Jurnal Sistem Teknik Industri
Journal homepage: **https://talenta.usu.ac.id/jsti**

The Use of Machine Learning Algorithms for Supply Chain Optimization at PT. XYZ

Diomen Syahputra Manik*[1], Nazaruddin[1], Nismah Panjaitan[1]
[1]*Industrial Engineering Department, Faculty of Engineering, Universitas Sumatera Utara, Jln. Dr. T. Mansyur No. 9 Padang Bulan, Medan, 20222, Indonesia*
*Corresponding Author: diomen.manik13@gmail.com

ARTICLE INFO
Article history:
Received 23 July 2025
Revised 9 September 2025
Accepted 16 September 2025
Available online 23 February 2026

E-ISSN: 2527-9408
P-ISSN: 1411-5247

How to cite:
D.S. Manik, Nazaruddin, and N. Panjaitan, "The use of machine learning algorithms for supply chain optimization at PT. XYZ," *J. Sist. Tek. Ind.*, vol. 28, no. 1, pp. 11–21, Feb. 2026.

ABSTRACT

Increased demand fluctuations pose a major challenge in supply chain management, particularly in the fast-food beverage industry like PT. XYZ. This research aims to build and evaluate a demand forecasting model based on machine learning, considering multivariate variables such as product price, seasonal trends, weather, per capita income, population, and historical sales data. The three algorithms used are Random Forest Regressor, Gradient Boosting Regressor, and Prophet Time Series Model. This research method employs a quantitative approach with descriptive-predictive analysis based on time-series data. Model evaluation was conducted using MAE, MSE, RMSE, and MAPE metrics. The research results indicate that Prophet has the highest accuracy (MAPE: 2.33%) and excels in capturing seasonal trends, while Random Forest ranks second (MAPE: 2.47%) with an advantage in comprehensively handling multivariate variables. Gradient Boosting yields the lowest accuracy (MAPE: 2.70%). The conclusion of this study recommends the use of Prophet for short-term seasonal-based predictions, while Random Forest is more suitable for medium to long-term strategic planning. The combination of the two has the potential to become an accurate and adaptive hybrid approach for optimizing the demand forecasting system at PT. XYZ.

Keyword: demand forecasting, machine learning, random forest, prophet, gradient boosting regressor.

ABSTRAK

Fluktuasi permintaan yang tinggi menjadi tantangan utama dalam pengelolaan rantai pasokan, khususnya di industri minuman cepat saji seperti PT. XYZ. Penelitian ini bertujuan untuk membangun dan mengevaluasi model peramalan permintaan berbasis machine learning dengan mempertimbangkan variabel multivariat seperti harga produk, tren musiman, cuaca, pendapatan per kapita, jumlah penduduk, dan data penjualan historis. Tiga algoritma yang digunakan adalah Random Forest Regressor, Gradient Boosting Regressor, dan Prophet Time Series Model. Metode penelitian ini menggunakan pendekatan kuantitatif dengan analisis deskriptif-prediktif berbasis data time-series. Evaluasi model dilakukan dengan menggunakan metrik MAE, MSE, RMSE, dan MAPE. Hasil penelitian menunjukkan bahwa Prophet memiliki akurasi tertinggi (MAPE: 2,33%) dan unggul dalam menangkap tren musiman, sementara Random Forest menempati posisi kedua (MAPE: 2,47%) dengan keunggulan dalam menangani variabel multivariat secara komprehensif. Gradient Boosting menghasilkan akurasi terendah (MAPE: 2,70%). Kesimpulan dari penelitian ini merekomendasikan penggunaan Prophet untuk prediksi jangka pendek berbasis musiman, sedangkan Random Forest lebih sesuai untuk perencanaan strategis jangka menengah-panjang. Kombinasi keduanya berpotensi menjadi pendekatan hybrid yang akurat dan adaptif untuk optimalisasi sistem peramalan permintaan di PT. XYZ.

Kata kunci: peramalan permintaan, machine learning, random forest, prophet, gradient boosting regressor.

1. Introduction

This research focuses on the fast-food beverage sector, using PT. XYZ as the case study. The industry context is important because demand fluctuations in this sector often show unique characteristics compared to other fast-moving consumer goods. Dynamic consumer preferences, influenced by seasons, lifestyle trends, and marketing activities, make demand patterns highly volatile [1]. If companies fail to anticipate these changes, imbalances between production, distribution, and market needs can easily occur, resulting in either overstock or stockout conditions [2]. Therefore, accurate demand forecasting is essential not only to maintain supply chain efficiency but also to ensure that market expectations are met in a timely and effective manner.

Although there have been many technological advancements, forecasting practices in the ready-to-drink beverage industry generally still rely on conventional methods such as moving averages and simple linear regression. This approach has limitations because it only utilizes historical sales data without considering other external variables that play a crucial role in shaping demand [3]. As a result, the predictions generated are unable to fully reflect market dynamics, leading to recurring errors in production and distribution planning over time [4].

Field observations indicate that consumer demand for ready-to-drink products does not solely depend on historical sales patterns, but is also significantly influenced by price changes. Price increases or decreases directly affect consumer purchasing power and demand volume. In addition, seasonal trends such as long school holidays, religious holidays, and changes in seasons (rainy or dry) also contribute to creating demand patterns that are non-linear and difficult to anticipate with only basic statistical approaches [5]. In addition to price and season, environmental factors such as weather conditions play an important role in influencing beverage demand [6]. For example, hot weather conditions often lead to a significant increase in the demand for bottled beverages. On the other hand, rain and cold weather tend to decrease the consumption of certain drinks. These weather changes are dynamic and vary by region, so they need to be systematically considered in the forecasting model. However, these variables have not yet been fully accommodated in the company's current forecasting system [5].

