systems

MDPI

Article

An Approach for Multi-Item Product Sales Forecasting Based on Advancing the BCG Matrix with Matrix-Clustering and Time Modeling Techniques

Che-Yu Hung and Chien-Chih Wang *

Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City 243303, Taiwan; cheyu219@mail.mcut.edu.tw
* Correspondence: ieccwang@mail.mcut.edu.tw; Tel.: +886-2-2908-9899

Abstract: Customized production has greatly diversified product categories, which has altered product life cycles and added complexity to business management. This paper introduces a matrix-clustering technique that integrates k-means clustering with the BCG Matrix, enhanced by time modeling, to offer a comprehensive framework for multi-item product sales forecasting. The approach builds upon existing BCG Matrix outcomes, re-clustering high-selling products more precisely and redefining their relationship with other product lines more objectively. This method addresses the challenge of forecasting situations with limited historical data, providing more accurate sales predictions. Using Taiwan's sales data, an empirical study on integrated circuit tray products demonstrated the effectiveness of the matrix clustering technique. The results showed improved data utilization, increasing from 35.93% with the original BCG analysis to 52.43% with the combined matrix-clustering and time modeling methods. This study contributes to academic research by presenting a portfolio analysis approach rooted in matrix clustering, which systematically enhances traditional BCG Matrix methods. The proposed framework is adaptable to the unique traits of different portfolios, offering businesses workflows that are efficient, reliable, sustainable, and scalable.

Keywords: BCG matrix; clustering technology; product forecasting; sustainable manufacturing

check for updates

Citation: Hung, C.-Y.; Wang, C.-C. An Approach for Multi-Item Product Sales Forecasting Based on Advancing the BCG Matrix with Matrix-Clustering and Time Modeling Techniques. *Systems* **2024**, *12*, 388. https://doi.org/10.3390/systems12100388

Academic Editor: Mingche Lee

Received: 25 July 2024
Revised: 17 September 2024
Accepted: 22 September 2024
Published: 25 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In today's dynamic market environment, companies providing customized products maintain finished goods inventories to handle urgent orders and manage temporary returns [1,2]. While this approach supports short-term demand fulfillment, it also introduces significant forecasting risks. Specifically, overestimating demand may result in overproduction, leading to surplus inventory and financial strain, whereas underestimating demand can cause stockouts and customer dissatisfaction. To mitigate these risks, firms typically lean toward demand underestimation [3]. However, this short-term focus on forecasting accuracy neglects the model's sustainability, creating a critical gap in achieving long-term reliability and efficiency.

This gap has become increasingly apparent with the recent surge in demand for integrated circuit (IC) trays, essential for packaging and testing precision IC components. Global disruptions such as the COVID-19 pandemic have amplified these challenges, exacerbated by volatile raw material costs, geopolitical instability, and rapid technological advancements. In this context, accurate demand forecasting is crucial for reducing operational costs, improving profitability, and enhancing customer satisfaction. Moreover, the growing trend of recycling disposable packaging and retrofitting older products for resale has further complicated the forecasting process [4]. This emerging business model disrupts traditional product lifecycle assumptions, presenting new challenges for demand forecasting systems.

Given these complexities, this study seeks to develop a sustainable and practical forecasting framework tailored to multi-item product portfolios. Conventional forecasting models often focus on single products, but the proposed framework addresses the more intricate dynamics of multi-product environments. Using IC tray sales as a case study, this research aims to overcome the limitations of traditional forecasting methods. The framework includes real-world constraints such as aggregating daily data into weekly intervals, applying fixed unit costs, anonymizing customer information, and assuming no product substitution.

The primary objective is to design a robust and adaptable sales forecasting framework capable of managing the complexities of multi-item portfolios. By integrating K-means clustering, the BCG Matrix, and time-series modeling, the framework seeks to enhance forecasting accuracy and scalability, particularly in industries with scarce historical data or volatile market conditions. This research extends existing portfolio analysis techniques and time-series forecasting models, offering a comprehensive approach to prediction in dynamic markets.

The study is guided by two main research questions: First, what portfolio characteristics can be identified for top-selling products using K-means clustering? Second, how do matrix-based and cluster-based product portfolios compare regarding forecasting performance? The research demonstrates that K-means clustering can improve sales forecasting by aligning inventory with market demand, reducing inventory pressure, and increasing forecasting accuracy. The methodology combines the BCG Matrix, K-means clustering, and time-series modeling to create a robust forecasting framework. The BCG Matrix categorizes products based on market growth and relative market share, while K-means clustering dynamically refines these categories by grouping similar products. Time-series modeling further enhances the framework by predicting future sales trends, even with limited or incomplete historical data.

The paper is structured as follows: Section 2 reviews existing literature on time-series forecasting, the BCG Matrix, portfolio analysis, K-means clustering, and multi-item sales forecasting. Section 3 details the research methodology, highlighting the integration of K-means clustering with the BCG Matrix. Section 4 presents the empirical results, with the findings discussed in Section 5. Finally, Section 6 concludes by summarizing key insights and proposing directions for future research.

2. Literature Review

This section organizes the literature review under key subheadings, covering classic and modern time-series forecasting methods, the BCG Matrix and product portfolio analysis, K-means clustering and market segmentation, and multi-item product sales forecasting. The tools and approaches discussed in these studies provide a foundation for developing the specific framework proposed in this research for forecasting multi-item product sales.

2.1. Classic and Modern Time Series Forecasting Methodologies

Time series forecasting includes both traditional statistical methods and modern machine learning approaches. Traditional techniques, such as Naïve forecasting, exponential smoothing, moving average, autoregressive moving average (ARMA), and ARIMA, are grounded in statistical learning principles [5–9]. In contrast, modern approaches utilize advanced algorithms like support vector regression, k-nearest neighbor, artificial neural networks, recurrent neural networks (RNN), long short-term memory (LSTM), and hybrid models that integrate exponential smoothing with RNNs [10–16]. This wide range of methods enables practitioners to select the most appropriate technique based on the specific characteristics of the time series data and the forecasting objectives.

2.2. BCG Matrix and Product Portfolio Analysis

Barksdale and Harris (1982) emphasized the importance of the product life cycle and the BCG Matrix as critical tools for developing market strategies [17]. Combining

these two models, they introduced a comprehensive product-lifecycle portfolio matrix, identifying market growth as a shared factor between the product life cycle and the BCG Matrix. However, they did not provide a specific formula or empirical evidence to support their claims on market growth. Stummer and Heidenberger (2003) proposed a three-stage portfolio analysis approach: project identification, portfolio filtering, and balancing resource and benefit categories. Their evaluation of nine research and development projects suggested that revisions were needed in the third stage [18]. Udo-Imeh, Edet, and Anani analyzed various portfolio models, such as the BCG Matrix, GE Matrix, Shell Directional Policy Matrix, and Arthur D. Little Strategic Condition Matrix, noting the unique metrics each model employs [19].

In his article "An Analysis of BCG Growth Sharing Matrix", Mohajan (2017) [20] argued that the BCG Matrix, also known as the Product Portfolio Matrix, is an effective business planning tool. It helps companies assess the competitive position of their product portfolios, focusing on strategy, cash flow, and profitability. Mohajan provided a formula for calculating the market growth rate and clarified the assumptions of the BCG Matrix using business case studies. For example, he classified Apple's products into different BCG Matrix categories: iPhones as Stars, iPods as Cash Cows, and the MacBook Air as a Question Mark. He also noted the necessity of products in the Dogs category, as Apple continuously updates its popular items. Despite its simplicity, the traditional BCG Matrix has limitations, particularly in data acquisition and static nature. To address these issues,

Nowak et al. (2020) proposed two novel methodologies for product portfolio analysis [21]. The initial methodology is the Grey Portfolio Analysis method, which employs Grey System Theory and expert scoring on market growth rate, relative market share, and turnover. This is exemplified through an IT company's case study on cloud computing service products. This method offers a dynamic approach to decision-making in complex market environments. In their second study, entitled "Product Portfolio Analysis Towards Operationalizing Science-Based Targets," the authors proposed the TEI-PE matrix as an environmentally sustainable alternative to the BCG Matrix. This alternative uses total environmental impact and profit as metrics rather than the traditional ones. This research emphasizes the necessity of aligning product strategies with science-based targets and sustainability, particularly for high-margin products, while recommending eliminating low-margin ones. An empirical study involving four Apple products employed the average growth rate as a reference line in the BCG Matrix to illustrate this approach.

Chiu and Lin (2020) introduced an innovative adaptation of the traditional BCG growth-share matrix by integrating a rule-based system for product portfolio analysis [22]. Using large datasets, they demonstrated a modern approach that leverages principles of artificial intelligence and software engineering. This system automates product classification and remains flexible, allowing companies to adjust rules and standards as market conditions evolve.

Kader and Hossain (2020) further discussed the BCG growth-share matrix as a strategic tool for evaluating product portfolios and making informed decisions [23]. A well-constructed matrix aids in resource allocation, strategic planning, and lifecycle management. However, the matrix oversimplifies complex market dynamics. It assumes market growth is the sole indicator of attractiveness, overlooking other factors such as technological changes, regulatory environments, customer preferences, and product interdependencies.

2.3. k-Means Clustering and Market Segmentation

K-means clustering, originally introduced in signal processing, is a technique used to divide a set of observations into distinct groups or "clusters". Each cluster is defined by its center point, known as the "centroid", and observations are grouped based on their proximity to these centroids. Various methods exist for determining the optimal number of clusters, including industry standards, the elbow method [24], the Baltic Dry Index, or specific cross-validation techniques. Once the optimal number is established, K-means clustering minimizes intra-cluster variation while maximizing inter-cluster variation. The

initial selection of centroids and distance metrics can affect the results. Syakur, Khotimah, Rochman, and Satoto (2018) recommended combining K-means clustering with the elbow method to identify the optimal number of customer profile clusters [25]. They emphasized the significance of considering business transactions and customer interactions when defining the number of clusters. As new data are introduced, the value of k and the centroids must be updated. While K-means clustering is an unsupervised learning method that does not directly predict outcomes, it can be combined with supervised learning techniques to enhance predictive models' usability and long-term effectiveness [26,27].

Cluster-based analysis is frequently applied in business development, customer management, and marketing, with existing literature predominantly focusing on structured data in these fields. For instance, Hosseini, Maleki, and Gholamian (2010) employed K-means clustering to analyze customer loyalty [14], developing a model based on product availability, purchase frequency, and purchase volume. Their findings revealed that this approach improved classification accuracy and enhanced customer relationship management. More recently, cluster-based analysis has been extended to unstructured image data, finding applications in fields such as medical diagnostics [28–31] and agricultural product identification [32,33].

Recent research has explored the integration of time-series methods and clustering. In 2020, Guijo-Rubio et al. introduced a novel approach to clustering time-series data, focusing on characterizing segment typologies to enhance the clustering process by considering the inherent patterns within time-series segments [34]. A key challenge in applying this approach is ensuring the effectiveness of segment typology patterns, which may not hold under real-world intermittent demand. Bandara, Bergmeir, and Smyl (2020) proposed an innovative time series forecasting method that combines clustering with recurrent neural networks (RNNs) [35]. By clustering similar time series and training RNNs on these clusters, their method improved accuracy and scalability, making it a valuable tool for businesses handling large-scale time series data. Alqahtani et al. (2021) reviewed deep learning-based time-series clustering methods, highlighting advancements and applications in this area [36]. Their work demonstrates the potential of these techniques to address the unique challenges posed by time-series data and improve clustering outcomes across diverse domains.

Hung et al. (2022) conducted a BCG Matrix classification of hot-selling products, addressing the challenges of missing data caused by intermittent demand [37]. They categorized these products into four classes, Stars, Cash Cows, Dogs, and Problem Children, providing actionable insights for managing product portfolios.

2.4. Multi-Item Sales Forecasting

Recent literature on multi-item sales forecasting can be categorized into three main areas. The first addresses the influence of seasonal factors and external variables, primarily related to model and feature selection. The second examines forecast errors and their impact on system performance. The third focuses on the connection between forecasting and inventory management, particularly regarding model application and execution.

Bunn and Vassilopoulos (1999) analyzed short-term trends for multiple commodities and found that grouping similar items improved forecasting accuracy [38]. They proposed and tested three seasonal decomposition tools using UK retail data, reducing forecast errors and enhancing model precision. Xie, Lee, and Zhao (2004) investigated how forecast errors affect a multi-item production system's performance using computer simulation data. Their findings indicated that forecast errors significantly impacted total costs, schedule stability, system service levels, and the selection of master production schedule freezing parameters [39]. Additionally, they noted that capacity tightness and cost structure influenced the extent of forecasting error effects on system performance.

