turnitin 2

by Febriyanti Panjaitan

Expert Systems With Applications 207 (2022) 118043

Contents lists available at ScienceDirect

Expert Systems With Applications

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

An approach for sales forecasting

Rendra Gustriansyah[a,1], Ermatita Ermatita[b,*,2], Dian Palupi Rini[b,3]

[a] Sriwijaya University Faculty of Engineering, Palembang 30139, Indonesia
[b] Sriwijaya University School of Computer Science, Palembang 30139, Indonesia

A R T I C L E I N F O

Keywords:
Data mining
Decision-making method
k-Means
RFM
Sales forecasting method

A B S T R A C T

Inaccurate sales forecasting (SF) leads to over-stock or stock-out, in which can increase inventory costs, thereby reducing profits and return on investment. Furthermore, stock-out can eliminate customer loyalty, as well as opportunities to acquire new customers, and maximize sales or competitive advantage. This study proposed a retail SF model called the SalesKBR, which integrates the decision-making (Best-Worst Method/BWM) and the data mining methods (k-Means) into the Recency-Frequency-Monetary (RFM) model. The BWM was used to extract criteria that have a significant influence on retail SF. The k-Means method and six validity indices were applied to improve the quality of the product clustering results, while the RFM model was used for assessment. The extraction results from BWM that correlate with the retail company databases showed the criteria that significantly affect SF are frequency, quantity, and monetary. The results also showed that SalesKBR is a retail SF model with a reasonable level of accuracy. Therefore, it can be utilized to forecast retail sales, and it has the potential to be an alternative solution for making scientific and valuable management decisions.

1. Introduction

Sales forecasting (SF) is one of the important aspects of business management process because it plays a major role in the allocation of resources, marketing, and finance. Furthermore, it influences decisions to provide products and services to consumers in the customer relationship management horizon (Hyndman & Athanasopoulos, 2019). Forecasting the number of sales can optimize inventory, as well as accurately provide products and services to the clients (Levy, Weitz, & Grewal, 2019), optimize customer experience (Gao & Fan, 2021), and increase satisfaction (Wang, Chang, & Zhou, 2019). However, inaccurate SF leads to over-stock or stock-out (Kilimci et al., 2019), which can increase inventory costs, hence reduce profits and investment returns (Gustriansyah, Suhandi, Antony, & Sanmorino, 2019). Running out of stock can lead to companies losing customers' trust/loyalty, missing out on new customers, opportunities to maximize sales, as well as competitive advantages (Chandriah & Naraganahalli, 2021; Islam & Amin, 2020).

Historically, studies on the retail SF models that are mostly used in the last five years is the classical statistical methods (Time Series, xARIMA, and xRegression) (Gustriansyah, Ermatita, Rini, & Malik, 2020a). This shows the classical statistical method still dominates the retail SF. It relies on sales quantity criteria extrapolated from historical trends and seasonal effects to predict future sales where companies

Abbreviations: SF, Sales Forecasting; RFM, Recency-Frequency-Monetary; BWM, Best-Worst Method; AHP, Analytical Hierarchy Process; FAHP, Fuzzy Analytical Hierarchy Process; ANP, Analytical Network Process; ID3, Iterative Dichotomiser 3; ARIMA, Autoregressive Integrated Moving Average; LASSO, Least Absolute Shrinkage and Selection Operator; TS, Traditional Statistics; RF, Random Forest; IRF, Incremental Random Forest; DL, Deep Learning; NN, Neural Network; ANN, Artifcial Neural Network; GAN, Generative Adversarial Networks; LSTM, Long Short-Term Memory; SA-LSTM, Simulated Annealing-LSTM; DNN, Deep Neural Network; BPN, Back-propagation Neural Network; SKU, Stock Keeping Unit; Centroid, Cluster center; SI, Silhouette Index; RLI, Ratkowsky-Lance Index; DI, Dunn Index; KLI, Krzanowski-Lai Index; DBI, Davies-Bouldin Index; CHI, Calinski-Harabasz Index; BGSS, Between-Groups Sum of Square; TSS, Total Sum of Square; WSSk, Within-Clusters Sum of Squares in k clusters; BSS, Between-Clusters Sum of Square; MAPE, Mean Absolute Percentage Error; SMAPE, Symmetric Mean Absolute Percentage Error; B2O, Best criterion against Other criteria; O2W, Other criteria against the Worst criterion; ξ^*, The optimal consistency ratio of the comparisons; Cro, Consistency Ratio; FQM, Frequency-Quantity-Monetary.

* Corresponding author at: Jl. Srijaya Negara, Bukit Besar, Palembang 30139, Indonesia.
E-mail addresses: rendra@uigm.ac.id (R. Gustriansyah), ermatita@unsri.ac.id (E. Ermatita), dprini@unsri.ac.id (D.P. Rini).
[1] ORCID: 0000-0001-7600-1147.
[2] ORCID: 0000-0002-5267-0589.
[3] ORCID: 0000-0003-4131-9686.

https://doi.org/10.1016/j.eswa.2022.118043
Received 9 December 2021; Received in revised form 14 June 2022; Accepted 30 June 2022
Available online 8 July 2022
0957-4174/

require to adjust to manual forecasting or rely on experts. Therefore, this study proposed a novel retail SF model called SalesKBR, which is expected to minimize the error rate in the results.

The proposed SalesKBR is an SF model that integrates the decision-making (Best-Worst Method) and the data-mining methods (k-Means) into the Recency-Frequency-Monetary (RFM) model. The BWM is used to extract criteria that have a significant influence on SF. Furthermore, it is a decision-making method that is more efficient than other similar ones (Liang, Brunelli, & Rezaei, 2020). The use of this method to extract forecasting criteria is a new approach.

Meanwhile, the k-Means and validity indices are used to improve the quality of the product clustering results. This is because the k-Means has a good speed in selecting the cluster center (centroid) with a simple algorithm (Abdelaziz & Lu, 2019; Hossain, Akhtar, Ahmad, & Rahman, 2019). The RFM is used to assess the product and is a component of the forecasting model.

The expected contribution from the results is the innovation of the RFM concept application in the SF model. Generally, this model is used to measure the value of consumers according to their purchase history and is widely used in the marketing field (Anitha & Patil, 2019; Zhou, Wei, & Xu, 2021). In this study, the RFM concept was used to assess the product.

