COSTING: Journal of Economic, Business and Accounting
Volume 7 Nomor 6, Tahun 2024
e-ISSN : 2597-5234

DEMAND FORECASTING BASED ON MACHINE LEARNING TO DETERMINE ORDER QUANTITY: A CASE STUDY OF BAHAGIA KOPI BANDUNG

PERAMALAN PERMINTAAN BERDASARKAN MACHINE LEARNING UNTUK MENENTUKAN JUMLAH PESANAN: STUDI KASUS BAHAGIA KOPI BANDUNG

Farel Nanda Rosya
Master of Business Administration, School of Business and Management, Institut Teknologi Bandung
farel_nanda@sbm-itb.ac.id

ABSTRACT

Coffee shops are becoming increasingly popular in Indonesia, and they are regarded as one of the business sectors that contribute to the country's industrial development. Difficulty to estimate sales and demand, disrupting coffee bean inventory management. Forecasting with machine learning models could provide a solution to these issues. The data used in this study is coffee bean demand from a POS (Point-of-Sales) system, which is calculated by converting coffee menu sales data to coffee bean demand. The data is time-series, spanning from. To improve model effectiveness, several external variables such as weather and event are included. The exploratory data analysis of these factors reveals the influence and pattern that affects the dynamics of coffee bean demand. Prediction models employed in this study include Multiple Linear Regression (MLR), Decision Tree (DT), Support Vector Regressor (SVR), and Neural Network (NN). Model training results demonstrate that models with all variables outperform models with simply date variables. The DT model produces the best forecast based on its pattern and error measurement. The prediction result is executed by constructing a dashboard that assists the businessman in determining the amount of coffee beans to order in the next months. These are the implementations that could be used to improve inventory management.
Keywords: *Demand Forecasting, Machine Learning, Determine Order Quantity, Bahagia Kopi Bandung.*

ABSTRAK

Kedai kopi menjadi semakin populer di Indonesia, dan dianggap sebagai salah satu sektor bisnis yang berkontribusi terhadap perkembangan industri di Indonesia. Kesulitan dalam memperkirakan penjualan dan permintaan, mengganggu manajemen persediaan biji kopi. Peramalan dengan model pembelajaran mesin dapat memberikan solusi untuk masalah ini. Data yang digunakan dalam penelitian ini adalah permintaan biji kopi dari sistem POS (Point-of-Sales), yang dihitung dengan mengubah data penjualan menu kopi menjadi permintaan biji kopi. Data tersebut merupakan data runtun waktu, dengan rentang waktu dari. Untuk meningkatkan efektivitas model, beberapa variabel eksternal seperti cuaca dan acara dimasukkan. Analisis data eksplorasi terhadap faktor-faktor ini mengungkapkan pengaruh dan pola yang mempengaruhi dinamika permintaan biji kopi. Model prediksi yang digunakan dalam penelitian ini meliputi Regresi Linier Berganda (MLR), Decision Tree (DT), Support Vector Regressor (SVR), dan Neural Network (NN). Hasil pelatihan model menunjukkan bahwa model dengan semua variabel mengungguli model dengan variabel tanggal saja. Model DT menghasilkan prediksi terbaik berdasarkan pola dan pengukuran kesalahan. Hasil prediksi dieksekusi dengan membuat dashboard yang membantu pengusaha dalam menentukan jumlah biji kopi yang harus dipesan di bulan-bulan berikutnya. Ini adalah implementasi yang dapat digunakan untuk meningkatkan manajemen persediaan.
Kata Kunci: Peramalan Permintaan, Machine Learning, Penentuan Jumlah Pesanan, Bahagia Kopi Bandung.

INTRODUCTION

Coffee shops, also known as cafes, have grown in popularity and demand in recent years, becoming an important sector in the growth of Indonesia's modern industry (Pramelani, 2020). This development has increased rivalry among coffee shops, necessitating excellent inventory management of raw ingredients to meet daily client demands.

Bahagia Kopi is a coffee shop in Bandung that serves a variety of coffee-based menu items. Due to a big customer base, the business owner has had difficulties in managing the supply of critical supplies, particularly coffee beans.

Machine learning has emerged as an efficient tool for solving a variety of commercial problems, including forecasting future sales. Machine learning for future predictions is one method for estimating demand and managing the supply of critical commodities (Tangtisanon, 2018; Zhao and Setyawan, 2020; Cetinkaya and Erdal, 2019).

Research by Cetinkaya and Erdal (2019) and Zhao and Setyawan (2020) examined include national holidays and special occasions in the demand prediction model. These factors were shown to have an impact on the outcome of the food stock demand projection. This study intends to develop an appropriate machine learning model to generate accurate forecasts utilising holiday data, as in studies by Zhao and Setyawan (2020) and Cetinkaya and Erdal (2019) as well as weather data, with the goal of enhancing the machine learning model's performance in predicting coffee beverage stock.

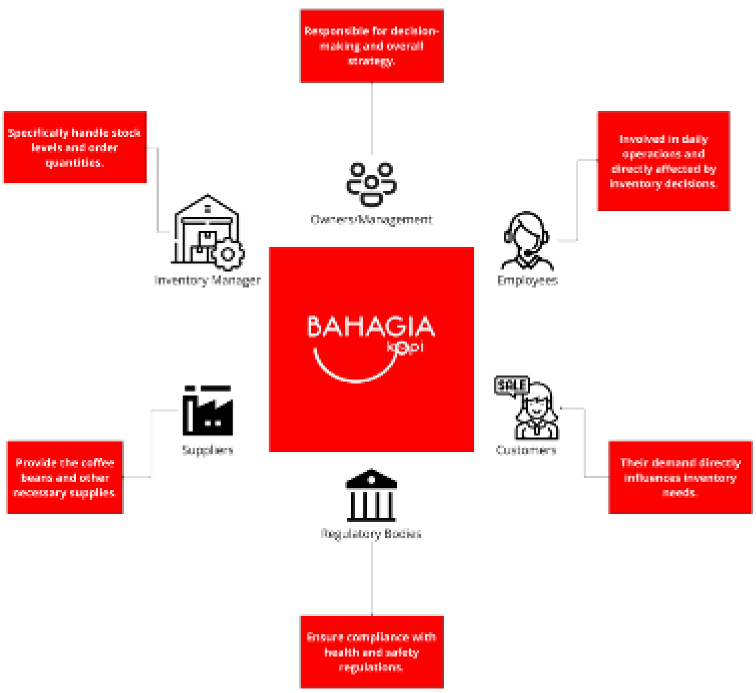

Figure 1. Rich Picture

Scenario

The thesis bases its research on Scenario 3, which is specifically developed to meet the actual issues that Bahagia Kopi faces in inventory management and demand forecasting. Scenario 3 entails creating and implementing a machine learning-based demand forecasting model that takes into account both internal sales data from the Point of Sale (POS) system and external elements such as weather and national holidays. This scenario is especially essential since it tries to optimise inventory management by properly estimating future sales and establishing suitable order quantities, hence minimising losses owing to the perishability of coffee. Scenario 3 offers a thorough solution to Bahagia Kopi's inventory management difficulties by emphasising real-world data and relevant external variables.

