p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 9 September 2024 http://jist.publikasiindonesia.id/

Cluster Analysis and Forecasting on Local Shoe Products: Case Study for Ventela in Indonesia

Vies Sata Zullah[1*], Raden Mohamad Atok[2]
Institut Teknologi Sepuluh Nopember, Indonesia
Email: vszullah@gmail.com[1*], moh_atok@statistika.its.ac.id[2]

*Correspondence

ABSTRACT

Keywords: k-means cluster; exponential smoothing; ventela shoe models; e-commerce transactions; best-selling cluster forecasting.

Shoes are a secondary need that is in demand by all age groups. Each shoe brand has many models, which means a store must provide complete stock to meet consumer needs. The available models have different purchasing power or demand, which creates difficulties for stores in determining shoe products that are often sold and shoe products that are not in demand by customers. The data used in cluster formation includes three variables recorded and collected from e-commerce transactions, namely Number of Visitors, Number of Buyers, and Total Sales. Based on these variables, Ventela shoe models are grouped into three clusters, namely low-selling, normal, and best-selling. Next, the variable number of transactions for Ventela shoe models in the best-selling cluster is taken to be predicted using the Exponential Smoothing method. The forecasts obtained are used to determine future demand to maximize profits. Based on the results of the clustering analysis, it was found that the number of shoe models included in the best-selling cluster was six, including (1) Ventela Ethnic Low All Black, (2) Ventela Ethnic Low Black Natural, (3) Ventela Public Low Black Natural, (4) Ventela Public Low Cream, (5) Ventela Republic Low Black Natural, and (6) Ventela Republic Low White. Referring to the sample of this study, which only spanned less than three years, several shoe models produced a forecast value of zero. It means, based on the forecast results for the next 12 months, there may be no sales.

Introduction

Locally made shoes are increasingly appearing for sale on social media and various e-commerce sites. The number of enthusiasts has also increased significantly. Tokopedia recorded that sales transactions for local shoe brands have almost doubled. This is driven by the increasing level of awareness of Indonesian consumers who are proud to use local products. Apart from that, many people also really appreciate locally-made shoe products (Yusditara et al., 2022). E-commerce is a platform to make shopping easier for customers. E-commerce itself facilitates the ability to make transactions from anywhere; customers can also make purchases directly, there are cuts in distribution channels, and there are cost savings (Alwiyah & Gata, 2019).

The increasing interest in local shoes has become a concern for local shoe sellers to pay more attention to the inventory or stock of each model (or what can be called articles). Some inventory problems that often occur are excess stock (overstock) or running out/short of stock (stockout) (Rachmawati & Lentari, 2022). Overstock or stockout can cause several losses. When there is too much stock, it causes reduced profitability due to the accumulation of goods. This equates to tying up cash that can be used to buy stock that comes out more frequently. Apart from that, excess stock can also pressure sellers to sell their products at prices below the margin to get rid of excess stock.

Meanwhile, stock shortages can lead to lost sales opportunities because customers cannot buy the products they need and result in loss of income for sellers. When customers cannot buy the products they need, it can cause dissatisfaction and potentially damage a business's reputation (Rachmawati & Lentari, 2022). Overstock and stockouts often occur in shoe products that have quite a lot of articles. There are many shoe brands in Indonesia, one of which is Ventela which is a locally made shoe and is currently quite famous in Indonesia.

Ventela is a local shoe brand with a casual type, which was introduced in 2017 by William Ventela, a vulcanized shoe factory owner, in 1989 in Bandung, West Java. With abundant resources, Ventela Shoes can produce canvas shoes in large quantities and of the best quality so that all groups can have high-quality shoes at affordable prices. Ventela does not have an official store that sells shoes retail to end-consumers but instead distributes them to resellers. So, resellers play an important role in retail sales. There are quite a lot of articles published by Ventela; there are more than 100 shoe articles. Each article has different selling power, so we need to know which articles should have more stock and which are not in demand by the market. These articles need to be grouped or formed into clusters using k-means clustering and forecasting demand so as not to disappoint customers and also allow shoe turnover to dash.

Clustering is one of the main methods for organizing a set of data into clusters so that the elements in each cluster have similarities and differences with other clusters (Pérez-Ortega et al., 2019). This clustering is used to create a report regarding the general characteristics of the groups formed, including shoe models ranging from those that sell poorly, normally and best-selling. One of the most widely used clustering algorithms currently is K-means because of its ease of interpreting the results and implementation (Pérez-Ortega et al., 2019).

Previous research related to clustering using the K-Means method was carried out by (Pratiwi & Marizal, 2022) with the title Application of "K-means Algorithm for Grouping and Least Square Method for Predicting Goods Sales (Case Study: Buana Mart Kendari)". This research obtained results that the grouping of initial and sold stock, as well as predictions of goods sold at the Buana Mart Kendari Store, were successfully developed by applying the K-Means algorithm and the Least Square method. Apart from that, (Jananto, 2022) have also conducted research using the ¬K-Means method with the title "Application of Data Mining to Product Sales Using the K-Means Clustering Method (Case Study of Kakikaki Shoe Store)". This research explains that the advantages of using the K-Means Clustering method are that it is easy to run, fast, and flexible, and the results are easy to understand and can be explained to many people.

In this research, the demand for Ventela shoe models in the best-selling cluster was predicted using the Exponential Smoothing method. The exponential smoothing method is a forecasting method for time series data by giving weight to previous data to predict the value of the following data. This model was chosen because it can accommodate stationary, trend and seasonal patterns simultaneously, even if the amount of data (sample) is limited. There are three types of exponential smoothing methods, namely single, double and triple. Single is used for data that has a stable fluctuating pattern, double is used for data that has a trend pattern, and triple is used for data that has a trend and seasonal pattern (Maricar, 2019).

Fahrudin et al. (2022) carried out previous research that applied the exponential smoothing method, titled "Demand Forecasting of Automobile Sales Using Least Square, Exponential Smoothing, and Double Exponential Smoothing." (Zunaidi, 2022) also conducted research related to "Analysis of Inventory Management in Order to Reduce Overstock (Case Study of TVF Footwear)." This research uses the Exponential Smoothing and Trend Analysis method to predict demand for the three research products.

