International Journal of Economics, Business and Accounting Research (IJEBAR)
Peer Reviewed – International Journal
Vol-8, Issue-2, 2024 (IJEBAR)
E-ISSN: 2614-1280 P-ISSN 2622-4771
https://jurnal.stie-aas.ac.id/index.php/IJEBAR

DESIGN OF AN INFORMATION SYSTEM FOR FORECASTING FABRIC DEMAND (CASE STUDY AT CAHAYA SANDANG)

Geraldy Setiawan[1], Dedy Suryadi[2]
Industrial Engineering Faculty, Parahyangan Catholic University, Bandung, Indonesia[1,2]
E-mail: 8132201002@student.unpar.ac.id[1], dedy@unpar.ac.id[2]

Abstract: In the course of weekly fabric procurement from suppliers, Cahaya Sandang frequently encounters challenges in determining the optimal quantity of fabrics to be ordered. The uncertainty surrounding order quantities has led to a loss of consumer confidence for Cahaya Sandang. To address this issue, a research initiative was undertaken, focusing on the development of an information system for forecasting fabric sales at the Cahaya Sandang Store. The fabric sales forecasting system was designed using Python, incorporating various forecasting methodologies such as moving average, exponential smoothing, linear regression, Autoregressive Integrated Moving Average (ARIMA), and Long Short Memory Term (LSTM). These designated forecasting methods are exclusively applied to predict cloth sales for the subsequent week. The determination of forecasting methods was based on the Mean Absolute Percentage Error (MAPE) values derived from the existing forecasting models. The findings affirm the efficacy of the five employed methods, demonstrating a maximum error value below 10% for the prediction of Cahaya Sandang fabric sales. By implementing this fabric sales forecasting system, the decision-making process regarding fabric orders from suppliers is expected to become more streamlined and targeted.

Keywords: *Moving Average, Exponential Smoothing, Linear Regression, ARIMA, LSTM*

Submitted: 2024-04-06; Revised: 2024-06-04; Accepted: 2024-06-01

1. Introduction

Cahaya Sandang is a textile store located in the Cigondewah Textile District, specializing in trade and offering a diverse range of fabrics for clothing and trousers. Positioned within a textile-district area, Cahaya Sandang faces substantial competition from numerous rivals. The presence of competitors necessitates Cahaya Sandang to effectively contend with its peers by providing a wide array of fabric types and colors in demand by consumers. Intense competition often results in the loss of customers for Cahaya Sandang when the desired fabric types and colors are not available.

Based on interviews, the owner of Cahaya Sandang frequently encounters challenges in determining the optimal quantity of fabric to be ordered each week. The uncertainty in determining weekly fabric orders often leads to stockouts, preventing Cahaya Sandang from fulfilling consumer demands. The owner of Cahaya Sandang observes that a substantial number of customers have been lost due to the stockouts. The owner anticipates that by implementing weekly demand forecasting, Cahaya Sandang can better meet consumer demands, particularly for the most popular fabric types and colors each week.

According to Heizer and Render (2005), forecasting is a crucial activity employed to anticipate future information by utilizing historical data as a reference. Forecasting holds significant importance in business operations, particularly in determining future demand projections. The anticipation of future demand stands as a pivotal component within the competitive landscape of business activities. The ability to forecast future demand facilitates business owners in making informed decisions regarding the quantity of goods to be ordered, ensuring the fulfillment of consumer demands. Addressing the concerns of the owner of Cahaya Sandang, engaging in forecasting activities becomes imperative. Through the implementation of weekly demand forecasting, the proprietor of Cahaya Sandang can more effectively decide on the quantity of fabric to be ordered from suppliers each week, thereby enhancing decision-making processes and ensuring the timely and sufficient provision of goods to meet consumer demands.

In light of the aforementioned issues, this study aims to assist the store owner by developing a fabric sales forecasting information system. The design of the fabric sales forecasting information system at Cahaya Sandang Store will be centered around fabrics with the highest sales, specifically Maxmara and Wolly Crepe with the codes MX01, MX02, MX03, MX04, WC01, WC02, WC03, and WC04. The implementation of this fabric sales forecasting system is expected to streamline and guide the decision-making process regarding the weekly quantity of fabric to be ordered from suppliers.

2. Literature Review

According to Prasetya and Lukiastuti (2019), forecasting is an endeavor aimed at predicting future conditions through the examination of past circumstances. Forecasting holds significant importance as a foundation for decision-making. As stated by Chase, Aquilano, and Jacobs (2000), the availability of effective inventory management tools necessitates forward planning. Forecasting provides crucial information regarding future demand, enabling decisions related to production capacity, inventory, budgeting, and supply chain management.