For retail companies, the result is expected to be input in efforts to consolidate inventory management performance because it can guarantee the product availability, enhance services to consumers (Levy et al., 2019) and improve the company's economy. One of the policies is when to delay repurchase transactions for slow-moving product stocks, thereby minimizing over-stock and determining the value of safety stock through a rational method. The implication is the efficiency of the inventory costs which have an impact on maximizing company profits (Huang, Xiao, Dai, & Yan, 2019; Islam & Amin, 2020).

2. Literature review

SF is interesting and challenging as evidenced by the comparative studies of various methods (Athiyarath, Paul, & Krishnaswamy, 2020; Ensafi, Amin, Zhang, & Shah, 2022; Kilimci et al., 2019; Pavlyshenko, 2019). Many retail SF methods have been studied and developed in recent years to improve performance as summarized in Table 1. These findings showed that the classical sales forecasting method still dominates study in retail SF. Therefore, innovation is needed to develop a new method that minimizes forecasting errors, including case study diversification.

The majority of the classical SF methods only rely on one forecasting criterion, which is the number of sales extrapolated from historical trends and seasonal effects to predict future sales (Fridley, 2018; Gustriansyah et al., 2019; Pavlyshenko, 2019). The company needs to manually make adjustments to the statistical forecasts or rely on an expert. Also, several studies used various forecasting criteria obtained from the existing literature (Ilseven & Gol, 2019; Narayanan et al., 2019; Pavlyshenko, 2019). The studies extracted various significant SF criteria (Fridley, 2018; Huang et al., 2019; Verstraete et al., 2020), hence, the model does not use only one criterion. Accuracy in the selection of these SF criteria can enhance the forecasting results, accelerate the training process, and decrease overall computing costs (Chaudhuri et al., 2021).

Other studies utilized computer algorithms to extract the SF criteria or used deep learning methods, hence, the extraction is carried out automatically (Aci & Doğansoy, 2022; Chandriah & Naraganahalli, 2021; Chaudhuri et al., 2021; Lopez-Martin, Sanchez-Esguevillas, Hernandez-Callejo, Arribas, & Carro, 2021; Vallés-Pérez et al., 2022; Verstraete et al., 2020). This method is practical, however, studies found difficulty controlling the extraction results. Therefore, study teams need to initialize/turn hyper-parameters during model generation (Chaudhuri et al., 2021). When the results of SF are less accurate, only the forecasting method can be controlled and verified, which is the main component of the model. Meanwhile, in (Fridley, 2018; Kilimci et al., 2019), the extraction of the criteria is not automatic and the results are controlled and validated. This method is less practical but more reliable. When the results are less accurate, the extraction method which is the initial component of the model can be verified first before (without) the SF method.

The (Fridley, 2018) approach in extracting SF criteria has inspired authors. However, the extraction in this study does not need to first build a data set, indicating it is simpler. This study utilizes a decision-making method (Best-Worst) (Gustriansyah, Ermatita, et al., 2020a) to extract the criteria, which is more efficient (Liang et al., 2020).

Table 1
The summary of relevant work in sales forecasting.

Authors	Method	Advantages	Disadvantages
(Chaudhuri, Gupta, Vamsi, & Bose, 2021)	DNN	Improved purchase prediction accuracy improve	Only for certain e-commerce platforms, the model cannot compare the prediction accuracy in real-time when the data is generated, is more resource-intensive than conventional ML techniques, and is difficult to predict when observational data is missing or at random intervals.
(Aci & Doğansoy, 2022; Chandriah & Naraganahalli, 2021)	LSTM	Faster convergence, efficient	Lack of feature engineering
(Kao & Chueh, 2022)	BPN	Effectively reduce the fluctuation of errors in purchasing	Limited product scalability, lack of feature engineering
(Vallés-Pérez et al., 2022)	RNN -Transformers	No need to retrain the model every time a prediction is needed.	In a real case, feature engineering may be useful in finding better representations.
(Verstraete, Aghezzaf, & Desmet, 2020)	LASSO Regression, Seasonal Naive	Extraction of features by combining expert and automated selection.	Lack of external validity due to using only mape.
(Gustriansyah et al., 2019; Kolade, 2019)	TS	Can predict multiple variables	No feature extraction, lack fit for various attributes
(Narayanan, Sahin, & Robinson, 2019)	TS	Sharing retailer order information is cheaper and easier than sharing POS data	No feature extraction, lack fit for various attributes
(Huang et al., 2019)	RF, ARIMA	Identify the important features, RF is better than ARIMA	Lack of external validity due to using only MAPE.
(Colón, 2019)	RF	A good model to forecast sales data	No feature extraction. It can take hours to train and have a result.
(Husein, Arsyal, Sinaga, & Syahputa, 2019)	GAN (DL)	Good performance	No feature extraction
(Wang et al., 2019)	SA-LSTM	Improved accuracy, reduced number of iterations, and good predictive effect.	No feature extraction, examination only on single product

The resulting criteria will be correlated with database fields. Subsequently, a new data set is built based on these criteria and is clustered using the k-Means. Six validity indices are applied to improve the quality of the clustering results. Finally, the concept of RFM is utilized to measure the value of each product. The integration of these three methods has been developed into a novel SF model called the SalesKBR.

3. Material and methods

3.1. The proposed model

Fig. 1 show the proposed SF model consisting of four steps. The first step is to determine the weight and priority of the criteria be utilized in the forecasting process. This step is the foundation and the most critical because the next stage depends on the relevance of the criteria obtained in this first step. Based on the literature review, the forecasting model generally uses only one criterion, namely the number of sales. However, in this proposed model, several criteria were used that significantly affect SF. This aims to minimize forecasting errors and the method utilized to extract these criteria is the BWM (Gustriansyah, Ermatita, Rini, & Malik, 2020b). This method was selected because it is simpler than AHP, FAHP, ANP, and other similar ones (Liang et al., 2020).

The second step is data collection, where significant criteria obtained from the first step were correlated with the database. Furthermore, the database was generated based on these significant criteria and preprocessing was carried out to clean up incomplete, noise, and outlier data, hence, the knowledge discovery becomes easier and more accurate.

The third step is the core part of this forecasting model. At this stage, the RFM concept that is integrated with the data mining method (k-Means) for data clustering was applied, and the determination of the best number of clusters used six validity indices (Gustriansyah, Ermatita, et al., 2020a). The result is the final score for each product, which is a component of the multiplier in forecasting the number of sales.

