model is advocated to establish a seasonality trend. Holiday effects $h(t)$ is also incorporated by generating a matrix of regressors. More details about Prophet and the above procedure can be found in Taylor & Letham [43].

3.3. Testing phase

The objective of testing phase is to ensemble predicted values of the forecasting methods using BMA. In this phase, firstly, the above reviewed clustering methods are implemented for test dataset and the results are ensembled by a majority voting. At the end of clustering section, three customer segments (low, mid, and high) are identified. Secondly, a pre-trained forecasting method (LSTM and Prophet) is applied to test data in each cluster. Finally, ensembling the results from different forecasting methods is done by combining them into a single predictor. The BMA method weighs these forecasts according to their posterior model probabilities. It, therefore, forms an averaged model with better-performing predictions having higher weights than worse-performing predictions [50]. BMA is implemented using widely applicable information criterion (WAIC), or leave-one-out cross-validation (LOO) approaches to estimate the above-mentioned weights as follows:

$$w_i = \frac{e^{-\frac{1}{2} dIC_i}}{\sum_j^M e^{-\frac{1}{2} dIC_j}} \tag{14}$$

where dIC_i is the difference between the ith information criterion value and the lowest value. In general, a lower dIC is better. This approach is called pseudo-Bayesian model averaging, or Akaike-like weighting, and it is a heuristic way to calculate the relative probability of each model (given a fixed set of models) from the information criteria values. The denominator ensures that the weights sum up to one [23,64].

3.4. Assumptions

- RFM method assumes that customers are sensitive to marketing campaigns but react differently to marketing strategies. RFM aims at ranking the customers based on their purchasing habits, in order for those with high ranks being targeted with marketing campaigns [40]. This concept is considered applicable to the dataset presented in this study. Thus, an RFM-based clustering is employed for demand prediction by focusing on customers with similar RFM ranks.
- It is assumed that no unprecedented events will happen in the future (such as pandemics, global recession, international shipping problems, crashes in shopping platforms, etc.).
- The demand is predicted on a daily basis.

4. Results analysis and discussion

4.1. Data description and performance criteria

The time-series data was collected from an open-source dataset available in [65]. It consisted of the supply chain data of three products: clothing, sports, and electronics supplies. The focus of this study is on sports products as the corresponding dataset was of larger size and with fewer missing values. Although, this dataset consisted of 52 features (Fig. 2), most of these features were not valuable (needed) for demand forecasting. The dataset contains of records from January 2015 to September 2017. Fig. 3 shows the daily demand for sports products corresponding to the last three months of dataset. The demand varied from 250 to 500 depending on time, sales price, and discounts.

The accuracy of the predictions was evaluated using Mean absolute percentage error (MAPE), Mean absolute error (MAE), and Root mean square error (RMSE). In doing so, a lower value indicates a better

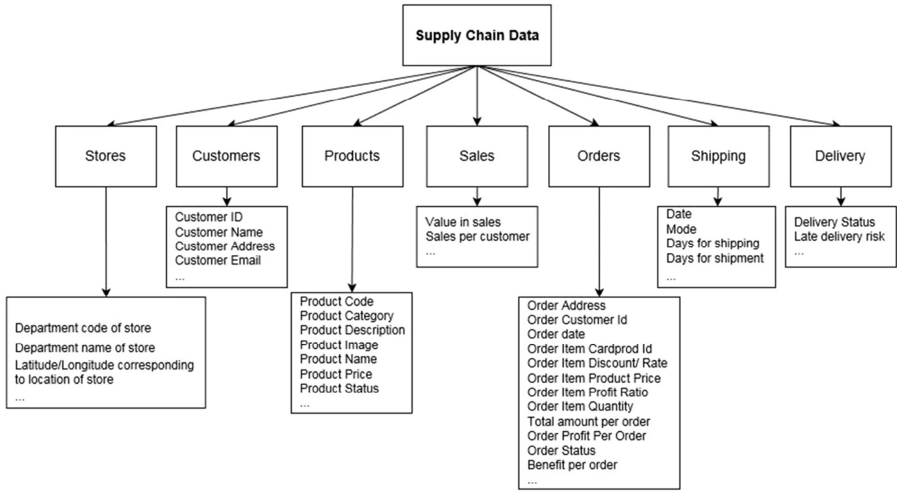

Fig. 2. Taxonomy of features in sports dataset. *Note*: Demand, customer ID, purchase date, sales price, amount purchased, frequency of customer visits, and holidays were used or extracted from the features [8] (Seyedan & Mafakheri, 2020).