Taylor (2011) introduced a seasonal exponential smoothing method to forecast monthly sales for a publishing company [40]. His results showed that this approach outperformed other models in cases with strong seasonality, and that simpler models were effective

after accounting for outliers. Spedding and Chan (2000) explored predictive modeling for time series with short product life cycles and long lead times, focusing on inventory management for manufacturers in Singapore [41]. They developed a Bayesian theory-based model and compared it to ARIMA, finding that their model better handled non-stationarity and seasonality. However, they also highlighted the need for a broader range of historical data to represent inventory conditions effectively.

3. Materials and Methods

This study proposes an integrated forecasting framework to address the complex challenges of multi-item product sales prediction by combining K-means clustering, the BCG Matrix, and time-series modeling. The framework is designed to improve the precision of sales forecasts by leveraging both portfolio analysis and time-based sales patterns, thereby enabling more informed decision-making for product management across diverse markets. Each framework component plays a distinct role, and their combined application ensures a comprehensive approach to forecasting.

The BCG Matrix serves as the foundation for product portfolio categorization. As a widely recognized strategic tool, the BCG Matrix classifies products based on two critical metrics: Market Growth Rate (MGR) and Relative Market Share (RMS). Products are grouped into four categories—Stars, Cash Cows, Dogs, and Question Marks—each reflecting different levels of market performance and growth potential. This categorization provides a strategic overview that helps companies prioritize investments and allocate resources according to each product's market position. The BCG Matrix offers insights into a product's competitive standing, which guides the initial grouping of products and supports long-term strategic planning.

Following this broad categorization, K-means clustering is employed to refine further segmentation within each BCG category. K-means clustering uses MGR and RMS values to create more specific clusters based on market performance and sales characteristics. This approach allows for a more granular segmentation of products, identifying clusters of similar products within each BCG Matrix category. By grouping products with comparable market behaviors, K-means clustering ensures that forecasting models are tailored to clusters of products with homogeneous characteristics. This stage enhances the accuracy of product classification by moving beyond the broader strategic categories of the BCG Matrix to a more detailed, data-driven segmentation.

After segmentation, time-series modeling is applied to project future sales patterns for each product cluster. Advanced time-series techniques, such as ARIMA, LSTM, and naïve forecasting, predict sales trends based on historical data. Model selection depends on the nature of the sales data, which may be regular, intermittent, or highly seasonal. A representative product from each cluster, selected based on proximity to the cluster centroid, is the predictor for the entire group. Time-series models are applied to the representative product's sales data, allowing the framework to generate accurate forecasts for the whole cluster. This method ensures scalability and adaptability, accommodating different data types and sales behaviors.

By integrating the BCG Matrix, K-means clustering, and time-series modeling, this framework provides a robust and flexible solution for multi-item sales forecasting. The BCG Matrix offers high-level strategic classification, K-means clustering refines product segmentation into more precise clusters based on performance, and time-series modeling enhances predictive accuracy by incorporating historical sales patterns for representative products. This integrated approach improves the reliability of sales forecasts, particularly in complex product portfolios with limited historical data. The framework enhances forecast accuracy and provides a scalable solution that adapts to shifting market conditions and varying product types. It is a valuable tool for businesses aiming to optimize their sales forecasting capabilities.

3.1. Forecasting Method

The Naïve forecast, a simple and intuitive method for time series prediction, is one of the most used forecasting tools in business [42]. ARIMA, a statistical learning model, has been widely applied in time series forecasting for decades [8]. LSTM, introduced by Hochreiter et al. (1997), is a deep learning approach that has gained significant popularity for time series forecasting over the past decade [43]. This study employs these three forecasting methods for the empirical analysis.

3.1.1. Naïve Forecast

A naive forecast is an intuitive and quick-to-apply time forecaster that uses previous valid data to predict the next. In general, a naive forecast is formed as follows:

$$y_t = \hat{y}_{t+1} \tag{1}$$

where y_t is the observation at time t and $\hat{y}_{t+1}$ is the forecast at time $t+1$. This method works remarkably well for many economic and financial time series [42].

3.1.2. Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model, which combines autoregressive and moving average models, requires the series to be stationary and involves a selection of series transformations. It is defined by three primary parameters: p for the autoregressive model, d for the number of differencing steps, and q for the moving average model. A nonseasonal ARIMA model combines differencing with an autoregressive and a moving average model. It is expressed as ARIMA (p, d, q), and the formula is:

$$y'_t = c + \delta_1 y'_{t-1} + \delta_2 y'_{t-2} + \cdots + \delta_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{2}$$

where y'_t represents the forecast of the transformed series after differencing. The model's predictors encompass lagged values of y_t alongside the corresponding lagged errors. The parameters p and q signify the order of the autoregressive and moving average models, respectively, while d represents the number of differencing steps necessary to integrate these two models. First, in constructing the ARIMA model, it is essential to execute both the augmented Dickey–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests to validate that the transformed series (post-differencing) exhibits stationarity. Subsequently, determining the optimal combination of p and q is paramount, followed by confirming whether the residuals adhere to a normal distribution.

3.1.3. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a variant of recurrent neural networks that features components such as input, output, and forget gates, along with a cell. These gates regulate the flow of information into and out of the cell, with values ranging from 0 to 1. A value of 0 indicates complete information disregards, while 1 signifies complete information retention. Four operations are carried out in a single cell from the input (x_t) to the output (h_t). These operations are divided into three stages and then compiled, using addition and multiplication, until the output (h_t) is generated.

In the initial step, the function determines the information that needs to be omitted from the cell state. This standard sigmoid function encompasses the input, the previous cell's output h_{t-1}, the weighted decay cap W_t, and the bias b_f. The output of this function is a value between 0 and 1, which is then multiplied by the cap C_{t-1} of a previous cell and forwarded. The f_t is expressed as follows:

$$f_t = \sigma\Big(W_f \cdot [h_{t-1}, x_t] + b_f\Big), 0 \le f_t \le 1 \tag{3}$$

In Step 2, two functions are employed, with i_t representing the retained information and $\overline{C}_t$ denoting a new candidate vector. These functions utilize the hyperbolic tangent

(tanh) to generate the vector. Step 2 governs the determination of additional information to be retained in the cell state, which is then combined with the output from Step 1 and progresses further. The expressions for i_t, $\overline{C}_t$, and C_t are as follows:

$$\widetilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_C) i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \quad (4)$$

In Step 3, the output function o_t is determined by a sigmoid function that incorporates the input x_t, the previous cell h_{t-1} output, the weighted decay W_o, and the bias b_o. The function o_t is executed to determine the information exiting the cell and is described as follows:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

The final output h_t of the cell is determined by taking the last update of the cell state, denoted as C_t, from Step 2. The cell state update cap C sub t is initially processed by applying a hyperbolic tangent (tanh) function to produce values from -1 to 1. These values are then multiplied by the output function C_t to yield the cell's ultimate output h sub t.

$$h_t = o_t \times tanh(C_t) \quad (6)$$

In this study, combining the three time-series forecasters (Naive forecast, ARIMA, LSTM) and the two methods (zero-filling, mean-impute) for handling missing data yields a total of six models for training. The six combined models are Naive forecast + zero-filling (N + Z), Naive forecast + Mean-impute (N + M), ARIMA + zero-filling (A + Z), ARIMA + mean-impute (A + M), LSTM + zero-filling (L + Z), and LSTM + mean-impute (L + M).

3.2. BCG Matrix and Product Portfolio

Mohajan et al. (2017) highlighted the BCG Matrix as a vital Product Portfolio Matrix, playing a pivotal role in business planning by assessing a company's competitive position in the market [20]. This tool helps companies strategically manage and balance cash flow, profitability, and overall strategy. The BCG Matrix employs a 2×2 framework to classify products into four categories—Stars, Cash Cows, Dogs, and Problem Child—based on their Relative Market Share (RMS) and Market Growth Rate (MGR). It enables firms to evaluate their products' or brands' competitive standing in local and global markets through the Product Portfolio Matrix.

3.2.1. Portfolio Category and Market Strategy

These four product categories in the BCG Matrix possess distinct characteristics that are closely aligned with the company's investment plans and market strategies.

- Stars: These products lead in high-growth markets, exhibiting both high growth rates and market share. While they theoretically generate cash, Stars require substantial investment to maintain their growth advantage. The primary goal for Stars is to maintain a balanced net cash flow, and their market strategy focuses on building and sustaining market share.
- Cash Cows: Cash Cows dominate in low or negative-growth markets and are characterized by high margins, resulting in strong positive cash flow and minimal investment requirements. These products can finance their own growth, and surplus cash can be redirected to support Stars or Problem Child products that could potentially evolve into future Cash Cows. The strategy here is to maintain market share.
- Dogs: These products have low market share in declining markets and neither generate significant cash nor justify further investment. In some cases, if Dogs show potential to evolve into Problem Child or Cash Cow products, the strategy might involve repositioning. However, if they fail to meet this criterion, the next step is typically to harvest or liquidate these products by discontinuing them and removing them from the product line.

- Problem Child: These products exhibit high growth rates but low market share. They require significant investment yet provide minimal short-term returns. Since Problem Child products have not yet achieved market dominance, they do not generate substantial cash. To gain market leadership, the company must invest heavily to build and sustain market share.

Given the uncertainty of market competition, companies must continuously update their BCG Matrix to reflect changing competitive positions and adjust their market strategies accordingly. The formulation and execution of these strategies are closely linked to customer demand, production scheduling, inventory management, and cash flow planning.

- Build: This strategy focuses on increasing market share by driving product sales. Since market share is a long-term objective, short-term profits are often sacrificed, with any generated cash reinvested into the market.
- Hold/Maintain: This strategy aims to preserve the current market share while continuing to generate significant cash flow, which is typically invested in other products.
- Harvest: This strategy is applied to weaker Cash Cows, Problem Child, and Dog products that lack future potential. The aim is to maximize short-term cash flow, even if this involves actions like raising prices or cutting costs, potentially at the expense of long-term benefits.
- Divest/Liquidation: This strategy applies to products with no future, such as Dogs with rapidly declining sales or Problem Child products with little chance of becoming Stars. The objective is to eliminate these products, negatively affecting the company's financial performance. Resources allocated to these products can then be redirected to more promising opportunities.

3.2.2. Relative Market Share and Market Growth Rate

Relative Market Share (RMS) and Market Growth Rate (MGR) are the primary indicators used to construct a BCG Matrix, with RMS plotted on the x-axis and MGR on the y-axis. This 2×2 matrix aids in determining a firm's competitiveness or a product's market position, both locally and globally. From a practical perspective, RMS reflects cash generation capabilities, while MGR represents cash usage performance. These two indicators can also be used to assess overall business health. Products with a high market share or strong growth potential are more likely to generate substantial profit margins.

RMS is always a positive value, ranging from 0.000 to 1.000. A midpoint of 0.500 is often used to indicate market share status, serving as a reference line when constructing the BCG Matrix. Specifically, an RMS of $\geq$0.500 denotes a high market share, while an RMS below 0.500 indicates an average market share. The equation for calculating RMS is as follows:

$$\text{RMS} = \frac{\text{Firm or Brand Sales this year}}{\text{Leading competitor Sales this year}} \times 100\% \quad (7)$$

Market Growth Rate (MGR) can take on both significantly positive and negative values. However, sales data from the previous fiscal year are required for calculation. Given the variety of products being compared, the standard practice is to use MGR = 0.000 as a reference line to distinguish between growing (MGR $\geq$ 0.000) and shrinking (MGR < 0.000) products. The equation for calculating MGR is as follows:

$$\text{MGR} = \frac{\text{Firm or Brand's Sales this year} - \text{Firm or Brand's Sales last year}}{\text{Firm or Brand's Sales this year}} \times 100\% \quad (8)$$

3.2.3. Mean Absolute Scaled Error (MASE)

Mean Absolute Scaled Error (MASE) was first introduced by Hyndman and Koehler in 2005 as a scale-independent measure of forecast accuracy [44]. This metric can handle zero values directly and applies equal weight penalties to both positive errors (overestimation) and harmful errors (underestimation), addressing the major limitations of the commonly

used Mean Absolute Percentage Error (MAPE). For nonseasonal time series data, MASE is calculated using the following equation:

$$\text{MASE} = mean\left(\frac{|e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|Y_i - Y_{i-1}|}\right) = \frac{\frac{1}{J}\sum_j |e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|Y_i - Y_{i-1}|} \quad (9)$$

The denominator in the MASE equation represents the mean absolute error of a one-step Naïve forecast applied to a training set with n data points. When the time series exhibits seasonality, the period t of the training set is adjusted accordingly. A MASE value below 1 indicates that the proposed model produces fewer errors than the one-step Naïve forecast, reflecting superior performance. Additionally, a smaller MASE value corresponds to better forecasting accuracy. In contrast, a MASE value approaching 1 suggests performance comparable to the Naïve forecast, while a value greater than 1 indicates poorer performance relative to the Naïve forecast.