Furthermore, socio-economic variables such as per capita income and population size also have a significant contribution to consumption patterns [7]. Regions with high income levels tend to have more stable and higher demand for premium products such as packaged beverages, while regions with low income show demand patterns that are more sensitive to price. Similarly, the population size in a region will determine the market potential and the volume of product distribution required. These demographic dimensions have not yet been fully integrated into the demand forecasting model at PT. XYZ.

The challenges related to demand forecasting, such as overstock and stockout, can be grouped into two categories: (1) internal operational limitations (reliance on conventional forecasting, lack of integration of external data) [8], and (2) external market dynamics (price sensitivity, weather, seasonal demand, and socio-economic variations) [9]. Grouping these challenges helps clarify the problem structure.

The failure to integrate these complex variables into the forecasting system has resulted in inaccurate demand estimates, leading to overstock on one side and stockout on the other [10]. Overstock causes high storage cost accumulation, resource wastage, and potential product damage, especially for products with a limited shelf life. On the other hand, stockouts can lead to lost sales opportunities, damage to distributor trust, and disruptions to customer loyalty [11]. Both situations indicate that the forecasting approach used so far has not been able to anticipate the rapidly changing market realities. In addressing that challenge, a machine learning-based approach has become a promising solution. Machine learning algorithms such as Random Forest, Gradient Boosting, and Prophet are capable of processing large and complex amounts of data by considering the relationships between variables, both historical and external [12]. in contrast to standard statistical models, machine learning has the ability to learn from non-linear patterns, detect anomalies, and adapt to changes in real-time [13]. Through integrating variables such as price, season, weather, demographics, and historical sales data into a single predictive model, the company can obtain more accurate and contextual demand projections.

With the implementation of a machine learning-based forecasting model, PT. XYZ can improve efficiency in production planning, minimize storage costs, and enhance the accuracy of product distribution to the right market at the right time. In addition, this system also enables more data-driven decision making, which will ultimately strengthen the company's supply chain resilience against market volatility [14] [15]. Therefore, the integration of machine learning into demand forecasting systems is not only a technological innovation but also a strategic necessity in supporting the competitiveness of companies in the era of Industry 4.0.

2. Method

This research uses a quantitative approach with a descriptive predictive method [16]. This approach was chosen to systematically describe the characteristics of product demand data while simultaneously building an accurate forecasting model based on machine learning algorithms. The main focus of this research is to develop a product demand prediction model that considers internal and external factors influencing market fluctuations [17]. Because the results are aimed at direct application in an industrial context, this research falls into the category of applied research, with the ultimate goal of providing practical solutions for optimizing supply chain planning at PT. XYZ.

In forecasting research that uses a machine learning approach, there are several data processing stages that must be carried out to ensure the model works optimally. The stages include processing raw data for input into Python, writing machine learning code, and simulating forecasting results using the trained model. The stages in more detail are as follows:

1. Data Collection and Preparation

The initial stage of this research begins with the collection and preparation of data consisting of primary and secondary data [18]. Primary data is obtained through informal interviews and direct observations of the forecasting and distribution systems implemented by the company, to understand the operational processes and relevant variables in demand forecasting. Secondary data is obtained from official and reliable sources, including: historical sales data (2022–2024), weekly or monthly product prices, information on seasonal trends and national holidays, weather data (temperature and rainfall) from BMKG [19], as well as economic data such as per capita income and population from BPS [20].

Primary data obtained from informal interviews and direct observations were transformed into quantitative indicators. For example, findings regarding the impact of holiday seasons were encoded into the “Trend” variable (0 = normal period, 1 = holiday/peak season). Observations related to environmental conditions, such as weather and seasonal changes, were aligned with the “Weather/Season” variable (1 = rainy/cold season, 2 = dry/hot season). Additional notes on operational bottlenecks were categorized as binary indicators and synchronized with the time-series dataset.

2. Data Plotting

After the data is collected and prepared, the next step is data plotting to see the patterns that emerge in the time series, such as trends, seasonality, or random fluctuations [21]. This visualization is important for understanding the characteristics of the data and helping to determine the appropriate forecasting approach. Plotting is done using line graphs and other relevant diagrams to illustrate the relationships between variables in the data [22].

3. Data Preprocessing for Machine Learning

Before being used in training the Machine Learning model, the data is first processed to be ready for analysis [23]. This stage includes data cleaning, such as handling missing data, and normalizing or standardizing numerical values to have a uniform scale. Next, data from various important variables such as price, weather, holidays, and economic factors are combined, and then rearranged into a time series format. After all the data is ready, it is divided into training and testing data to measure the model's performance in making predictions [24].

4. Machine Learning Model Development

The model development applies three Machine Learning algorithms, namely Random Forest Regressor, Gradient Boosting, and Prophet, because all three can handle time series data and follow trend and seasonal patterns. Random Forest works by combining many decision trees to improve accuracy, while Gradient Boosting builds the model gradually to correct errors from the previous model [25]. Prophet, on the other hand, is specifically designed for forecasting business data that has seasonal patterns, trends, and holidays. The three models were trained using training data and tested with test data, then evaluated using MAE, RMSE, and MAPE metrics to determine which model was the most accurate in predicting product demand [26].