The fourth step is number of sales forecast for the following month which can be obtained by multiplying the sales of the current month by the final score of a product. At this step, an evaluation of the forecasting results were carried out to determine the feasibility of implementing this model. The results evaluation was done by measuring the level of error of the forecasting model. Furthermore, the SMAPE metric was utilized to measure deviations from the results. The evaluation outcomes will demonstrate whether the model is feasible to be applied in actuality or needs to be revised by incorporating causal factors. Each method will give different results and a good one is considered to produce a small error rate or high forecasting accuracy.

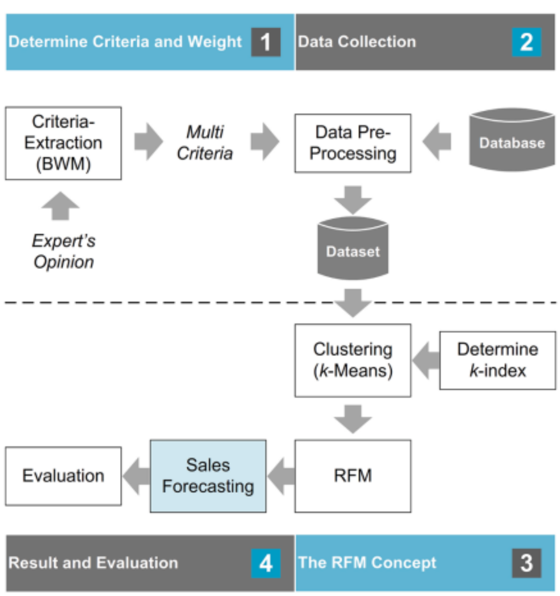


Fig. 1. The proposed model.

3.2. Criteria extraction

The Best-Worst Method was used to extract the forecasting criteria (BWM) (Gustriansyah, Ermatita, et al., 2020b). It generates results by comparing the best and the worst criterion against all others (Mohammadi & Rezaei, 2022). The best criterion is one that has the most significant influence on SF, while the worst has no effect. Experts were asked for their opinion regarding the priority level for the best criterion against others (Best-to-Others) and priority degrees for other criteria against the worst criterion (Others-to-Worst) using numbers 1 to 9.

In simple terms, the stages of BWM are as follows (Mohammadi & Rezaei, 2022):

- Identification of decision criteria (C_j);
- Determine the best (C_B) and the worst (C_W) criterion;
- Determine the priority scale of the best criterion against other criteria (a_{Bj}) using numbers 1–9;
- Determine the priority scale of all other criteria against the worst criterion (a_{jW}) using numbers 1–9;
- Calculate the optimal weight of all criteria (w_j) using Eq. (1) where $w_j > 0$ and $j = 1,2,3,\ldots n$;

$$\min\max_{j}\left\{\left|\frac{w_B}{w_j}-a_{Bj}\right|,\left|\frac{w_j}{w_W}-a_{jW}\right|\right\} \tag{1}$$

- Calculate the output-based Consistency Ratio (CR_o) using Eq. (2) (Liang et al., 2020). ξ^* is the optimal output value of Eq. (1), ξ_{max} is the maximum value of ξ, and $CR_o \leq 0.10$. This is indicates the smaller the CR_o value, the more consistent the comparison vector. Therefore, when the CR_o value $>$ 0.10, the expert assessment is inconsistent and should be revised.

$$CR_o = \frac{\xi^*}{\varepsilon_{max}} \tag{2}$$

3.3. Data collection

This study used a pharmaceutical retail company database containing 2,077 product items, 399,738 sales transactions, and 3,956,683 data on products sold from January 1 to December 31, 2015 (Gustriansyah et al., 2019). The database was generated based on forecasting criteria. Furthermore, preprocessing was carried out to transform the data into a certain format, indicating the dataset becomes easier and more effective. The results obtained are accurate, and it can reduce computation time for large-scale sets. The main activities carried out in preprocessing are:

- Integration (data integration) of several tables into one per product
- Data cleaning involving incomplete, noise (data-errors or outliers), and inconsistent data, and duplication of records
- Data reduction based on the extraction criteria
- Data transformation using natural logarithm (log base 10).

3.4. Recency, frequency, and monetary

Stone and Bob (1989) first proposed the idea of a Recency, Frequency, and Monetary (RFM) model. RFM is a behavior-based model to analyze a consumer and make predictions based on the database (Khalilinezhad, Fazlollahtabar, Minaei-Bidgoli, & Nosratabadi, 2021). Hughes (1994) defined the model using information about past consumer buying behavior (Zhou et al., 2021). It is a simple but effective model that can segment products. This is based on the similarity of data attributes by examining when (recency), how often (frequency), and money spent (monetary) on goods or services (Gustriansyah, Suhandi, & Antony, 2020).

The three RFM criteria are easily computed on any database based on purchase history, easy to understand, and are very good at predicting. The detailed definitions are as follows (Gustriansyah, Suhandi, et al., 2020):

- Recency (R) is the time interval between the customer's last purchase. The smaller the interval, the higher the recency
- Frequency (F) is the number of purchases during the same period. The more the transactions in the time interval, the higher the frequency;
- Monetary (M) is the total money spent on all purchases in a given period. The larger the money, the higher the monetary value.

3.5. K-Means Method

Clustering is the process of grouping large datasets into segments based on their similarities. When the number of transactions becomes larger, the management process becomes a difficult task. However, a clustering method can solve this problem by dividing all products into appropriate clusters based on their similarities. The values of the various clusters can then be estimated and provide informed decisions that are useful for management in rationally using resources.

K-Means is a non-hierarchical clustering method that divides n data into k clusters where the intra-cluster similarity is high (the sum of intra-cluster squares is minimal) and the inter is low (the sum of inter-cluster squares is maximal) (Gustriansyah, Suhandi, et al., 2020). Furthermore, k-Means is one of the most widespread clustering methods because of the speed of determining the center of the cluster (centroid) and its simple algorithm (Abdelaziz & Lu, 2019; Hossain et al., 2019). This method utilizes the Euclidean distance rule to discover the similarity of data.

The stages of data clustering with the k-Means method consist of: (Gustriansyah, Suhandi, et al., 2020).

- Determine the number of k clusters;
- Initialize the value of k as the center of the cluster (centroid) randomly;
- Group each into the nearest cluster. Use the Euclidean distance rule to determine the closeness of two data points;
- Re-calculate each centroid by averaging all data with the current cluster;
- Regroup each of the data using all-new centroids (return to stage 3);
- Data clustering is complete when the centroid no longer changes.