METHOD

This chapter describes the research methodology for conducting research and the stages that are passed. The research methodology begins with identifying the problems faced at the Coffee Shop. After that, review theories about machine learning predictions and stock ordering methods. Then collect the data needed to do the modeling. The data collected is pre-processed so that it can be trained by the machine learning model. In the machine learning model training stage, there are algorithms and parameters that are trained which are explained in the machine learning model training subchapter. The trained models are evaluated to determine the feasibility of the model to make predictions. Models that have been trained and compared in performance.

The model that is considered the best in making predictions based on analysis will be used to make predictions for the next two months. The prediction results are used to determine the order quantity with the method described in the Implementation of model prediction results subchapter. At the end of the research, conclusions are made based on the entire research process as well as suggestions for future research.

Problem Identification

The research begins with the problem identification phase where the problem focused on in this research is to make food inventory management through the prediction of the prediction model. Inventory management in food places can make losses if not managed properly because it can cause oversupply and undersupply. To overcome this, literature studies were conducted to support the research.

Literature Review

The literature review aims to support solutions to problem identification. The things needed in the literature review are basic theories about predictive models and machine learning. The machine learning models used are regression models that can produce continuous numbers because the data studied are time-series. Then proceed with methods of scheduling the ordering of materials.

Data Collection

The required data is used as input for the prediction model used, a regression model that takes into account several predictor variables. There are several parts to data collection.

1. POS Data
   Point of Sales (POS) is a system that regulates the sale and entry of ingredients. In the POS system, every sales transaction and entry of raw materials is recorded. The data obtained from the POS system used in this study are as follows:
   a. Sales of espresso-based coffee menu every day
   b. The ingredients needed for the coffee menu
   This data is needed to predict raw material demand. Coffee sales data is needed to determine the amount of raw material demand at a certain time. The amount of raw materials needed is obtained by multiplying the number of coffee menus sold by the raw materials needed.
2. Weather Data
   In this study, weather data which includes daily weather conditions such as temperature and rainfall are

used as predictor variables as elateral input. In restaurants or coffee shops, weather conditions are a factor that determines the arrival of customers to the coffee shop. Therefore, by adding weather data as input, it can be believed that it will improve the prediction results of the model.

3. Holiday Data
   There are several big days in a year such as holidays, national holidays and collective leave, these dates can trigger the arrival of customers to the coffee shop due to work or school holidays and several other things. Information on holidays and red dates is obtained from the Indonesian national calendar in 2023 and 2024.

In this study, weather data which includes daily weather conditions such as temperature and rainfall are used as predictor variables as elateral input. In restaurants or coffee shops, weather conditions are a factor that determines the arrival of customers to the coffee shop. Therefore, by adding weather data as input, it can be believed that it will improve the prediction results of the model.

RESULT AND DISCUSSION

Data Input

The data to be used in the research consists of internal data and external data. Internal data is data obtained from the company itself, while external data is data sourced from outside the company. External data is used to enrich the analysis and improve the model's accuracy by adding additional variables that influence the predicted value.

Internal Data

Internal data is the transaction history obtained from the POS system and the recipes or lists of staple ingredients that have been registered in the POS system for each menu selected as the object of research.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Item Name	Item Variant Name	Category Name	SKU	Item Sold	Item Refunded	Gross Sales	Discount	Refund	Net Sales	COGS	Gross Profit	Gross Margin
2	BMTH	Ice	NON COFFEE		2	0	44346	0	0	44346	0	44346	100%
3	Bagel Only	Double Choco Cheese Bagel	BAGEL		1	0	22173	0	0	22173	0	22173	100%
4	Bagel Only	Chocochips Bagel	BAGEL		1	0	13304	0	0	13304	0	13304	100%
5	Bagel Only	Rainbow Bagel	BAGEL		1	0	13304	0	0	13304	0	13304	100%
6	Bagel Only	Cheese Bagel	BAGEL		2	0	26608	0	0	26608	0	26608	100%
7	Bagel Only	Garlic Cheese Bagel	BAGEL		1	0	17738	0	0	17738	0	17738	100%
8	Bagel Only	Cranberry Cheese Bagel	BAGEL		1	0	22173	0	0	22173	0	22173	100%
9	Bagel Sandwich	Chicken Sandwich	BAGEL		3	0	93126	0	0	93126	0	93126	100%
10	Bahagia	Ice	COFFEE		1	0	22173	0	0	22173	0	22173	100%
11	Bento Cake DYC	Original	DYC		3	0	159645	0	0	159645	0	159645	100%
12	Black	Ice	COFFEE		1	0	15965	0	0	15965	0	15965	100%
13	Bottled Water	500ml	ADDS ON		1	0	8869	0	0	8869	0	8869	100%
14	Cake of the Weekend	1 Slice	Cake Of The Weekend		1	0	26608	0	0	26608	0	26608	100%
15	Chocolate	Hot	NON COFFEE		1	0	22173	0	0	22173	0	22173	100%
16	Chocolate	Ice	NON COFFEE		3	0	66519	0	0	66519	0	66519	100%
17	Cinnamon Roll	1 Pcs	SNACKS		1	0	8869	0	0	8869	0	8869	100%
18	Crack Drink		SUPRISINGLY GOOD!	BMTH	2	0	53216	0	0	53216	0	53216	100%
19	Kopi Susu Bu'le	Ice	COFFEE		4	0	88692	-55432	0	33260	0	33260	100%
20	Kopi Susu Madam	Ice	COFFEE		2	0	44346	0	0	44346	0	44346	100%
21	MP w/ Ice Cream		MINI PANCAKE		2	0	53216	0	0	53216	0	53216	100%
22	MP w/ Topping		MINI PANCAKE		1	0	22173	0	0	22173	0	22173	100%
23	Matcha	Ice	NON COFFEE		2	0	44346	0	0	44346	0	44346	100%
24	Ngopi Pagi	Ice White	PROMO		1	0	13304	0	0	13304	0	13304	100%

Figure 1. Data Extraction from the POS System

Internal data taken from the POS system is as follows:

1. Date
2. Daily sales of espresso-based coffee menu
3. Recipe for espresso-based coffee menu

External Data

The external data used in this research are data obtained from other sources, including weather data and national holidays. The weather data is sourced from Visualcrossing.com, which provides paid weather data from various parts of the world.