To reduce the risk of overfitting, the models were trained and validated using a k-fold cross-validation approach in addition to the conventional train-test split. Hyperparameter tuning was performed through grid search to optimize model performance [27].

3. Result and Discussion

3.1. Result

The volume of glass bottled tea product sales for the period 2022 to 2024 is presented in the form of raw data. Data is collected weekly or monthly and serves as the basis for trend analysis and demand prediction model development. Table 1 shows the actual demand data used in this study.

Table 1. Data on drinks product demand for the years 2022–2024

Period	Product Price	Tren	Weather / Season	Consumer Per Capita Income	Population Count	Historical Sales
01/01/2022	58000	0	2	5910000	15588500	27981
09/01/2022	58000	0	2	5910000	15588500	27110
16/01/2022	58000	0	2	5910000	15588500	28110
23/01/2022	58000	0	2	5910000	15588500	28480
01/02/2022	58000	0	2	5910000	15588500	28973
06/02/2022	58000	0	2	5910000	15588500	27187
13/02/2022	58000	0	2	5910000	15588500	28210
20/02/2022	58000	0	2	5910000	15588500	27560
01/03/2022	58000	0	2	5910000	15588500	26345
06/03/2022	58000	0	2	5910000	15588500	27356
13/03/2022	58000	1	2	5910000	15588500	27015
20/03/2022	58000	1	2	5910000	15588500	28032
27/03/2022	58000	1	2	5910000	15588500	29120
01/04/2022	58000	1	1	5910000	15588500	27981
03/04/2022	58000	1	1	5910000	15588500	29110
10/04/2022	58000	1	1	5910000	15588500	27110
17/04/2022	58000	1	1	5910000	15588500	28760
24/04/2022	58000	1	1	5910000	15588500	29820
01/05/2022	58000	1	1	5910000	15588500	29851
08/05/2022	58000	1	1	5910000	15588500	29050
15/05/2022	58000	1	1	5910000	15588500	29800
22/05/2022	58000	1	1	5910000	15588500	29926
01/06/2022	58000	0	1	5910000	15588500	27503
05/06/2022	58000	0	1	5910000	15588500	27500
12/06/2022	58000	0	1	5910000	15588500	27250
19/06/2022	58000	0	1	5910000	15588500	28887
26/06/2022	58000	0	1	5910000	15588500	29400
01/07/2022	58000	0	1	5910000	15588500	28202
10/07/2022	58000	0	1	5910000	15588500	28104
17/07/2022	58000	1	1	5910000	15588500	29135
24/07/2022	58000	1	1	5910000	15588500	29845
01/08/2022	58000	1	1	5910000	15588500	29680
07/08/2022	58000	0	1	5910000	15588500	26600
14/08/2022	58000	0	1	5910000	15588500	26915
21/08/2022	58000	0	1	5910000	15588500	27790
01/09/2022	58000	0	1	5910000	15588500	26880
11/09/2022	58000	1	1	5910000	15588500	29730
18/09/2022	58000	1	1	5910000	15588500	29550
25/09/2022	58000	1	1	5910000	15588500	30160
01/10/2022	58000	0	2	5910000	15588500	28440

Period	Product Price	Tren	Weather / Season	Consumer Per Capita Income	Population Count	Historical Sales
09/10/2022	58000	0	2	5910000	15588500	28645
16/10/2022	58000	0	2	5910000	15588500	28055
23/10/2022	58000	0	2	5910000	15588500	27792
01/11/2022	58000	0	2	5910000	15588500	27811
06/11/2022	58000	0	2	5910000	15588500	29150
13/11/2022	58000	1	2	5910000	15588500	28960
20/11/2022	58000	1	2	5910000	15588500	28757
01/12/2022	58000	1	2	5910000	15588500	29244
11/12/2022	58000	1	2	5910000	15588500	29055
18/12/2022	58000	1	2	5910000	15588500	29728
25/12/2022	58000	1	2	5910000	15588500	29157
01/01/2023	58000	1	2	6250000	15386640	30068
09/01/2023	58000	0	2	6250000	15386640	28094
16/01/2023	58000	0	2	6250000	15386640	28615
23/01/2023	58000	0	2	6250000	15386640	29383
01/02/2023	58000	1	2	6250000	15386640	29843
06/02/2023	58000	1	2	6250000	15386640	28536
13/02/2023	58000	0	2	6250000	15386640	27328
20/02/2023	58000	0	2	6250000	15386640	27449
01/03/2023	60000	0	2	6250000	15386640	28345
06/03/2023	60000	1	2	6250000	15386640	29356
13/03/2023	60000	1	2	6250000	15386640	29615
20/03/2023	60000	1	2	6250000	15386640	29632
01/04/2023	60000	0	1	6250000	15386640	27020
10/04/2023	60000	1	1	6250000	15386640	29312
17/04/2023	60000	1	1	6250000	15386640	29250
24/04/2023	60000	1	1	6250000	15386640	30003
01/05/2023	60000	1	1	6250000	15386640	29012
08/05/2023	60000	1	1	6250000	15386640	29200
15/05/2023	60000	0	1	6250000	15386640	28100
22/05/2023	60000	0	1	6250000	15386640	27056
01/06/2023	60000	0	1	6250000	15386640	29843
12/06/2023	60000	0	1	6250000	15386640	28536
19/06/2023	60000	0	1	6250000	15386640	28328
26/06/2023	60000	0	1	6250000	15386640	27449
01/07/2023	60000	0	1	6250000	15386640	28940
10/07/2023	60000	0	1	6250000	15386640	28356
17/07/2023	60000	0	1	6250000	15386640	27615
24/07/2023	60000	0	1	6250000	15386640	28020
01/08/2023	60000	0	1	6250000	15386640	29632
07/08/2023	60000	0	1	6250000	15386640	28920
14/08/2023	60000	0	1	6250000	15386640	29886
21/08/2023	60000	0	1	6250000	15386640	28268
01/09/2023	60000	0	1	6250000	15386640	27100
11/09/2023	60000	0	1	6250000	15386640	27635
18/09/2023	60000	0	1	6250000	15386640	28947
25/09/2023	60000	0	1	6250000	15386640	28792
01/10/2023	60000	0	1	6250000	15386640	28843
09/10/2023	60000	0	1	6250000	15386640	29423
16/10/2023	60000	0	2	6250000	15386640	29053
23/10/2023	60000	0	2	6250000	15386640	27223
01/11/2023	60000	0	2	6250000	15386640	29440
06/11/2023	60000	0	2	6250000	15386640	28945
13/11/2023	60000	0	2	6250000	15386640	28055
20/11/2023	60000	0	2	6250000	15386640	28792