3.6. Validity index

The best number of k clusters for the k-Means method is determined using six validity indices, namely Silhouette, Ratkowsky-Lance, Dunn, Krzanowski-Lai, Davies-Bouldin, and Calinski-Harabasz as shown in Table 2.

3.7. Sales forecasting

The sales of each product for the following months can be predicted by multiplying the sales in the previous month, which can be actual or

Table 2
The formula of validity indices.

No	Validity indices	Note
1.	Silhouette Index (SI) (Rousseeuw, 1987): $SI = \frac{1}{n}\sum_{i=1}^{n} s(i)$ Where: $s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$ $a(i) = \frac{1}{\lvert C_i \rvert - 1}\sum_{j \in C_i,\, i \neq j} d(i,j)$ $b(i) = \min_{i \neq j} \frac{1}{\lvert C_j \rvert}\sum_{j \in C_j} d(i,j)$	Let j denotes another object in the cluster $\lvert C_i \rvert$, $d(i,j)$ denotes the square of Euclidean distance between objects i and j in cluster C_i, and $b(i)$ denotes the distance the average of object i with all objects in other clusters. The SI value close to 1 is the best number of k clusters.
2.	Ratkowsky-Lance Index (RLI) (Ratkowsky & Lance, 1978): $RLI = \frac{\overline{S}}{\sqrt{k}}$ Where: $\overline{S}^2 = \frac{1}{p}\sum_{j=1}^{p} \frac{BGSS_j}{TSS_j}$ $BGSS_j = \sum_{q=1}^{k} n_q (c_{qj} - \overline{x}_j)^2$ $TSS_j = \sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2$	$BGSS$ denotes the between-groups sum of square for each data and TSS is the total sum of square of each data within the cluster. The maximal RLI value is the best number of k clusters.
3.	Dunn Index (DI) (Dunn, 1973): $DI = \frac{\min_{1 \le i \le j \le k} d(C_i, C_j)}{\max_{1 \le m \le k} d(C_m)}$ Where: $d(C_i, C_j) = \min_{1 \le i \le j \le k} d_{ij}$ $d(C_m) = \max_{i,j \in C_m} d_{ij}$	$d(C_i, C_j)$ denotes the square of Euclidean distance between pairs of objects in cluster C_i and C_j (inter-cluster). The $d(C_m)$ is the square of Euclidean distance between objects in one cluster (intra-cluster), and the d_{ij} is the square of Euclidean distance between object i and j. The maximal DI value is the best number of k clusters.
4.	Krzanowski-Lai Index (KLI) (Krzanowski & Lai, 2006): $KLI(k) = \left\lvert \frac{diff(k)}{diff(k+1)} \right\rvert$ Where: $diff(k) = (k-1)^{2/D} WSS_{k-1} - k^{2/D} WSS_k$ for $k = 2, 3, \ldots$	$diff(k)$ denotes the differentiation of the function. D is the feature dimension of the input (number of attributes), WSS_k is the Within-Clusters Sum of Squares in k clusters. The maximal $KLI(k)$ value is the best number of k clusters.
5.	Davies-Bouldin Index (DBI) (Davies & Bouldin, 1979): $DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j} R_{ij}$ Where: $R_{ij} = \frac{WSS_i + WSS_j}{BSS_{ij}}$ $WSS_i = \frac{1}{m_i}\sum_{j=1}^{m_i} d(x_j, c_i)$ $BSS_{i,j} = d(c_i, c_j)$	BSS is the Between-Clusters Sum of Square, $d(x,y)$ is the Euclidean distance between x and y, x_j denotes the cluster j, and c_j is the centroid of cluster x_j. The minimal DBI value is the best number of k clusters ($DBI \ge 0$).
6.	Calinski-Harabasz Index (CHI) (Calinski & Harabasz, 1974): $CHI(k) = \frac{BSS/(k-1)}{WSS/(n-k)}$	The maximal $CHI(k)$ value is the best number of k clusters.

forecast values by the final score of each product as an external factor. In this study, the final RFM score was determined by the weights added to the criteria R, F, and M as shown in Eq. (3) (Gustriansyah, Ermatita, et al., 2020a), where α, β, and γ are the weights of each R, F, and M. These weights were obtained from the extraction criteria. Furthermore, the final RFM score of each product is a multiplier factor in predicting the sales of each product (Monalisa, Nadya, & Novita, 2019).

$$\text{Final score} = \alpha^{*}\text{R} + \beta^{*}\text{F} + \gamma^{*}\text{M} \quad (3)$$

3.8. Evaluation

Measuring the deviation level from the forecasting results against the actual data is a way of evaluating the model performance. This indicates the smaller the error rate generated by the method, the higher the level of accuracy and performance.

The metric used to measure the error rate of the forecasting method was the Symmetric Mean Absolute Percentage Error (SMAPE). This is a variant of MAPE introduced by Armstrong, Collopy, and Makridakis

(1993), and it measures the deviation level of the forecasting model based on the relative error. This metric has an error limit, hence, it is better able to deal with outliers compared to other metrics. The SMAPE error limit used in this study is 0–100% (Maiseli, 2019). It has an advantage over MAPE because it can solve the 'infinite issue' due to zeros in the denominator (Jierula, Wang, OH, & Wang, 2021). The SMAPE formula is shown in Eq. (4). Assuming that the actual value of the n number of data for the t-th is R_t and the forecast value is F_t. Table 3 presents the error rate and accuracy of the metric.

$$SMAPE = \frac{\sum_{t=1}^{n}|R_t - F_t|}{\sum_{t=1}^{n}(R_t + F_t)} \quad (4)$$

4. Results and discussion

Three experts who are responsible for the retail sales determine the significant criteria that affect the process. They create and select the criteria that most influence the process. Initially, 20 criteria were obtained, but only 10 were selected and assessed for their priority level. This begins from the priority level for the best criterion against all others (B2O) and is followed by the priority for all other against the worst (O2W). The assessment results from these experts were determined using Eq. (1).

The average weight of each criterion and the resulting optimal value of ξ^* are presented in Table 4. This shows the criteria 'Frequency', 'Quantity' and 'Monetary' are the three that most significantly affect SF because they have the highest weights.

The value of ξ^* generated by BWM in Table 4 is close to zero. This shows the consistency ratio (CR_o) of the criteria comparison vector is high, hence, the assessment results from the experts is reliable and accepted. Subsequently, the weights of these FQM criteria will be a multiplier factor in determining the final score for the SF process.