Based on the previous description, researchers want to conduct research like (Zullah, 2024) using k-means to obtain clusters of Ventela shoe models. However, the Ventela shoe models from one of the cluster results (bestselling) obtained will be predicted using both methods using exponential smoothing to produce demand forecasts for the next period.

Research Methods

K-Means Cluster

K-Means is the simplest and most common clustering method. It can group quite large amounts of data with fast and efficient computing time (Santosa, 2007). The clustering process begins by identifying data with the Euclidean formula, as in Equation (1), and is illustrated in Figure 1.

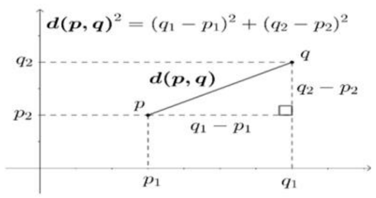


Figure 1 Euclidean formula illustration

Based on Figure 1, in general the Euclidean formula can be stated as follows.

$$d(p,q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \quad (1)$$

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

Where $d(q,p)$ is the distance from point p to point q, q_i indicates the i-th point q, p_i indicates the center point of the i-th cluster p, and i indicates the number of attributes. A data becomes a member of the kth cluster if the distance of the data to the centre of the kth cluster is the smallest compared to the distance to the centres of other clusters. Next, form groups of data that are members of each cluster. The new cluster centre value can be calculated by finding the average value of the data that is a member of the cluster using the formula in Equation (2).

$$d\mu_k = \frac{1}{N_k}\sum_{t=1}^{N_k} x_t \qquad (2)$$

Where μ_k is the centroid point of the k-th cluster, N_k shows the amount of data in the k-th cluster, and x_t is the t-th data in the k-th cluster

In general, the K-Means Cluster initializes centroids randomly. According to (Dwiarni & Setiyono, 2020), the steps for clustering using the K-Means Cluster method are as follows:

1. Select the number of clusters k.
2. Initialize k cluster centers. Most often, this is done randomly.
3. Allocate all data/objects to the nearest cluster. The distance between them determines the closeness of two objects, and the distance between the data and the cluster centre determines the proximity of data to a particular cluster.
4. Recalculate the cluster centre with the current cluster membership. Reassign each object using the new cluster centre. If the cluster centre does not change again, then the clustering process is complete.

One method commonly used to determine the number of clusters is the elbow method. By the elbow method, it can be observed how the sum squared error (SSE) value changes as the number of clusters increases. In the SSE graph against the number of clusters, one can look for points where the decrease in inertia slows down and resembles an elbow shape, which can provide clues about the optimal number of clusters. This point indicates that adding clusters afterwards does not provide significant benefits in reducing SSE (Shi et al., 2021).

Data

The data used in this research is secondary data obtained from historical sales transactions carried out by one of the Ventela shoe resellers in Surabaya. Sales data used starts from August 2021 to March 2024. This research applies two statistical methods, namely clustering and forecasting. Therefore, there are two different variable definitions according to each method. The following research variables used in this research are explained in Table 1 and Table 2.

Table 1
Variables for the clustering

Variable	Symbol	Unit	Information
Number of visitors	X_1	People	The number of individuals who access product pages on e-commerce.
Number of buyers	X_2	People	The number of individuals who purchased the product.
Total sales	X_3	Rupiah	Total sales are income generated from e-commerce transactions.

Table 2
Variables for the forecasting

Variable	Type	Symbol	Unit	Information
Number of product sold	Dependent	Y	Piece	Number of products sold on e-commerce.

Analysis Steps

The next step is clustering analysis based on the K-Means method. This analysis was carried out to group the data into three clusters. In this step, the data is grouped based on three variables, namely the number of visitors, number of buyers, and total sales. The stages carried out for this analysis include:

1. Selection of the number of clusters (K)

At this stage, the number of clusters (K) is determined. Choosing the right K is one of the key stages in K-Means. Indicators that can be used to determine the number of clusters are usually Elbow or Silhouette. The number of clusters in the research was determined to be 3 to represent the low-selling, normal and best-selling sales clusters.

2. Determination of Cluster Centre

Determining the cluster centre is carried out to obtain the cluster centre based on the Euclidean method, as in Equation (1). Determining the cluster centre uses random numbers (arbitrary) as initialization. The initial cluster centre is the initial reference in the clustering process.

3. Cluster Centre Iteration Based on K-Means

Iterations are carried out until convergence; that is, there are no significant changes in the placement of data into clusters. Each iteration involves several steps, including calculating the distance of a data point to the nearest cluster centre, determining cluster members, calculating a new cluster centre based on the average of its members (K-Means), and finally repeating these steps until it converges.

4. Determination of Cluster Labels

After obtaining three clusters, the next step is to determine each cluster's label or name. A best-selling cluster is defined as the cluster with the highest cluster centre value, while a low cluster is defined as the cluster with the lowest cluster centre value.

Forecasting Steps

The experience process is carried out by applying the exponential smoothing method. Exponential smoothing allows us to identify trends and patterns in shoe sales data in best-selling groups. By forecasting these Ventela shoe models, companies or sellers can better prepare for changes in demand and can optimize inventory and sales strategies in hot groups. The following are the stages of Exponential Smoothing forecasting in detail.

1. Selection of the Exponential Smoothing Model

Select the Exponential Smoothing (ES) model that is most appropriate for the data used. Several types of ES models can be used, such as Single Exponential Smoothing (SES), Double Exponential Smoothing (Holt), and Triple Exponential Smoothing (Holt-Winters). The choice of model depends on whether the data has a trend, seasonal component, or both. If the use of visuals does not sufficiently confirm the presence of trends, seasonality, or both, then decomposition is required.

2. Initialize Model Parameters

Initialization of the initial model parameters determines the initial parameters to be applied. For example, in the Single Exponential Smoothing (SES) model, it is necessary to initialize the alpha (α) value, which is the exponential weight parameter for the latest observations. This initial value can be selected using the grid search technique.

3. Model Prediction Calculation

At this stage, prediction calculations are carried out based on the selected model. These calculations are needed to obtain estimated (fitted) values , which are then used to validate the goodness of the model.

4. Measurement of Prediction Error

Prediction error measurements validate the model's ability to predict actual data. The evaluation of prediction quality for time series data usually uses Mean Absolute Percentage Error (MAPE).