Period	Product Price	Tren	Weather / Season	Consumer Per Capita Income	Population Count	Historical Sales
01/12/2023	60000	0	2	6250000	15386640	29273
11/12/2023	60000	1	2	6250000	15386640	27800
18/12/2023	60000	1	2	6250000	15386640	28000
25/12/2023	60000	1	2	6250000	15386640	29927
01/01/2024	60000	1	2	6550000	15588500	29588
08/01/2024	60000	1	2	6550000	15588500	29560
15/01/2024	60000	0	2	6550000	15588500	27090
22/01/2024	60000	0	2	6550000	15588500	27997
01/02/2024	60000	0	2	6550000	15588500	29500
05/02/2024	60000	0	2	6550000	15588500	28455
12/02/2024	60000	0	2	6550000	15588500	29650
19/02/2024	60000	1	2	6550000	15588500	29074
01/03/2024	60000	1	2	6550000	15588500	28090
10/03/2024	60000	1	2	6550000	15588500	29105
17/03/2024	60000	1	2	6550000	15588500	29188
24/03/2024	60000	1	2	6550000	15588500	29608
01/04/2024	60000	0	1	6550000	15588500	29135
07/04/2024	60000	0	1	6550000	15588500	27800
14/04/2024	60000	0	1	6550000	15588500	28114
21/04/2024	60000	0	1	6550000	15588500	27951
01/05/2024	60000	1	1	6550000	15588500	29692
05/05/2024	60000	1	1	6550000	15588500	30024
12/05/2024	60000	1	1	6550000	15588500	29360
19/05/2024	60000	1	1	6550000	15588500	29695
26/05/2024	60000	1	1	6550000	15588500	30800
01/06/2024	60000	0	1	6550000	15588500	29545
10/06/2024	60000	0	1	6550000	15588500	28950
17/06/2024	60000	0	1	6550000	15588500	27880
24/06/2024	60000	0	1	6550000	15588500	28417
01/07/2024	60000	0	1	6550000	15588500	28451
07/07/2024	60000	0	1	6550000	15588500	29796
14/07/2024	60000	0	1	6550000	15588500	29060
21/07/2024	60000	0	1	6550000	15588500	29231
01/08/2024	60000	0	1	6550000	15588500	27850
11/08/2024	60000	0	1	6550000	15588500	29590
18/08/2024	60000	0	1	6550000	15588500	28521
25/08/2024	60000	0	1	6550000	15588500	29517
01/09/2024	60000	0	1	6550000	15588500	28179
08/09/2024	60000	0	1	6550000	15588500	28610
15/09/2024	60000	0	1	6550000	15588500	27800
22/09/2024	60000	0	1	6550000	15588500	28787

Note that the “Trend” column represents holiday or seasonal effects, where 0 = normal period and 1 = holiday/peak season. The “Weather/Season” column represents climate conditions (1 = rainy/cold season, 2 = dry/hot season).

To facilitate understanding of demand fluctuations over time, the data in Table 1 was then visualized in the form of a time series graph. This visualization aims to identify trend patterns, seasonality, and potential anomalies in product demand. Figure 1 presents the pattern of changes in demand for bottled tea products during the period from 2022 to 2024.