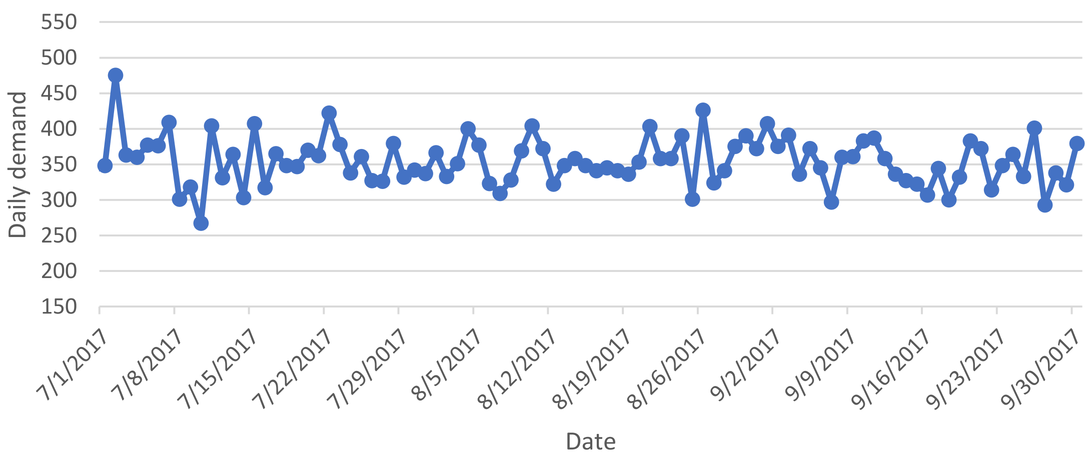

Fig. 3. Daily demand (quantity) for sports products.

prediction performance (or less inaccuracy). The definitions of these criteria are provided as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{17}$$

where y_i and $\hat{y}_i$ show the actual and predicted values at day i, respectively. Also, n is the total number of testing days (30 days in this case).

4.2. Clustering and predictions

To calculate recency, for each transaction, we subtracted the snapshot day (the current day) from the date the transaction was performed. Frequency was also calculated as the number of transactions made by each customer. For monetary value (of a customer), we summed up all transactions by each customer. In order to establish the above features, it is essential to identify the optimum number of clusters to ensure minimization of each cluster's within-cluster variance. The elbow method [60] was employed to identify the optimum value of k for each feature (i.e. optimal number of clusters). Results are presented in Fig. 4, showing that $k = 4$ is optimal for recency, frequency, and monetary clusters. For each cluster, one score (from 0 to 3) is assigned according to their cluster centroid values. In recency, this score is 0 for the highest cluster centroid and 3 for the lowest cluster centroid, while in case of frequency and monetary; the situation is in opposite. Finally, by calculating the overall scores of customers (summing up all scores for recency, frequency, and monetary), they are clustered into three customer segments: low-value customers, mid-value customers, and high-value customers. Fig. 5 shows these customer segments based on the two-dimensional features and K-means clustering. High-value customers are those who purchased more recently, tend to buy more frequently, and spend more than other clusters. After applying different clustering methods, using majority voting, we combined the outcomes.

In the next step, we aggregated daily demand and average daily sale prices for each customer segment. We then used them as input to predict the next month's daily demand. We extracted two additional features (besides historical daily demand and sale prices) – of dates and holidays – because they strongly affect future demand. The average daily demand was calculated for holidays (364) and non-holidays (362). This shows that customers were more inclined to buy products during holidays rather than non-holidays. To predict the future daily demand, we turned to Prophet and LSTM algorithms as described in Section 3.2.2.

Using an exhaustive search approach, the optimal parameters chosen for our LSTM were established as follows: number of hidden layers = 50, number of epochs = 50, batch size = 72, with *adam* selected as the