3.2.4. Within-Mean Difference

Hung et al. (2022) introduced the Within-mean Difference (WD) indicator, which quantifies the difference between the mean forecasts and the actual mean as a percentage [37]. This measure is applied to the validation set and is particularly relevant for companies managing specific product inventories. When no additional stock is added, the effect of underestimated forecasts is amplified, while the impact of overestimated forecasts is diminished. This metric, also referred to as inherent deployment (WDI), is calculated using the following equation:

$$\text{WD}_{\text{Validation}} = \text{WD}^{\text{I}} = \frac{\overline{\text{F}} - \overline{\text{A}}}{\overline{\text{F}}} \times 100\% \quad (10)$$

where $\overline{\text{F}}$ and $\overline{\text{A}}$ are the means of the forecasts and the actual values for the validation set, respectively. A WD value nearing zero indicates a marginal disparity between the average forecast and the actual mean, signifying that the models effectively capture the validation set's data. A positive WD value signifies an overestimation by the models, while a negative WD value suggests an underestimation. Models with a WD exceeding 100% or falling below −100% are advised against due to their inadequate performance.

3.3. Cluster-Based Forecasting for Multi-Item Products

This section discusses the conceptual implications of establishing an executable framework for multi-item product forecasting. The framework builds on insights from a journal article published in 2022. Using the same dataset introduced by Hung et al. (2022), this study re-clusters and bundles hot-selling and lower-ranked products to enhance the forecasting accuracy for lower-ranked items, thereby reducing the uncertainty and prediction instability associated with limited data [37].

Cluster analysis is a statistical technique that categorizes objects or individuals into groups (clusters) based on their similarity. It operates on the premise that items within the same group share more similarities than those in other groups. This approach can be applied to products, grouping them into specific portfolios exhibiting high internal similarity (homogeneity) and significant differences between groups (heterogeneity).

The BCG Matrix is a strategic tool that classifies products based on their relative market share (RMS) along the horizontal axis and market growth rate (MGR) along the vertical axis. The resulting four categories—Stars, Cash Cows, Dogs, and Question Marks—define different product portfolios. By incorporating cluster analysis into the BCG Matrix methodology, minimizing variation among products within each category and establishing clear distinctions between categories becomes possible. This approach relies on the principle of minimum variation within clusters, which is determined by the cluster centroid—the geometric center of the group. Ideally, the distance from any product in the cluster to the

centroid should be uniform. However, since the centroid's position can shift when new products are added, careful initial selection of the centroid is critical.

The proposed framework is divided into three distinct phases with ten specific tasks and is supported by two iterative feedback mechanisms to ensure continuous evaluation and improvement. These feedback loops allow flexibility to revisit earlier phases, enabling reassessment, adjustment, or optimization of portfolios based on evolving results. As shown in Figure 1, the framework illustrates the sequential tasks across the three phases. The integration of K-means clustering with the BCG Matrix forms the foundation of the forecasting process. At the same time, the feedback mechanisms facilitate iterative adjustments, continuously improving the model's accuracy and reliability.

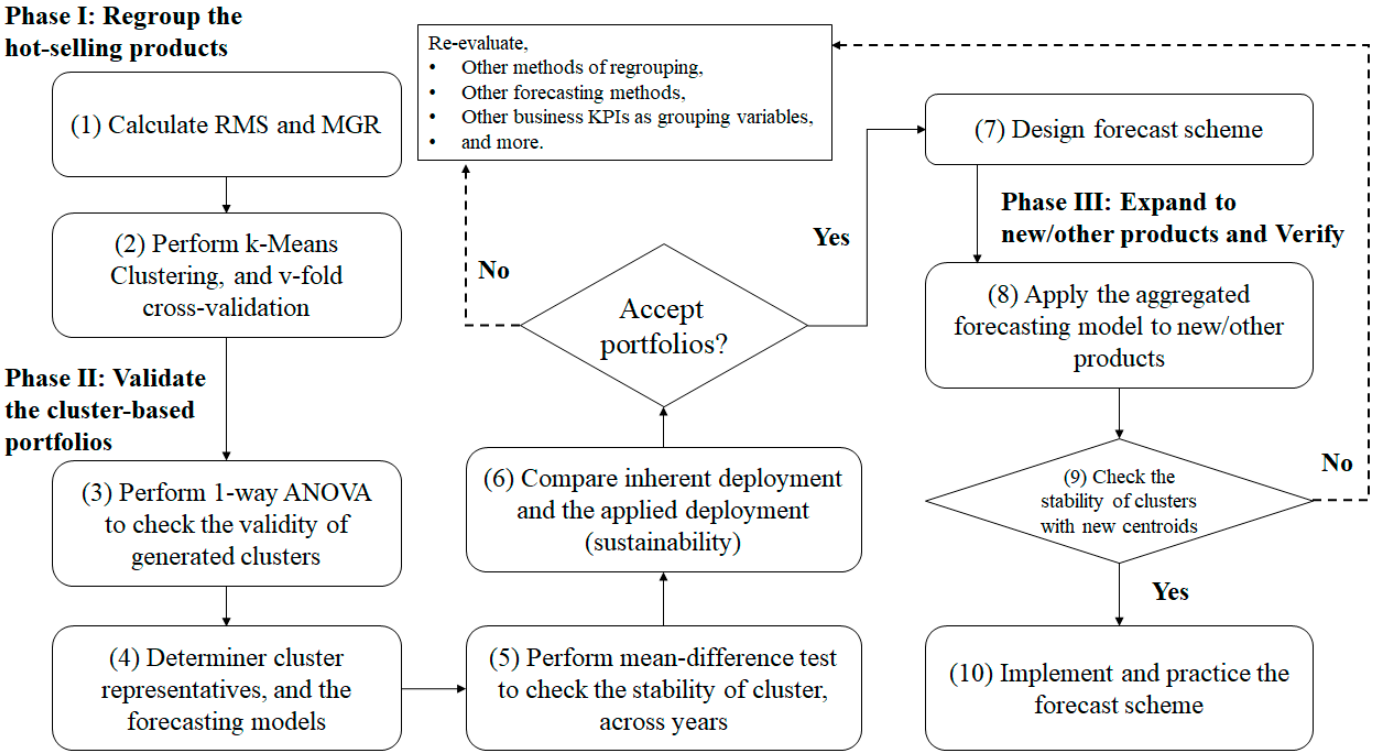


Figure 1. Framework for building a multi-item product sales forecasting.

3.3.1. Phase I: Regroup the Products

The first phase aims to regroup the products using the BCG Matrix and K-means clustering. This phase includes two critical tasks:

- Task 1: Calculate the RMS and MGR of the products.

Relative Market Share (RMS) and Market Growth Rate (MGR) are calculated for each product. RMS provides a comparative analysis of a product's market share relative to its competitors, while MGR reflects the growth or decline of the market for each product. This step focuses on identifying the top products in the portfolio for further analysis.

- Task 2: Perform k-means Clustering and use the v-fold cross-validation tool.

Once RMS and MGR are calculated, K-means clustering is used to segment products into clusters. The v-fold cross-validation method optimizes the clustering process, ensuring the clusters are distinct and internally homogeneous. The process stops when cross-validation achieves convergence, avoiding overfitting and ensuring optimal clusters are created.

3.3.2. Phase II: Validate the Cluster-Based Portfolios

This phase focuses on validating the portfolios formed in Phase I using several statistical methods and stability checks. The phase includes five tasks:

- Task 3: Perform a one-way Analysis of Variance to check the validity of clusters.

The one-way Analysis of Variance (ANOVA) is performed to validate whether the variables used in clustering have sufficient discriminating power. This test confirms that the clusters are distinct and suitable for further forecasting.

- Task 4: Determine cluster representatives and forecasting models.

The product closest to the centroid is selected for each cluster as the representative product. Based on its sales pattern, the time-series forecasting model—ARIMA, LSTM, or Naïve forecasting—is chosen for the representative product. This model will be used to predict future sales for the entire cluster.

- Task 5: Perform a Mean difference test to check the stability of clusters.

A mean difference test is conducted to assess the stability of each cluster over time. This test ensures that the products within the cluster maintain consistent performance, which is critical for forecasting accuracy.

- Task 6: Compare inherent deployment and applied deployment.

Two key indicators are compared: the inherent deployment (forecast mean value from the cluster representative) and the applied deployment (mean of actual values from each cluster member). If the conclusions from these indicators align, the portfolio is accepted. If they diverge, the portfolio is rejected, and the process returns to Phase I for re-evaluation using Turnback Mechanism 1.

- Task 7: Design a forecasting scheme.

This scheme comprises the portfolio outline, the forecasting scheme, and the proposed market strategy. The portfolio outline includes details about the product code, sales data, the chosen representative, and other portfolio members. The forecasting scheme provides information about the implemented model and the expected forecast gap. The market strategy follows the guidelines of the BCG Matrix. Three forecast formulas are provided for different scenarios: baseline forecast, optimism forecast (for underestimated cases), and preserved forecast (for overestimated cases).

$$\text{Baseline forecast} = \text{Actual sales} \times (1 + \text{Average MGR}) \tag{11}$$

$$\text{Optimism forecast} = \text{Actual sales} \times (1 + |\text{Forecast Gap}|) \tag{12}$$

$$\text{Preserved forecast} = \text{Actual sales} \times (1 - |\text{Forecast Gap}|) \tag{13}$$

The incorporation of three forecasting formulas—Baseline forecast, Optimism forecast, and Preserved forecast—enhances the adaptability of the forecasting model to varying market conditions, thereby improving its flexibility and reliability. The Baseline forecast represents the standard projection, assuming market growth aligns with historical trends, and is typically used for general production planning. The Optimism forecast is applied when actual sales surpass expectations, allowing businesses to adjust forecasts upward and avoid underproduction, capitalizing on favorable market conditions. In contrast, the Preserved forecast is employed when actual sales fall below projections, enabling a downward adjustment to prevent overproduction and excess inventory. By utilizing these three forecasting methods, the framework ensures that businesses can dynamically respond to positive and negative market fluctuations, optimizing inventory management and mitigating financial risk.

3.3.3. Phase III: Expend to New/Other Products and Verify

The final phase expands the forecasting scheme to include new products, ensuring the framework's scalability and reliability. This phase includes three tasks.

- Task 8: Apply the aggregated forecasting model to new or other products.

The forecasting model developed in the earlier phases is applied to new products. These products are categorized into existing clusters based on their RMS and MGR, ensuring that the model can accommodate new additions.

- Task 9: Check cluster stability with new centroids.

Including new products changes the cluster centroids, potentially affecting within-cluster variance. This task assesses whether the clusters remain stable after adding new products, ensuring that the expanded clusters can still provide reliable forecasts.

- Task 10: Implement the forecasting scheme and practice.

After passing the stability check, the expanded cluster is integrated into real-world operations. This step ensures the forecasting scheme remains adaptable and robust, even as market conditions and product portfolios change.

4. Analysis and Results

The original dataset used in this study is based on the research by Hung et al. (2022) [37]. Building on the 2022 results, this study introduces the K-means clustering method into the existing BCG Matrix analysis to more scientifically re-cluster hot-selling products and integrate lower-ranked products into the established clusters. This approach enables the case company to objectively redefine and apply the sales portfolio across both hot-selling and lower-ranked products. The following sections will present and interpret the results of real-world sales data analysis.

4.1. Background of the Used Data

The data for this study were obtained through an industry-academia collaboration with a Taiwan-based IC tray manufacturer. The feasibility of the proposed method was evaluated using real-world sales data from this collaboration. The dataset consists of sales records for plastic injection trays used as external packaging for consumer electronics chips. The dataset covers 1 January 2017 to 30 September 2019, offering a comprehensive sales overview.