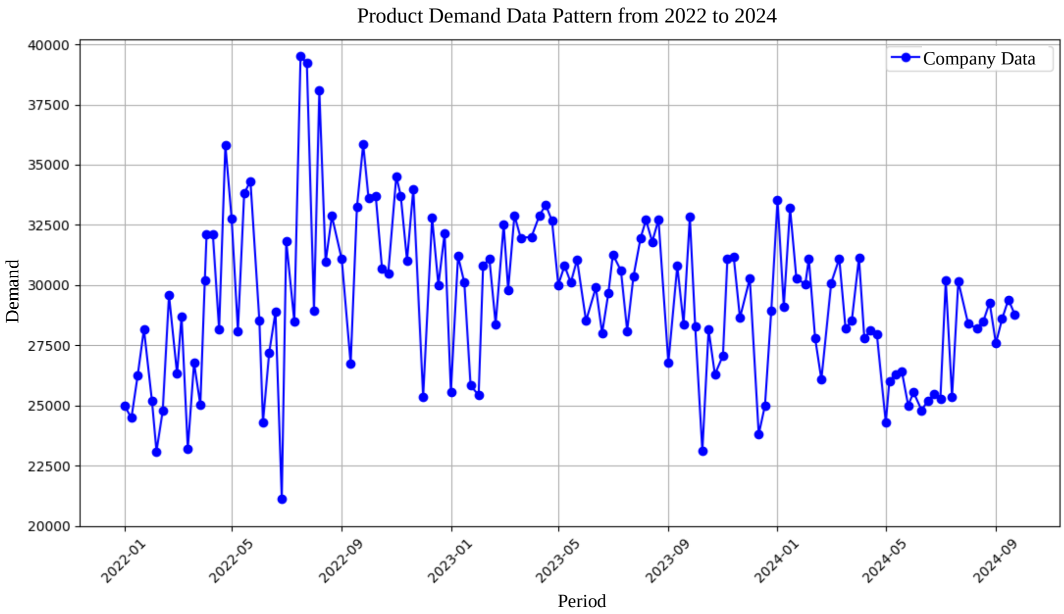

Figure 1. Product demand data pattern for the years 2022-2024

Figure 1 shows the demand pattern for beverage products from 2022 to 2024 based on the company's historical data. This graph illustrates demand fluctuations with an upward trend in mid-2022, peaking around the third quarter, before experiencing a decline and stabilization in the following years. The early 2022 period shows an upward trend, likely influenced by seasonal factors or increased consumption. However, after reaching the peak, there was a significant decline, followed by a more stable demand pattern with slight fluctuations throughout 2023 and 2024.

Based on the analysis of the data pattern, which tends to be linear and has a relatively small number of features (simple data), the chosen machine learning model is Random Forest. This model is used to predict weekly demand based on historical variables such as product prices, seasonal trends, weather, per capita income, and population size. The model displays the input, process, and output components used in the system. The following programming code is part of the implementation used in this research.

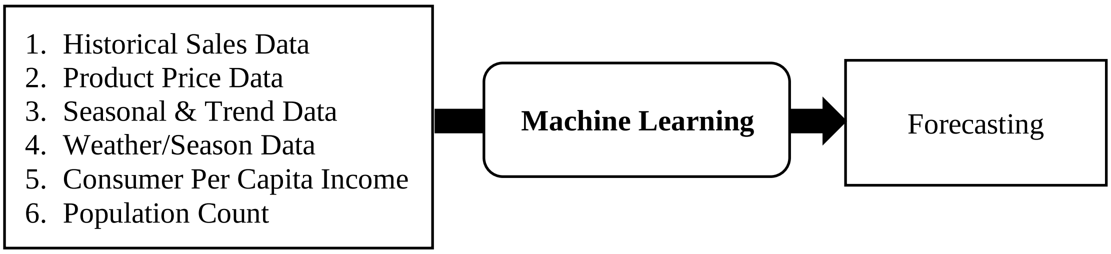

Figure 2. Model forecasting drinks product demand with machine learning

The three algorithms used are Random Forest Regressor, Gradient Boosting Regressor, and Prophet Time Series Model. The data used covers the period from January 2022 to September 2024, with supporting features such as product price, seasonal factors, weather/season, per capita income, population, and historical sales data.

To determine the accuracy of each model in predicting demand, a comparison was made between the forecast results and the actual weekly sales data during the testing period, from June to September 2024. The table below presents the actual sales data as well as the predictions from three machine learning models: Random Forest Regressor, Gradient Boosting Regressor, and Prophet Time Series Model. In addition to the predicted values, the prediction error rates in percentage form for each model are included to facilitate the analysis of the performance and precision of each algorithm.

Table 2. Comparison Results of Machine Learning Model Forecasting against Actual Sales Data

Period (Week)	Date	Actual Data	Random Forest	Gradient Boosting	Prophet	RF Error (%)	GBM Error (%)	Prophet Error (%)
1	01/06/2024	29545	28327,64	28440,43	28727,23	4,12	3,74	2,77
2	10/06/2024	28950	28475,75	28421,07	28543,76	1,64	1,83	1,40
3	17/06/2024	27880	28935,08	29483,19	28628,42	3,78	5,75	2,68
4	24/06/2024	28417	28727,21	29062,65	28871,63	1,09	2,27	1,60
5	01/07/2024	28451	28908,85	29292,93	29017,20	1,61	2,96	1,99
6	07/07/2024	29796	28743,05	28736,39	28956,80	3,53	3,56	2,82
7	14/07/2024	29060	28610,43	28904,25	28877,48	1,55	0,54	0,63
8	21/07/2024	29231	28375,11	28527,54	29024,99	2,93	2,41	0,70
9	01/08/2024	27850	28476,08	28696,95	29332,49	2,25	3,04	5,32
10	11/08/2024	29590	28498,27	28916,54	29393,46	3,69	2,28	0,66
11	18/08/2024	28521	28737,65	28834,08	28888,46	0,76	1,10	1,29
12	25/08/2024	29517	28397,48	28926,91	28013,98	3,79	2,00	5,09
13	01/09/2024	28179	28724,26	28997,45	27399,62	1,93	2,90	2,77
14	08/09/2024	28610	28280,32	28153,77	27548,88	1,15	1,59	3,71
15	15/09/2024	27800	28746,00	28880,21	28373,58	3,40	3,89	2,06
16	22/09/2024	28787	28136,87	27803,67	29290,80	2,26	3,42	1,75

Table 2 displays the results of forecasting the weekly sales volume of bottled tea products from June 1 to September 22, 2024. The "Actual Data" column shows the actual sales volume recorded by the company, while the "Random Forest", "Gradient Boosting" and "Prophet" columns represent the prediction results of each respective machine learning model.