Moving Average is one of the most commonly used forecasting methods and represents the simplest forecasting technique. The primary objective of Moving Average is to smooth out randomness inherent in past data.Moving Average computes a new value by calculating the average of a set of historical data. Once a new data point is obtained, it is incorporated into the calculation, and the oldest data point is excluded (Makridakis, 1999). The simple moving average, or single moving average, can be computed using the following formula:

$$F_{i+1} = \frac{D_i + D_{i-1} + \cdots + D_{i-N+1}}{N}$$

Information:
D_i = Actual Demand Period-i
F_{i+1} = Demand Forecast Period-i+1
N = Total Number of Demand

Exponential smoothing is a forecasting method closely resembling the moving average. According to Render and Heizer (2005), exponential smoothing is a technique of weighted moving averages where data is assigned weights by an exponential function. This method is typically employed for data lacking discernible trends and seasonality. The primary distinction from the moving average lies in the application of exponentially weighted values to historical data. Render and Heizer (2005) provide the mathematical formula for forecasting using this method, as indicated below.

$$F_{t+1} = F_t + a(X_t - F_t)$$

Information:

F_{t+1} = Demand Forecast Period-t+1
X_t = Actual Demand Period-t
a = *Smoothing alpha* ($0 < a < 1$)

Linear Regression, alternatively termed Linear Regression, is a forecasting method involving two variables. According to Sugiyono (2011), this method focuses on identifying a linear relationship between the dependent and independent variables. In the context of sales demand forecasting, the demand variable is designated as the dependent variable, while the time variable serves as the independent variable. Linear regression is usually applied when data exhibits a discernible upward or downward trend. According to Sugiyono (2011), the mathematical equation for calculating forecasts using this method can be observed in the formula provided below.

$$Y = a + bX$$

Information:

Y = Dependent Variable
X = Independent Variable
a = Intercept
b = *Slope*

Autoregressive Integrated Moving Average, commonly abbreviated as ARIMA, stands as one of the most frequently utilized methods for forecasting time series data. ARIMA itself comprises three main components: AutoRegressive (AR), Integrated (I), and Moving Average (MA). In addition to ARIMA, there is the ARMA method, a combination of AR and MA models. According to Box and Jenkins (2008), a non-stationary time series model can be referred to as an Autoregressive Integrated Moving Average process (ARIMA). Meanwhile, for the other three methods, namely AutoRegressive (AR), Moving Average (MA), and AutoRegressive Moving Average (ARMA), the data used must be stationary. ARIMA is commonly known as the Box-Jenkins model, denoted by the notation ARIMA (p, d, q). Here, p represents the autoregressive order, d represents the integrated order, and q represents the moving average order. The number of past values used in the autoregressive order p indicates the level of that order, usually denoted as AR(p). The number of differentiations required determines the level of the integrated order d, typically denoted as I(d). The level within the moving average order is usually denoted as MA(q). According to Box, Jenkins, and Reinsel (2008), the mathematical equation for the ARIMA method can be expressed as follows.

$$\left(1 - \emptyset_1 B - \cdots - \emptyset_p B^p\right)(1 - B)^d y_t = c + (1 - \theta_1 B - \cdots - \theta_q B^q)\varepsilon_t$$

Information:

B = Operator Lag
$\emptyset_p$ = Autoregressive Coefficient
ε_t = Residual Error Period-t
θ_q = Moving Average Coefficient
y_t = Actual Demand Periode-t
d = Level of Differentiation

Long Short Term Memory, commonly abbreviated as LSTM, is one of the deep learning algorithms resembling a neural network inspired by the human brain. LSTM can connect information from past data due to the presence of hidden states in each cell of the network. Each LSTM cell also has recurrent connections, allowing for the processing of sequences of input data. The structure of the LSTM algorithm resembles a neural network, consisting of various functions such as sigmoid and hyperbolic tangent, enabling LSTM to be employed for time series data prediction. In contrast to statistical methods that require specific assumptions before processing time series data, LSTM can be used for time series data without the need for such assumptions. LSTM structure can be seen in Figure 1.

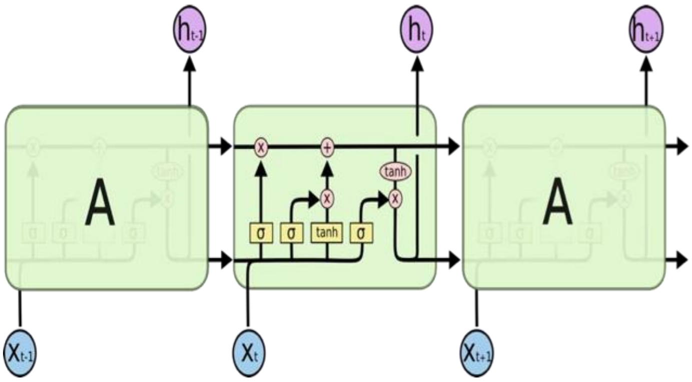


Figure 1. LSTM Cell Structure
(Source: Olah, 2015)

The formula for the LSTM algorithm is as follows.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$f_t = \sigma(U_f . h_{t-1}, + W_f . x_t + b_f)$$
$$i_t = \sigma(U_i . h_{t-1} + W_i . x_t + b_i)$$
$$\tilde{C}_t = \tanh(U_c . h_{t-1} + W_c . x_t + b_c)$$
$$O_t = \sigma(U_o . h_{t-1}, + W_o . x_t + b_o)$$
$$h_t = O_t \odot \tanh(C_t)$$