The dataset comprises 60,090 daily records representing 576 distinct product items, with orders recorded on regular working days. Due to the large volume of data, weekly aggregation was conducted to streamline the analysis and ensure a more effective examination of trends over time. The aggregated data were then divided into three subsets, training, testing, and validation, each serving a specific role in the model's development and evaluation.

As shown in Table 1, the training set includes data from 1 January 2017 to 31 December 2018, covering 104 weeks and accounting for 72.72% of the total data. This set trains the models, enabling them to identify patterns and relationships within the data. The test set spans 1 January 2019 to 30 June 2019, comprising 26 weeks and representing 18.18% of the total data. It evaluates the model's ability to process unseen data and detect potential issues like overfitting. The validation set, covering the period from 1 July 2019 to 30 September 2019, includes 13 weeks, accounting for 9.09% of the data. The validation set assesses the model's capacity to generalize and perform accurately on data it has not previously encountered, providing an additional layer of confirmation regarding its robustness before deployment.

Table 1. Definition of Training, Test, and Validation sets.

Set	Period	# Of Weeks	The Use
Training	1 January 2017–31 December 2018	104 (72.72%)	Train the models
Test	1 January 2019–30 June 2019	26 (18.18%)	Test if the trained models are appropriate
Validation	1 July 2019–30 September 2019	13 (9.09%)	Validate the performance of trained models deployed on unused data

This approach ensures the robustness of the proposed forecasting model by training it on historical data, testing its performance on a separate dataset, and validating its forecast accuracy on a third dataset.

Table 2 summarizes the cumulative sales proportion for the top ten best-selling items in 2017, 2018, and from January to September 2019. The analysis reveals that a limited number of products consistently dominate the company's sales performance. Six products appeared on the top ten bestsellers list across all three periods, contributing significantly to the company's overall sales. These six products accounted for over 20% of total annual sales each year, representing 26.619% of sales in 2017, 20.658% in 2018, and 22.011% in the first nine months of 2019. This consistent performance underscores the critical importance of these products within the company's portfolio.

Table 2. Cumulated sales proportion (%) of the top ten products by years.

	2017	**2018**	**2019 January–September**
(1) Top ten products	44.516%	35.928%	35.480%
(2) The recurring six	26.619%	20.658%	22.011%
Difference = (1) − (2)	17.897%	15.2780%	13.469%

Comparing the cumulative sales proportions of the top ten products with those of the six recurring products reveals the contribution of other, less consistent top sellers. In 2017, the difference between the top ten products and the six recurring items was 17.897%, which narrowed slightly to 15.278% in 2018 and decreased to 13.469% in 2019. This indicates that while the six recurring products provide a stable foundation for sales, the remaining products in the top ten contribute varying degrees of additional sales each year. Understanding this distribution helps the company maintain appropriate inventory and production levels for these high-demand items while also identifying opportunities to boost sales of other products.

Interestingly, the remaining four items within the top ten products vary from year to year. Despite this variability, these products consistently account for 13–18% of annual sales, underscoring the diverse needs of customers and the complexities of managing a multi-item inventory. The annual variation in these items suggests a dynamic market where customer preferences and demands shift over time. This highlights the need for an adaptable sales strategy and a robust forecasting methodology to effectively respond to changing market conditions.

4.2. Evaluating Time Series Forecasting Models for Hot-Sell Products

Hung et al. (2022) developed time-series forecasting models for the top ten plastic tray products as part of an empirical study to compare sales forecasting performance [37]. To evaluate their effectiveness, these models were applied to two datasets, the Test and Validation sets. The performance of both sets was measured using Mean Absolute Scaled Error (MASE), while Within-Difference (WD) was used exclusively for the Validation set to assess deployment performance. Models with lower MASE values generally demonstrated a better fit on the Test set, while models with MASE values exceeding 1 indicated reduced accuracy and required further refinement. In contrast, lower WD values reflected a closer alignment with actual data in practical terms, with negative WD values signaling under-forecasting.

Table 3 summarizes the MASE and WD results for the top ten products, indicating the recommended and backup models based on their performance during the test and validation phases. Lower MASE values indicate superior forecasting accuracy, while WD measures the discrepancy between projected and actual demand, with negative values suggesting under-forecasting. In most cases, the recommended models outperformed or matched the backup models in efficacy. For example, the BGA 8X13mm product showed an improvement in MASE from 0.82548 (test) to 0.72481 (validation) using the recommended model (A + Z), with the backup model (A + M) yielding similar results. Additionally, for

TSOP I 12X20, the recommended model (N + Z) exhibited enhanced performance, with a lower MASE and more stable WD compared to the alternative model (N + M). In the case of BGA 11.5X13, although the backup model (L + Z) had a slightly better MASE during the test phase, the recommended model (L + M) outperformed it in the validation phase, showing a smaller WD.

Table 3. Summary of model evaluation of the top ten products.

Product	Model	Test (Model Evaluation)	Validation (Deployment)	
		MASE	MASE	WD
BGA 8X13mm	A + Z (Recommended)	0.82548	0.72481	−16.817%
	A + M (Backup)	0.84080	0.74119	−16.915%
TSOP I 12X20	N + Z (Recommended)	0.60660	0.70566	9.248%
	N + M (Backup)	0.96284	0.69797	9.248%
BGA 8X12.5	A + Z (Recommended)	0.41726	0.39119	−2.172%
	A + M (Backup)	0.40284	0.39119	−2.172%
TSOP II 54/86P	N + Z (Recommended)	0.38653	0.56681	12.534%
	N + M (Backup)	0.46659	0.56826	12.534%
BGA 7.5X13mm	N + Z (Recommended)	2.10839	2.16694	−1.414%
	N + M (Backup)	1.67242	1.89118	−1.414%
TQFP 7X7X1.4MM	N + Z (Recommended)	0.41522	0.64403	2.243%
	N + M (Backup)	0.41782	0.63529	2.243%
QFN 9X9	A + Z (Recommended)	0.90446	0.93259	1.973%
	A + M (Backup)	0.92847	0.93259	1.973%
BGA 11.5X13	L + Z (Backup)	0.78889	1.13025	−23.588%
	L + M (Recommended)	0.73817	1.17246	−25.090%
TQFP 14X14X1.4	N + Z (Backup)	0.99793	1.84759	−3.049%
	N + M (Recommended)	0.95987	1.83079	−3.049%
TSOP II 54/86 135′C	L + Z (Backup)	0.62501	0.63586	−12.818%
	L + M (Recommended)	0.56571	0.62717	−13.230%

Table 3 provides clear guidance for selecting the most suitable model, ensuring optimal forecasting performance across different products.

4.3. Establishing the Cluster-Based Portfolios

Table 4 analyzes the top ten items sold in 2018 using the K-means Clustering method. The variables analyzed include RMS18 and MGR18, representing relative market shares and market growth rates, respectively, based on 2018 sales data. The optimal number of clusters was determined to be three, following a 10-fold cross-validation process.

Table 4. Summary for k-Means Clustering.

Algorithm	k-Means
Used variable	RMS18, MGR18
Distance method	Euclidean distances
Initial centers	Maximize initial distance
Cross-validation	10-folds
Training error	0.119240
Number of clusters	3

Table 5 summarizes the clustering results for the top ten products sold in 2018. It includes each cluster's ID, centroid values (average RMS and MGR), and cluster membership. For example, Cluster 1 comprises three products—BGA 8X13mm, TSOP I 12X20, and BGA 8X12.5—with centroid coordinates of RMS = 0.87401 and MGR = 0.02329. Similarly, Cluster 2 consists solely of BGA 7.5X13mm, while Cluster 3 includes five products with centroid values of RMS = 0.51511 and MGR = −0.01351. This clustering approach helps identify common patterns in product behavior, enabling more customized demand forecasting and improved inventory management.

Table 5. Summary of generated clusters and centroids for the 2018 top ten products.

Cluster-ID	RMS	MGR	Member
1	0.87401	0.02329	BGA 8X13mm, TSOP I 12X20, BGA 8X12.5
2	0.64181	11.76121	BGA 7.5X13mm
3	0.51511	−0.01351	TSOP II 54/86P, TQFP 7X7X1.4MM, QFN 9X9, BGA 11.5X13, TQFP 14X14X1.4, TSOP II 54/86 135′C

Table 6 provides a detailed overview of the three clusters identified for the top ten products of 2018. It includes the cluster members, their individual RMS and MGR values, the centroids of each cluster, and the selected cluster representatives, along with the criteria for their selection.

Table 6. Summary of generated clusters for the top ten products in 2018.

Cluster-ID	Member	RMS	MGR	Cluster Rep. and the criteria
1	BGA 8X13mm	1.00000	0.32710	BGA 8X13mm. • The benchmark learning object. • The same direction (positive RMS, positive MGR) as the cluster centroid.
	TSOP I 12X20	0.83506	−0.11552	
	BGA 8X12.5	0.78697	−0.14172	
	Centroid 1	0.87401	0.02329	
2	BGA 7.5X13mm	0.64181	11.76121	BGA 7.5X13mm. • The only member.
	Centroid 2	0.64181	11.76121	
3	TSOP II 54/86P	0.65612	−0.41113	BGA 11.5X13. • The same direction (positive RMS, negative MGR) as the cluster centroid. • The shortest distance to the cluster centroid.
	TQFP 7X7X1.4MM	0.56781	0.08242	
	QFN 9X9	0.53835	0.48619	
	BGA 11.5X13	0.46862	−0.14220	
	TQFP 14X14X1.4	0.44934	0.28608	
	TSOP II 54/86 135′C	0.41042	−0.38220	
	Centroid 3	0.51511	−0.01351	

The K-means clustering process yielded several key insights, including the identification of new product groupings that aligned with the BCG Matrix classification. For example, Cluster 1 is represented by BGA 8X13mm, which was chosen due to its alignment with the cluster centroid (positive RMS and MGR). Cluster 2 consists solely of BGA 7.5X13mm, while Cluster 3 is represented by BGA 11.5X13, selected based on its proximity to the cluster centroid and its alignment with the centroid's characteristics (positive RMS, negative MGR).

Table 7 presents the results of a one-way ANOVA test, showing that the newly formed clusters are statistically significant at the 0.05 significance level. The *p*-values for RMS (0.003767) and MGR (0.000000) provide strong statistical evidence supporting the clustering results, validating the use of K-means clustering. These results confirm that the differences between clusters are not due to random variation and that the clustering process effectively distinguishes between product groups based on RMS and MGR.

Table 7. ANOVA table for RMS18 and MGR18.

	Between SS	Degree of Freedom	Within SS	Degree of Freedom	F	*p*-Value
RMS	0.2577	2	0.065617	7	13.7436	0.003767
MGR	124.5226	2	0.797897	7	546.2219	0.000000

Table 8 summarizes the results of the mean difference test and variance homogeneity for the three clusters. The *p*-values for Clusters 1 and 3 indicate no statistically significant changes in sales between 2017 and 2018, confirming the consistency of their performance. In contrast, the two-means test for Cluster 2 ($p = 0.000000$) shows a significant increase in sales for BGA 7.5X13mm from 2017 to 2018. However, since Cluster 2 consists of a single product, a variance homogeneity test could not be performed. These findings demonstrate the reliability and robustness of the K-means clustering methodology over time, particularly between 2017 and 2018, with most clusters showing stable sales patterns.

Table 8. Summary of mean difference test for three clusters.

Cluster-ID	Year	Mean (Std. Dev)	*t*-Test for Mean *p*-Value	*t*-Test for Mean 95% Confidence Interval	Homogeneity of Variance Levene p.	Homogeneity of Variance B-F p.
1	2017	1,912,200 (226,192.2)	0.919693 (NS)	(−540,147, 529,227)	0.879153 (NS)	0.920500 (NS)
	2018	1,917,660 (245,148.8)				
2	2017	1051 (1627.8)	0.000000 (S)	N/A	N/A	N/A
	2018	2250 (2734.0)				
3	2017	1,302,233 (617,621.2)	0.530495 (NS)	(−417,897, 761,956)	0.130384 (NS)	0.240139 (NS)
	2018	1,130,204 (197,831.8)				

Note. Type I error (alpha) for significance is set to 0.05. NS = Not significant. N/A = Not Available. Levene p. = *p*-value of Levene's test. B-F p. = *p*-value of Brown and Forsythe's test.

Figure 2 visually demonstrates why certain products are grouped together while others are separated. The next step involves evaluating the stability of these clustering results to validate their robustness, thereby concluding the overall analysis process.