5. Error Value Checking

After measuring the model error, the next step is to check whether the error value meets the specified requirements. For example, MAPE is categorized as good if it is less than 20%. If it is not appropriate, then initialize the model parameters again using grid search.

6. Data Forecasting

The final stage is to forecast the data using the methods and parameters that are considered the most optimal. Forecasting is intended to derive some value in the future. The length of the data forecast will be determined later and will be adjusted to the researcher's needs.

Results and Discussion

Descriptive Statistics of Clustering Data

The data used in this research includes four variables, namely Number of Visitors, Number of Buyers, Total Sales, and Number of Products Sold. The first three variables, namely the Number of Visitors, Number of Buyers, and Total Sales, are used for grouping and the remainder, namely the Number of Products Sold, are used for forecasting. Therefore, the descriptive statistics of this research variable are divided into two types, namely grouping variables and forecasting variables. The results of descriptive analysis of the variables used for grouping are presented in Table 3.

Table 3
Descriptive statistics of clustering variables

Measures of variable data centering	Variable		
	Number of visitors (X_1)	Number of buyers (X_2)	Total Sales (X_3)
Minimum	457.00	2.00	Rp 659,481.00
Median	6,954.00	170.00	Rp 49,904,657.00
Maximum	45,697.00	1,115.00	Rp 273,968,637.00
Range	45,240.00	1,113.00	Rp 273,309,156.00
Mean	10,303.00	264.00	Rp 69,261,843.49
Standard deviation	9,786.00	282.00	Rp 67,539,149.93

Table 3 shows that the average number of visitors who only saw the Ventela shoe model was 10,303 people, of which only 2.56% of that number purchased Ventela shoes, with an average total sale of each shoe model of IDR 69,261,843.49. The standard deviation obtained from the three variables has a significant value, indicating that each shoe model has a very varied number of visitors, number of buyers and total sales, or there are very small or very large values for these variables. This is proven by the large range of values, where the most visited model is Republic Low White with 45,697 people, and the least visited is Reborn Reflective Navy with 457 people. The range of buyers is also quite large, where the most purchased model is Ethnic Low Black Natural by 1,115 people, and the least is New Public Low Dark Brown by two people. Republic Low White produced the largest total sales, namely IDR 273,968,637.00 and the lowest total sales, namely New Public Low Dark Brown, amounted to IDR 659,481.00. The condition of each variable can be clarified by visualization using a box plot to see the distribution of the data. The visualization results of the grouping variables are shown in Figure 2.

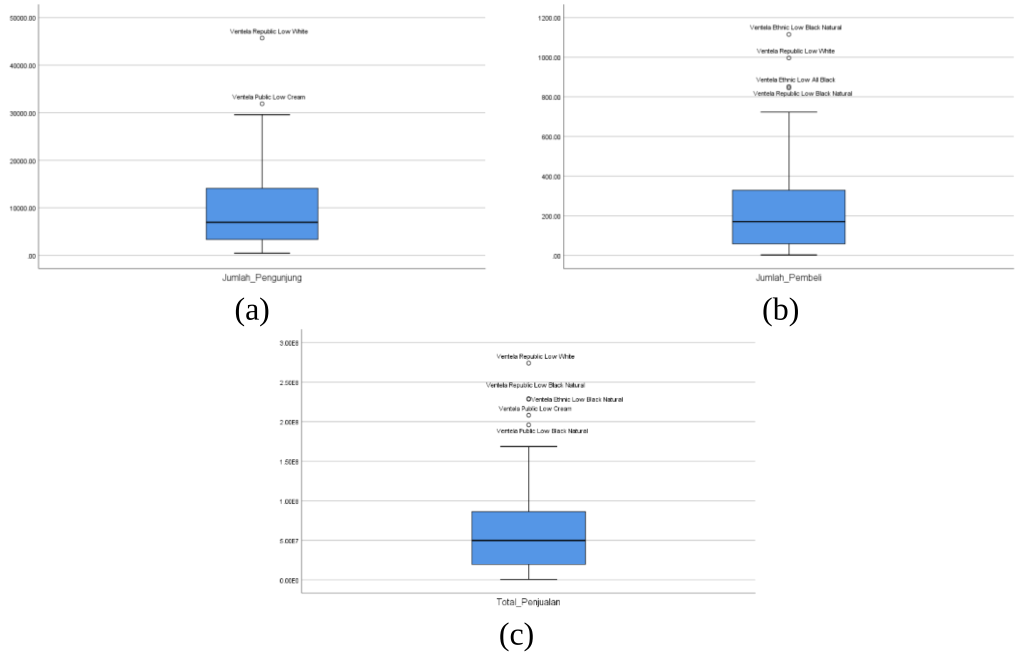


Figure 2. Box plot of clustering variables
(a) Number of visitors, (b) Number of buyers, and (c) Total sales

Figure 2 shows that there are several observations (Ventela shoe model) that are outliers in each variable. Some of the Ventela shoe models include Republic Low White, Public Low Cream, Ethnic Low Black Natural, Republic Low Black Natural, and Public Low Black Natural which have high sales and can be assumed to be the Ventela shoe models that are in the best-selling group, but this statement It is just a guess without any basis for analysis. Therefore, to ensure that the Ventela shoe models are grouped in the best-selling cluster, a cluster analysis was carried out based on three grouping variables using the k-means clustering method.

Clustering Shoe Models Using K-Means

This subchapter discusses the grouping results obtained from the clustering process using the k-means clustering method. In accordance with the principles used by k-means, the number of clusters in the grouping has been determined to be three: the low-selling, normal, and best-selling groups. The initial cluster centre applied in the first stage of the grouping process is presented in Table 4.

Table 4
Initial value of cluster centre

Variable	Cluster		
	1	2	3
Number of visitors	1,147.00	16,496.00	45,697.00
Number of buyers	2.00	314.00	996.00

Total sales	659,481.00	111,804,153.00	273,968,637.00

Table 4 shows that the order of groups that analogize the lowest to highest values in sequence are clusters 1, 2, and 3. This means that cluster 1 represents the group of Ventela shoe models that sell less well, cluster 2 represents the group of Ventela shoe models that sell normally, and cluster 3 represents the best-selling Ventela shoe models. Next, the cluster center will be iterated until it converges and gets the final cluster center. Changes that occur in the cluster center values at each iteration are presented in Table 5.