Information:
C_t = Cell State Period-t
$\tilde{C}_t$ = New Candidate
f_t = Forget Gate Period-t
i_t = Input Gate Period-t
tanh = Hiperbolic Tangent Function
σ = Sigmoid Function
h_t = Output Value Period-t
x_t = Input Value Period-t
W_c = Weight W for Cell State
U_c = Weight U for Cell State
w_f = Weight W for Forget Gate
U_f = Weight U forget gate
w_i = Weight W for Input Gate
U_i = Weight W for Input Gate
b_f = Bias Value for Forget Gate

b_i = Bias Value for Input Gate
b_c = Bias Value for Cell Gate
W_o = Autoregressive Coefficient
U_o = Residual Error Period-t

3. Research Method

The research methodology will encompass the sequential steps undertaken throughout the research process. The writing of the research methodology aims to delineate the research stages from inception to conclusion. Moreover, the articulation of the research methodology aims to provide a structured and directed approach to the conducted research. Figure 2 illustrates the procedural steps of the research methodology employed.

Cahaya Sandang is a textile store located in the Cigondewah Textile District, Bandung City, Indonesia, specializing in trade and offering a diverse range of fabrics for clothing and trousers. The initial data collection stage involves direct interviews and observations conducted by the researcher at Cahaya Sandang Store. The observations and interviews are aimed at acquiring information regarding Cahaya Sandang Store. Subsequently, a literature review is conducted to search for supporting literature and study theories relevant to the research. The literature review serves the purpose of enhancing knowledge pertaining to the theories to be utilized in the research. Based on the information gathered from the observation and interview phase, the identification of issues within Cahaya Sandang Store is performed. This identification is instrumental in formulating the objectives and benefits of the research concerning Cahaya Sandang Store.

This study aims to assist the store owner by developing a fabric sales forecasting information system. The design of forecasting information system will be centered in fabrics with the highest sales, specifically Maxmara and Wolly Crepe with the codes MX01, MX02, MX03, MX04, WC01, WC02, WC03, and WC04. Table 1 contains the fabric code along with the type and color. The implementation of the forecasting system is expected to streamline and guide the decision-making process regarding the weekly quantity of fabric to be ordered from suppliers.

Table 1 Fabric Code

Code	Type	Color
MX01	Maxmara	Gray
MX02	Maxmara	Blue
MX03	Maxmara	Brown
MX04	Maxmara	Black
MX05	Maxmara	Red
WC01	Wolly Crepe	Blue
WC02	Wolly Crepe	Brown
WC03	Wolly Crepe	Black
WC04	Wolly Crepe	Red

Upon establishing the research objectives, the research commences with the collection of necessary data for the investigation, such as the historical sales data of Cahaya Sandang Store. This historical data serves a crucial role in the data processing phase of this research. The data processing procedure initiates with data preprocessing, aimed at addressing the conditions of missing values and noise within the historical sales data. In this context, missing values are identified as instances where certain sales data for specific periods are recorded as zero due to Cahaya Sandang Store's inactivity, while noise in sales data arises from a specialized orders occurring at certain times. Noise and missing values within the historical data typically lead to outliers, which could potentially disrupt the forecasting accuracy of the forecasting models. The data identified as missing values and noise will subsequently be replaced with the average sales values for each week. Table 2 represent an example of 10 week hitorical sales data for each fabric code.

Table 2 Example of 10 Week Historical Sales Data (Meter)

Week	MX01	MX02	MX03	MX04	MX05	WC01	WC02	WC03	WC04
1	79,00	106,50	147,25	60,75	84,25	80,75	123,00	197,25	134,00
2	84,25	95,55	149,75	62,25	50,50	130,75	56,75	140,25	137,75
3	69,75	93,30	147,25	53,50	60,50	137,00	87,25	156,25	145,50
4	64,25	99,85	184,75	76,00	114,80	71,25	74,25	203,75	147,00
5	87,00	105,85	138,75	55,00	32,25	85,50	98,25	99,75	96,50
6	73,00	91,50	131,75	50,50	125,00	98,00	77,25	167,75	142,50
7	69,13	102,96	130,67	62,13	44,04	113,17	92,17	98,00	148,75

The data processing procedure proceeds with the conversion of daily sales data into weekly sales data by aggregating the daily sales data over a seven-day period. This conversion from daily to weekly sales data is conducted to align with the requirements of Cahaya Sandang Store's owner, where fabric orders from suppliers are placed on a weekly basis. The historical weekly sales data, cleansed of missing values and noise, will then be visualized in graphical form to discern the data patterns for each fabric type.