The clustering analysis was extended to include products ranked 11–20 to further explore multi-item product forecasting. Table 9 presents the mean difference test results for the top twenty products from 2017 to 2018, categorized by the three-cluster definition. The results for Clusters 1 and 3 remain consistent with those for the top ten products, showing no statistically significant changes in sales between the two years. In Cluster 2, the *p*-value for the mean difference test was 0.061747, which is close to the 0.05 significance level but does not meet the threshold for statistical significance. The Levene test produced a *p*-value of 0.023436, indicating heterogeneous variance within Cluster 2, reflecting its significant growth. Despite the observed variance, the mean sales difference among cluster members was not statistically significant. These results reinforce the stability of the clustering structure while acknowledging some variability in Cluster 2, likely due to its rapid growth.

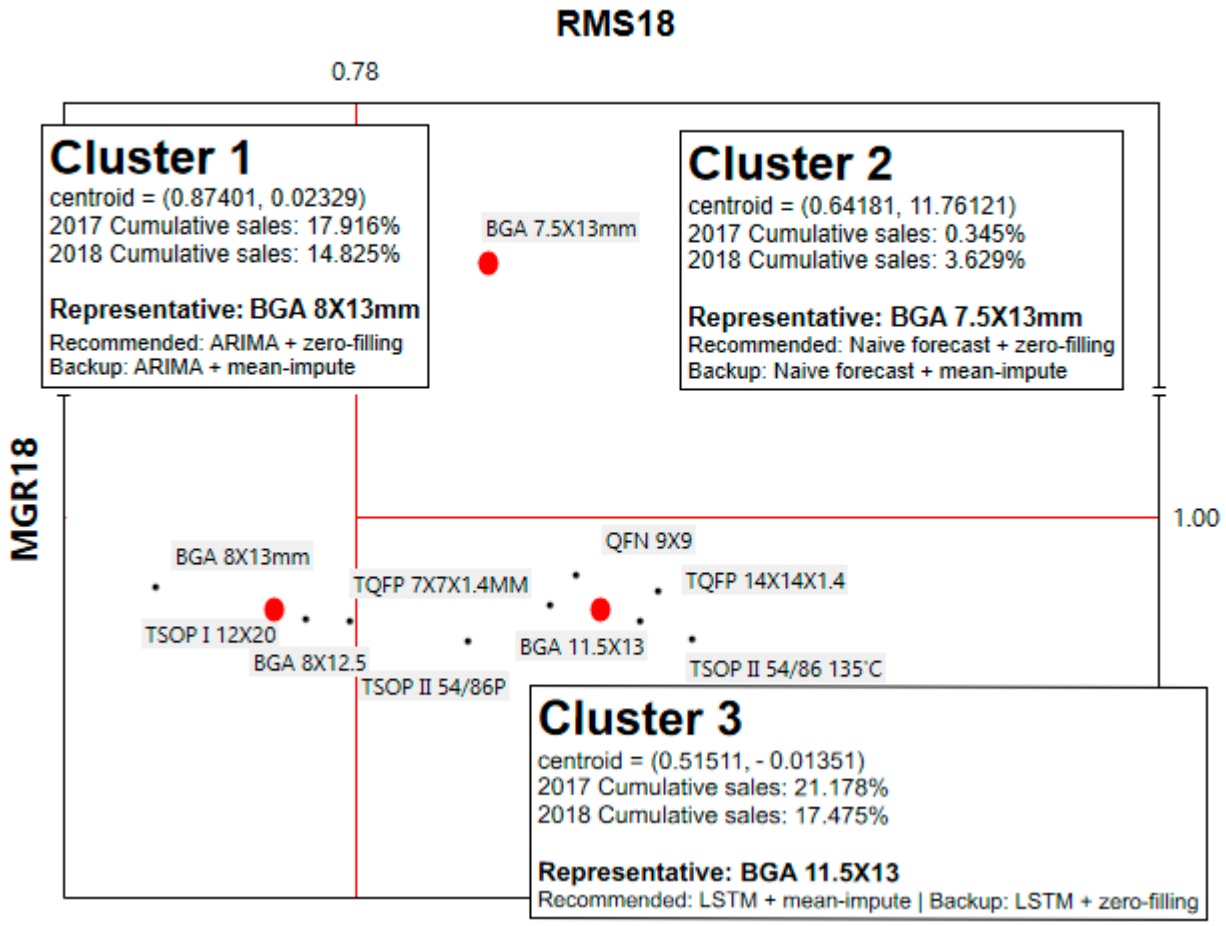


Figure 2. Illustration of the three-cluster portfolio for the top ten products.

Table 9. Summary of mean difference test for the top twenty products for three clusters.

Cluster-ID	Year	Mean (Std. Dev)	*t*-Test for Mean		Homogeneity of Variance	
			p-Value	95% Confidence Interval	Levene p.	B-F p.
1	2017	1,912,200 (226,192.2)	0.919693 (NS)	(−540,147, 529,227)	0.879153 (NS)	0.920500 (NS)
	2018	1,917,660 (245,148.8)				
2	2017	50,383 (52,350.4)	0.061747 (NS)	(−1,592,078, 60,378)	0.023436 (S)	0.409057 (NS)
	2018	816,233 (512,766.7)				
3	2017	867,056 (552,268.2)	0.997595 (NS)	(−341,789, 340,778)	0.119745 (NS)	0.355106 (NS)
	2018	867,561 (284,487.7)				

Note. Type I error (alpha) for significance is set to 0.05. NS = Not significant. N/A = Not Available. Levene p. = *p*-value of Levene's test. B-F p. = *p*-value of Brown and Forsythe's test.

Figure 3 illustrates Clusters 1 and 3. For Cluster 1, the characteristics and proposed market strategy are as follows:

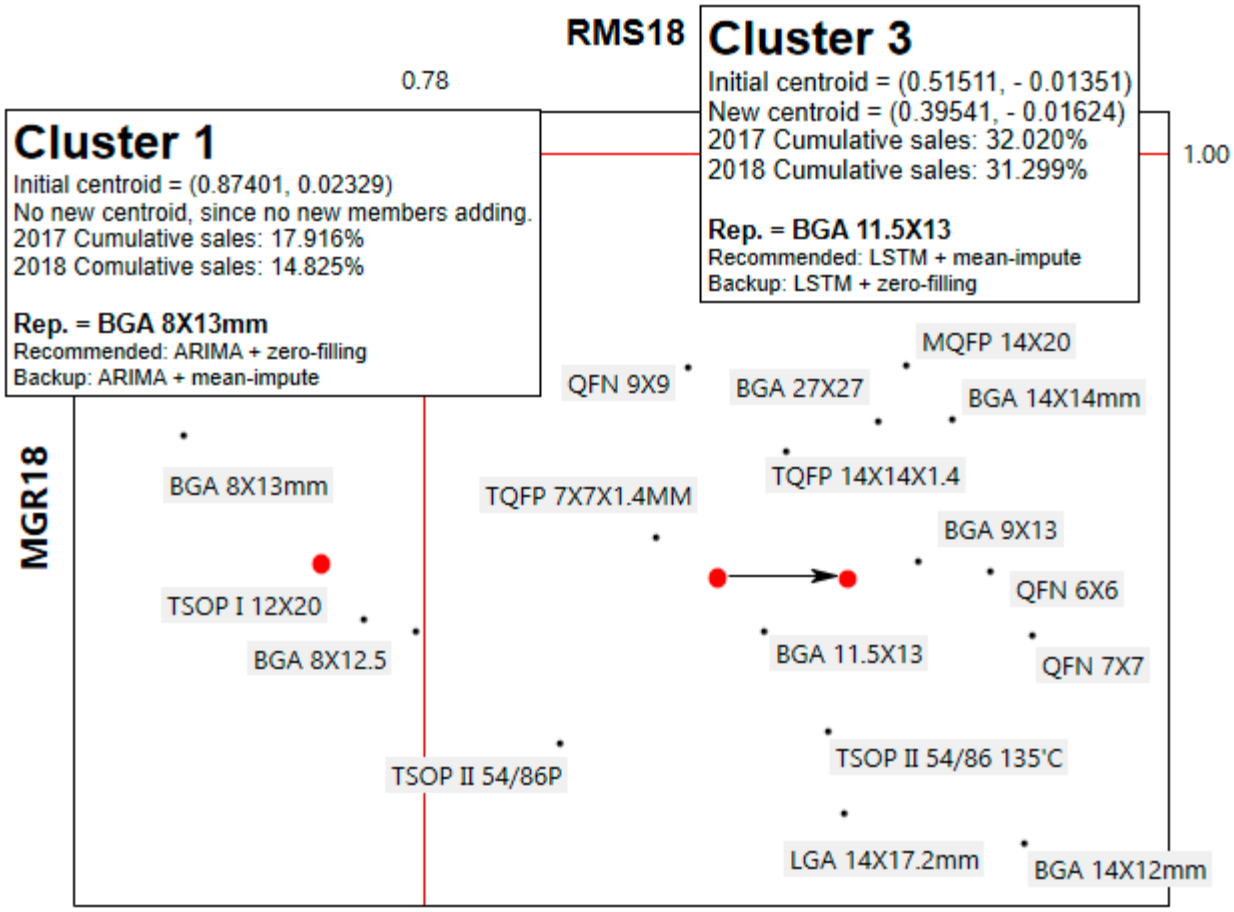


Figure 3. Illustration of the expanded Cluster 1 and expanded Cluster 3.

- This cluster, characterized by a high market share (0.87401) and a modest growth rate (0.023287), includes three products: BGA 8X13mm (classified as a 'Star') and TSOP I 12X20 and BGA 8X12.5 (both classified as 'Cash Cows') in the BCG Matrix.
- Despite a decline in annual sales proportion (2017: 17.9%, 2018: 14.8%, see Figure 3), total turnover has increased (see Table 9), indicating that Cluster 1 remains profitable.
- BGA 8X13mm, the product with the highest MGR in the cluster, should play a larger role in driving sales. The performance of the other two members should be closely monitored.
- The MGR for each product is critical for evaluation. If overall sales increase (or remain steady) due to the strong performance of the cluster representative while the other members underperform, gross margins may decline in the long run.

Table 10 outlines the forecasting scheme for Cluster 1, which includes products classified as "Stars" and "Cash Cows" according to the BCG Matrix. In 2018, actual sales for this cluster totaled 5,752,980 units, accounting for 14.825% of the year's total sales, with an average market growth rate (MGR) of 2.3287%. BGA 8X13mm serves as the cluster representative, with other members including TSOP I 12X20 and BGA 8X12.5. To forecast future sales for the cluster representative, the ARIMA + zero-filling model was applied, resulting in a forecast gap of −16.817%, indicating that the model has underestimated future sales. The baseline sales projection for 2019 is 5,881,437 units, with an optimistic scenario estimating 6,720,459 units.

Table 10. Forecasting Scheme of Cluster 1.

Outline	
Portfolio code:	Stars Cash-cows
2018 Actual Sale (unit):	5,752,980 (14.825% of the year)
2018 Avg. MGR:	2.3287%
Cluster representative:	BGA 8X13mm
Other members:	TSOP I 12X20, BGA 8X12.5
Forecasting Scheme	
Model of the cluster representative:	ARIMA + zero-filling
Forecast gap of the cluster representative (2018):	−16.817% (underestimated)
2019 Baseline (unit):	5,881,437 [=5,752,980 × (100 + 2.3287) %]
2019 Optimism (unit):	6,720,459 [=5,752,980 × (100 + |−16.817|) %]
2019 Preserved (unit):	N/A
Market Strategy	
• Build a market for BGA 8X13mm and maintain the current demand for TSOP I 12X20 and BGA 8X12.5.	

The market strategy, aligned with the BCG Matrix, focuses on expanding the market for BGA 8X13mm while maintaining stable demand for TSOP I 12X20 and BGA 8X12.5. Implementing a strategy that includes bundling sales and continued promotion of these products is recommended.

Table 11 outlines the forecasting scheme for the underestimated cases in Cluster 3, designated as the "Dream-chasing Child" portfolio. In 2018, total sales reached 4,095,800 units, accounting for 10.555% of the year's sales, with an average MGR of 6.197%. QFN 9X9 was chosen as the cluster representative due to its minimal missing data and alignment with the cluster centroid. Other products in this cluster include BGA 11.5X13, TQFP 14X14X1.4, and TSOP II 54/86 135 °C.

Table 11. Forecasting scheme of underestimated cases of expanded Cluster 3.