	A	B	C	D	E	F
1	Name	Date time	Maximum Temperature	Minimum Temperature	Temperature	Precipitation
2	Bandung, Jawa Barat, Indonesia	2022-07-01	28,2	16,7	21,8	0,1
3	Bandung, Jawa Barat, Indonesia	2022-07-02	28,2	19,1	22,7	0,008
4	Bandung, Jawa Barat, Indonesia	2022-07-03	27,2	19,9	22,6	0,01
5	Bandung, Jawa Barat, Indonesia	2022-07-04	27,9	20,1	23,4	0,002
6	Bandung, Jawa Barat, Indonesia	2022-07-05	28,2	19,4	22,8	4,982
7	Bandung, Jawa Barat, Indonesia	2022-07-06	26,9	20	22,3	1,134
8	Bandung, Jawa Barat, Indonesia	2022-07-07	27,9	18,8	22,9	0,018
9	Bandung, Jawa Barat, Indonesia	2022-07-08	28,2	18,3	22,4	0,6
10	Bandung, Jawa Barat, Indonesia	2022-07-09	28,7	17,7	21,9	0,8
11	Bandung, Jawa Barat, Indonesia	2022-07-10	25,4	16,8	21,1	0,9
12	Bandung, Jawa Barat, Indonesia	2022-07-11	27,7	16,9	22,1	0,5
13	Bandung, Jawa Barat, Indonesia	2022-07-12	28,6	20,2	22,8	0,029
14	Bandung, Jawa Barat, Indonesia	2022-07-13	25,7	19,5	21,6	6,117
15	Bandung, Jawa Barat, Indonesia	2022-07-14	27,9	19,2	22,1	11,943
16	Bandung, Jawa Barat, Indonesia	2022-07-15	26,2	20	22,1	4,068
17	Bandung, Jawa Barat, Indonesia	2022-07-16	22,9	18,8	20,7	33,051
18	Bandung, Jawa Barat, Indonesia	2022-07-17	25,4	18,9	21,9	13,923

Figure 2. Weather data extraction from Visualcrossing

Figure 2 shows weather data taken from Visualcrossing.com, with only a few variables selected from the data due to their relevance to conditions in Indonesia. The external data used in this research includes:

1. Maximum and minimum temperature
2. Rainfall
3. Holidays and national days

Data Cleaning

Data Cleaning is performed using Microsoft Excel and Jupyter Notebook. For internal data, the raw data as shown in Figure 4.1 is first processed using a Pivot Table to obtain the daily sales values of the menu items.

The data used is in the form of a time series where there are time and value variables, and the value will be predicted. The value used is the daily demand for coffee raw materials. The sales data from the POS records the time (hours, minutes, and seconds) and date of each transaction. Therefore, aggregation using a Pivot Table is necessary to obtain the daily sales values.

	A	B	C	D	E
1	Date	Affogato	Affogato - Single - Arabica	Affogato - Single - Robusta	Affogato - Double - Arabica
2	01/07/22	0	1	0	0
3	02/07/22	5	2	0	0
4	03/07/22	4	1	0	0
5	04/07/22	2	1	0	0
6	05/07/22	1	0	0	0
7	06/07/22	3	1	0	0
8	07/07/22	0	0	0	0
9	08/07/22	2	2	0	1
10	09/07/22	1	2	0	0
11	10/07/22	2	0	0	1
12	11/07/22	0	0	1	0
13	12/07/22	2	0	0	0
14	13/07/22	0	2	0	0
15	14/07/22	0	1	0	0
16	15/07/22	2	0	0	1
17	16/07/22	2	3	0	0

Figure 3. Transformation of coffee sales data

The sold menus listed as categories have their 'Items' column transformed into columns with the names of the menus, and their values represent the quantity sold on that date. The 'Date' column, which initially had several identical values, is changed to unique values.

	A	B	C	D	E
1	**Date**	**Arabica**	**Robusta**	**Filter**	**Coffee_Beans**
2	01/07/22	270	1290	15	1575
3	02/07/22	910	1680	30	2620
4	03/07/22	720	1390	75	2185
5	04/07/22	410	1150	45	1605
6	05/07/22	470	1310	30	1810
7	06/07/22	570	1350	30	1950
8	07/07/22	230	1050	15	1295
9	08/07/22	740	1240	15	1995
10	09/07/22	450	1160	30	1640

Figure 4. Coffee bean data transformation

Since the case discussed in this research involves raw materials, further conversion is needed to change the daily sales data into the amount of coffee beans required each day. The operation performed to convert daily menu sales into coffee beans in grams is to sum the sales figures of all coffee menus for each observation and multiply by the grams of coffee beans needed according to the recipe.

After that, the processed POS data is combined with weather data until it forms a single table or dataframe.

Data Exploration

The Influence of Day

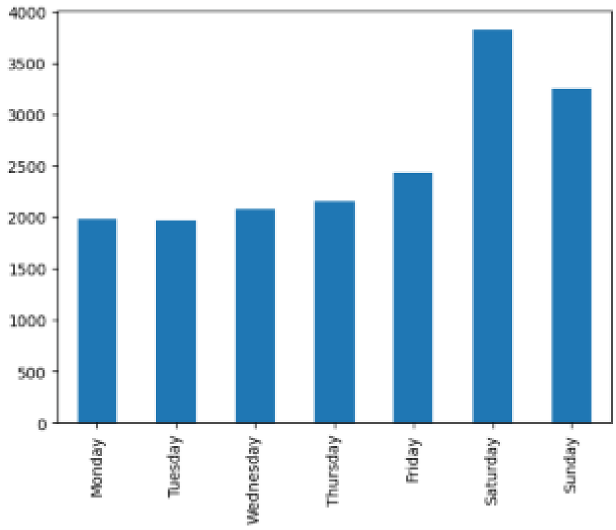

Figure 5. Average demand by day

To see more specifically the factors affecting the fluctuations in coffee bean demand, it can be observed from the average demand by day as shown in Figure IV.5. On weekends, the demand shows the highest levels, especially on Saturdays.

The Infuence of Rain

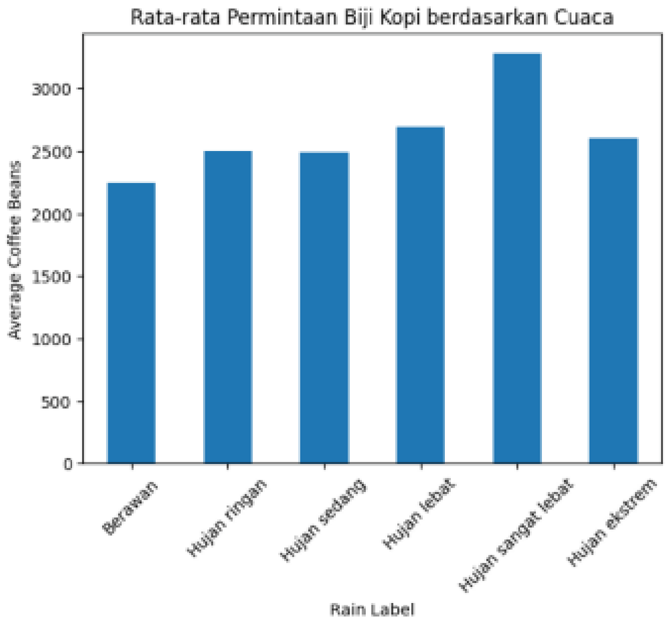

Figure 6. Average demand based on rainfall

External data can help analyze changes in the value of coffee bean demand, such as rainfall data as shown in Figure IV.6. The categorization of rainfall based on the amount of rainfall in mm/day as determined by BMKG is as follows:

- 0 mm/day: Overcast (no rain)
- 0.5 - 20 mm/day: Slight rain
- 20 - 50 mm/day: Moderate rain
- 50 - 100 mm/day Heavy Rain
- 100 - 150 mm/day: Very heavy rain
- >150 mm/day: Extreme rain

The demand for coffee lowest point on days when there is no rain and tends to increase as the rain becomes heavier. The demand for coffee beans reaches its highest point on days when there is very heavy rain. This becomes s a hidden pattern that can determine the value of coffee bean demand.