Table 5
Change in cluster center value

Iteration	**Cluster**		
	1	**2**	**3**
1	23.240.594,687	16.919.588,118	46.870.502,108
2	1.159.838,129	2.347.244,801	9.749.313,140
3	0,000	0,000	0,000

After the iteration to get the cluster center has converged, then the most optimal cluster center value is obtained. The optimal cluster center results are shown in Table 6.

Table 6
Optimal cluster center value

Variable	**Cluster**		
	1	**2**	**3**
Jumlah Pengunjung	4.046,21	14436,33	29.170,50
Jumlah Pembeli	81,00	364,53	869,17
Total Penjualan	25.059.913,64	92.537.320,20	217.348.824,33

As in Table 6, the final cluster center values in Table 4.4 give the same cluster labels. The order of cluster center values from low to high is clustered 1, 2, and 3 so that cluster 1 represents the group of Ventela shoe models that sell less well, cluster 2 represents the group of Ventela shoe models that sell normally, and cluster 3 represents the Ventela shoe models that sell well. The number of Ventela shoe models grouped into each cluster is shown in Table 7 below.

Table 7
Number of cluster members

Cluster	**Number of Members**
1 (low-selling)	28
2 (normal-selling)	15
3 (best-selling)	6

Based on Table 7, there are only 6 Ventela shoe models included in the best-selling cluster (cluster 3). These shoe models are presented in Table 8.

Table 8. Ventela shoe model in the best-selling cluster

No	Model	Color
1	Ventela Ethnic Low	All Black
2	Ventela Ethnic Low	Black Natural
3	Ventela Public Low	Black Natural
4	Ventela Public Low	Cream
5	Ventela Republic Low	Black Natural
6	Ventela Republic Low	White

Based on the Ventela shoe model contained in Table 8, these results indirectly confirm observations that are categorized as outliers in the boxplot for the variables used in the grouping process.

Descriptive Statistics of Forecasting Data

After obtaining members from the best-selling cluster group, the next step is to forecast the number of product sales of the Ventela shoe models included in the cluster. The data that will be predicted is the total products sold for each of the six Ventela shoe models, according to Table 8. The main objective of this forecasting is to determine the potential product sales of each model of Ventela shoes that sell well to avoid the possibility of stockouts. An illustration of the movement in the number of shoe product sales in the best-selling cluster is shown in Figure 3 below.

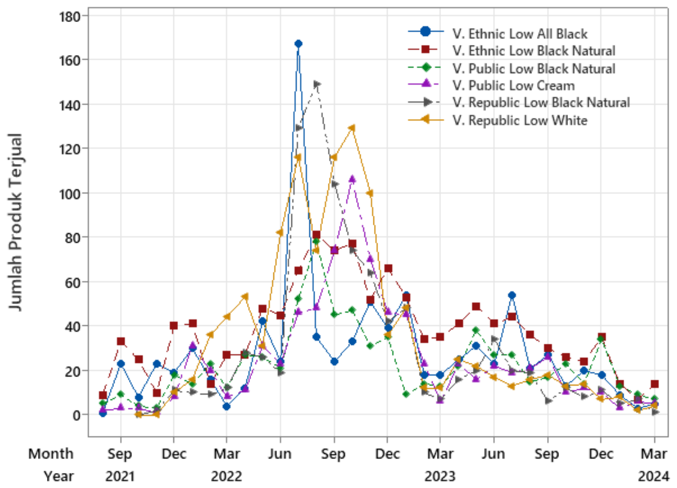


Figure 3. Illustration of the sales number of best-selling cluster shoe products

Figure 3 shows that the movement pattern of the number of products sold for the six shoe models has a trend that tends to be irregular. At the beginning of 2022, all shoe models experience a relatively high increase in sales until their peak in mid-2022. After that, the graph of total products sold will decline until they show a sloping pattern and tend to decline. The pattern of an increase followed by a decrease (up then down) shows that there are two possible forecasting models, namely Double Exponential Smoothing (DES) and Triple Exponential Smoothing (TES).

Data patterns can also be identified using boxplots. Boxplots are created based on data in the same month. For example, the number of products sold in January means that the boxplot is built from data on the number of products sold each January in all years of observation. Boxplot data on the number of products sold is illustrated in Figure 4.