Outline	
Portfolio code:	Dream-chasing Child
2018 Actual Sale (unit):	4,095,800 (10.555% of the year)
2018 Avg. MGR:	6.197%
Cluster representative:	QFN 9X9 (Fewer missing data; same direction to centroid)
Other members:	BGA 11.5X13, TQFP 14X14X1.4, TSOP II 54/86 135′C
Forecasting Scheme	
Model of the cluster representative:	ARIMA + zero-filling
Forecast Gap of the cluster representative (2018):	1.973% (overestimated)
2019 Baseline (unit):	4,349,617 [=4,095,800 × (100 + 6.197) %]
2019 Optimism (unit):	N/A
2019 Preserved (unit):	4,014,990 [=4,095,800 × (100 − 1.973) %]
Market Strategy	
• Build and expand market share as much as possible without worrying about being unprofitable in the short term.	

The ARIMA + zero-filling model produced a forecast gap of 1.973% (overestimation) for 2018. The baseline sales projection 2019 is 4,349,617 units, with a preserved estimate of 4,014,990 units. The cluster's centroid places it in the 'Problem Child' category of the BCG Matrix, but the label 'Dream-chasing Child' is more fitting given the current growth rate and its market share of less than 0.5. It is recommended that the company pursue an aggressive strategy to expand market share, focusing on long-term growth rather than short-term profitability.

Table 12 outlines the forecasting scheme for the overestimated cases in Cluster 3, labeled the "Stable Office Workers" portfolio. In 2018, total sales reached 8,050,054 units, accounting for 20.744% of the year's sales, with an average MGR of −4.753%. TSOP II 54/86P was selected as the cluster representative due to its minimal missing data and alignment with the cluster centroid. Other products in this group include TQFP 7X7X1.4 MM, LGA 14X17.2mm, and BGA 27X27.

Table 12. Forecasting scheme of overestimated cases of expanded Cluster 3.

Outline	
Portfolio code:	Stable Office Workers
2018 Actual Sale (unit):	8,050,054 (20.744% of the year)
2018 Avg. MGR:	−4.753%
Cluster representative:	TSOP II 54/86P (Fewer missing data; same direction to centroid)
Other members:	TQFP 7X7X1.4MM, LGA 14X17.2mm, BGA 27X27, MQFP 14X20, BGA 9X13, BGA 14X14mm, QFN 6X6, BGA 14X12mm, QFN 7X7
Forecasting Scheme	
Model of the cluster representative:	Naïve forecast + zero-filling
Forecast Gap of the cluster representative (2018):	12.534% (overestimated)
2019 Baseline (unit):	7,667,435 [=8,050,054 × (100 − 4.753) %]
2019 Optimism (unit):	N/A
2019 Preserved (unit):	7,041,060 [=8,050,054 × (100 − 12.534) %]
Market Strategy	
• Take a wait-and-see strategy. If the decline becomes apparent, postpone or even terminate production.	

The application of the Naïve forecast plus zero-filling model resulted in a forecast gap of 12.534% for 2018, indicating an overestimation. The baseline projection for 2019 is 7,667,435 units, with a preserved estimate of 7,041,060 units. The cluster's centroid aligns with the "Dogs" category in the BCG Matrix, suggesting the need for a conservative strategy, such as a wait-and-see approach. If the sales decline continues, there may be a need to delay or even terminate production.

Table 13 outlines the forecasting scheme for Cluster 2, designated as "Grayed Loose Diamonds." In 2018, total sales reached 2,448,700 units, accounting for 6.310% of the year's sales, with an average MGR of 2236.759%. The cluster representative, BGA 7.5X13mm, was selected, alongside other members such as BGA 11.4X11mm and BGA 7.5X12mm.

Table 13. Forecasting scheme of the expanded Cluster 2.

Outline	
Portfolio code:	Grayed Loose Diamonds
2018 Actual Sale (unit):	2,448,700 (6.310% of the year)
2018 Avg. MGR:	2236.759%
Cluster representative:	BGA 7.5X13mm
Other members:	BGA 11.4X11mm, BGA 7.5X12mm
Forecasting Scheme	
Model of the cluster representative:	Naïve forecast + zero-filling
Forecast Gap of the cluster representative (2018):	−1.414% (underestimated)
2019 Baseline (unit):	57,220,218 [=2,448,700 × (100 + 2236.759) %]
2019 Optimism (unit):	2,483,325 [=2,448,700 × (100 + |−1.414|) %]
2019 Preserved (unit):	N/A
Market Strategy	
• Build and expand market share as much as possible without worrying about being unprofitable in the short term.	

The Naïve forecast with zero-filling yielded a forecast gap of −1.414%, indicating an underestimation. The baseline sales projection for 2019 is 57,220,218 units, reflecting the significant MGR, with an optimistic estimate of 2,483,325 units also calculated. The market strategy for this cluster focuses on aggressive market expansion, prioritizing long-term growth over short-term profitability by leveraging the portfolio's rapid growth potential.

Figure 4 illustrates Cluster 2. With the addition of new members, the cluster's average MGR becomes more pronounced, even as the average RMS decreases. This high-growth characteristic is evident in the annual cumulative sales performance, which rose from 0.472% in 2017 to 6.310% in 2018.

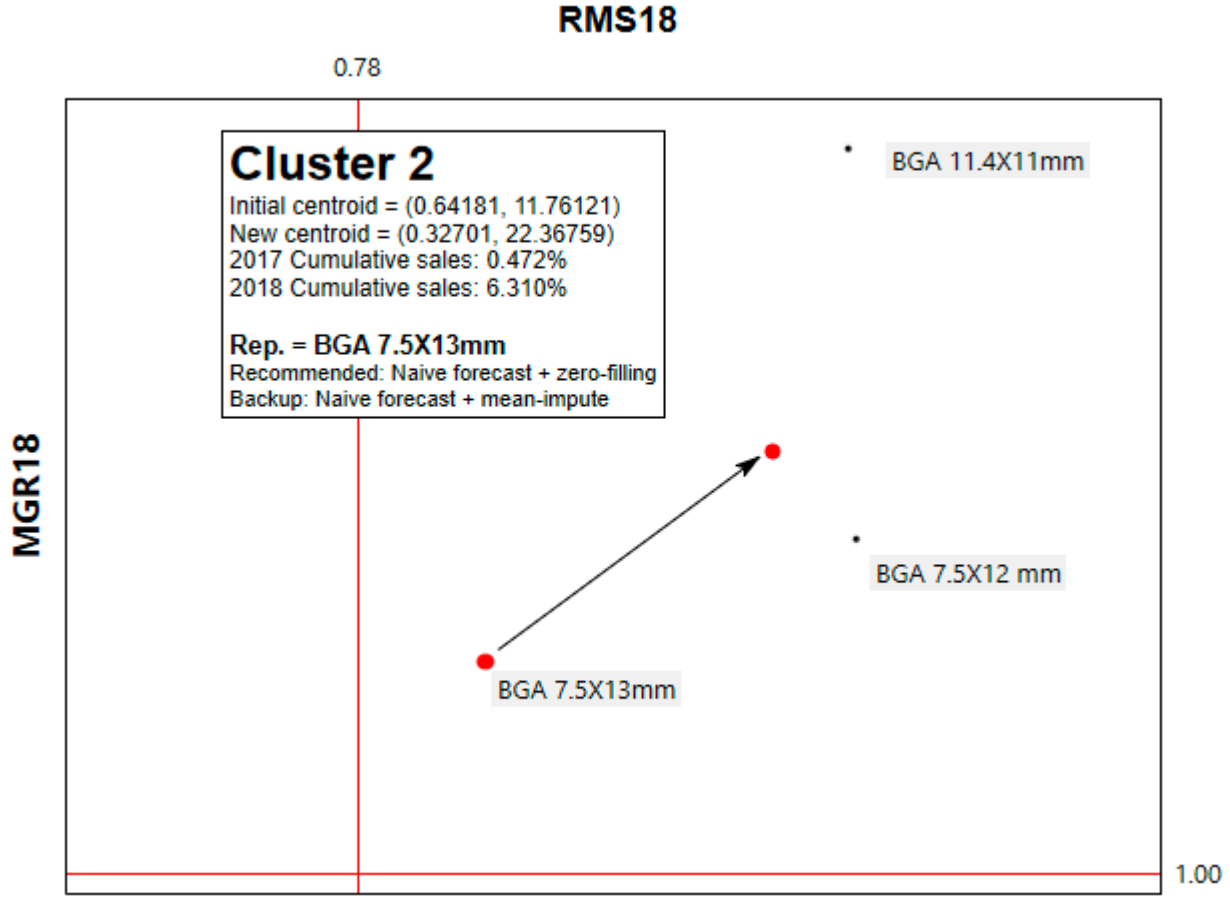


Figure 4. Illustration of the expanded Cluster 2.

5. Discussion

As this article aims to develop a specific framework for multi-item product forecasting, it integrates individual methods, such as the BCG Matrix and K-means Clustering. The following section summarizes two key findings from the previous analysis. First, the matrix-based portfolio results derived from the BCG Matrix are demonstrated. Table 14 outlines the managerial recommendations for each group's recommended and backup models. Second, a general comparison is made between the classical matrix-based approach and the proposed cluster-based approach.

Table 14. Combined summary of BCG Matrix and the forecasting models.

Group	Product	Recommended Model	Managerial Recommendation for the Recommended Model	Backup Model	Managerial Recommendation for the Backup Model
Stars	BGA 8X13mm	ARIMA + zero-filling	Increase quarterly capacity by 16.817%.	ARIMA + mean-impute	Increase quarterly capacity by 16.915%
	BGA 7.5X13mm	Naïve forecast + zero-filling	Increase quarterly capacity by 1.414%. Alternatively, do not do any capacity adjustment activities.	Naïve forecast + mean-impute	Increase quarterly capacity by 1.414%. Alternatively, do not do any capacity adjustment activities.
	TQFP 7X7X1.4MM	Naïve forecast + zero-filling	Decrease quarterly capacity by 2.243%. Alternatively, do not do any capacity adjustment activities.	Naïve forecast + mean-impute	Decrease quarterly capacity by 2.243%. Alternatively, do not do any capacity adjustment activities.
	QFN 9X9	ARIMA + zero-filling	Decrease quarterly capacity by 1.973%. Alternatively, do not do any capacity adjustment activities.	ARIMA + mean-impute	Decrease quarterly capacity by 1.973%. Alternatively, do not do any capacity adjustment activities.
Cash-cows	BGA 8X12.5	ARIMA + zero-filling	Increase quarterly capacity by 2.172%. Alternatively, do not do any capacity adjustment activities.	ARIMA + mean-impute	Increase quarterly capacity by 2.172%. Alternatively, do not do any capacity adjustment activities.
	TSOP II 54/86P	Naïve forecast + zero-filling	Decrease quarterly capacity by 12.534%.	Naïve forecast + mean-impute	Decrease quarterly capacity by 12.534%.
	TSOP I 12X20	Naïve forecast + mean-impute	Decrease quarterly capacity by 9.248%.	Naïve forecast + zero-filling	Decrease quarterly capacity by 9.248%.
Dogs	TSOP II 54/86 135′C	LSTM + mean-impute	Increase quarterly capacity by 13.230%.	LSTM + Zean-filling	Increase quarterly capacity by 12.818%.
	BGA 11.5X13	LSTM + mean-impute	Increase quarterly capacity by 25.090%.	LSTM + zero-filling	Increase quarterly capacity by 23.588%.
Problem-Child	TQFP 14X14X1.4	Naïve forecast + mean-impute	Increase quarterly capacity by 3.049%. Alternatively, do not do any capacity adjustment activities.	Naïve forecast + zero-filling	Increase quarterly capacity by 3.049%. Alternatively, do not do any capacity adjustment activities.

5.1. The Matrix-Based Portfolios for the Top Ten Products

In an empirical comparison of sales forecasts for plastic tray products, Hung et al. (2022) concluded the following [37]:

- A combined model of LSTM and zero-filling is suitable for "Dogs" products.
- The zero-filling method handles missing data for high market share products (Stars and Cash Cows).
- The mean imputation method is appropriate for addressing missing data in general market share products (Dogs and Problem Child).

Figure 5 presents the BCG Matrix for the top ten selling products using RMS = 0.500 and MGR = 0.000 as reference lines. This matrix divides the top ten products into four groups—Stars, Cash Cows, Dogs, and Problem Child—providing detailed information on cumulative sales percentages, average RMS and MGR, and standard deviations for both 2017 and 2018.