Data Preparation

Regression models generally make predictions based on numbers in the predictor variables. Categorical variables and date variables need to be converted into numerical forms (integer or float) to be processed by the model. There are two types of data that need to be transformed, namely date data and categorical data. For date-type data, it is transformed into 6 variables, namely dayofweek, month, year, dayofyear, dayofmonth, and weekofyear.

Table 1. Transformation of date variables

Date		dayofweek	month	year	dayofmonth	weekofyear
01/07/2022		5	7	2022	1	26
02/07/2022	→	6	7	2022	2	26
03/07/2022		7	7	2022	3	26

Tabel 1 shows the transformation technique into variables that have integer data types. With the transformation of these variables, the machine learning model can process the training because the variables are in numerical form.

Table 2. Explanation of predictor variables

Category	Explanatory Variable	Definition
Date	Year	Year on the date
	Month	Month of the year (1-12)
	Day	Days of the week (1-7)
	Day of the month	Day of the month (1-31)
	Week of the year	Week of the year (1-52)
Holiday	Holiday	Public holiday on the calendar
	Before Holiday	D-1 before the holiday
	No Holiday	No holiday
Event	Ascension day	18 May 2023 9 May 2024
	Christmas	Every December 25th
	Eid Al-Fitr	22 April 2023 10 April 2024
	Eid al-Adha	10 July 2022 29 June 2023 17 June 2024
	Good Friday	7 April 2023 29 March 2024
	Independence Day	Every August 17th
	Hijri new year	30 July 2022 19 July 2023 7 July 2024
	Isra Miraj	18 February 2023 8 February 2024
	Chinese New Year	22 January 2023 10 February 2024
	Labor Day	Every May 2nd
	Mass Leave	24 – 26 April 2023 12 – 15 April 2024
	Pancasila Day	Every June 1st
	Vesak	4 June 2023 23 May 2024
	Prophet's Birthday	8 October 2022 28 September 2024 16 September 2024
	Easter	9 April 2023 31 March 2024

After separating the date variables and categorical variables, the definitions of each variable are explained in Table 4.4. Date variables such as month and day are considered predictor variables because they have the potential to significantly impact the prediction results. Then additional information on that day, such as holidays and festive days (events), is also considered to be variables that have an influence based on the results of the data exploration conducted.

In general, training a machine learning model requires training data and test data for data testing. Train data is the data used by the model for training, and after the model is trained using the train data, the model will make predictions using the test data, where the target variable values in the test data will be compared with the predicted values to obtain the model's error. The case faced in this research is predicting the demand for coffee beans for the next month. Thus, the data proportion for the data division is that the training data is the data from January 1, 2020, to May 31, 2021, and the test data is the data from June 1, 2021, to June 30, 2021.

Model Training

Multiple Linear Regression

The training of the MLR model was conducted by performing feature selection using the step-wise method. Two types of step-wise methods were used, namely Forward and Backward.

The Forward Method starts by adding predictor variables one by one. The feasibility of the variables included in the MLR model is measured by the significance value (p-value) of the predictor variables against the target variable. The Backward method starts by including all variables initially, then removing the insignificant variables one by one. The significance value used as a reference is 0.05.