The three significant criteria which were the results of the extraction were correlated (relevant) with the fields from the company's actual database and all the data were imported into a new dataset. Meanwhile, the column containing the number of sales transactions for each product was correlated with the frequency criterion (F). The number of sales was correlated with the quantity criterion (Q), and the value (total money paid by consumers) was correlated with monetary (M). The new partial dataset is presented in Table 5.

The value interval for criterion F obtained from the data set in Table 5 is 1 – 1.726. This means the greater the value of F, the more often the product is sold. Also, the value interval for the Q criteria is 1 – 20,419, which means the greater Q, the higher the number of sales. The value interval for the M criteria is Rp. 1,250 – Rp. 220,588,445, which showed the greater the M value, the higher the total money from selling the product from January 1 to December 31, 2015.

The next step was data preprocessing which started by importing and integrating three tables, namely the product, *trans*-sales, and sales_details into a new dataset. Meanwhile, unrelated fields, such as no receipt, category, and others were reduced. A column (month) was created by separating the data from the TransRec field into a date, month, and year. The data types of the various fields were converted before being used in the model. Furthermore, the dataset was cleaned of 53 products consisting of 34 incomplete data, 4 that was not sold, 3 incorrect data, and 12 outliers, leaving 2,024 products to be clustered using the k-Means method. Table 6 presents a visualization of data based on the number of sales, and it shows a trending and random patterns.

The last step of data-preprocessing was transformation using a logarithmic scale (log base 10) (Zumel & Mount, 2019). Fig. 2a shows all the FQM data, and 2b shows those transformed to lognormal, hence, they were normally distributed.

Table 3
The error rate and accuracy of the metric.

Error rate	Accuracy
< 10%	High
10–20 %	Good
20–50%	Reasonable
> 50%	Incorrect

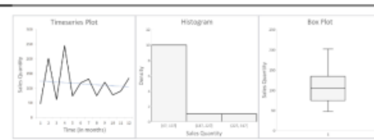
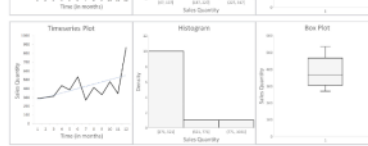
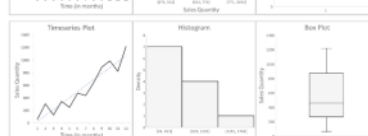
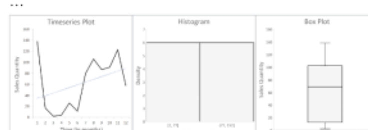
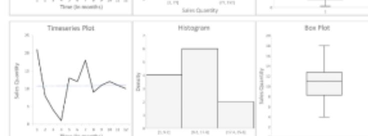
Table 4
The weight of each criterion from the experts.

Criteria		Expert 1	Expert 2	Expert 3	Average
C_1*	Monetary	0.154645	0.114239	0.262750	0.177211
C_2*	Lead time	0.077323	0.114239	0.066013	0.085858
C_3*	Quantity	0.103097	0.274174	0.165033	0.180768
C_4*	Season	0.154645	0.085679	0.110022	0.116782
C_5*	Service	0.061858	0.057120	0.066013	0.061664
C_6*	Check stock	0.051548	0.068543	0.047152	0.055748
C_7*	Frequency	0.252316	0.171359	0.165033	0.196236
C_8*	Discount	0.044184	0.048960	0.041258	0.044801
C_9*	Management	0.078679	0.042840	0.055011	0.058843
C_{10}*	Speculation	0.021705	0.022848	0.021715	0.022089
	ξ*	0.056975	0.068543	0.067316	0.064278

Table 5
The new partial dataset from 1 Jan – 31 Dec 2015.

Month	Product ID	Frequency	Quantity	Monetary
1	AI-0024	5	47	14,545
1	AI-0025	16	292	118,166
⋮	⋮	⋮	⋮	⋮
4	AI-0024	16	245	63,000
⋮	⋮	⋮	⋮	⋮
8	AI-0539	77	201	13,455,685
⋮	⋮	⋮	⋮	⋮
12	R-5154	202	5,014	19,340,000
12	R-5226	1	30	463,650

Table 6
Visualization of a partial dataset based on the number of sales.

Product ID	Plot data
AI-0024	
AI-0025	
AI-0027	
…	…
R-4494	
R-4495	

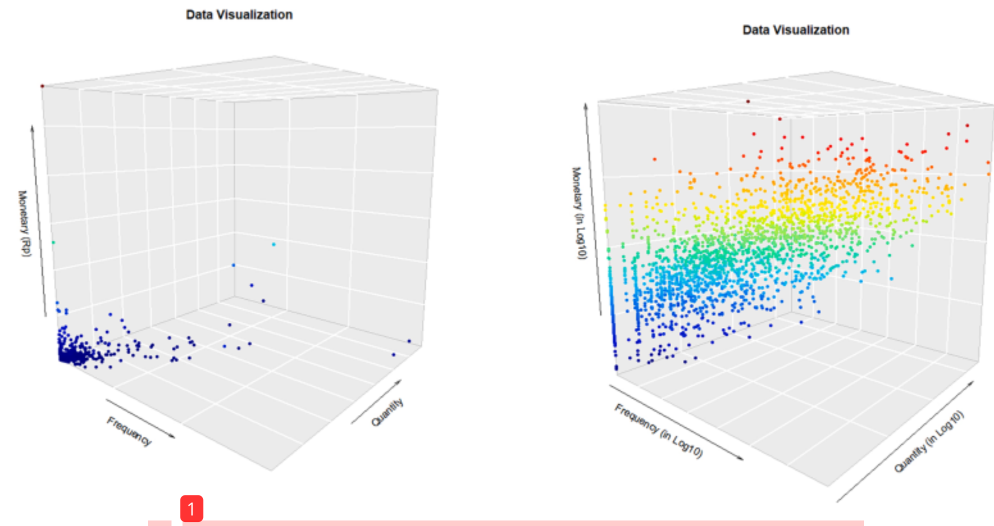


Fig. 2. Visualization of FQM data distribution before and after transformation with log base 10.

Fig. 3b shows that the product with a high FQM value was red and at the top of the cube. In contrast, those with low values were dark blue and at the bottom. Before the new data set was clustered using the k-Means method, the best number of k clusters were determined in advance. In this study, the six validity indices determined the best number of k clusters with R programming software version 3.5.3. The best from the dataset were measured with the Silhouette, Ratkowsky, Dunn, Krzanowski-Lai, Calinski-Harabasz, and Davies-Bouldin indices (Gustriansyah, Suhandi, et al., 2020). The clusters measured start from $k = 1$, 2, 3, …, 10, and the results are as demonstrated in Fig. 3.