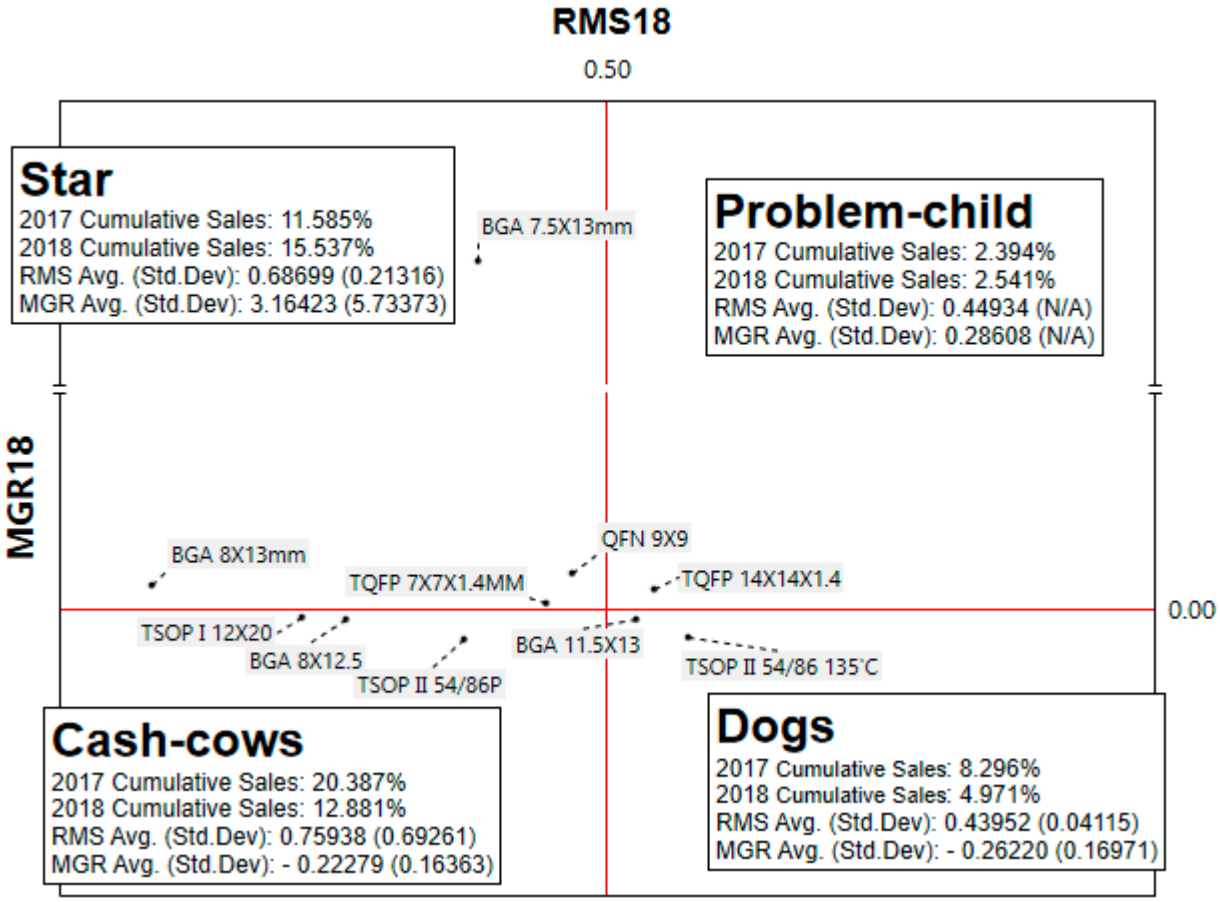


Figure 5. BCG Matrix for 2018 Top Ten Products.

Table 14 summarizes the BCG Matrix classification and forecasting models for the ten most prominent products, along with recommended and alternative models and associated managerial recommendations.

- For Stars products like BGA 8X13mm, the recommended ARIMA + zero-filling model suggests increasing quarterly capacity by 16.817%. The alternative model, ARIMA + mean-imputation, offers a similar recommendation. For BGA 7.5X13mm, both the recommended and backup models indicate no need for significant capacity adjustments.
- For high-profitability products like BGA 8X12.5, the ARIMA + zero-filling model advises a 2.172% increase in capacity. Conversely, TSOP II 54/86P should see a 12.534% capacity reduction based on the naïve forecast + zero-filling model.
- For Dogs products, such as TSOP II 54/86 135 °C, the LSTM + mean-imputation model recommends a substantial 13.230% capacity increase, with a slightly lower recommendation from the backup model.
- For problematic products like TQFP 14X×14X1.4, both the recommended and backup models suggest a modest 3.049% capacity increase, or no adjustment if deemed unnecessary.

5.2. A Comparison of Matrix-Based and Cluster-Based Portfolios

Table 15 presents a comparative analysis of matrix-based (BCG Matrix) and cluster-based (K-means Clustering) portfolio designs, summarizing the results from Tables 10–13. This comparison highlights key differences between the two methodologies regarding underlying principles, variables used, advantages and disadvantages, management strategies, and market implications.

Table 15. General comparison of portfolio design, matrix- and cluster-based.

Type	**Matrix-Based, Top Ten Only**	**Cluster-Based, Top Twenty**
Method	BCG Matrix	k-Means Clustering
Used variable	RMS, MGR	RMS, MGR.
Pros	Easy to build, fast to read	Science-based and data-driven
Cons	Lack of technical indicators	Complicated computation
Management	Top-down	Cross and interactive
Reference Line	RMS = 0.500, and MGR = 0.000	RMS = 0.780, and MGR = 1.000
Code	Stars	Stars Cash-cows
Criteria	RMS $\geqq$ 0.500, and MGR < 0.000	RMS $\geqq$ 0.780
2018 Sales (%)	15.537%	14.825%
Market Strategy	Build and then maintain	Build for the representative and maintain the current market for the Others.
Code	Cash-cows	Stable Office Workers
Criteria	RMS $\geqq$ 0.500, and MGR $\leqq$ 0.000	0.780 > RMS $\geqq$ 0.000, and M.G.R. < 0.000
2018 Sales (%)	12.881%	20.744%
Market Strategy	Maintain as prior	Take a wait-and-see strategy. If the decline becomes apparent, postpone or even terminate production.
Code	Dogs	Dream-chasing Child
Criteria	0.500 > RMS $\geqq$ 0.000, and M.G.R. < 0.000	0.780 > RMS $\geqq$ 0.000, and MGR $\geqq$ 0.000
2018 Sales (%)	4.970%	10.555%
Market Strategy	Harvest, even Liquid/Terminate	Build and expand market share as much as possible without worrying about being unprofitable in the short term.
Code	Problem-child	Grayed Loose Diamonds
Criteria	0.500 > RMS $\geqq$ 0.000, and MGR $\geqq$ 0.000	0.780 > RMS $\geqq$ 0.000, and MGR $\geqq$ 1.000
2018 Sales (%)	2.541%	6.310%
Market Strategy	Build and maintain	Build and expand market share as much as possible without worrying about being unprofitable in the short term.
Used Data Amount	Top ten products (35.928%) only	Top twenty products (52.434%)

As exemplified by the BCG Matrix, the matrix-based approach applies to the top ten products, using RMS (Relative Market Share) and MGR (Market Growth Rate) as core variables. Its simplicity and ease of interpretation allow for rapid decision-making, making it an attractive option for strategic planning. However, its limitations include the lack of technical indicators and the subjective nature of reference lines (RMS = 0.500, MGR = 0.000). Additionally, the limited data usage (35.928%) may lead to classification errors and issues with variance, mainly when dealing with a narrow data set.

In contrast, the cluster-based approach (K-means Clustering) extends to the top twenty products, using the same RMS and MGR variables but with a more data-driven, scientifically grounded methodology. While this method is more computationally complex, it minimizes within-group variation and allows for broader applicability. It provides greater practical value by accommodating newly added products (ranked 11–20) and dividing them into subgroups based on deployment results (e.g., overestimated and underestimated cases). This ensures portfolio validity and stability. Furthermore, this approach significantly increases data utilization (52.434%), resulting in more robust and comprehensive outcomes.

From a management perspective, matrix-based designs typically follow a top-down approach, where predefined strategies are applied to each product category—Stars, Cash Cows, Dogs, and Problem Children—based on their market performance. For instance, the plan for "Stars" is typically to "build and maintain" market share, while "Dogs" are often

managed through harvesting or liquidation. However, this approach may oversimplify the complexity of product behavior and market dynamics.

In contrast, cluster-based designs support a more interactive and cross-functional management approach. For example, in the case of "Stars," the strategy is to expand the cluster representative's market share while maintaining the other products' current position. Cash Cows were reclassified as "Stable Office Workers"; a more conservative, wait-and-see strategy is recommended, with the option of postponing or halting production if market decline is evident. Additionally, for products classified as "Dream-chasing Child," the cluster-based approach recommends aggressive market expansion, even at the expense of short-term profitability, due to their long-term growth potential.

Regarding product classification, the matrix-based approach relies on simple RMS and MGR thresholds, which can result in overly conservative recommendations, particularly for categories like Problem Child. The cluster-based approach refines these classifications through more detailed data analysis, enabling more targeted market strategies. For example, products in the "Dream-chasing Child" category are advised to pursue aggressive market expansion in contrast to the more cautious approach recommended for "Dogs" in the matrix-based method.

This study demonstrates that integrating K-means clustering, the BCG Matrix, and time-series modeling significantly improves sales forecasting for product portfolios. The framework enhances forecast accuracy by adapting to product and market changes, mainly when there is limited historical data. This approach can be applied across various industries to help businesses manage various products.

The proposed framework is versatile and can be implemented in multiple industries. It is especially beneficial in sectors like consumer electronics, where products are released rapidly and demand fluctuates significantly between models. In the fast-moving consumer goods industry, where consumer preferences and market conditions shift frequently, the framework supports retailers in managing inventory more effectively. The pharmaceutical industry can also leverage this framework to improve demand forecasting and optimize resource allocation. Beyond demand forecasting, the framework can be applied to other business management areas, such as supply chain optimization. Companies can reduce costs and enhance operational efficiency by aligning production and procurement with demand forecasts. The framework is also useful for forecasting new products with limited historical data, enabling businesses to compare products and predict sales more accurately. Additionally, it helps companies minimize waste and conserve resources by ensuring they produce the right products at the right time.

While the framework effectively improves forecasting accuracy, there are opportunities for further enhancement. One potential improvement involves incorporating additional data sources such as market trends, customer feedback, and economic indicators, which could enhance the model's ability to anticipate market shifts. Another avenue for future research is automated parameter tuning. Implementing automated machine learning could streamline the optimization of model parameters, such as the number of clusters in K-means or ARIMA configurations, making the framework more scalable and reducing manual intervention.

Finally, combining traditional time-series methods with advanced machine-learning algorithms could make the framework more robust. Ensemble learning techniques may help address complex sales patterns, improving forecasting accuracy and reliability.

6. Conclusions

This paper introduces a framework for forecasting sales across multiple products, with matrix clustering as a central component. By applying data science to real-world data, the framework overcomes the limitations of traditional forecasting tools, particularly their challenges in adapting to shifts in product life cycles.

6.1. Outline Product Portfolios by BCG Matrix

An empirical study was conducted on integrated circuit tray products using sales data from Taiwan. The BCG Matrix was used to categorize product portfolios, classifying the top ten products into four distinct categories and revealing their sales patterns. The concept of a cluster representative was integrated with individual product sales forecasting models to emphasize collective sales trends. The BCG Matrix was also updated, and portfolio forecasting methods were developed.

6.2. Revise BCG Matrix and Build Forecasting Schemes for Specific Portfolios

The top ten product portfolios were restructured using K-means Clustering based on data-driven attributes. The analysis resulted in three clusters that more accurately represented business characteristics than the BCG Matrix analysis. Sales forecasting methods were developed to account for the expected gaps, yielding unified forecasting performance and a suggested market strategy. These initial forecasting methods were then extended to lower-ranking products, identifying four distinct portfolios through consolidated forecasting and cluster stability analysis.

This paper offers two critical scholarly insights. First, it introduces cluster-based portfolio analysis, a more scientific and data-driven approach that refines traditional BCG Matrix analysis. Second, it develops a practical framework for forecasting multiple product sales, incorporating modeling, performance evaluation, and corrective mechanisms. In terms of business practice, this paper provides specific forecasting methods tailored to each portfolio's business and market characteristics, offering practitioners a reliable, stable, sustainable, and scalable workflow and decision-making model.