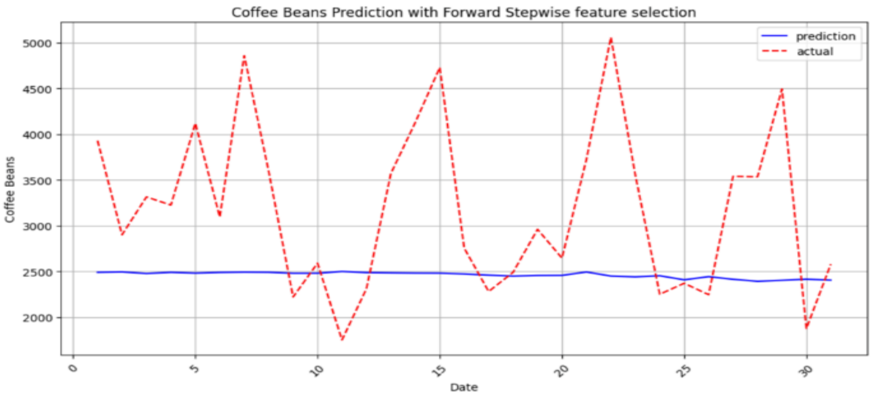

Figure 7. MLR prediction result with the step-wise forward method

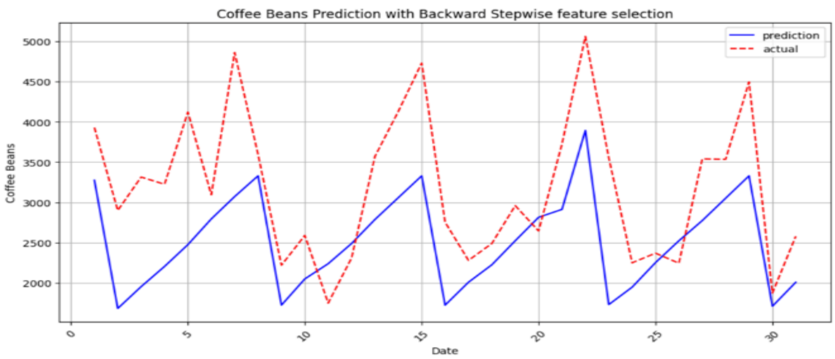

Figure 8. MLR prediction result with the step-wise backward method

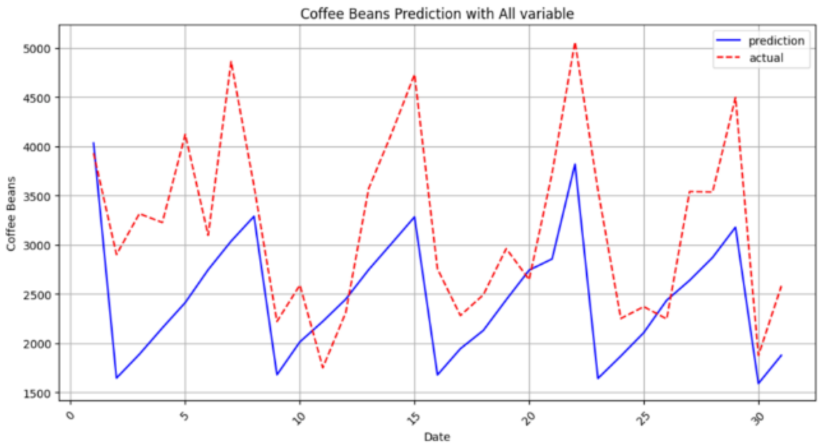

Figure 9. MLR prediction result using all variables

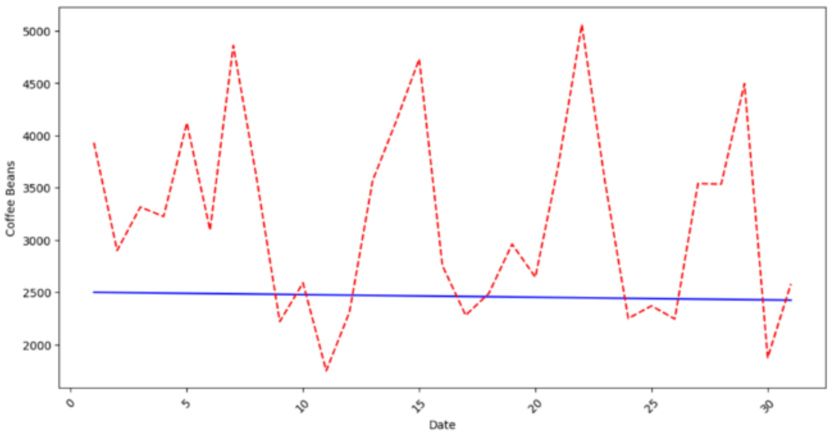

Figure 10. MLR prediction result using only date variables

Table 2. Measurement of prediction error from the MLR model

	Model	RMSE	R2	MAE	MAPE
0	MLR All variables	829.800024	0.375281	635.050082	6.900081e+16
1	MLR Date Only	995.035959	-0.000812	803.145775	3.799659e+16

Looking at the errors produced by the trained MLR models, the MLR All variables generated the smallest RMSE and MAE. Meanwhile, the MLR date only method produced the smallest MAPE. The R^2 value also shows a higher proportion of variance in the MLR All variables compared to the other models. In the MLR model, the addition of other variables besides the date variable does not help reduce the error value but helps create a more dynamic pattern and follow the trend.

Decision Tree

In conducting Decision Tree training, hyperparameter tuning of the model is necessary. Two models were produced during the training of the DT model, namely the model using all variables and the model using only the date variable. A parameter that most influences the occurrence of the overfitting phenomenon is Max Depth or the depth of the model tree.

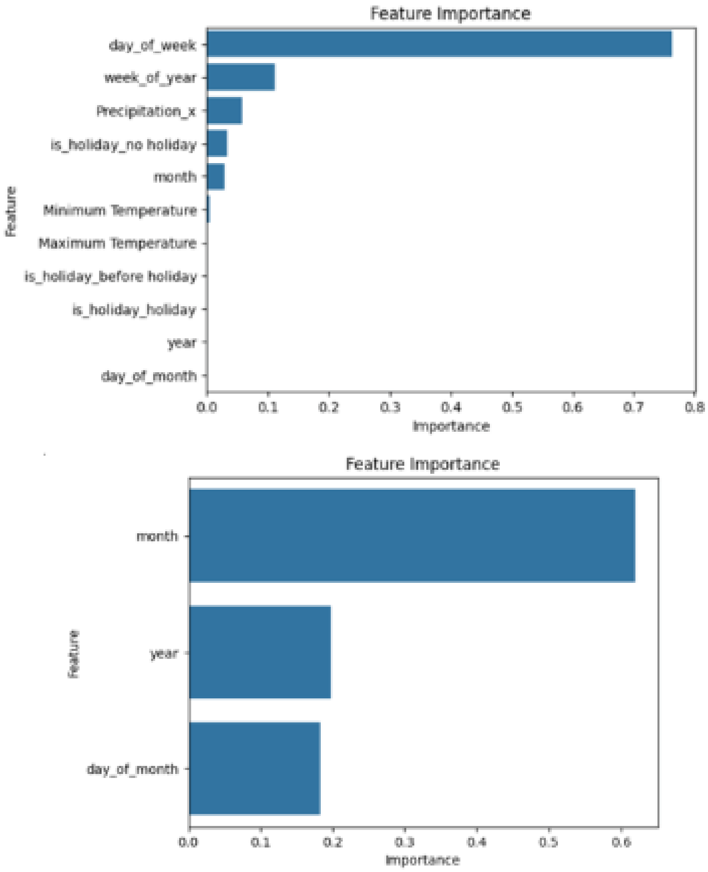

Figure 11. Feature importance ranking in the DT model

The feature importance ranking results from both models are displayed in Figure 11. For the model using all variables, the day of week are considered important features that determine the prediction results due to their influence on changes in coffee bean demand. Whereas for the model that uses only date variables, the only feature considered important is the month.

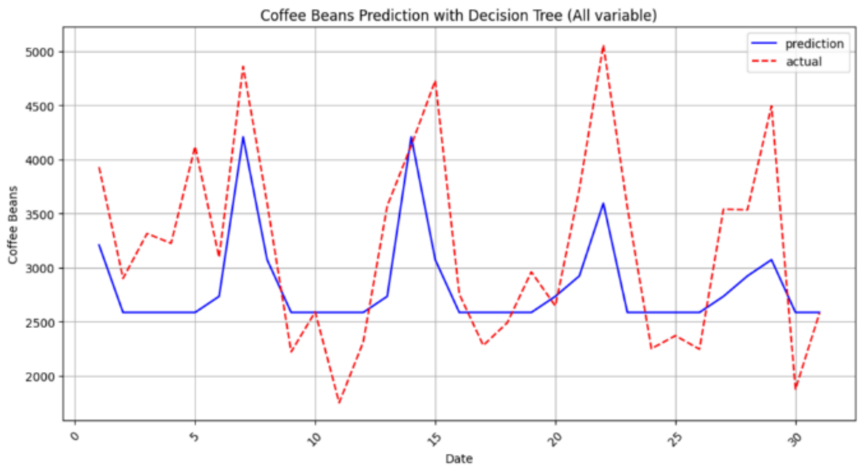

Figure 12. The prediction results of DT using all variables

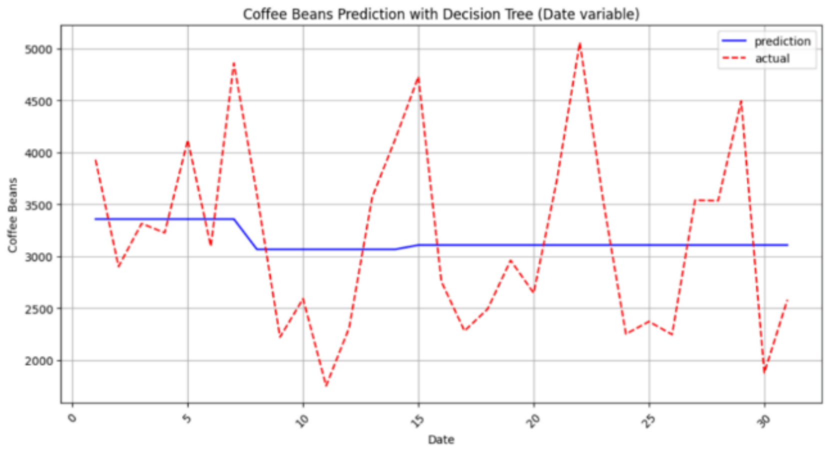

Figure 13. The DT prediction results only use the date variable

Table 3. Measurement of prediction error from the Decision Tree model

	Model	RMSE	R2	MAE	MAPE
0	DT All variables	786.464715	0.438827	554.225174	6.916905e+16
1	DT Date Only	995.035959	-0.000812	803.145775	3.799659e+16

The error produced by the model using all variables appears significantly lower compared to using only the date variable in all error measurements. The R2 value in the model using all variables is significantly higher compared to the model using only the date variable. The addition of weather variables, and holidays contributes to the variance of the predictor variables. Based on the error values and the prediction patterns shown, adding variables beyond the date variable results in predictions that are closer to the actual values.