The k-value result from Fig. 3 showed the best number of k clusters for the k-Mean method was $k = 2$. Furthermore, the dataset was clustered using k-Means with $k = 2$, and it utilizes RStudio software. Table 7 presents the FQM clustering and its value intervals.

Based on Table 7, the scores for each cluster were justified by a logical value scale for each criterion. The frequency score will be higher when the sales are larger. Furthermore, the quantity criteria score will be higher when the number of product sales is bigger. The score of the monetary criteria will also be higher when the total money spent is larger. Table 8 presents a score scale for each FQM criterion based on the scoring justification.

Table 7
FQM criteria and their value intervals for each cluster.

Criteria	Cluster 1	Cluster 2
Frequency	0 – 243	F > 243
Quantity	0 – 2,059	Q > 2,059
Monetary	0 – 11,067,439	M > 11,067,439

Table 8
Scales and scores for each FQM Criterion.

Score	Frequency	Quantity	Monetary
1	0 – 243	0 – 2,059	0 – 11,067,439
2	F > 243	Q > 2,059	M > 11,067,439

The data scoring aimed to convert the value of each product into a number, hence, it can be utilized in calculations. The conversion score for each product was based on Table 8. Meanwhile, the final score was determined using Eq. (3), where the weights of the F, Q, M criteria were 0.196236, 0.180768, and 0.177211 respectively obtained from Table 4.

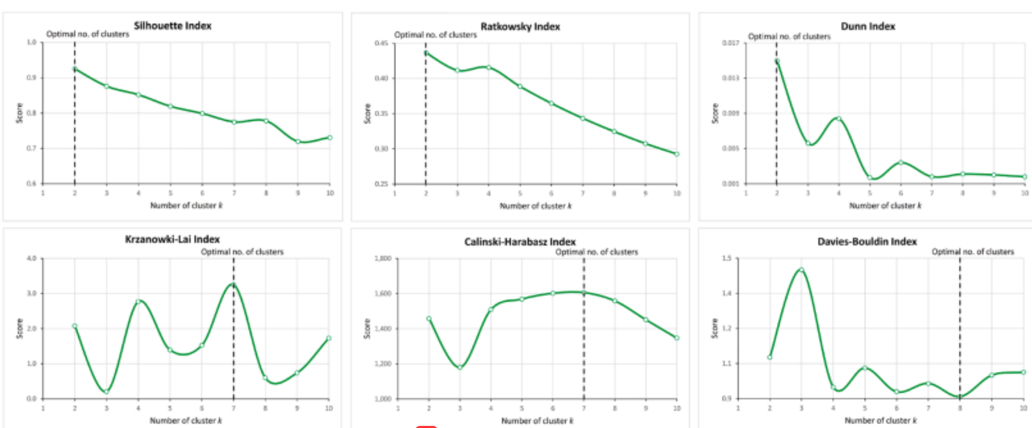


Fig. 3. The best number of k clusters of each index.

Table 9
The monthly partial data set of FQM scoring and final score for each product.

Month	Product ID	Frequency	Quantity	Monetary	Score			Final score
					F	Q	M	
1	AI-0024	5	47	14,545	1	1	1	0.5542
1	AI-0025	16	292	118,166	1	1	1	0.5542
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4	AI-0024	16	245	63,000	1	1	1	0.5542
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
8	AI-0539	77	201	13,455,685	1	1	2	0.7314
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12	R-5154	202	5,014	19,340,000	1	2	2	0.9122
12	R-5226	1	30	463,650	1	1	1	0.5542

Table 9 presents the FQM scoring results and the final score. In addition, the final score is a multiplier factor for forecasting the number of sales of each product.

SF for the following month (period t + 1) is obtained by multiplying the number of sales in the current month (period t) by the final score of each product (Table 9). Table 10 presents the monthly partial forecasting results for each product.

The error rate of the SF model was evaluated using the SMAPE metric. Fig. 4 shows the SMAPE value for all the products monthly, which was between 20% and 30%, and the average value for a year was 27.12%. Therefore, the accuracy of the forecasting model was at a reasonable level. This result showed that the SalesKBR can be utilized as a retail SF model, and it has the potential to be further developed.

5. Conclusions

A novel sales forecasting approach called SalesKBR has been developed in this study. This approach integrates the Best-Worst Method and k-Means into the Recency-Frequency-Monetary concept. The SalesKBR is multi-criteria-based and applies the concept of RFM to products. Furthermore, it presents a new approach to valuing products using the RFM model that is typically used for consumer assessment. This study identified Frequency, Quantity, and Monetary as the most influential criteria in retail SF based on the RFM concept.

The results showed the average accuracy level of sales forecasting was reasonable. This means the SalesKBR is a retail SF model with a high level of product diversity. However, it has a relatively small error rate (27.12%), implying it can potentially be an alternative solution for stakeholders to make scientific and effective management decisions. It was found that most of the error rates were obtained from over sales, not under sales. This means these deviations do not significantly affect the level of service to consumers.

The limitation of this study include, first, the scope of retail sales is limited to a particular platform. Therefore, the generalizability of the findings is limited even though they contextually represent retail sales. Further studies could examine the approach on different platforms to increase the generalizability. Secondly, SF criteria used are limited, and in situations where the criteria are complex, the approach may not be appropriate. Deep learning methods can also be explored to adopt complex criteria and compare them with existing benchmarks. Thirdly, the dataset used is quite large, and when it is much larger, this approach may require adjustments to improve the forecasting performance. Therefore, the performance of the models that involve different dataset sizes can be compared.

CRediT authorship contribution statement

Rendra Gustriansyah: Conceptualization, Data curation, Investigation, Software, Validation, Writing – original draft, Resources. **Ermatita Ermatita:** Supervision, Methodology, Visualization, Writing – review & editing. **Dian Palupi Rini:** Supervision, Methodology, Validation, Writing – original draft, Formal analysis.

Table 10
The monthly partial dataset of actual and forecast sales.