In addition, the prediction error is also calculated as a percentage to measure how far the deviation of the prediction results from the actual data. The error value is calculated using the absolute percentage error formula for each model. The smaller the error value, the more accurate the predictions made by the model.

To provide a clearer visual representation, the comparison results between the predictions and the actual data are presented in the following graph. This graph shows the tendency of each model in following the actual demand pattern and identifies which model is closest to the actual sales realization.

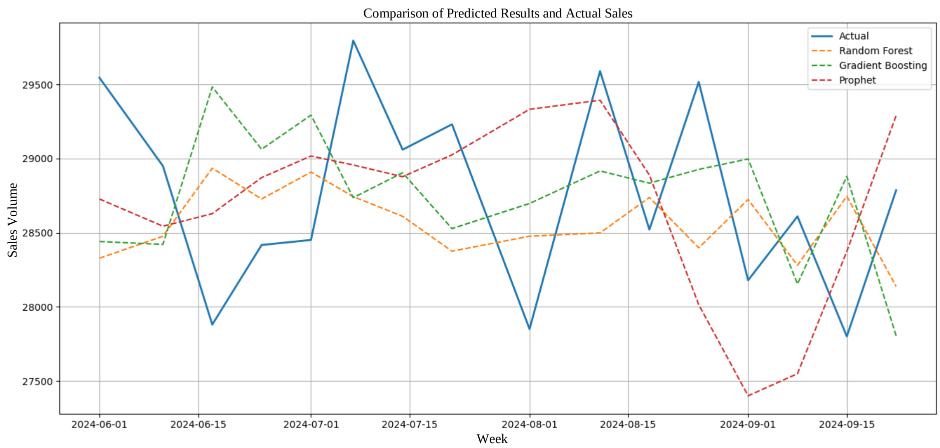

Figure 3. Comparison of predicted results and actual weekly sales data by three machine learning models

Figure 3 shows a comparison between actual sales volume and the predictions generated by three machine learning algorithms, namely Random Forest Regressor, Gradient Boosting Regressor, and Prophet Time Series Model, during the period from June to September 2024. The blue line represents the actual data, while the other dashed lines show the predictions of each model. Prophet shows the prediction pattern that most closely aligns with the actual trend, particularly in capturing sharp weekly fluctuations, compared to the Random Forest and Gradient Boosting models.

In addition to comparing weekly prediction results with actual data, a quantitative evaluation of each model's performance was also conducted using four evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These four metrics are used to measure the magnitude of prediction errors in absolute, squared, square root, and percentage forms. The following table presents the overall performance evaluation results of each model based on the test set data from June to September 2024. The model with the smallest error value is considered to have the best performance in forecasting product demand.

Table 3. Results of the forecasting model performance evaluation

Model	MAE (unit)	MSE (unit)	RMSE (unit)	MAPE (%)
Random Forest Regressor	712.39	611,061.30	781.70	2.47
Gradient Boosting	775.30	716,658.37	846.56	2.70
Prophet	668.02	602,993.03	776.53	2.33

From the evaluation results, it was found that the Prophet model outperformed the other two models. Prophet recorded the lowest MAE and RMSE values, at 668.02 and 776.53 units respectively, as well as a MAPE of 2.33%. This MAPE value indicates a very low prediction error rate and still falls within the very good category for demand forecasting needs.

The Random Forest Regressor model ranks second, with an MAE of 712.39 units and an RMSE of 781.70 units. Although its performance is quite good and close to Prophet, this model is less optimal in capturing seasonal patterns and long-term trends that are the main characteristics of beverage sales data. Meanwhile, the Gradient Boosting Regressor shows the lowest performance among the three models. Nevertheless, the MAPE produced is still within an acceptable accuracy range (2.70%), indicating that this model remains usable, but is less efficient compared to Prophet and Random Forest.

These results indicate that time-series-based approaches like Prophet are superior in handling sales data with seasonal and fluctuating characteristics. Prophet automatically identifies and accommodates annual, weekly trends, and holiday effects without requiring the creation of complex manual features.

While the Mean Absolute Percentage Error (MAPE) indicates Prophet's superiority, further statistical testing (paired t-test on forecast errors) suggests that the differences between Prophet and Random Forest are practically significant but not statistically significant at $\alpha = 0.05$. This means that although Prophet consistently shows slightly better accuracy, Random Forest remains competitive.

3.2. Discussion

The results indicate that the Prophet model provides the highest forecasting accuracy for weekly beverage sales at PT. XYZ, with MAE of 668.02, RMSE of 776.53, and MAPE of 2.33%. Its ability to automatically capture seasonal and long-term trends makes Prophet particularly effective in reflecting actual sales fluctuations, as also shown by the close alignment between its prediction curve and the historical data. [28].