Support Vector Regressor

SVR training is conducted with hyperparameter tuning using GridSearchCV. GridSearchCV works by trying one by one the combinations of parameters that have been registered and selecting the best one based on the MSE value.

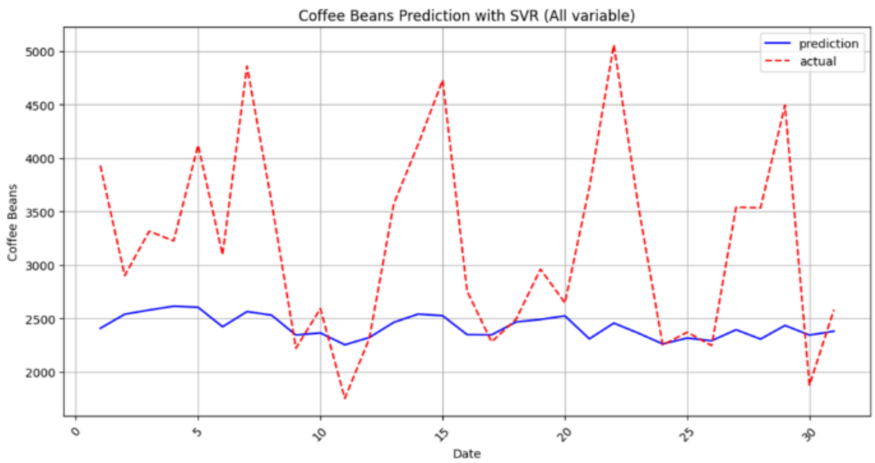

Figure 14. The SVR prediction results using all variables

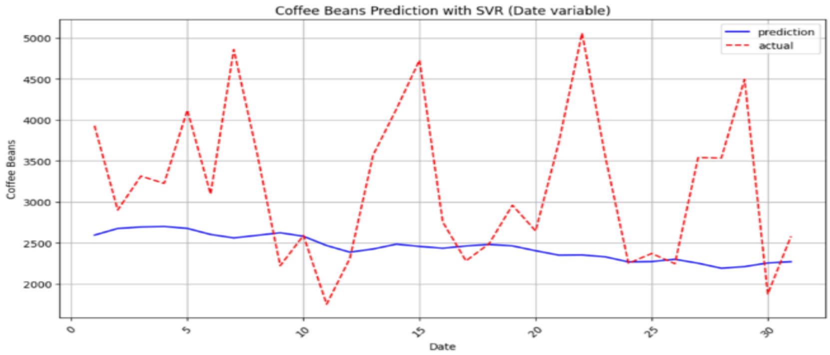

Figure 15. The SVR prediction results using only the date variable

Model training was conducted using the established parameters, and the prediction results of both models are shown in Figures IV.15 and IV.16. The prediction patterns of both models appear to move horizontally. The model that uses all variables has a smooth and non-constant change pattern due to the numerous influences of the variables. But the model can still follow the trend and weekly pattern. Whereas the pattern that only uses the date variable tends to be constant and closely follows the weekly pattern and cannot follow the trend.

Table 4. Measurement of prediction error from the SVR model

	Model	RMSE	R2	MAE	MAPE
0	SVR All variables	1029.477358	0.038451	772.158887	5.747760e+16
1	SVR Date Only	1046.760853	0.005894	791.304774	5.376854e+16

Looking at the errors produced by both models, the model that uses all variables yields smaller MAE, MAPE and RMSE values. The R2 value in the indicating that the predictions generated are in line with the ongoing trend. This proves that for the SVR model, the addition of variables beyond the existing ones improves the quality of the predictions.

Neural Network

The Neural Network model was trained using an MLP architecture with only one hidden layer in the model to ensure comparability with other models. The training of the MLP model also uses GridSearchCV to determine the best combination, measured by the MSE value. Two models were trained: one model using variables and one model using only date variables.

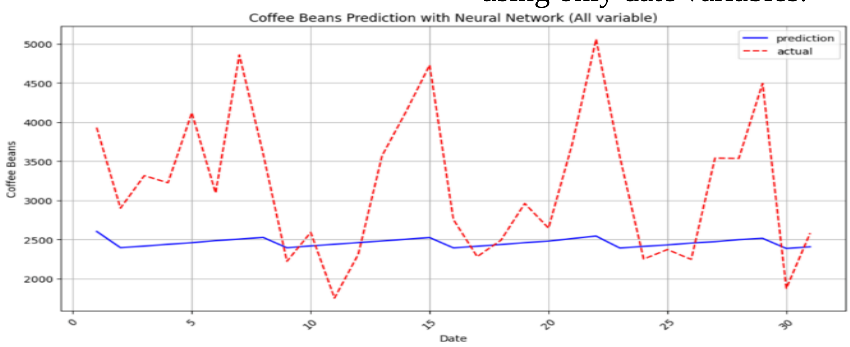

Figure 16. The prediction results of MLP using all variables

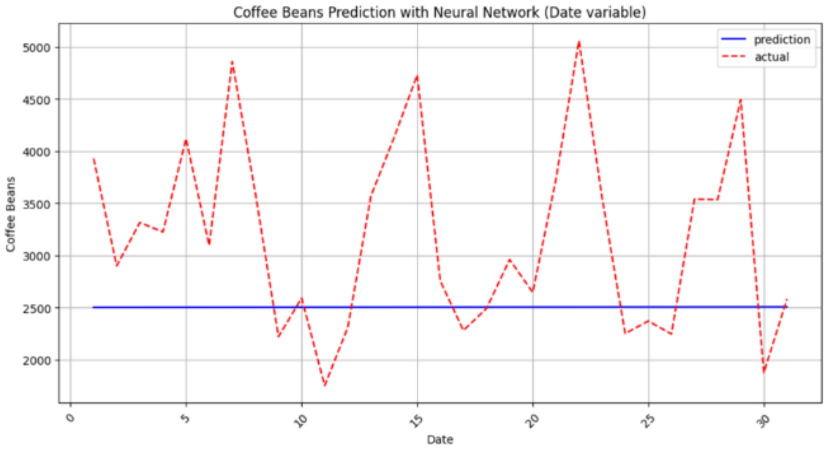

Figure 17. The MLP prediction results using only the date variable

The MLP prediction results using all variables that are flat do not experience significant increases and decreases but can follow the trend pattern that is currently declining. The model that uses only the date variable follows a weekly pattern and incorrectly interprets the ongoing downward trend, predicting it as an upward trend.