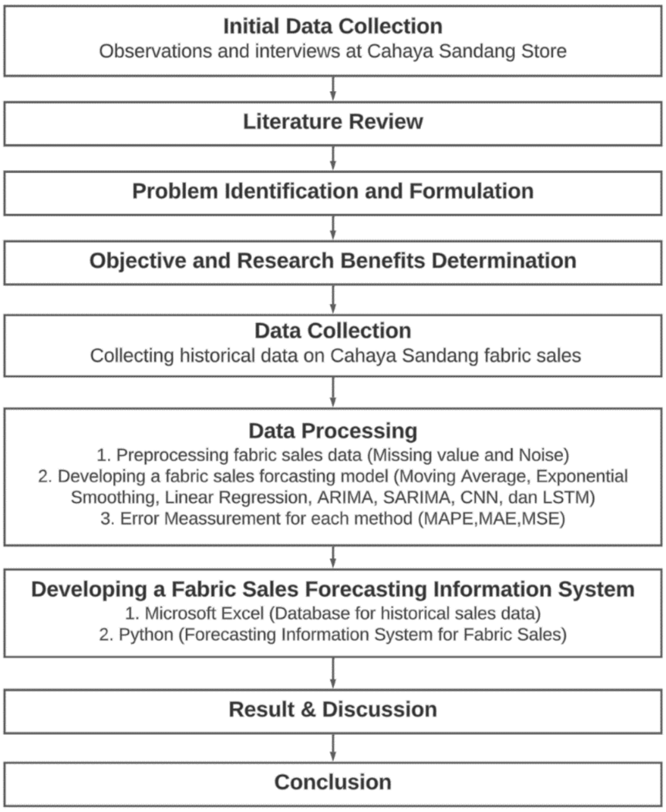


Figure 2 Research Framewok

The next step involves the design of forecasting models. The creation of forecasting models will entail employing several forecasting methods such as moving average, exponential smoothing, linear regression, Auto Regression Integrated Moving Average (ARIMA), and

Long Short Term Memory (LSTM). In the design process, the moving average and Long Short Term Memory (LSTM) forecasting methods will utilize the historical data divided into three parts: training data, validation data, and test data, with respective allocation percentages of 80%, 10%, and 10%. The training data will serve as the basis for model formation, the validation data will be used to identify the best parameters for the forecasting model, and the test data will be utilized to determine the error of the forecasting methods. The division of data for the exponential smoothing, linear regression, and ARIMA methods will consist of two parts, with 90% allocated for training data and 10% for test data. The training data will be used for parameter determination in the model, while the test data will be employed to assess the error of these three forecasting methods.

The design of the forecasting model using the Moving Average method commences by determining the parameter n periods. This parameter search entails calculating the average of the training data from 2 periods to N periods to forecast the validation data. Here, N represents the total number of available training data. Subsequently, the forecasted results are compared with the actual validation data to ascertain the error value. The parameter n periods with the smallest average will be designated as the moving average parameter utilized for forecasting the test data. The forecasting outcomes are then compared with the actual test data to evaluate the error value of the Moving Average forecasting model.

The design of the forecasting model using the Exponential Smoothing method begins with identifying the differentiation value of the training data. This differentiation involves subtracting the data for a particular period from the data for the subsequent period. The purpose of this data differentiation process is to render the data stationary. To determine the optimal smoothing constant alpha parameter, the 'statsmodels' package in Python is utilized. With the optimal smoothing constant alpha, the test data is forecasted to determine the error of the Exponential Smoothing forecasting model.

The design of the forecasting model using the Linear Regression method starts with calculating the slope and intercept using the training data. These obtained slope and intercept values will be used to forecast the test data. The forecasting results are then compared with the actual test data to assess the error of the Linear Regression forecasting model.

The design of the forecasting model using the ARIMA method commences with identifying the parameters or orders of ARIMA (x, y, z). The optimal parameters or orders are determined using the 'pmdarima' package in Python, where the value x represents the Autoregressive order, the value y represents the differentiation order, and the value z represents the moving average order. The optimal parameters or orders are then utilized to forecast the test data. The forecasting outcomes are compared with the actual test data to determine the error of the ARIMA forecasting model.

Lastly, for the design of the forecasting model using the LSTM method, the process begins with normalizing the training data to a range between 0 and 1. The normalized data is then used to create the LSTM forecasting model with the assistance of the 'Keras' package in Python. The parameters required for the LSTM model include the batch size and lag of the training data. The batch size denotes the number of data groups entered into the model at each iteration, while the lag represents the number of data in one batch to be inputted into the model. The determination of the optimal batch and lag values from the training data is subjective, where the batch values used are 1, 5, and 7, while the lag values used are 1, 2, 3, 4, and 5. The optimal parameter determination involves employing all combinations of these parameters to forecast the validation data. The parameters with the smallest error value are chosen as the best parameters, subsequently used to forecast the test data to ascertain the error value of the LSTM forecasting model.

4. Results and Discussion

4.1. Results

Moving Average Model

The creation of the moving average model begins with the identification of the optimal parameters to be used for the nine fabric code categories. To determine the best period parameter, forecasting of the validation data is initially conducted using a 2-week to 132-week moving average. The forecast results are then compared with the actual validation data, and the Mean Absolute Percentage Error (MAPE) is calculated. The period that yields the smallest MAPE value in Table 3 is the optimal value for moving average period. This parameter is selected for constructing the moving average model for each fabrics code.