Several research constraints were set to conduct the empirical study and validate the research questions effectively. Daily data were aggregated into weekly data to address the data insufficiency caused by intermittent orders. The assumption of fixed unit costs and no product substitutability was made to reduce uncertainty in product transactions, which could affect the model's prediction accuracy.

The empirical analysis confirmed the feasibility and stability of the proposed framework. Future research will focus on two main areas: testing the framework's applicability in other industries and gradually relaxing the research constraints to allow for broader general application. This study also makes three recommendations for further research to validate and expand the framework:

1. Add more key performance indicators to enrich the forecast plan. The BCG Matrix's static nature assumes that market growth is the sole indicator of attractiveness. This study validated that the cluster-based approach can more effectively forecast multi-item product portfolios. Incorporating other market factors will improve the generalizability of this approach.
2. Introduce more time series forecasting and clustering techniques. Recent research has shown that deep-learning methods can outperform classical time series forecasters. Integrating deep-learning forecasters into the proposed framework could enhance the analysis process, yielding more accurate results.
3. Gather more sales data to validate findings. Table 15 shows that 52.434% of data from the top twenty products were utilized, demonstrating that the cluster-based approach is more effective in improving data usage than the traditional matrix-based method. Increasing the data available for forecasting multi-item products through clustering could lead to more stable and reliable forecasting outcomes.

6.3. Summary

This study introduces a novel approach to multi-item product sales forecasting, integrating the traditional BCG Matrix with K-means clustering and time-series modeling techniques. This integration enhances the static nature of the BCG Matrix, providing a more dynamic and data-driven method for product re-clustering. Compared to previous models, such as the Grey Portfolio Analysis, which assumes static data, our proposed framework is

more adaptable to changing market conditions, offering businesses a more accurate and scientific tool for sales forecasting.

Moreover, our model addresses the challenge of sparse historical data, a limitation inherent in traditional forecasting models like ARIMA and LSTM. By employing zero-filling and mean-imputation techniques, the proposed method significantly improves forecast accuracy in scenarios with limited data, advancing beyond prior studies that rely heavily on abundant historical data. This improvement is particularly crucial for businesses managing multi-item product portfolios, where data scarcity often hinders precise forecasting.

An empirical study applying the framework to real-world IC tray sales data demonstrated its practical effectiveness. Unlike earlier theoretical models, such as those by Bunn and Vassilopoulos (1999) [38], empirical testing has validated our framework. This real-world validation not only strengthens the robustness of our findings but also underscores the practicality of integrating K-means clustering with the BCG Matrix for sales forecasting, instilling confidence in its application.

Additionally, incorporating an iterative feedback mechanism ensures that the proposed framework remains adaptable and sustainable. This dynamic feature distinguishes our model from static approaches, providing businesses with a continuously evolving tool for managing product portfolios. The model's continuous evolution not only addresses the shortcomings of classical time-series forecasting and static portfolio analysis techniques but also paves the way for future enhancements, making this approach a significant advancement in the field.

In conclusion, this study presents a comprehensive and innovative solution for multi-item sales forecasting. By comparing our findings with existing research, we have demonstrated the superiority of this approach in tackling critical challenges such as data sparsity, model adaptability, and empirical validation. Future research can build on these insights to improve sales forecasting models' accuracy and scalability across various industries.

Author Contributions: Conceptualization, C.-Y.H. and C.-C.W.; methodology, C.-Y.H. and C.-C.W.; validation, C.-Y.H. and C.-C.W.; formal analysis, C.-Y.H. and C.-C.W.; data curation, C.-C.W.; writing—original draft preparation, C.-Y.H. and C.-C.W. writing—review and editing, C.-C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Council, R.O.C, and grant number is 110-2221-E-131-027-MY3.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Du, B.; Li, Z.; Yuan, J.; Zheng, J.; Shu, W.; Jin, Y. Customer's Channel Selection Behavior on Purchasing Standardized and Customized Products: Optimized Prices and Channel Performances. *Front. Psychol.* **2022**, *13*, 1634. [CrossRef] [PubMed]
2. Yang, C.; Yao, J.; Lou, W.; Xie, S. On demand response management performance optimization for microgrids under imperfect communication constraints. *IEEE Internet Things J.* **2017**, *4*, 881–893. [CrossRef]
3. Zsidisin, G.A.; Panelli, A.; Upton, R. Purchasing organization involvement in risk assessments, contingency plans, and risk management: An exploratory study. *Supply Chain Manag. Int. J.* **2000**, *5*, 187–198. [CrossRef]
4. Cooper, D.R.; Gutowski, T.G. The environmental impacts of reuse: A review. *J. Ind. Ecol.* **2017**, *21*, 38–56. [CrossRef]
5. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018.
6. Wongoutong, C. Imputation methods in time series with a trend and a consecutive missing value pattern. *Thail. Stat.* **2021**, *19*, 866–879.
7. Borges, C.E.; Kamara-Esteban, O.; Castillo-Calzadilla, T.; Andonegui, C.M.; Alonso-Vicario, A. Enhancing the missing data imputation of primary substation load demand records. *Sustain. Energy Grids Netw.* **2020**, *23*, 100369. [CrossRef]
8. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2015.

9. Wang, C.C.; Chien, C.H.; Trappey, A.J. On the application of ARIMA and LSTM to predict order demand based on short lead time and on-time delivery requirements. *Processes* **2021**, *9*, 1157. [CrossRef]
10. Yuan, X.; Zhang, X.; Zhang, D. Analysis of the impact of different forecasting techniques on the inventory bullwhip effect in two parallel supply chains with a competition effect. *J. Eng.* **2020**, *2020*, 2987218. [CrossRef]
11. Abbasimehr, H.; Khodizadeh Nahari, M. Improving demand forecasting with LSTM by taking into account the seasonality of data. *J. Appl. Res. Ind. Eng.* **2020**, *7*, 177–189.
12. Babu, C.N.; Reddy, B.E. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. *Appl. Soft Comput.* **2014**, *23*, 27–38. [CrossRef]
13. Kiefer, D.; Grimm, F.; Bauer, M.; van Dinther, C. Demand forecasting intermittent and lumpy time series: Comparing statistical, machine learning and deep learning methods. In Proceedings of the 54th Hawaii International Conference on System Sciences, Maui, HI, USA, 5–8 January 2021; HICSS. Grand Wailea: Maui, HI, USA, 2021; pp. 1425–1434.
14. Hosseini, S.M.S.; Maleki, A.; Gholamian, M.R. Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Syst. Appl.* **2010**, *37*, 5259–5264. [CrossRef]
15. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. A comparison of ARIMA and LSTM in forecasting time series. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications, Orlando, FL, USA, 17–20 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1394–1401.
16. Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [CrossRef]
17. Barksdale, H.C.; Harris, C.E., Jr. Portfolio analysis and the product life cycle. *Long Range Plan.* **1982**, *15*, 74–83. [CrossRef]
18. Stummer, C.; Heidenberger, K. Interactive R&D portfolio analysis with project interdependencies and time profiles of multiple objectives. *IEEE Trans. Eng. Manag.* **2003**, *50*, 175–183.
19. Udo-Imeh, P.T.; Edet, W.E.; Anani, R.B. Portfolio analysis models: A review. *Eur. J. Bus. Manag.* **2012**, *4*, 101–120.
20. Mohajan, H.K. An Analysis on BCG Growth Sharing Matrix. *Noble Int. J. Bus. Manag. Res.* **2017**, *2*, 1–6.
21. Nowak, M.; Mierzwiak, R.; Wojciechowski, H.; Delcea, C. Grey portfolio analysis method. *Grey Syst. Theory Appl.* **2020**, *10*, 439–454. [CrossRef]
22. Chiu, C.C.; Lin, K.S. Rule-based BCG matrix for product portfolio analysis. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 17–32.
23. Hossain, H.; Kader, M.A. An analysis on BCG growth sharing matrix. *Int. J. Contemp. Res. Rev.* **2020**, *11*, 21899–21905. [CrossRef]
24. Pradana, M.G.; Ha, H.T. Maximizing strategy improvement in mall customer segmentation using k-means clustering. *J. Appl. Data Sci.* **2021**, *2*, 19–25. [CrossRef]
25. Syakur, M.A.; Khotimah, B.K.; Rochman, E.M.S.; Satoto, B.D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *336*, 012017. [CrossRef]
26. Wu, S.; Yau, W.C.; Ong, T.S.; Chong, S.C. Integrated churn prediction and customer segmentation framework for telco business. *IEEE Access* **2021**, *9*, 62118–62136. [CrossRef]
27. Xiahou, X.; Harada, Y. B2C e-commerce customer churn prediction based on k-means, S.V.M. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 458–475. [CrossRef]
28. Abdulla, S.H.; Sagheer, A.M.; Veisi, H. Breast cancer segmentation using K-means clustering and optimized region-growing technique. *Bull. Electr. Eng. Inform.* **2022**, *11*, 158–167. [CrossRef]
29. Altini, N.; De Giosa, G.; Fragasso, N.; Coscia, C.; Sibilano, E.; Prencipe, B.; Hussain, S.M.; Brunetti, A.; Buongiorno, D.; Guerriero, A.; et al. Segmentation and identification of vertebrae in CT scans using CNN, k-means Clustering and k-NN. *Informatics* **2021**, *8*, 40. [CrossRef]
30. Jebarani, P.E.; Umadevi, N.; Dang, H.; Pomplun, M. A novel hybrid k-means and GMM machine learning model for breast cancer detection. *IEEE Access* **2021**, *9*, 146153–146162. [CrossRef]
31. Khan, A.R.; Khan, S.; Harouni, M.; Abbasi, R.; Iqbal, S.; Mehmood, Z. Brain tumor segmentation using k-means Clustering and deep learning with synthetic data augmentation for classification. *Microsc. Res. Tech.* **2021**, *84*, 1389–1399. [CrossRef] [PubMed]
32. Tian, K.; Li, J.; Zeng, J.; Evans, A.; Zhang, L. Segmentation of tomato leaf images based on adaptive clustering number of k-means algorithm. *Comput. Electron. Agric.* **2019**, *165*, 104962. [CrossRef]
33. Zhang, H.; Peng, Q. PSO and K-means-based semantic segmentation toward agricultural products. *Future Gener. Comput. Syst.* **2022**, *126*, 82–87. [CrossRef]
34. Guijo-Rubio, D.; Durán-Rosal, A.M.; Gutiérrez, P.A.; Troncoso, A.; Hervás-Martínez, C. Time-series clustering based on the characterization of segment typologies. *IEEE Trans. Cybern.* **2020**, *51*, 5409–5422. [CrossRef]
35. Bandara, K.; Bergmeir, C.; Smyl, S. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert Syst. Appl.* **2020**, *140*, 112896. [CrossRef]
36. Alqahtani, A.; Ali, M.; Xie, X.; Jones, M.W. Deep time-series clustering: A review. *Electronics* **2021**, *10*, 3001. [CrossRef]
37. Hung, C.Y.; Wang, C.C.; Lin, S.W.; Jiang, B.C. An empirical comparison of the sales forecasting performance for plastic tray manufacturing using missing data. *Sustainability* **2022**, *14*, 2382. [CrossRef]
38. Bunn, D.W.; Vassilopoulos, A.I. Comparison of seasonal estimation methods in multi-item short-term forecasting. *Int. J. Forecast.* **1999**, *15*, 431–443. [CrossRef]

39. Xie, J.; Lee, T.S.; Zhao, X. Impact of forecasting error on the performance of capacitated multi-item production systems. *Comput. Ind. Eng.* **2004**, *46*, 205–219. [CrossRef]
40. Taylor, J.W. Multi-item sales forecasting with total and split exponential smoothing. *J. Oper. Res. Soc.* **2011**, *62*, 555–563. [CrossRef]
41. Spedding, T.A.; Chan, K.K. Forecasting demand and inventory management using Bayesian time series. *Integr. Manuf. Syst.* **2000**, *11*, 331–339. [CrossRef]
42. Gopagoni, D.R.; Lakshmi, P.V.; Chaudhary, A. Evaluating machine learning algorithms for marketing data analysis: Predicting grocery store sales. *Commun. Softw. Netw.* **2021**, *134*, 155–163.
43. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735. [CrossRef]
44. Hyndman, R.J.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. Prediction intervals for exponential smoothing using two new classes of state space models. *J. Forecast.* **2005**, *24*, 17–37. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.