Table 5. Measurement of prediction error from the NN model

	Model	RMSE	R2	MAE	MAPE
0	MLP All variables	1028.468251	0.040335	820.475633	6.342855e+16
1	MLP Date Only	1051.690368	-0.003492	846.194330	6.342396e+16

According to the error calculation results displayed in Table 5. The MAE and RMSe of the model with all variables show a smaller error, although the MAPE is. The R^2 value in the model using all variables is higher, compared to the model using only the date variable. This indicates that the addition of other variables beyond the date variable affects the proportion of variance in the predictor variable.

Model Prediction Evaluation

The trained prediction models are compared with each other to further analyze which model is the most effective in predicting coffee bean demand. The prediction models use all variables compared to each other, as well as the model that only uses the date variable.

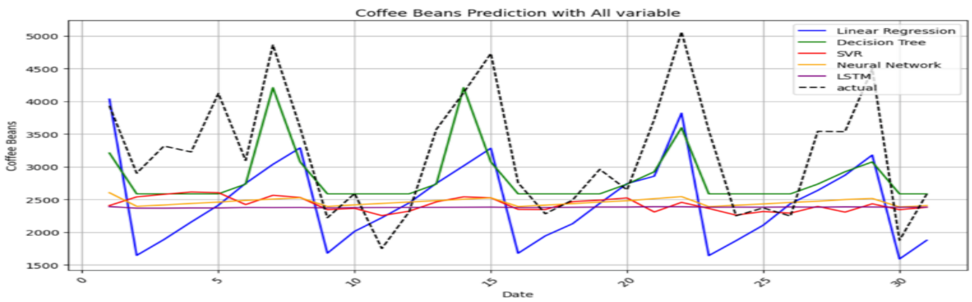

Figure 18. Comparison of predictions from all models using all variables

The prediction results of the four models using all variables are shown in Figure IV.19 along with the actual values. The DT model shows a prediction pattern that is most similar to the actual values during the weekend. The MLP and SVR models predict values that are not far apart between the weekend and the beginning of the week, although they still follow the weekly pattern and have more dynamic value changes. The Linear Regression or MLP model produces prediction patterns that are most similar to the weekly patterns, experiencing linear increases and decreases and can follow a declining trend.

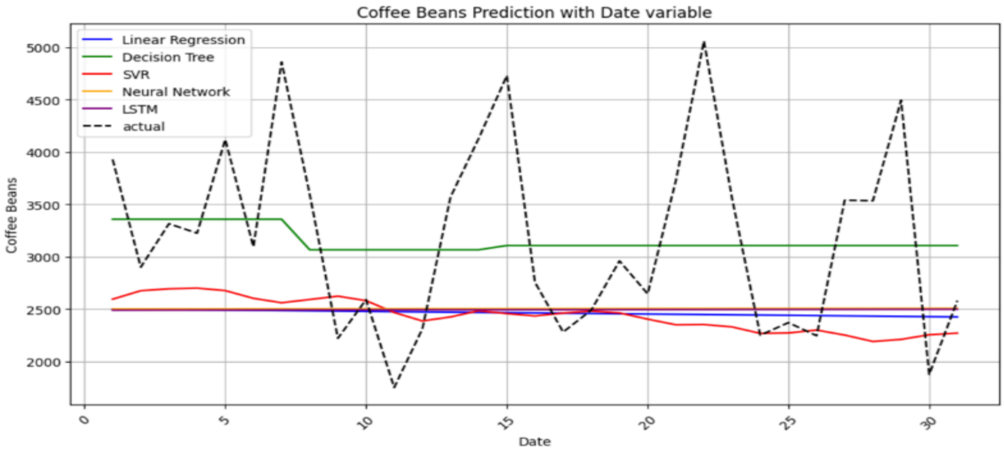

Figure 19. Comparison of predictions from all models using date variable

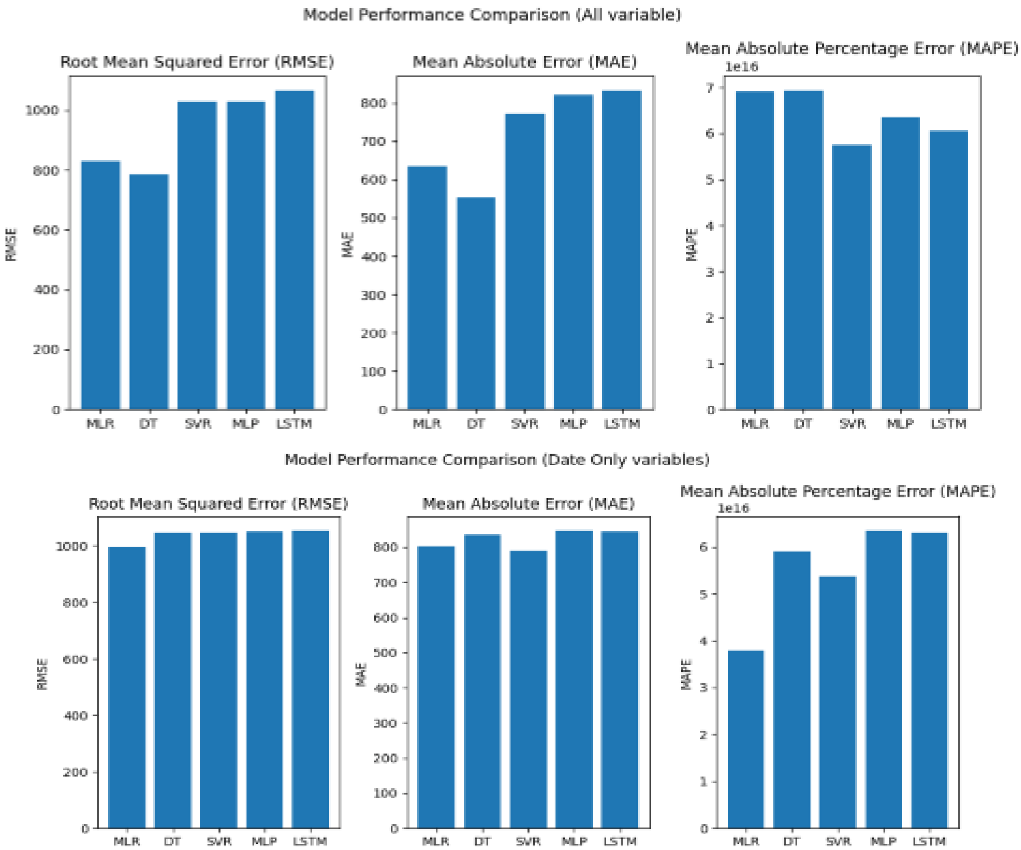

Figure 20. Comparison of errors for all models

Figure 20 shows a comparison of error measurements between all the trained models. Overall, the model that uses all variables produces smaller errors compared to the model that uses only the date variable.

Implementation of Prediction Results

The prediction results of the DT model using all variables are used for the implementation of stock quantity management. The DT model was chosen because it produced the smallest error value and the pattern most similar to the actual data. Table 6 shows the prediction results using the DT model for December 2024 to March 2025.