Table 3 Optimal Parameter MA

Code	Moving Average Parameter
MX01	33
MX02	21
MX03	2
MX04	7
MX05	2
WC01	4
WC02	3
WC03	2
WC04	3

Exponential Model

The development of the exponential smoothing model begins with the initial identification of the optimal parameters to be utilized for the nine fabric code categories. The determination of the best smoothing constant alpha parameter is facilitated with the assistance of the statsmodel package in Python. Table 4 illustrates the optimal exponential smoothing parameters that yields the smallest MAPE Value. This parameter is utilized in constructing the exponential smoothing forecasting model for the nine fabric code categories.

Table 4 Optimal Exponential Smoothing Parameters

Code	Smoothing Alpha Parameter
MX01	0,000000015
MX02	0,000000015
MX03	0,005443989
MX04	0,000000015
MX05	0,048500007
WC01	0,000000015
WC02	0,132273758
WC03	0,066662792
WC04	0,005000000

This parameter is utilized in constructing the exponential smoothing forecasting model for the nine fabric code categories.

Linear Regression

The linear regression forecasting model is constructed by determining the slope and intercept values for each fabric code category. The parameters of slope and intercept are derived using the training data from the historical fabric sales. Table 5 represents the parameters and the model for the forcasting model.

Table 5 Linear Regression Parameters and Model

Code	Slope	Intercept	Linear Regression Model
MX01	0,603	58,086	58,086 + 0,603 X
MX02	0,602	81,677	81,677 + 0,602 X
MX03	0,727	121,308	121,308 + 0,727 X
MX04	0,567	48,958	48,958 + 0,567 X
MX05	0,983	37,798	37,798 + 0,983 X
WC01	0,826	98,044	98,044 + 0,826 X
WC02	1,108	55,573	55,573 + 1,108 X
WC03	0,798	128,051	128,051 + 0,798 X
WC04	1,218	94,842	94,842 + 1,218 X

ARIMA

The determination of parameters in creating the ARIMA forecasting model is carried out with the assistance of the pmdarima package in Python. By utilizing the pmdarima package, the ARIMA order along with the best coefficients for each fabric code is obtained. The parameters are derived using the training data from the historical fabric sales. Table 6 represents the optimal ARIMA parameters for the nine fabric codes used in the forecasting model.

Table 6 ARIMA Parameters

Code	Orde (p,d,q)	AR1	AR2	AR3	MA1	MA2	Constant
MX01	(0,1,1)	0	0	0	-6,006	0	0
MX02	(0,1,1)	0	0	0	-0,530	0	0
MX03	(0,1,1)	0	0	0	-0,578	0	0
MX04	(0,1,1)	0	0	0	-0,7078	0	0,4912
MX05	(3,1,2)	-0,416	0,5203	0,3654	-0,2252	-0,6625	0
WC01	(0,1,1)	0	0	0	-0,8003	0	10,333
WC02	(0,1,2)	0	0	0	-0,8092	0,25	0
WC03	(0,1,1)	0	0	0	-0,7504	0	0
WC04	(1,1,2)	-0,817	0	0	0,0837	-0,7918	2,312

LSTM

The determination of LSTM parameters is conducted by experimenting with lag values ranging from 1 to 5 and batch sizes of 1, 5, and 7 to predict validation data. The prediction results are then compared with the actual validation data, and the Mean Absolute Percentage Error (MAPE) is calculated to determine the selected lag and batch values for each fabric code. Table 7 represents the selected LSTM parameters.

Table 7 LSTM Parameters

Code	*Lag*	Batch
MX01	5	7

MX02	1	7
MX03	4	7
MX04	5	7
MX05	4	5
WC01	1	1
WC02	4	7
WC03	1	7
WC04	5	1

The LSTM parameters specified above will be utilized to design the forecasting model using the LSTM method with the assistance of the Keras package in Python. In order to provide a clearer understanding of the LSTM model, a visualization of the LSTM forecasting model for fabric MX01 will be presented in Figure 3.