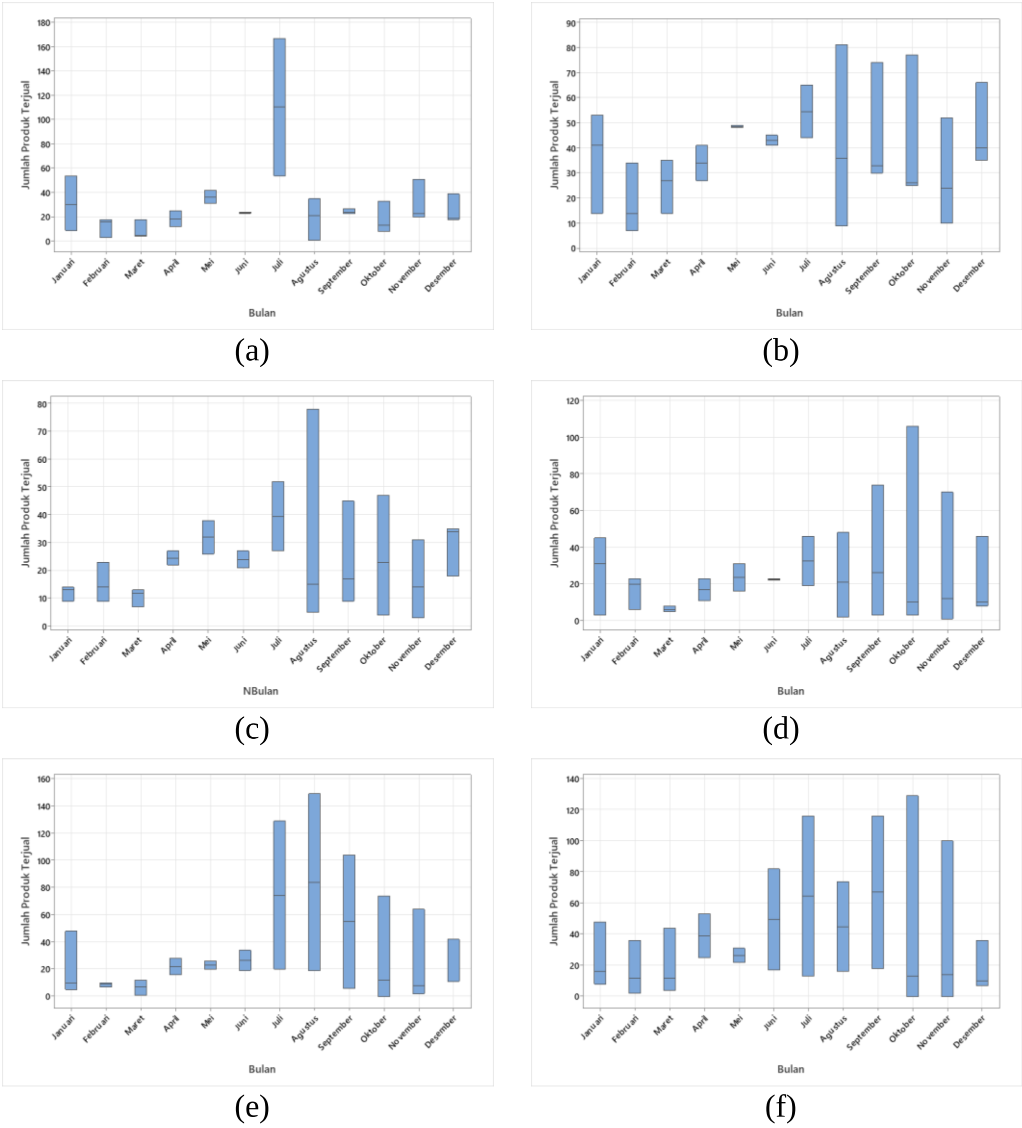


Figure 4. Monthly boxplot visualization of the number of products sold

(a) Ventela Ethnic Low All Black, (b) Ventela Ethnic Low Black Natural, (c) Ventela Public Low Black Natural, (d) Ventela Public Low Cream, (e) Ventela Republic Low Black Natural, dan (f) Ventela Republic Low White

The Boxplot illustrated in Figure 4 shows that the sales trend for all shoe models in large quantities tends to occur at the end of the year, starting from July to October and then sloping again (Putra, 2022). This information implies a seasonal trend likely to occur in data on the number of products sold. However, a reasonably long boxplot body also means that the alleged seasonality does not strongly influence the number of products sold. Therefore, a good ES model to use for forecasting has yet to be determined with certainty, and modeling simulations are needed to try both possible ES models that can be used.

Data Forecasting Using Exponential Smoothing

The data patterns for the six shoe models tend to be irregular, resulting in no direct determination regarding the most suitable Exponential Smoothing (ES) model. Therefore, experiments are carried out in the forecasting process to apply the possible types of ES models, namely DES and TES. The simulated TES model is additive, because the seasonal pattern does not show any multiplication (increase over time). The absence of provisions regarding the selection of weight values in ES modeling makes it necessary to determine which weights represent small (0.25), medium (0.50), and large (0.75). The entire model is built by adjusting the three predetermined weights.

Double Exponential Smoothing Modelling

The next data modeling was done using the Double Exponential Smoothing (DES) method. This model involves two weights, namely α and β, which are coefficients or smoothing weights for level and trend. The values α and β are determined according to the previous provisions so that nine different models are formed for each shoe model. The best model for the six shoe models in the best-selling cluster is presented in Table 9.

Table 9
Best DES model for each shoe model

No	Shoe model	MAPE (%)	Model equation
1	Ventela Ethnic Low All Black	162.25	$alL_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})CapT_t = 0{,}25(L_t - L_{t-1}) + (1 - 0{,}25)T_{t-1}F_{t+1} = L_t + T_t$
2	Ventela Ethnic Low Black Natural	49.48	$equalL_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})CapT_t = 0{,}5(L_t - L_{t-1}) + (1 - 0{,}5)T_{t-1}F_{t+1} = L_t + T_t$
3	Ventela Public Low Black Natural	63.21	$equalL_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})T_t = 0{,}75(L_t - L_{t-1}) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$

No	Shoe model	MAPE (%)	Model equation
4	Ventela Public Low Cream	110.96	$equalL_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})T_t = 0{,}75(L_t - L_{t-1}) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$
5	Ventela Republic Low Black Natural	78.99	$equalL_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})CapT_t = 0{,}75(L_t - L_{t-1}) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$
6	Ventela Republic Low White	69.53	$L_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})CapT_t = 0{,}75\left(L_t - L_{t-1equal}\right) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$

Based on Table 9, the DES model MAPE value for all shoe models is large, greater than 20%. This is thought to be caused by irregular data patterns. Even though the complexity of the DES model has increased with the βParameter compared to the SES model, it is still unable to capture the data patterns we want to predict well. Therefore, these models cannot be considered good models for predicting data.

Triple Exponential Smoothing Modelling

The final data modeling used the Triple Exponential Smoothing (TES) method. This model is the most complex of the other ES models, which involves three weights, namely. α, β, and γ, Each of which is a smoothing coefficient or weight at level, trend, and seasonality. The values α, β, and γ are determined according to the previous provisions so that 27 different models are formed for each shoe model. The best model for the six shoe models in the best-selling cluster is presented in Table 10.