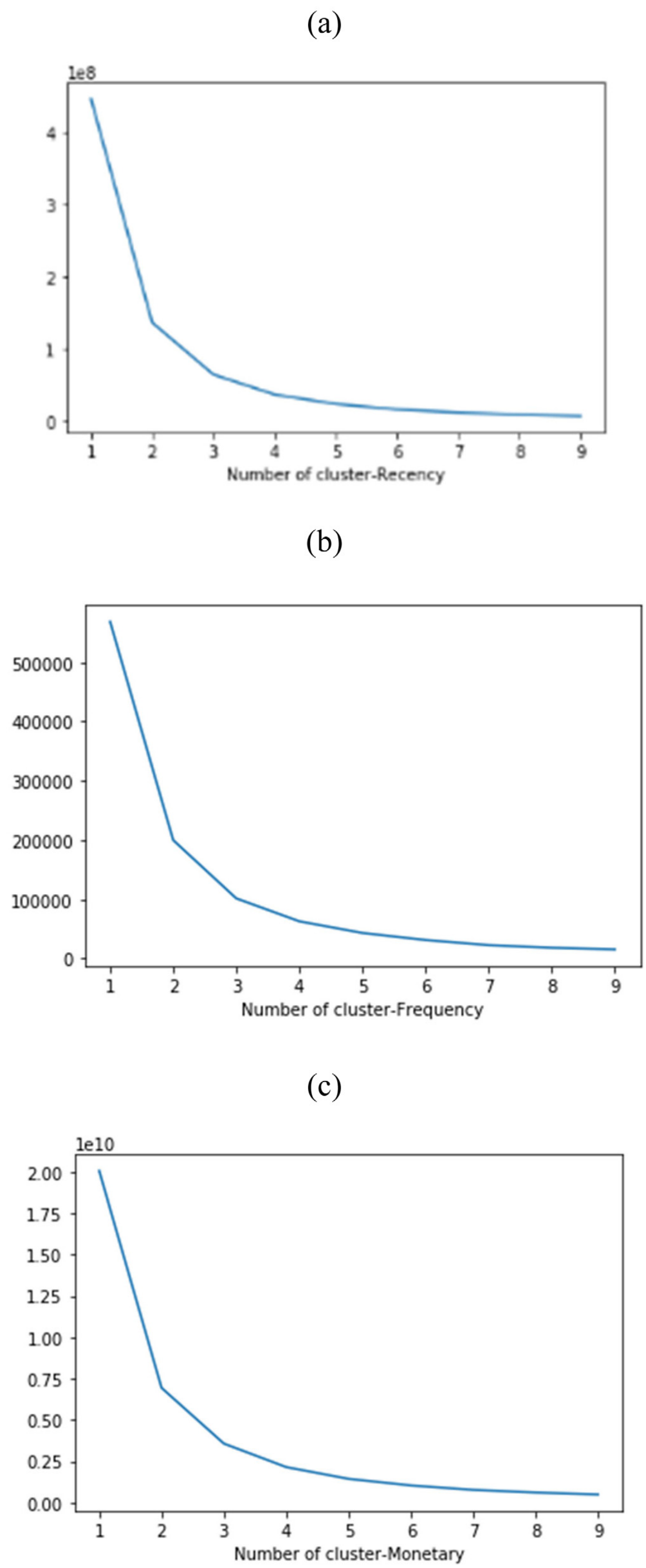

Fig. 4. Optimal number of clusters for (a) Recency, (b) Frequency, & (c) Monetary.

accuracy optimizer function. In addition, for Prophet, we considered daily, weekly, and yearly seasonality data forming a multiplicative seasonality mode [58].

After applying LSTM and Prophet, each method established a predicted value for the daily demand for next (following) month. To ensemble the two forecasting results, we used simple average and BMA with the results described in the following sections.

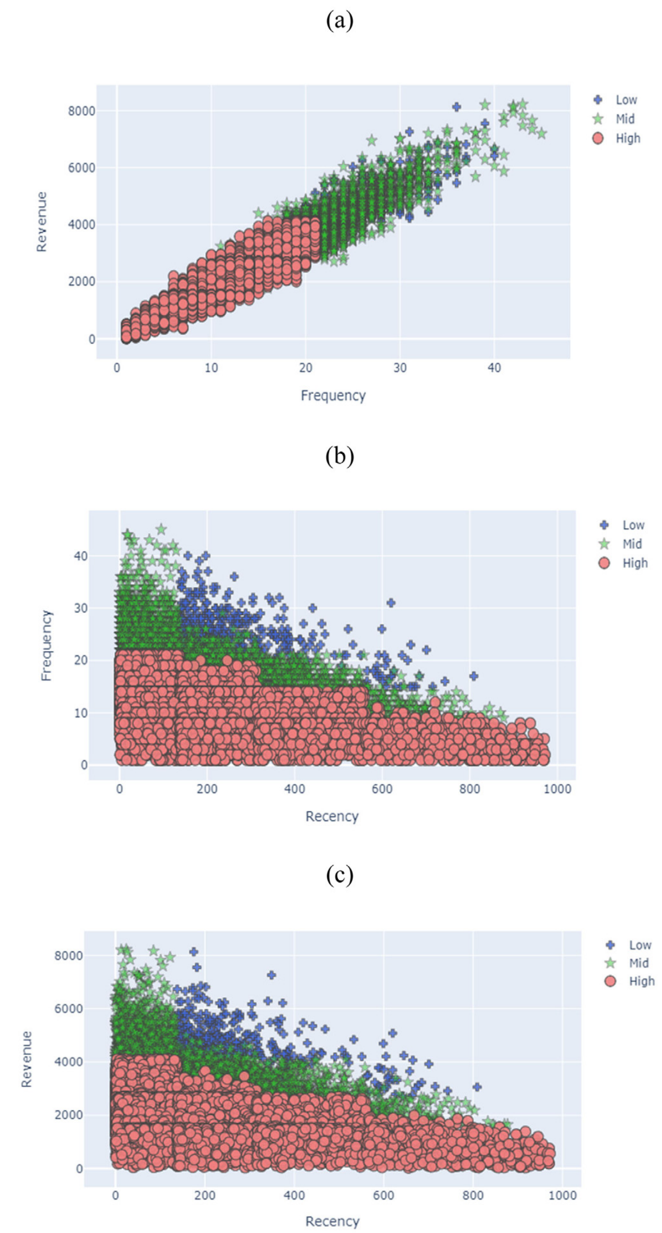

Fig. 5. Customer segmentations using K-means: (a) Revenue vs Frequency, (b) Frequency vs Recency, (c) Revenue vs Recency.