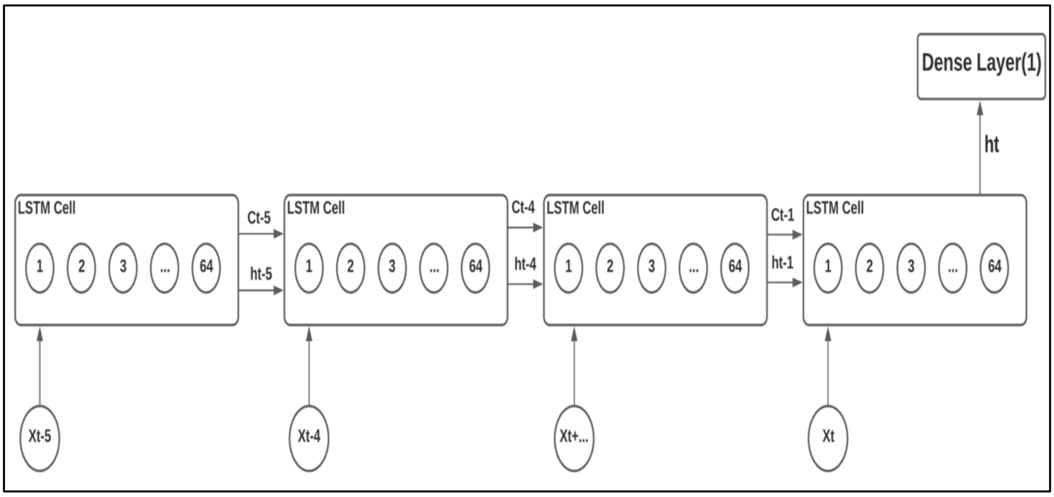


Figure 3 LSTM Model for fabric MX01

The LSTM model for fabric forecasting comprises a single LSTM cell that processes input data sequentially. In Figure 3, the LSTM cell is depicted separately for each input, although the processed LSTM cell remains the same for all inputs. When processing a single input value, the LSTM cell utilizes the Ct and ht values obtained from the previous process to process the next input. Inside the LSTM cell, there is a 64 circular shape representing an LSTM unit. Each of the LSTM unit contains the LSTM structure as seen in Figure 1. Each LSTM unit within the LSTM cell operates independently and does not have a connection with others.

The Final Model

The design of the fabric sales forecasting system is carried out in Python using several forecasting methods such as moving average, exponential smoothing, linear regression, Autoregressive Integrated Moving Average (ARIMA), and Long Short Memory Term (LSTM). The calculation of errors is conducted on the forecasted results of the test data using the model with the selected parameters for each fabric code. This error calculation is performed to determine the suitability of parameters for forecasting sales one week ahead and to ascertain the success of the developed forecasting model. Table 8 presents the results of the forecast model error calculation.

E-ISSN: 2614-1280 P-ISSN 2622-4771
https://jurnal.stie-aas.ac.id/index.php/IJEBAR

Table 8 Forecast Model Error Calculation

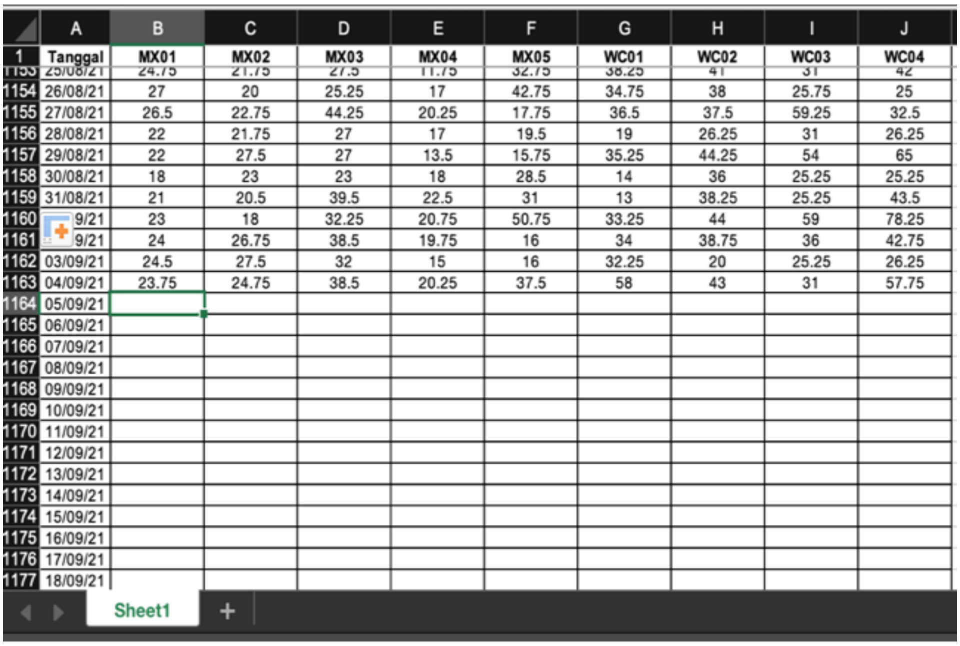
Code	MA	*Exponential Smoothing*	*Linear Regression*	ARIMA	LSTM
MX01	0,054	0,059	0,038	0,045	0,051
MX02	0,052	0,064	0,099	0,053	0,062
MX03	0,120	0,146	0,104	0,117	0,064
MX04	0,051	0,055	0,103	0,065	0,035
MX05	0,075	0,106	0,069	0,077	0,118
WC01	0,090	0,127	0,074	0,093	0,110
WC02	0,072	0,085	0,140	0,074	0,143
WC03	0,113	0,132	0,105	0,091	0,121
WC04	0,144	0,159	0,141	0,150	0,026

Based on Table 8 above, the forecasting models yielding the smallest MAPE values are highlighted in yellow for each fabric code. For MX01, MX05, and WC01, the linear regression model exhibits the lowest MAPE values. MX02 and WC02 exhibit the lowest MAPE values under the moving average forecasting model. WC03 shows the lowest MAPE value under the ARIMA forecasting model. MX03, MX04, and WC04 display the lowest MAPE values under the LSTM forecasting model. Forecasting for the upcoming week will be conducted using these methods. However, it is important to note that method selection may vary over time, as the choice of method will fluctuate according to the MAPE results obtained each time the model is executed weekly.