Table 10
Best TES model for each shoe model

No	Shoe model	MAPE (%)	Model equation
1	Ventela Ethnic Low All Black	133.73	$L_t = 0{,}5(Y_t - S_{t-s}) + (1 - 0{,}5)(L_{t-1} + T_{t-1}), T_t = 0{,}5(L_t - L_{t-1}) + (1 - 0{,}5)T_{t-1}, S_t = 0{,}25(Y_t - L_t) + (1 - 0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t,$

No	Shoe model	MAPE (%)	Model equation
2	Ventela Ethnic Low Black Natural	29.17	$L_t = 0{,}5(Y_t - S_{t-s}) + (1-0{,}5)(L_{t-1} + T_{t-1}), T_t = 0{,}5(L_t - L_{t-1}) + (1-0{,}5)T_{t-1}, S_t = 0{,}25(Y_t - L_t) + (1-0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t,$
3	Ventela Public Low Black Natural	53.08	$L_t = 0{,}75(Y_t - S_{t-s}) + (1-0{,}75)(L_{t-1} + T_{t-1}), T_t = 0{,}25(L_t - L_{t-1}) + (1-0{,}25)T_{t-1}, S_t = 0{,}25(Y_t - L_t) + (1-0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t,$
4	Ventela Public Low Cream	119.88	$L_t = 0{,}5(Y_t - S_{t-s}) + (1-0{,}5)(L_{t-1} + T_{t-1}), CapT_t = 0{,}25(L_t - L_{t-1}) + (1-0{,}25)T_{t-1}, S_t = 0{,}25(Y_t - L_t) + (1-0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t,$
5	Ventela Republic Low Black Natural	138.26	$L_t = 0{,}75(Y_t - S_{t-s}) + (1-0{,}75)(L_{t-1} + T_{t-1}), T_t = 0{,}25(L_t - L_{t-1}) + (1-0{,}25)T_{t-1}, CapCapS_t = 0{,}25(Y_t - L_t) + (1-0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t,$
6	Ventela Republic Low White	96.62	$L_t = 0{,}75(Y_t - S_{t-s}) + (1-0{,}75)(L_{t-1} + T_{t-1}), T_t = 0{,}25(L_t - L_{t-1}) + (1-0{,}25)T_{t-1}, CapS_t = 0{,}25(Y_t - L_t) + (1-0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t,$

Table 10 shows that the MAPE value for the TES model is still not much different from the DES model, where for all shoe models, it has a large value, namely greater than 20%. In some shoe models, the complexity of the TES model worsens the prediction results obtained. Therefore, these models cannot be considered as good models for predicting data.

Selection of the Best Model Based on MAPE Accuracy

After modeling using two ES models, the best model is selected globally, namely the best model from the best models in each model (DES and TES). Based on the MAPE obtained in the three models, the best model globally is obtained, as presented in Table 11.

Table 11
Best ES model for each shoe model

No	Shoe model	MAPE (%)	Model equation
1	Ventela Ethnic Low All Black	TES (0,5; 0,5; 0,25)	$L_t = 0{,}5(Y_t - S_{t-s}) + (1 - 0{,}5)(L_{t-1} + T_{t-1})$, $T_t = 0{,}5(L_t - L_{t-1}) + (1 - 0{,}5)T_{t-1}$, $S_t = 0{,}25(Y_t - L_t) + (1 - 0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t$,
2	Ventela Ethnic Low Black Natural	TES (0,5; 0,5; 0,25)	$L_t = 0{,}5(Y_t - S_{t-s}) + (1 - 0{,}5)(L_{t-1} + T_{t-1})$, $T_t = 0{,}5(L_t - L_{t-1}) + (1 - 0{,}5)T_{t-1}$, $S_t = 0{,}25(Y_t - L_t) + (1 - 0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t$,
3	Ventela Public Low Black Natural	TES (0,75; 0,25; 0,25)	$L_t = 0{,}75(Y_t - S_{t-s}) + (1 - 0{,}75)(L_{t-1} + T_{t-1})$, $T_t = 0{,}25(L_t - L_{t-1}) + (1 - 0{,}25)T_{t-1}$, $S_t = 0{,}25(Y_t - L_t) + (1 - 0{,}25)S_{t-s}, F_{t+1} = L_t + T_t + S_t$,
4	Ventela Public Low Cream	DES (0,25; 0,75)	$L_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})$ $T_t = 0{,}75(L_t - L_{t-1}) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$
5	Ventela Republic Low Black Natural	DES (0,25; 0,75)	$L_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})$ $T_t = 0{,}75(L_t - L_{t-1}) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$
6	Ventela Republic Low White	DES (0,25; 0,75)	$L_t = 0{,}25Y_t + (1 - 0{,}25)(L_{t-1} + T_{t-1})$ $T_t = 0{,}75(L_t - L_{t-1}) + (1 - 0{,}75)T_{t-1}F_{t+1} = L_t + T_t$

The MAPE value obtained from the best model in Table 11 still has a large value (>20%), but the best model must still be selected for forecasting. Figure 5 below illustrates the prediction results obtained based on the best model for each shoe model.