Nevertheless, the dataset used in this study is multivariate, incorporating factors such as product price, seasonal events, weather, per capita income, and population. In this context, the Random Forest Regressor offers methodological advantages because it can process diverse input features and evaluate their contribution to sales predictions. With a MAPE of 2.47%, Random Forest delivers competitive accuracy while also revealing the relative influence of external drivers. Feature importance analysis showed that product price and weather conditions were the most impactful predictors, followed by per capita income and holiday indicators, underscoring the relevance of multivariate modeling [29].

By contrast, the Gradient Boosting Regressor recorded the lowest accuracy (MAPE 2.70%) and demonstrated higher sensitivity to overfitting, particularly when handling seasonal data with high variability. Although still within an acceptable range, its lower robustness makes it less suitable for long-term operational use.

This study contributes novelty by providing a comparative evaluation of three widely used machine learning algorithms while integrating multiple external variables into a single demand forecasting framework. The application of Prophet in forecasting FMCG products in Indonesia is also still relatively rare, offering additional methodological contributions to the literature on supply chain forecasting.

In practice, Prophet can be applied for short-term operational planning, especially in anticipating seasonal peaks, while Random Forest is more suitable for medium- to long-term strategic decision-making such as pricing scenarios, economic simulations, and demographic projections. A hybrid approach combining both models offers a promising pathway toward building an integrated forecasting system at PT. XYZ to improve supply chain performance.

It should also be noted that the dataset (2022–2024) may reflect unusual events such as post-pandemic recovery effects and atypical weather patterns, which could influence the results and limit generalizability. Future research may expand the dataset and test additional algorithms to enhance robustness and applicability.

4. Conclusion

This research aims to compare the performance of three machine learning algorithms, namely Random Forest Regressor, Gradient Boosting Regressor, and Prophet Time Series Model, in predicting the weekly sales volume of beverage products at PT. XYZ by utilizing multivariate data. Based on the evaluation results using MAE, MSE, RMSE, and MAPE metrics, the Prophet model shows the highest accuracy level with a MAPE value of 2.33%, followed by Random Forest with 2.47%, and Gradient Boosting with 2.70%.

Although Prophet has advantages in terms of accuracy and the ability to capture seasonal patterns, this model is limited to processing time series data and does not fully accommodate the influence of non-temporal external variables. On the other hand, the Random Forest Regressor has an advantage in handling multivariate data because it can process various types of input variables such as product prices, weather, per capita income, and population, and provides flexibility in analyzing the influence of each feature on sales predictions.

Considering the data structure and the need for medium to long-term predictions in the context of operational and strategic decision-making, the Random Forest Regressor is deemed a more relevant and adaptive model. Meanwhile, Prophet can still be used as a short-term prediction tool focusing on seasonal fluctuations. The combination of the two models can serve as a strong hybrid approach to build a more accurate, comprehensive, and adaptive demand forecasting system to support the optimization of PT. XYZ's supply chain.

In practical terms, the implementation of this hybrid forecasting system should be integrated into the company’s existing SCM/ERP software, enabling automated updates of demand predictions. Prophet can be used for short-term operational planning (weekly/monthly adjustments), while Random Forest supports medium-to-long-term strategic decisions, such as production capacity planning and distribution route optimization. By embedding the hybrid system into daily operations, PT. XYZ can reduce stock imbalances, optimize inventory costs, and enhance resilience in responding to demand fluctuations.

References

[1] L. Li, T. Chi, T. Hao, and T. Yu, “Customer Demand Analysis Of The Electronic Commerce Supply Chain Using Big Data,” *Ann. Oper. Res.*, vol. 268, pp. 113–128, 2018.

[2] T. Hines, *Supply Chain Strategies: Demand Driven and Customer Focused*. Routledge, 2024.

[3] C. W. Chase, *Demand-driven Forecasting: A Structured Approach To Forecasting*. John Wiley & Sons, 2013.

[4] W. Albrecht and M. Steinrücke, “Coordinating Continuous-time Distribution and Sales Planning of Perishable Goods with Quality Grades,” *Int. J. Prod. Res.*, vol. 56, no. 7, pp. 2646–2665, 2018.

[5] S. M. Akid, A. Ishak, and S. Sinulingga, “Jurnal Sistem Teknik Industri The Design Of A Demand Forecasting Model Of Glass Bottled Tea Products With Machine Learning Approach,” *Jurnal Sistem Teknik Industri*, vol. 27, no. 2, pp. 57–65, 2025.

[6] B. Keleş, P. Gómez-Acevedo, and N. I. Shaikh, “The Impact of Systematic Changes in Weather on The Supply and Demand of Beverages,” *Int. J. Prod. Econ.*, vol. 195, pp. 186–197, 2018.

[7] E. Beard, J. Brown, R. West, E. Kaner, P. Meier, and S. Michie, “Associations Between Socio-economic Factors and Alcohol Consumption: A Population Survey of Adults in England,” *PLoS One*, vol. 14, no. 2, p. 209, 2019.

[8] N. Fose, A. R. Singh, S. Krishnamurthy, M. Ratshitanga, and P. Moodley, "Empowering Distribution System Operators: A Review of Distributed Energy Resource Forecasting Techniques," *Heliyon*, vol. 10, no. 15, 2024.

[9] P. K. Tripathi, C. K. Singh, R. Singh, and A. K. Deshmukh, "A farmer-centric agricultural decision support system for market dynamics in a volatile agricultural supply chain," *Benchmarking: An International Journal*, vol. 30, no. 10, pp. 3925–3952, 2023.