The primary highlight of this forecasting system design is its capability to forecast sales using different methods for each fabric code on a weekly basis. The forecasting system is not only designed to compare between forecasting methods but also excels in flexibility by determining different forecasting methods for each fabric type every week. Table 8 presents only an example of the MAPE forecasting results for the upcoming week, where each fabric code has the lowest MAPE value using a different method. These lowest MAPE forecasting results are only for the next week, as the method selection will vary in the following weeks for each fabric code based on the lowest MAPE value each week.

Microsoft Excel will also be utilized for inputting daily historical demand data during the design process. A Microsoft Excel spreadsheet is created to store daily historical fabric sales data. Figure 3 depicts the interface of the Excel file created for storing daily historical fabric sales data.

E-ISSN: 2614-1280 P-ISSN 2622-4771
https://jurnal.stie-aas.ac.id/index.php/IJEBAR



	A	B	C	D	E	F	G	H	I	J
1	Tanggal	MX01	MX02	MX03	MX04	MX05	WC01	WC02	WC03	WC04
1153	25/08/21	24.75	21.75	27.5	11.75	32.75	38.25	41	31	42
1154	26/08/21	27	20	25.25	17	42.75	34.75	38	25.75	25
1155	27/08/21	26.5	22.75	44.25	20.25	17.75	36.5	37.5	59.25	32.5
1156	28/08/21	22	21.75	27	17	19.5	19	26.25	31	26.25
1157	29/08/21	22	27.5	27	13.5	15.75	35.25	44.25	54	65
1158	30/08/21	18	23	23	18	28.5	14	36	25.25	25.25
1159	31/08/21	21	20.5	39.5	22.5	31	13	38.25	25.25	43.5
1160	9/21	23	18	32.25	20.75	50.75	33.25	44	59	78.25
1161	9/21	24	26.75	38.5	19.75	16	34	38.75	36	42.75
1162	03/09/21	24.5	27.5	32	15	16	32.25	20	25.25	26.25
1163	04/09/21	23.75	24.75	38.5	20.25	37.5	58	43	31	57.75
1164	05/09/21									
1165	06/09/21									
1166	07/09/21									
1167	08/09/21									
1168	09/09/21									
1169	10/09/21									
1170	11/09/21									
1171	12/09/21									
1172	13/09/21									
1173	14/09/21									
1174	15/09/21									
1175	16/09/21									
1176	17/09/21									
1177	18/09/21									

Figure 3 Microsoft Excel Spreadsheet Design

4.2. Discussion

Preprocessing Data

The pre-processing of historical sales data from Toko Cahaya Sandang is aimed at addressing missing value conditions within the dataset. The occurrence of missing values in the historical sales data of Toko Cahaya Sandang is attributed to the non-operation of the store during certain periods, resulting in zero values for historical data in those periods. It is imperative to mitigate these missing value conditions to prevent adverse effects on cloth sales forecasting outcomes. Missing values can affect the accuracy of the forecasting model utilized, as they tend to lower the data averages and do not disrupt the existing forecasting model development process. Based on research conducted by Kang (2013), missing values significantly influence conclusion drawing and increase bias in parameter estimation. The pre-processing of data to address missing values is undertaken due to the nature of historical sales data being a time series highly dependent on the sequence of time periods, ensuring that the missing values that occur are not discarded from the historical sales data. The replacement of missing values with weekly average sales is conducted to ensure that these averages still represent the prevailing conditions of each week.

Additionally, data pre-processing is also performed to mitigate noise from the historical sales data of cloth at Toko Cahaya Sandang. Noise in the historical sales data can be identified by outliers in the sales data. Preprocessing to eliminate this noise is carried out to ensure that the historical data used truly represents the sales conditions of cloth at Toko Cahaya Sandang. Addressing noise in the sales data is essential as it can distort the historical data and affect the accuracy of the forecasting model to be employed. As per Singh (1999), noise in data will reduce the accuracy of forecasting models. Special orders, contributing to noise, can inflate sales significantly in certain periods, thus impacting cloth sales forecasting. Noise stemming specifically from these special orders needs to be addressed by reducing historical data containing such orders within those periods. Other sales data not associated with special orders will still be utilized as it accurately reflects the actual conditions at Toko Cahaya Sandang.