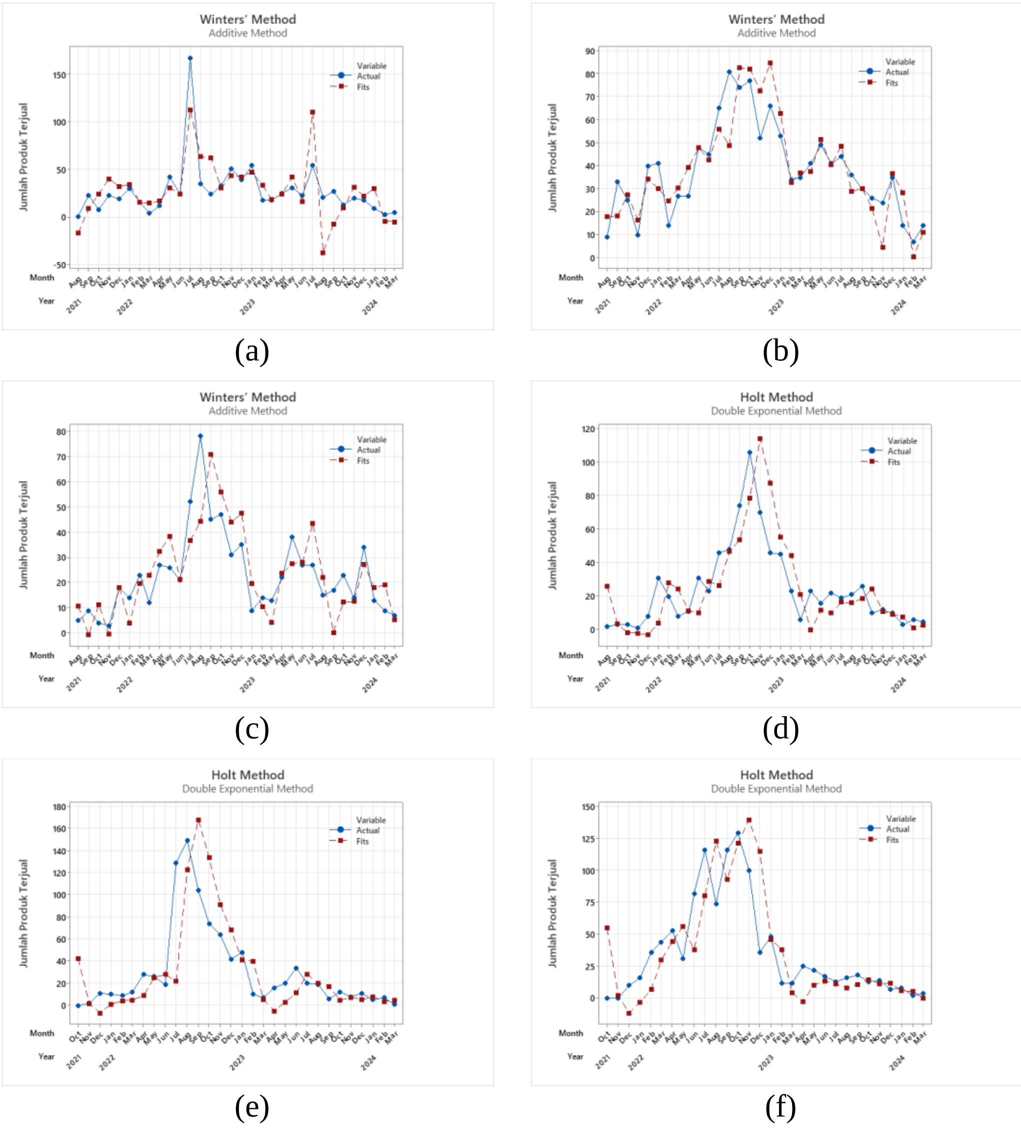


Figure 5 Visualization of Actual Comparison and Best ES Model Predictions

(a) Ventela Ethnic Low All Black, (b) Ventela Ethnic Low Black Natural, (c) Ventela Public Low Black Natural, (d) Ventela Public Low Cream, (e) Ventela Republic Low Black Natural, dan (f) Ventela Republic Low White

Figure 5. shows that the model predictions obtained for each shoe model follow the actual data pattern. Thus, visually, the model can be considered capable of predicting the model well and can be used to predict the number of products for each shoe model.

Forecasting Results and Illustrations

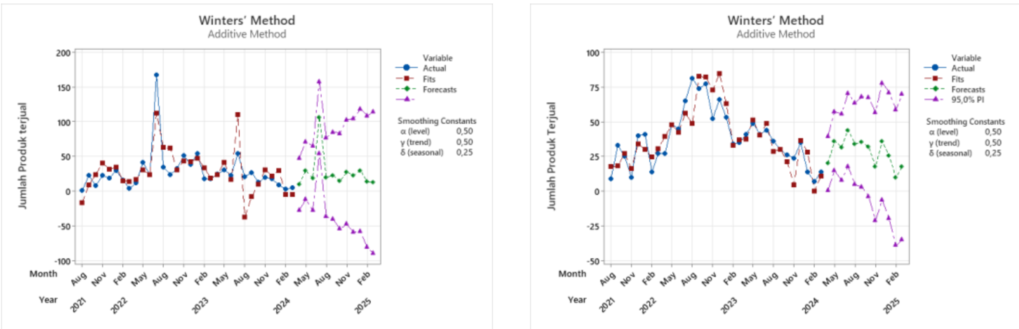
Based on the best model explained in the previous subsection, forecasting is carried out to obtain a projected number of product sales for each shoe in the next few months. In this study, the number of predictions was determined to be 12, which refers to the frequency of data used, namely monthly. Forecasting results for each shoe model are presented in Table 12.

Table 12
Forecasting result for each shoe model

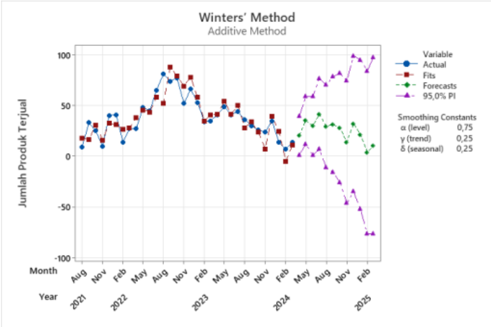
Periode	V. Ethnic Low All Black	V. Ethnic Low Black Natural	V. Public Low Black Natural	V. Public Low Cream	V. Republic Low Black Natural	V. Republic Low White
April 2024	10	20	20	3	0	1
Mei 2024	30	36	27	1	0	0
Juni 2024	19	32	19	0	0	0
Juli 2024	106	44	34	0	0	0
Agustus 2024	20	34	28	0	0	0
September 2024	23	35	17	0	0	0
Oktober 2024	15	32	17	0	0	0
November 2024	28	18	8	0	0	0
Desember 2024	23	36	21	0	0	0
Januari 2025	30	26	3	0	0	0
Februari 2025	14	10	6	0	0	0
Maret 2025	13	18	1	0	0	0

* V. stands for Ventela

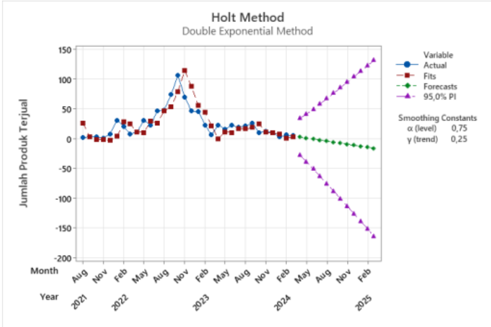
To clarify the forecast movement pattern obtained, an illustration is carried out as in Figure 6.