4.3. Accuracy test

To validate the proposed clustering-based forecasting approach with ensemble learning, we considered three different scenarios. In the first scenario, the forecasting methods were applied to raw data without clustering. The results of three performance evaluation criteria (MAPE, MAE, and RMSE) were presented in Table 2, showing that the lowest RMSE was achieved in forecasting results ensembled by BMA.

In the second scenario (Table 3), five clustering methods were considered, and the results were shown for each customer segment. The LSTM and Prophet results were ensembled using simple averaging and BMA. These results were shown in Table 4. We combined the clustering results using a majority voting. Then, we forecasted the daily demand by applying alternative multivariate time-series forecasting. Finally,

Table 2
Forecasting performance criteria—Without clustering.

Methods of forecasting	MAPE %	MAE	RMSE
LSTM	8.169	67.147	111.927
Prophet	9.920	304.366	316.661
Simple average	8.753	29.041	35.813
BMA	8.177	27.393	32.887

Table 3
Forecasting performance criteria—Using clustering methods.

Methods of clustering	Methods of forecasting	MAPE %	MAE	RMSE
K-means clustering	LSTM	8.213	205.340	248.496
	Prophet	8.296	155.665	214.737
	Simple average	8.153	27.775	33.339
	BMA	8.252	28.234	34.836
Single linkage clustering	LSTM	8.434	75.153	124.783
	Prophet	10.016	293.855	309.561
	Simple average	8.967	29.738	36.266
	BMA	8.494	28.433	33.675
Complete linkage clustering	LSTM	8.680	103.018	160.688
	Prophet	10.720	267.247	293.250
	Simple average	9.501	31.590	37.914
	BMA	8.666	29.140	34.271
Average linkage clustering	LSTM	8.506	65.114	109.038
	Prophet	9.898	307.646	321.007
	Simple average	9.074	30.217	36.523
	BMA	8.446	28.373	33.890
Ward's linkage clustering	LSTM	8.885	143.755	202.902
	Prophet	8.622	212.575	258.640
	Simple average	8.703	29.224	34.849
	BMA	8.655	29.293	34.948

Table 4
Forecasting performance criteria—With majority voting in clustering.

Methods of forecasting	MAPE %	MAE	RMSE
LSTM (univariate)	8.126	125.300	150.835
Prophet (univariate)	8.951	216.550	225.696
Simple average	8.464	28.524	33.739
BMA	8.219	27.760	32.984
LSTM (multivariate)	8.076	62.723	104.985
Prophet (multivariate)	9.928	306.254	319.201
Simple average	8.842	29.314	35.815
BMA	8.120	27.106	32.801

we ensembled the results using BMA. The error was less than other scenarios with MAPE = 8.12%, MAE = 27.1, and RMSE = 32.8. The results showed that customer segmentation did increase the accuracy of the predictions. In addition, adopting multivariate forecasting, we could integrate the impact of additional features on daily demand predictions, making them more accurate compared to univariate forecasting. Furthermore, use of ensemble learning and combining the results of different clustering and forecasting models, by means of a majority voting in the clustering section and BMA in the prediction section, has further improved the accuracy of predictions making them more representative of reality.

4.4. Sensitivity analysis

This section discusses the influence of prediction variables and how they contribute to demand forecast

s. The effects of holidays and average daily sales on demand were presented in Fig. 6, with average daily sales and holiday variables changing across their ranges (minimum and maximum values derived based on historical data) while keeping the rest of the variables fixed. Fig. 6 shows that:

- There is an increasing trend for sales in both holidays and weekdays. When the daily demand rises, the average price of the products rises as well leading to higher sales. Average price of products increases, daily demand rises as well. This incrimination shows that when the market is hot (high demand), people have higher preference for online shopping (with willingness to pay a higher price in such a hot market). This is reflecting the demand and sale direct dependency.
- The graph also shows that buyers shop online more on holidays than weekdays when the average daily sale is less than a certain threshold (about $200). This outcome may be due to the customers' preference to buy from retail stores in person and avoid online shopping when the prices are higher.

Table 5
Performance improvement by ensemble clustering (for electronic products dataset)

Methods of clustering	Methods of forecasting	Performance improvement by ensemble clustering (%)
Without clustering	BMA (LSTM & Prophet)	5.65
K-means clustering		6.08
Single linkage clustering		4.58
Complete linkage clustering		9.12
Average linkage clustering		6.75
Ward's linkage clustering		2.84

In addition, to test the robustness of the methodology proposed in Section 3, it is applied to an alternative dataset of Electronic products [65]. The proposed methodology also was well-performed with the electronic products dataset. For instance, the proposed majority voting method with ensemble forecasting results showed a 5.65% accuracy improvement compared to the scenario without clustering. The results obtained for the alternative dataset are shown in Table 5.