International Journal of Economics, Business and Accounting Research (IJEBAR)
Peer Reviewed – International Journal
Vol-8, Issue-2, 2024 (IJEBAR)
E-ISSN: 2614-1280 P-ISSN 2622-4771
https://jurnal.stie-aas.ac.id/index.php/IJEBAR

Demand Forecasting System Model

The forecasting model construction process begins with dividing the historical demand data into several parts: train, validation, and test sets. The division percentage in this study adheres to the Pareto principle of 80/20, where 80% of the historical data is allocated to the train set. Applying the Pareto principle for data allocation is motivated by its assertion that 20% of the data can represent 80% of the remaining data. This division aims to determine the forecasting error values used in parameter determination for each forecasting method utilized.

Validation data is essential in constructing the moving average and LSTM forecasting models as parameter determination for these methods is manually conducted by comparing error values between the validation and train data. Conversely, validation data is not required for exponential smoothing, linear regression, and ARIMA models as parameter search is automated using Python packages. According to Bruni (2021), utilizing validation data for parameter determination in forecasting models leads to high-quality parameter values, enhancing the accuracy of the forecasting model.

Parameter determination in the LSTM forecasting model involves using multiple batch and lag values to limit the running time of the demand forecasting system at Toko Cahaya Sandang. Varying batch values aim to explore the update frequency of the LSTM model's weight, where smaller batch values result in more frequent weight updates, increasing the model's sensitivity to input data changes. Conversely, larger batch values lead to reduced model sensitivity to input data changes. Determining multiple lag values aims to explore variations in the length of data inputs. According to Liu, Gherbi, Li, and Cheriet (2019), lag values significantly influence the accuracy of LSTM forecasting models. LSTM demonstrates significantly lower MAPE results compared to other methods for fabrics with codes MX03, MX04, and WC04. The selection of LSTM as the method for forecasting demand for the upcoming week indicates its superiority in capturing data patterns for code MX04 compared to other statistical forecasting methods.

The average MAPE value for the established forecasting models is below 10%. According to Lewis (1982), forecasting models with MAPE values below 10% are categorized as highly accurate. This suggests that the flexibility of fabric demand forecasting information system for Toko Cahaya Sandang, which combines statistical forecasting methods and machine learning, has become an effective forecasting system capable of assisting future decision-making for the store owner.

5. Conclusion

Based on the research findings and discussions, the author draws the following conclusions:

1) The historical sales data from Toko Cahaya Sandang contains missing values and outliers, thus necessitating the preprocessing data by replacing these values with weekly average sales to improve the accuracy of the forecasting model.
2) The combination of statistical forecasting methods and machine learning has created an effective fabric demand forecasting system for Toko Cahaya Sandang Fabric demand with the aim of achieving average MAPE value below 10%.

Fabric demand forecasting information system for Toko Cahaya Sandang was developed with the aim of achieving accuracy above 90%. The design of this forecasting information system was carried out using Microsoft Excel and Python. The design in Microsoft Excel serves as the repository for historical data used in forecasting cloth demand, while the model design was conducted in Python using statistical forecasting and machine learning methods such as moving average, exponential smoothing, linear regression, Auto Regressive Integrated Moving Average (ARIMA), and Long Short Term Memory (LSTM). The forecasting system were designed not only to compare between forecasting methods but also excels in flexibility by determining different forecasting methods for each fabric type every week for accommodating potential changes in data patterns in the future.

References

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis Forecasting and Control* (4th ed.). New Jersey: Pearson.

Bruni, E. (2021). The Importance of Data Validation for Data Science. Downloaded from: https://fullstack.net/the-importance-of-data-validation-for-data-science/.

Chase, R. B., Aquilano N. J., & Jacobs, F. R. (2000). *Operations Management For Competitive Advantage* (9th ed.). New York: McGraw-Hill.

Heizer, J., & Render, B. (2005). *Operations Management*. Jakarta: Salemba Empat.

Kang, H. (2013). The prevention and handling of the missing data. *Korean J Anesthesiol*. *64*(5), 402–406. doi: 10.4097/kjae.2013.64.5.402.

Lewis, C. D. (1982). Int*ernational and Business Forecasting Methods*. London: Butterworths.

Liu, X., Gherbi, A., Li, W., & Cheriet, M. (2019). Multi Features and Multi-time steps LSTM Based Methodology for Bike Sharing Availability Prediction. *Procedia Computer Science.* 155. 394-401. doi: 10.1016/j.procs.2019.08.055.

Makridakis, S., Wheelwright, S. C., & McGee, V. E. (1999). *Metode dan Aplikasi Peramalan*. Jakarta: Erlangga.

Olah, C. (2015, 27 Agustus). Understanding LSTM Networks. Downloaded from: https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf.

Prasetya, H., & Lukiastuti F. (2019). *Manajemen Operasi*. Yogyakarta: Media Presindo.

Sugiyono. (2011). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Afabeta.

Singh, S. (1999). Noise Impact on Time-Series Forecasting using an Intelligent Pattern Matching Technique. *Pattern Recognition*. *32*(8), p.1389-1398. doi: 10.1016/S0031-3203(98)00174-5.