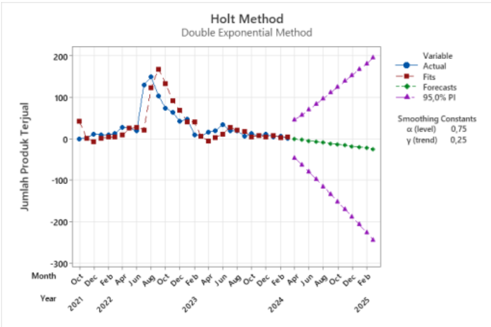
(a) (b)



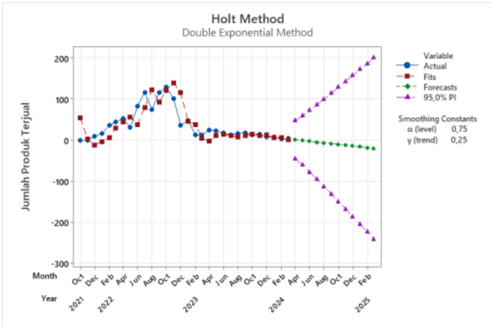
(c)



(d)



(e)



(f)

Figure 6. Illustration of the best es model forecast results

Ventela Ethnic Low All Black, (b) Ventela Ethnic Low Black Natural, (c) Ventela Public Low Black Natural, (d) Ventela Public Low Cream, (e) Ventela Republic Low Black Natural, dan (f) Ventela Republic Low White

Based on Figure 6, it can be seen that there are forecast results that show fluctuating value movements for forecasts using the TES model. However, forecasts using the DES model show a constant decreasing trend. However, the constant forecast results can still be used as a reference because they have a range of forecast intervals or Prediction Intervals (PI) that could occur. Therefore, to maximize the sales potential of each shoe model, sellers must prepare a stock of the maximum value of the PI range obtained so that stockouts do not occur.

Managerial Implications

The managerial implications of this research relate to the forecasting results for each shoe model in the best-selling cluster. Several shoe models, such as Public Low Cream, Republic Low Black Natural, and Republic Low white, show no sales for the next few months or have a forecast result 0. A shoe model with a forecast result of 0 means that the model no longer has sales power or has not been sought after by buyers. Sellers can consider not re-stocking this model or only having a maximum of 1 stock as an opportunity to make a profit. Shoe models that previously had high sales will not necessarily have high sales in the future.

Sellers can cut the selling price of shoes to reduce the stock in the warehouse or shelves so that they do not only sit for a short time because selling power has decreased. The sales proceeds can be used as capital to replace it with the latest model. Sellers may also consider adding other brands while Ventela produces new models. This can fill empty warehouses or shelves by replacing models that have expired so that sellers still have the opportunity to make a profit. The shelf life of each shoe model is different. Therefore, sellers can estimate the shelf life from forecast results to avoid overstock.

Conclusion

Based on the results of grouping using the K-Means clustering method on shoe models with three variables (number of visitors, number of buyers, and total sales), it was found that the number of shoe models included in the best-selling cluster was six. The six shoe models include (1) Ventela Ethnic Low All Black, (2) Ventela Ethnic Low Black Natural, (3) Ventela Public Low Black Natural, (4) Ventela Public Low Cream, (5) Ventela Republic Low Black Natural, and (6) Ventela Republic Low White.

The ES model's accuracy value (MAPE) in forecasting obtained is above 20%, which shows that the Exponential Smoothing model cannot model the number of shoe products sold for each shoe model well. Referring to the sample of this study, which only spans less than three years, several shoe models produce a forecast value of zero (especially for Ventela Public Low Cream, Ventela Republic Low Black Natural, and Ventela Republic Low White). Based on the forecast results for the next 12 months, the three shoe models included in the best-selling cluster will have no sales.

Bibliography

Alwiyah, M., & Gata, G. (2019). E-Commerce Untuk Meningkatkan Penjualan Sepatu Sneakers Bintaro. *IDEALIS: InDonEsiA JournaL Information System*, *2*(6), 345–350.

Dwiarni, B. A., & Setiyono, B. (2020). Akuisisi dan Clustering Data Sosial Media Menggunakan Algoritma K-Means sebagai Dasar untuk Mengetahui Profil Pengguna. *Jurnal Sains Dan Seni ITS*, *8*(2), A65–A70.

Jananto, A. (2022). Perbandingan Analisis Cluster Algoritma K-Means Dan AHC Dalam Perencanaan Persediaan Barang Pada Perusahaan Manufaktur. *Progresif: Jurnal Ilmiah Komputer*, *18*(2), 257–270.

Maricar, M. A. (2019). Analisa perbandingan nilai akurasi moving average dan exponential smoothing untuk sistem peramalan pendapatan pada perusahaan xyz. *Jurnal Sistem Dan Informatika (JSI)*, *13*(2), 36–45.

Pérez-Ortega, J., Almanza-Ortega, N. N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., & Martínez-Rebollar, A. (2019). The K-means algorithm evolution. *Introduction to Data Science and Machine Learning*, 69–90.

Pratiwi, W. A., & Marizal, M. (2022). Penerapan Metode Eksponential Smoothing Dalam Memprediksi Hasil Pencapaian Kinerja Pelayanan Perangkat Daerah Dinas Pendidikan Provinsi Riau. *Indonesian Council of Premier Statistical Science*, *1*(1), 4–14.

Putra, Y. Y. (2022). *Perlindungan Hukum Pemegang Hak Merek Deenay Sebagai Merek Terkenal Dari Produk Yang Dipasarkan Melalui Platform E-Commerce Berdasarkan Uu No. 20 Tahun 2016 Tentang Merek Dan Indikasi Geografis Juncto Uu No 19 Tahun 2016 Tentang Informasi Dan Transaksi Elektronik*. Fakultas Hukum Universitas Pasundan.

Rachmawati, N. L., & Lentari, M. (2022). Penerapan metode Min-Max untuk Minimasi Stockout dan Overstock persediaan bahan baku. *Jurnal INTECH Teknik Industri Universitas Serang Raya*, *8*(2), 143–148.

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking*, *2021*, 1–16.

Yusditara, W., Damanik, H. M., Purba, M. L., & Samosir, H. E. S. (2022). Pengaruh Brand Image Dan Promosi Terhadap Keputusan Pembelian Sepatu Lokal Merek Compass. *Jurnal Ilmiah Simantek*, *6*(2), 27–32.

Zullah, V. S. (2024). *Analisis Klaster dan Peramalan Pada Produk Sepatu Lokal*. Institut Teknologi Sepuluh Nopember.

Zunaidi, R. A. (2022). Market Simulator Konsep Produk Baru Sepatu Wanita Bagi UMKM Klastik Footwear. *Journal of Advances in Information and Industrial Technology*, *4*(1), 29–36.