5. Conclusions

Demand forecasting highly affects business decisions, such as production planning, inventory management, financial planning, and sales. As a result, enhancing forecasting accuracy can improve a firm's decision-making efficacy, reduce risk of unmet demand, and lower operational costs by preventing over production. This study proposed a methodology for demand forecasting using ensemble learning in case of sports retail industry to improve the accuracy of future daily demand forecasting.

The proposed forecast framework provided a cluster-based demand prediction using time-series forecasting methods of LSTM and Prophet. In addition, we used majority voting and BMA as ensemble learning techniques in clustering and forecasting, respectively. The aim was to achieve a higher forecast accuracy in demand prediction by intelligently combining different models' performances and assigning higher weights to the better-performing models. The prediction performance of the proposed forecasting framework was presented for sports products dataset.

The proposed framework was analyzed under different forecasting scenarios, including with clustered and without clustered customers as well as with or without ensemble learning. The sensitivity analysis of forecast results showed an improvement in prediction accuracy of the clustered-ensembled approach compared to use of single models, resulting in the minimum values of MAPE (8.12%), MAE (27.1%), and RMSE (32.8%) for daily forecasts. BMA effectiveness was also observed in terms of forecast error reduction. The proposed cluster-based forecasting framework had considerably increased prediction accuracy across various seasonal and monthly cases.

The proposed framework has a number of limitations. We made efforts to develop a generic methodology applicable to different sets of supply chain data while using a minimum number of variables. Access to more information about customers such as income rate,

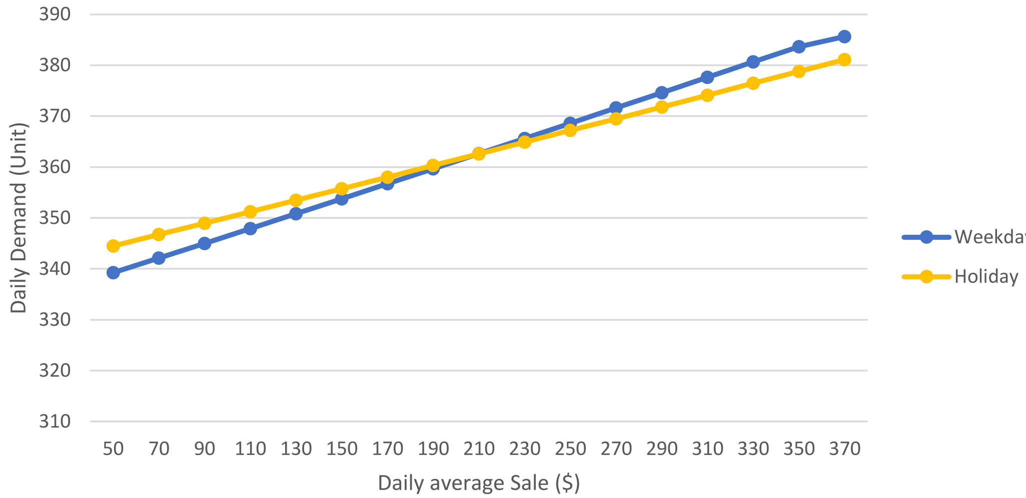

Fig. 6. Effect of holiday and average sales on demand forecasts.

age, location, monthly budget, and similar personalized data could significantly improve the segmentation.

Access to a larger dataset could also help better project a multi-period demand forecast and improve forecasting accuracy. We used a dataset of sports products. However, having multiple retail data from different sources could enhance the performance of predictions. Just to present the robustness of our proposed approach, it was also tested on an alternative dataset (electronic products).

Future avenues of research could include combining the proposed forecasting model with an optimization model with prescriptive capabilities in response to future demand scenarios and expectations. The literature has indicated that while predictive analytics have been utilized in demand management and procurement, prescriptive analytics have rarely been applied to demand forecasting.

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.

References

[1] T. Nguyen, L. Zhou, V. Spiegler, P. Ieromonachou, Y. Lin, Big data analytics in supply chain management: A state-of-the-art literature review, Comput. Oper. Res. 98 (2018) 254–264, http://dx.doi.org/10.1016/J.COR.2017.07.004.

[2] G. Wang, A. Gunasekaran, E.W.T. Ngai, T. Papadopoulos, Big data analytics in logistics and supply chain management: Certain investigations for research and applications, Int. J. Prod. Econ. 176 (2016) 98–110, http://dx.doi.org/10.1016/J.IJPE.2016.03.014.

[3] R. Fildes, S. Ma, S. Kolassa, Retail forecasting : Research and practice, Int. J. Forecast. (2019) http://dx.doi.org/10.1016/j.ijforecast.2019.06.004.