Table 6. DT model demand prediction results

	Date	Coffee_Beans
0	2024-12-01	3596
1	2024-12-02	2668
2	2024-12-03	2668
3	2024-12-04	2668
4	2024-12-05	2668
...	...	...
116	2025-03-27	1727
117	2025-03-28	2734
118	2025-03-29	2922
119	2025-03-30	3596
120	2025-03-31	1727

121 rows × 2 columns

Stock Supply Management

The predicted demand results for December 2024 to March 2025 serve as input to determine the variables that influence the ROP (Re-order point) value. The standard deviation (σ) and mean (μ) were obtained from the values in table IV.6. Meanwhile, for ordering coffee beans, it takes a maximum of 7 days from the order until the coffee beans arrive, so the lead time (LT) used is 7. Table 7 displays the variable values for the ROP calculation.

Table 7. Variables for calculation

Variable	Value
LT	7
Z	1.65
σ	672.12
μ	2612.36

Using the above variables, the value of Safety Stock (SS), the average demand during the order period (d_L), and ROP can be calculated.

$$SS = z\sigma\sqrt{LT} = 1.65(672.12)\sqrt{7} = 2934.15\ gram$$

$$d_L = \mu LT = 2612.36(7) = 18286.49\ gram$$

$$ROP = SS + d_L = 2934.15 + 18286.49 = 21,220.64\ gram \approx 22\ bags$$

The ROP value has been made into information on the dashboard so that users know when to place an order. The dashboard helps in displaying data in a form of visualization that is easy for users to understand. The results of future time predictions using a trained model are visualized through the dashboard, equipped with features needed by users to meet their analytical needs.

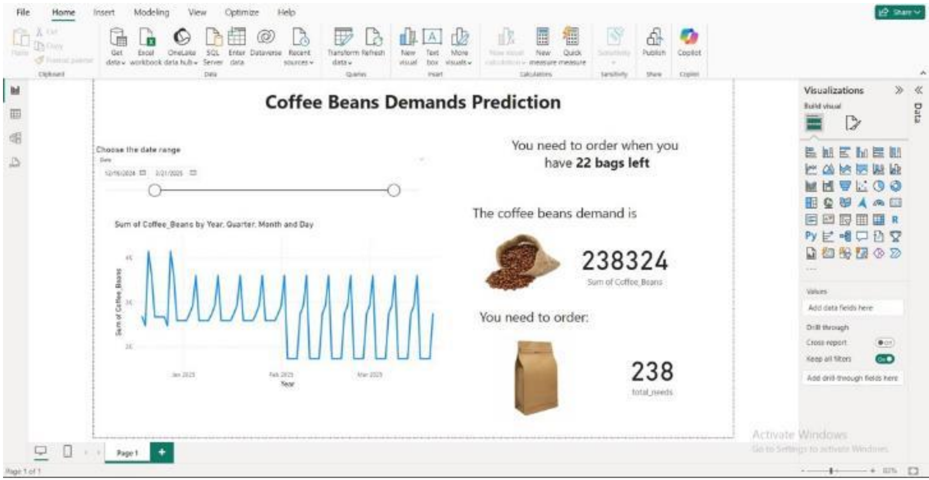

Figure 21. Coffee Bean Demand Dashboard Prototype

Figure 21 is a prototype dashboard display containing information on coffee bean demand predictions for December 2024 to March 2025, created using Microsoft PowerBI. The predicted values displayed are based on the prediction results of the Decision Tree model in Table IV.6. On the dashboard, there is a line graph that informs the amount of demand needed within the selected time range. The selection of the time range or date can be adjusted using the filter located at the top left. The quantity of coffee bean requests is displayed on the right, along with the number of coffee bags that need to be ordered. The ROP information is displayed in the top right corner to remind the user to order coffee beans when the stock of coffee beans is down to 22 bags.

Through this dashboard, users or business operators can estimate the amount of coffee beans that need to be ordered according to their needs. Considering user suggestions and needs to make the dashboard more informative and user-friendly.

CONCLUSION AND RECOMENDATION

Conclusion

Based on the research that has been conducted from modeling to implementation, it can be concluded that:

1. This research successfully conducted a machine learning-based prediction of coffee bean demand using external variables such as weather and holidays to test their impact on the prediction results.

2. Machine learning models can follow weekly patterns in their predictions, with an acceptable error value of RIMSE and MAE below 800, because a difference in demand of 800 grams per day is not significant in this case.
3. Overall, models that use all variables produce better prediction results than models that use only the date variable, indicating that weather and holidays contribute to improving the performance of the prediction model.
4. The DT model, using all variables, makes the best predictions based on the smallest error calculation and patterns similar to the prediction results.
5. All models can produce predictions according to the weekly pattern where demand decreases at the beginning of the week and increases at the end of the week.
6. The prediction results using the machine learning model can be applied to the business using a Dashboard to plan coffee bean orders.
7. Determining the order quantity of coffee beans using the ROP approaches can help business operators determine costs, order time, and stock quantity based on predictive results.

Recommendation

To support the effectiveness of the proposed method in this research, further evaluation with business practitioners is necessary to adjust the proposed method to field conditions, particularly in determining ROP; Dashboard, Expected Value Analysis, and FIFO. In addition, further research is needed that considers the approach of weather and holidays variables as predictor variables that influence the prediction results of machine learning models. Specifically, using the models that have been employed in this research, in other cases or using different models considering the variables of the weather and holidays.

DAFTAR PUSTAKA

Cetinkaya, Z., & Erdal, E. (2019, September). Daily food demand forecast with artificial neural networks: Kırıkkale University case. In *2019 4th international conference on computer science and engineering (UBMK)* (pp. 1-6). IEEE.

Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*.

J. Evans, “Business Analystics, EBook, Global Edition,” 2016, Accessed: Aug. 21, 2024. [Online]. Available: https://www.goodreads.com/book/show/36536095-business-analytics-global-edition.

J. Fitzsimmons and M. Fitzsimmons, *Service management,* Seventh Edition. 2006.

M. Taddy, “Business data science : combining machine learning and economics to optimize, automate, and accelerate business decisions,” Accessed Aug. 21, 2024. [Online]. Available: https://www.oreilly.com/library/view/business-data-science/9781260452785/.

Pramelani, P. (2020). Faktor Ketertarikan Minuman Kopi Kekinian Terhadap Minat Beli Konsumen Kalangan Muda. *Managament Insight: Jurnal Ilmiah Manajemen*, *15*(1), 121-129.

Tangtisanon, P. (2018, April). Web service based food additive inventory management with forecasting system. In *2018 3rd International Conference on Computer and Communication*

Systems (ICCCS) (pp. 448-452). IEEE.

Tanizaki, T., Hoshino, T., Shimmura, T., & Takenaka, T. (2019). Demand forecasting in restaurants using machine learning and statistical analysis. Procedia CIRP, 79, 679-683.

Zhào, M. A., & Setyawan, B. (2020, November). Sales forecasting for fresh foods: a study in Indonesian FMCG. In *2020 International Conference on Information Science and Communications Technologies (ICISCT)* (pp. 1-9). IEEE.