Row	Month	Product ID	Actual sales	Forecasting
1	2	AI-0024	202	26
2	2	AI-0025	301	161
⋮	⋮	⋮	⋮	⋮
4,612	6	R-0755	22	21
⋮	⋮	⋮	⋮	⋮
9,921	12	R-5145	104	106
9,922	12	R-5154	5,014	5,109

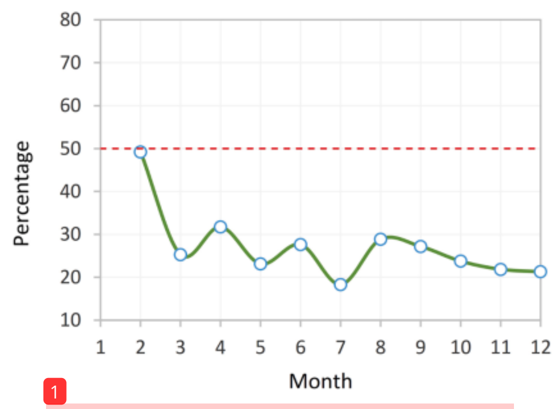


Fig. 4. The average SMAPE value for all the products monthly.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

Abdelaziz, S., & Lu, S. (2019). K-means algorithm with level set for brain tumor segmentation. *Indonesian Journal of Electrical Engineering and Computer. Science, 15*(2), 991–1000. https://doi.org/10.11591/ijeecs.v15.i2.pp991-1000

Aci, M., & Doğansoy, G. A. (2022). Demand forecasting for e-retail sector using machine learning and deep learning methods. *Journal of the Faculty of Engineering and Architecture of Gazi University, 37*(3), 1325–1339. https://doi.org/10.17341/gazimmfd.944081

Anitha, P., & Patil, M. M. (2019). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences.* https://doi.org/10.1016/j.jksuci.2019.12.011

Athiyarath, S., Paul, M., & Krishnaswamy, S. (2020). A Comparative Study and Analysis of Time Series Forecasting Techniques. *SN Computer Science, 1*(3), 175–182. https://doi.org/10.1007/s42979-020-00180-5

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1974.

Chandriah, K. K., & Naraganahalli, R. V. (2021). RNN / LSTM with modified Adam optimizer in deep learning approach for automobile spare parts demand forecasting. *Multimedia Tools and Applications, 80*(17), 26145–26159. https://doi.org/10.1007/s11042-021-10913-0

Chaudhuri, N., Gupta, G., Vamsi, V., & Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems, 149*, 1–10. https://doi.org/10.1016/j.dss.2021.113622

Colón, J. G. S. (2019). Data Mining Techniques and Machine Learning Model for Walmart Weekly Sales Forecast. *Retrieved from.* https://prcrepository.org/xmlui/handle/20.500.12475/174.

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2)*, 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics, 3*(3), 32–57. https://doi.org/10.1080/01969727308546046

Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights, 2*(1), 1–16. https://doi.org/10.1016/j.jjimei.2022.100058

Fridley, L. (2018). *Improving Online Demand Forecast using Novel Features in Website Data: A Case Study at Zara.*

Gao, W., & Fan, H. (2021). Omni-Channel Customer Experience (In)Consistency and Service Success: A Study Based on Polynomial Regression Analysis. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(6), 1997–2013. https://doi.org/10.3390/jtaer16060112

Gustriansyah, R., Ermatita, Rini, D. P., & Malik, R. F. (2020a). Integration of Decision-Making Method and Data-Mining Method as A Preliminary Study of Novel Sales Forecasting Method. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(4), 5730–5735. 10.30534/ijatcse/2020/227942020.

Gustriansyah, R., Ermatita, Rini, D. P., & Malik, R. F. (2020b). The Criteria That Have A Significant Effect on Forecasting the Number of Sales Using the Best-Worst Method. In *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 236–241). Jakarta: IEEE. 10.1109/ICIMCIS51567.2020.9354276.

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science, 18*(1), 470–477. https://doi.org/10.11591/ijeecs.v18.i1.pp470-477

Gustriansyah, R., Suhandi, N., Antony, F., & Sanmorino, A. (2019). Single Exponential Smoothing Method to Predict Sales Multiple Products. *Journal of Physics: Conference Series, 1175*(1), 1–7. https://doi.org/10.1088/1742-6596/1175/1/012036

Hossain, M. Z., Akhtar, M. N., Ahmad, R. B., & Rahman, M. (2019). A dynamic K-means clustering for data mining. *Indonesian Journal of Electrical Engineering and Computer Science, 13*(2), 521–526. https://doi.org/10.11591/ijeecs.v13.i2.pp521-526

Huang, W., Xiao, Q., Dai, H., & Yan, N. (2019). Sales Forecast for O2O Services - Based on Incremental Random Forest Method. In *15th International Conference on Service Systems and Service Management (ICSSSM)* (pp. 1–5). IEEE. 10.1109/ICSSSM.2018.8465121.

Husein, A. M., Arsyal, M., Sinaga, S., & Syahputa, H. (2019). Generative Adversarial Networks Time Series Models to Forecast Medicine Daily Sales in Hospital. *SinkrOn, 3*(2), 112. 10.33395/sinkron.v3i2.10044.

Hyndman, R. J., & Athanasopoulos, G. (2019). *Forecasting: principles and practice* (3rd ed.). Melbourne, Australia: OTexts. Retrieved from OTexts.com/fpp3.

Ilseven, E., & Gol, M. (2019). A comparative study on feature selection based improvement of medium-term demand forecast accuracy. In *IEEE Milan PowerTech* (pp. 1–6). IEEE. 10.1109/PTC.2019.8810598.

Islam, S., & Amin, S. H. (2020). Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques. *Journal of Big Data, 7*(1), 65. https://doi.org/10.1186/s40537-020-00345-2

Jierula, A., Wang, S., OH, T.-M., & Wang, P. (2021). Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data. *Applied Sciences, 11*(5), 2314. https://doi.org/10.3390/app11052314

Kao, C.-Y., & Chueh, H.-E. (2022). Deep Learning Based Purchase Forecasting for Food Producer-Retailer Team Merchandising. *Scientific Programming, 2022*, 1–9. https://doi.org/10.1155/2022/2857850

Khalilinezhad, S., Fazlollahtabar, H., Minaei-Bidgoli, B., & Nosratabadi, H. E. (2021). Detecting Valuable Customers Using the Trade Patterns of Financial Transactions Applying Integrated RFM and OLAP. Retrieved from *International Journal of Industrial Engineering & Production Research, 32*(3), 1–15 http://ijiepr.iust.ac.ir/article-1-1169-en.html.