[4] Y. Acar, E.S. Gardner, Forecasting method selection in a global supply chain, Int. J. Forecast. 28 (4) (2012) 842–848, http://dx.doi.org/10.1016/J.IJFORECAST.2011.11.003.

[5] A. Athlye, A. Bashani, Multivariate demand forecasting of sales data, Int. J. Res. Appl. Sci. Eng. Technol. 6 (10) (2018) 198–211.

[6] Q. Yu, K. Wang, J.O. Strandhagen, Y. Wang, Application of long short-term memory neural network to sales forecasting in retail—A case study, in: Lect. Notes Electr. Eng., vol. 451, 2018, pp. 11–17, http://dx.doi.org/10.1007/978-981-10-5768-7_2.

[7] M. Ŝtepnicka, J. Peralta, P. Cortez, L. Vavrícková, G. Gutierrez, Forecasting seasonal time series with computational intelligence: Contribution of a combination of distinct methods, in: Proc. 7th Conf. Eur. Soc. Fuzzy Log. Technol. EUSFLAT 2011 French Days Fuzzy Log. Appl. LFA 2011, Vol. 1, no. 1, 2011, pp. 464–471, http://dx.doi.org/10.2991/eusflat.2011.7.

[8] M. Seyedan, F. Mafakheri, Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities, J. Big Data 7 (1) (2020) http://dx.doi.org/10.1186/s40537-020-00329-2.

[9] A.S. Bozkir, E.A. Sezer, Predicting food demand in food courts by decision tree approaches, Procedia Comput. Sci. 3 (2011) 759–763, http://dx.doi.org/10.1016/J.PROCS.2010.12.125.

[10] G.-Y. Ban, N.B. Keskin, Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity, Manage. Sci. (2020) http://dx.doi.org/10.2139/ssrn.2972985.

[11] K. Venkatesh, V. Ravi, A. Prinzie, D. Van Den Poel, Cash demand forecasting in ATMs by clustering and neural networks, European J. Oper. Res. 232 (2) (2014) 383–392, http://dx.doi.org/10.1016/j.ejor.2013.07.027.

[12] K.L. López, C. Gagné, G. Castellanos-Dominguez, M. Orozco-Alzate, Training subset selection in hourly Ontario energy price forecasting using time series clustering-based stratification, Neurocomputing 156 (2015) 268–279, http://dx.doi.org/10.1016/J.NEUCOM.2014.12.052.

[13] C.-J. Lu, L.-J. Kao, A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server, Eng. Appl. Artif. Intell. 55 (2016) 231–238, http://dx.doi.org/10.1016/J.ENGAPPAI.2016.06.015.

[14] P.W. Murray, B. Agard, M.A. Barajas, Market segmentation through data mining: A method to extract behaviors from a noisy data set, Comput. Ind. Eng. 109 (2017) 233–252, http://dx.doi.org/10.1016/j.cie.2017.04.017.

[15] P.W. Murray, B. Agard, M.A. Barajas, Forecasting supply chain demand by clustering customers, IFAC-PapersOnLine 48 (3) (2015) 1834–1839, http://dx.doi.org/10.1016/J.IFACOL.2015.06.353.

[16] B. Jason, Deep learning for time series forecasting, Ml 1 (1) (2018) 1–50, http://dx.doi.org/10.1093/brain/awf210.

[17] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, PLoS One 13 (3) (2018) 1–26, http://dx.doi.org/10.1371/journal.pone.0194889.

[18] M. Galar, A. Fern, E. Barrenechea, H. Bustince, A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches, IEEE Trans. Syst. Man Cybern. Part C 42 (4) (2012) 463–484.

[19] J. Gastinger, S. Nicolas, D. Stepić, M. Schmidt, A. Schülke, A study on ensemble learning for time series forecasting and the need for meta-learning, no. Ijcnn, 2021, [Online]. Available: http://arxiv.org/abs/2104.11475.

[20] T. Alqurashi, W. Wang, Clustering ensemble method, Int. J. Mach. Learn. Cybern. 10 (6) (2019) 1227–1246, http://dx.doi.org/10.1007/s13042-017-0756-7.

[21] M. Nilashi, K. Bagherifard, M. Rahmani, V. Rafe, A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques, Comput. Ind. Eng. 109 (2017) 357–368, http://dx.doi.org/10.1016/j.cie.2017.05.016.

[22] R. Ghaemi, N. Sulaiman, H. Ibrahim, N. Mustapha, A survey: Clustering ensembles techniques, World Acad. Sci. Eng. Technol. 38 (2009) 644–653, http://dx.doi.org/10.5281/zenodo.1329276.

[23] M.Q. Raza, M. Nadarajah, C. Ekanayake, Demand forecast of PV integrated bioclimatic buildings using ensemble framework, Appl. Energy 208 (December) (2017) 1626–1638, http://dx.doi.org/10.1016/j.apenergy.2017.08.192.

[24] G. Li, J. Shi, Application of Bayesian model averaging in modeling long-term wind speed distributions, Renew. Energy 35 (6) (2010) 1192–1202, http://dx.doi.org/10.1016/j.renene.2009.09.003.

[25] D.K. Barrow, N. Kourentzes, Distributions of forecasting errors of forecast combinations: Implications for inventory management, Int. J. Prod. Econ. 177 (2016) 24–33, http://dx.doi.org/10.1016/J.IJPE.2016.03.017.

[26] E.R.S. Kone, M.H. Karwan, Combining a new data classification technique and regression analysis to predict the cost-to-serve new customers, Comput. Ind. Eng. 61 (1) (2011) 184–197, http://dx.doi.org/10.1016/J.CIE.2011.03.009.

[27] V. Sillanpää, J. Liesiö, Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions, Int. J. Prod. Res. 56 (12) (2018) 4168–4185, http://dx.doi.org/10.1080/00207543.2018.1431413.

[28] F. Turrado García, L.J. García Villalba, J. Portela, Intelligent system for time series classification using support vector machines applied to supply-chain, Expert Syst. Appl. 39 (12) (2012) 10590–10599, http://dx.doi.org/10.1016/J.ESWA.2012.02.137.

[29] R. Carbonneau, K. Laframboise, R. Vahidov, Application of machine learning techniques for supply chain demand forecasting, European J. Oper. Res. 184 (3) (2008) 1140–1154, http://dx.doi.org/10.1016/J.EJOR.2006.12.004.

[30] P.W. Murray, B. Agard, M.A. Barajas, Forecast of individual customer's demand from a large and noisy dataset, Comput. Ind. Eng. 118 (2018) 33–43, http://dx.doi.org/10.1016/J.CIE.2018.02.007.

[31] M. McDonald, M. Christopher, M. Bass, Market segmentation, in: Marketing. Palgrave, London, 2003.

[32] T. Qu, J.H. Zhang, F.T.S. Chan, R.S. Srivastava, M.K. Tiwari, W.Y. Park, Demand prediction and price optimization for semi-luxury supermarket segment, Comput. Ind. Eng. 113 (2017) 91–102, http://dx.doi.org/10.1016/j.cie.2017.09.004.

[33] Q. Lu, N. Liu, Pricing games of mixed conventional and e-commerce distribution channels, Comput. Ind. Eng. 64 (1) (2013) 122–132, http://dx.doi.org/10.1016/j.cie.2012.09.018.

[34] K.R. Kashwan, C.M. Velu, Customer segmentation using clustering and data mining techniques, Int. J. Comput. Theory Eng. 5 (6) (2013) 1–6, http://dx.doi.org/10.7763/IJCTE.2013.V5.811.

[35] J.A. McCarty, M. Hastak, Segmentation approaches in data-mining : A comparison of RFM, CHAID, and logistic regression, 60, J. Bus. Res. (2007) 656–662, http://dx.doi.org/10.1016/j.jbusres.2006.06.015.

[36] M. Espinoza, C. Joye, R. Belmans, B. De Moor, Short-term load forecasting, profile identification, and customer segmentation : A methodology based on periodic time series, IEEE Trans. Power Syst. 20 (3) (2005) 1622–1630.

[37] J. Wei, S. Lin, C. Weng, H. Wu, Expert systems with applications A case study of applying LRFM model in market segmentation of a children ' s dental clinic, Expert Syst. Appl. 39 (5) (2012) 5529–5533, http://dx.doi.org/10.1016/j.eswa.2011.11.066.

[38] P.W. Murray others, Forecasting supply chain demand by clustering customers, IFAC-PapersOnLine 48 (3) (2015) 1834–1839, http://dx.doi.org/10.1016/j.ifacol.2015.06.353.

[39] R.S. Collica, Customer Segmentation and Clustering using SAS Enterprise Miner, third ed., SAS Institute Inc, 2017.

[40] K. Coussement, F.A.M. Van den Bossche, K.W. De Bock, Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees, J. Bus. Res. 67 (1) (2014) 2751–2758, http://dx.doi.org/10.1016/J.JBUSRES.2012.09.024.

[41] A.X. Yang, How to develop new approaches to RFM segmentation, J. Targeting, Meas. Anal. Mark. 13 (1) (2004) 50–60, http://dx.doi.org/10.1057/palgrave.jt.5740131.

[42] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Appl. Soft Comput. J. 11 (2) (2011) 2664–2675, http://dx.doi.org/10.1016/j.asoc.2010.10.015.

[43] S. Taylor, B. Letham, Forecasting at scale, Am 72 (1) (2018) 37–45, http://dx.doi.org/10.1080/00031305.2017.1380080.

[44] H. Abbasimehr, M. Shabani, M. Yousefi, An optimized model using LSTM network for demand forecasting, Comput. Ind. Eng. 143 (2019) 106435, http://dx.doi.org/10.1016/j.cie.2020.106435.