Kilimci, Z. H., Akyuz, A. O., Uysal, M., Akyokus, S., Uysal, M. O., Atak Bulbul, B., & Ekmis, M. A. (2019). An Improved Demand Forecasting Model Using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain. *Complexity, 2019*, 1–15. https://doi.org/10.1155/2019/9067367

Kolade, O. J. (2019). Demand Forecasting and Measuring Forecast Accuracy in a Pharmacy. *Economic Development, Technological Change, and Growth, 15*(3), 157–169.

Krzanowski, W. J., & Lai, Y. T. (2006). A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics, 44*(1), 23. https://doi.org/10.2307/2531893

Levy, M., Weitz, B., & Grewal, D. (2019). *Retailing Management* ((10th ed.).). McGraw-Hill Education.

Liang, F., Brunelli, M., & Rezaei, J. (2020). Consistency issues in the best worst method: Measurements and thresholds. *Omega, 96*, 1–11. https://doi.org/10.1016/j.omega.2019.102175

Lopez-Martin, M., Sanchez-Esguevillas, A., Hernandez-Callejo, L., Arribas, J. I., & Carro, B. (2021). Additive Ensemble Neural Network with Constrained Weighted Quantile Loss for Probabilistic Electric-Load Forecasting. *Sensors, 21*(9), 1–37. https://doi.org/10.3390/s21092979

Maiseli, B. J. (2019). Optimum design of chamfer masks using symmetric mean absolute percentage error. *EURASIP Journal on Image and Video Processing, 2019*(1), 74. https://doi.org/10.1186/s13640-019-0475-y

Mohammadi, M., & Rezaei, J. (2022). Hierarchical Evaluation of Criteria and Alternatives Within BWM: A Monte Carlo Approach. In *Advances in Best-Worst Method* (pp. 16–28). 10.1007/978-3-030-89795-6_2.

Monalisa, S., Nadya, P., & Novita, R. (2019). Analysis for Customer Lifetime Value Categorization with RFM Model. *Procedia Computer Science, 161*, 834–840. https://doi.org/10.1016/j.procs.2019.11.190

Narayanan, A., Sahin, F., & Robinson, E. P. (2019). Demand and order-fulfillment planning: The impact of point-of-sale data, retailer orders and distribution center orders on forecast accuracy. *Journal of Operations Management, 65*(5), 468–486. https://doi.org/10.1002/joom.1026

Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data, 4*(1), 1–11. https://doi.org/10.3390/data4010015

Ratkowsky, D., & Lance, G. (1978). Criterion for determining the number of groups in a classification. Retrieved from *Australian Computer Journal, 10*(3), 115–117 http://hdl.handle.net/102.100.100/300266?index=1.

Rousseeuw, P. J. (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Vallés-Pérez, I., Soria-Olivas, E., Martínez-Sober, M., Serrano-López, A. J., Gómez-Sanchís, J., & Mateo, F. (2022). Approaching sales forecasting using recurrent neural networks and transformers. *Expert Systems with Applications, 201*, 1–9. https://doi.org/10.1016/j.eswa.2022.116993

Verstraete, G., Aghezzaf, E.-H., & Desmet, B. (2020). A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. *Computers & Industrial Engineering, 139*, 106–169. https://doi.org/10.1016/j.cie.2019.106169

Wang, Y., Chang, D., & Zhou, C. (2019). The Study of a Sales Forecast Model Based on SA-LSTM. *Journal of Physics: Conference Series, 1314*, Article 012215. https://doi.org/10.1088/1742-6596/1314/1/012215

Zhou, J., Wei, J., & Xu, B. (2021). Customer segmentation by web content mining. *Journal of Retailing and Consumer Services, 61*, Article 102588. https://doi.org/10.1016/j.jretconser.2021.102588

Zumel, N., & Mount, J. (2019). *Practical Data Science with R (Second)*. Shelter Island, New York: Manning Publications Co.

turnitin 2

ORIGINALITY REPORT

91%	38%	90%	22%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Rendra Gustriansyah, Ermatita Ermatita, Dian Palupi Rini. "An approach for sales forecasting", Expert Systems with Applications, 2022 Publication	79%
2	reunir.unir.net Internet Source	1%
3	Submitted to Tilburg University Student Paper	1%
4	Submitted to CSU, San Jose State University Student Paper	1%
5	spectrum.library.concordia.ca Internet Source	1%
6	Submitted to Coventry University Student Paper	1%
7	jurnal.polgan.ac.id Internet Source	1%
8	Submitted to University of Nottingham Student Paper	1%

9	radar.ibiss.bg.ac.rs Internet Source	1 %
10	Submitted to IUBH - Internationale Hochschule Bad Honnef-Bonn Student Paper	1 %
11	Submitted to Vilnius Gediminas Technical University Student Paper	<1 %
12	ueaeprints.uea.ac.uk Internet Source	<1 %
13	www.thefreelibrary.com Internet Source	<1 %
14	eprints.upnyk.ac.id Internet Source	<1 %
15	digitalcommons.usf.edu Internet Source	<1 %
16	bc.pollub.pl Internet Source	<1 %
17	link.springer.com Internet Source	<1 %
18	Submitted to Sim University Student Paper	<1 %
19	Submitted to University of Northumbria at Newcastle Student Paper	<1 %

20	iecscience.org Internet Source	<1 %
21	www.omicsonline.org Internet Source	<1 %
22	Submitted to Universitas Prima Indonesia Student Paper	<1 %
23	Submitted to Universiti Teknologi MARA Student Paper	<1 %
24	Tu Van Binh, Ngo Giang Thy, Ho Thi Nam Phuong. "Measure of CLV Toward Market Segmentation Approach in the Telecommunication Sector (Vietnam)", SAGE Open, 2021 Publication	<1 %
25	Submitted to Wright State University Student Paper	<1 %
26	Eddy Ibrahim, Maulana Yusuf, Yoga Deniswara. "Determination on Methane Gas in Place in Coal Seam by using Volumetric Method at X Field, East Kalimantan, Indonesia", Journal of Physics: Conference Series, 2020 Publication	<1 %
27	Submitted to University of Bristol Student Paper	<1 %

28	Submitted to Liverpool John Moores University Student Paper	<1 %
29	repository.tudelft.nl Internet Source	<1 %
30	revistas.udc.gal Internet Source	<1 %
31	Submitted to iGroup Student Paper	<1 %
32	www.researchgate.net Internet Source	<1 %
33	dokumen.pub Internet Source	<1 %
34	prcrepository.org Internet Source	<1 %
35	www.scilit.net Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		