[45] M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, Appl. Soft Comput. J. 11 (2) (2011) 2664–2675, http://dx.doi.org/10.1016/j.asoc.2010.10.015.

[46] A. Sagheer, M. Kotb, Time series forecasting of petroleum production using deep LSTM recurrent networks, Neurocomputing 323 (2019) 203–213, http://dx.doi.org/10.1016/j.neucom.2018.09.082.

[47] A.R.S. Parmezan, V.M.A. Souza, G.E.A.P.A. Batista, Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model, Inf. Sci. (Ny). 484 (2019) 302–337, http://dx.doi.org/10.1016/j.ins.2019.01.076.

[48] Y. Bai, J. Xie, D. Wang, W. Zhang, C. Li, A manufacturing quality prediction model based on AdaBoost-LSTM with rough knowledge, Comput. Ind. Eng 155 (January) (2021) 1–10, http://dx.doi.org/10.1016/j.cie.2021.107227.

[49] C. Lemke, B. Gabrys, Meta-learning for time series forecasting and forecast combination, Neurocomputing 73 (10–12) (2010) 2006–2016, http://dx.doi.org/10.1016/j.neucom.2009.09.020.

[50] G. Li, J. Shi, J. Zhou, BayesIan adaptive combination of short-term wind speed forecasts from neural network models, Renew. Energy 36 (1) (2011) 352–359, http://dx.doi.org/10.1016/j.renene.2010.06.049.

[51] N.C. Das Adhikari, R. Garg, S. Datt, L. Das, S. Deshpande, A. Misra, Ensemble methodology for demand forecasting, in: Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2017, No. Iciss, 2018, pp. 846–851, http://dx.doi.org/10.1109/ISS1.2017.8389297.

[52] K.I. Papageorgiou, K. Poczeta, E. Papageorgiou, V.C. Gerogiannis, G. Stamoulis, Exploring an ensemble of methods that combines fuzzy cognitive maps and neural networks in solving the time series prediction problem of gas consumption in Greece, Algorithms 12 (11) (2019) 1–27, http://dx.doi.org/10.3390/a12110235.

[53] K. Bandara, C. Bergmeir, S. Smyl, Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach, Expert Syst. Appl. 140 (2020) http://dx.doi.org/10.1016/j.eswa.2019.112896.

[54] H. Abbasimehr, M. Shabani, A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques, J. Ambient Intell. Humaniz. Comput. 12 (1) (2021) 515–531, http://dx.doi.org/10.1007/s12652-020-02015-w.

[55] M. Massaoudi, S.S. Refaat, I. Chihi, M. Trabelsi, F.S. Oueslati, H. Abu-Rub, A novel stacked generalization ensemble-based hybrid LGBM-XGB-mlp model for short-term load forecasting, Energy 214 (2021) http://dx.doi.org/10.1016/j.energy.2020.118874.

[56] S. Zhang, Y. Chen, W. Zhang, R. Feng, A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting, Inf. Sci. (Ny). 544 (2021) 427–445, http://dx.doi.org/10.1016/j.ins.2020.08.053.

[57] B. Rostami-Tabar, J.F. Rendon-Sanchez, Forecasting COVID-19 daily cases using phone call data, Appl. Soft Comput. 100 (2021) http://dx.doi.org/10.1016/j.asoc.2020.106932.

[58] C. Beneditto, A. Satrio, W. Darmawan, B.U. Nadia, N. Hanafiah, Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET, Procedia Comput. Sci. 179 (2020) (2021) 524–532.

[59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, Scikit-learn: Machine learning in python, J. OfMachine Learn. Res. 12 (2011) 2825–2830, http://dx.doi.org/10.1289/EHP4713.

[60] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, USA, 2013.

[61] I.F. Chen, C.J. Lu, Sales forecasting by combining clustering and machine-learning techniques for computer retailing, Neural Comput. Appl. 28 (9) (2017) 2633–2647, http://dx.doi.org/10.1007/s00521-016-2215-x.

[62] G. Carlsson, F. Mémoli, A. Ribeiro, S. Segarra, Hierarchical clustering of asymmetric networks, Adv. Data Anal. Classif. 12 (1) (2018) 65–105, http://dx.doi.org/10.1007/s11634-017-0299-5.

[63] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, Inc., 1988.

[64] Y. Zhou, F.J. Chang, H. Chen, H. Li, Exploring copula-based Bayesian model averaging with multiple ANNs for PM2.5 ensemble forecasts, J. Clean. Prod. 263 (2020) 121528, http://dx.doi.org/10.1016/j.jclepro.2020.121528.

[65] F. Constante, F. Silva, A. Pereira, Dataco smart supply chain for big data analysis, mendeley data, 2019, https://data.mendeley.com/datasets/8gx2fvg2k6/5.