[10] N. B. Bhavikatta, "AI-Driven Inventory Optimization in Supply Chains: A Comprehensive Review on Reducing Stockouts and Mitigating Overstock Risks," *Journal of Computer Science and Technology Studies*, vol. 7, no. 7, pp. 01–13, 2025.

[11] S. B. Dalimunthe, S. Sinulingga, and R. Ginting, "JSTI Jurnal Sistem Teknik Industri Implementation Of Machine Learning in Demand Forecasting: A Review of Method Used in Demand Forecasting with Machine Learning," *Jurnal Sistem Teknik Industri (JSTI)*, vol. 25, no. 1, p. 2023, 2023, doi: 10.32734/jsti.

[12] A. A. Nasution, N. Matondang, and A. Ishak, "Inventory Optimization Model Design with Machine Learning Approach in Feed Mill Company," *Jurnal Sistem Teknik Industri*, vol. 24, no. 2, pp. 254–272, 2022.

[13] E. El-Shafeiy, M. Alsabaan, M. I. Ibrahem, and H. Elwahsh, "Real-time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique," *Sensors*, vol. 23, no. 20, p. 8613, 2023.

[14] S. Balasubramanian, M. Vodenicharova, and C. Srinu, "From Data to Decisions Leveraging Machine Learning in Supply-Chain Management," *Journal of Propulsion Technology*, vol. 44, no. 4, pp. 4218–4225, 2023.

[15] N. Sembiring, W. B. Sari Harahap, F. A. Khalif, K. Chandra, and C. Herlim, "Forecasting Demand for Donuts Using a Machine Learning Approach," *Proceedings - ELTICOM 2024: 8th International Conference on Electrical, Telecommunication and Computer Engineering: Tech-Driven Innovations for Global Organizational Resilience*, pp. 159–164, 2024, doi: 10.1109/ELTICOM64085.2024.10864725.

[16] M. Zohdi, M. Rafiee, V. Kayvanfar, and A. Salamiraad, "Demand Forecasting Based Machine Learning Algorithms on Customer Information: An Applied Approach," *International Journal of Information Technology*, vol. 14, no. 4, pp. 1937–1947, 2022.

[17] G. B. R. Matahurila, F. P. Putra, L. F. Himmah, A. Maulana, S. Riyadi, and M. Y. R. Pandin, "Analisis Pengaruh Forecasting Demand Terhadap Efisiensi Manajemen Persediaan," *As-Syirkah: Islamic Economic & Financial Journal*, vol. 3, no. 3, pp. 1475–1489, 2024.

[18] N. Walliman, *Research Methods: The Basics*. Routledge, 2021.

[19] BMKG, "Data suhu dan curah hujan tahunan Indonesia tahun 2022–2024," Badan Meteorologi, Klimatologi, dan Geofisika. Accessed: Oct. 05, 2024. [Online]. Available: https://www.bmkg.go.id

[20] BPS, "Pendapatan per Kapita dan Jumlah Penduduk Indonesia 2022–2024," Badan Pusat Statistik. Accessed: Oct. 05, 2024. [Online]. Available: https://www.bps.go.id

[21] I. Winarto, R. H. A. Shiddieqy, and S. T. Rivai Wardhani, *Manajemen Operasi, Perencanaan Produksi, dan Pengendalian Persediaan*. Penerbit Andi, 2025.

[22] E. Arif, I. Suherman, and A. P. Widodo, *Revolusi Prediksi Saham: Pemanfaatan Machine Learning dan Analisis Sentimen dalam Dunia Keuangan*. Greenbook Publisher, 2025.

[23] M. A. M. R. Setiawan, K. Kusrini, and A. D. Hartono, "Menggunakan Metode Machine Learning Untuk Memprediksi Nilai Mahasiswa Dengan Model Prediksi Multiclass," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 190–204, 2025.

[24] N. Sunendar, H. P. Putro, and R. Hesananda, "Prediksi Penjualan Aerosol Menggunakan Algoritma ARIMA, LSTM Dan GRU," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 4, no. 1, pp. 113–126, 2025.

[25] M. A. Purnama, J. Ramadhani, Y. S. Anugraha, and L. Efrizoni, "Perbandingan Performa Algoritma Random Forest Dan Gradient Boosting Dalam Mengklasifikasi Churn Telco," *Techno.com*, vol. 23, no. 3, 2024.

[26] R. Kaestria and E. F. Himmah, "Implementasi Bahasa Pemrograman Python untuk Path analysis," *Jurnal Komputasi*, vol. 11, no. 2, pp. 105–117, 2023.

[27] T. Hassan, M. Majeed, and M. Ahmad, "Optimizing SVM Performance through Combinatorial Hyperparameter Tuning and Model Selection," *International Journal Bioautomation*, vol. 29, no. 2, p. 117, 2025.

[28] A. Wiejaya and I. Fenriana, "Prediksi Harga Saham Top 10 NASDAQ dengan Time Series Prophet," *bit-Tech*, vol. 7, no. 2, pp. 252–262, 2024.

[29] E. R. Susanto and A. Cahyana, "Penerapan Algoritma XGBoost untuk Prediksi Diabetes: Analisis Confusion Matrix dan ROC Curve," *Fountain of Informatics Journal*, vol. 10, no. 1, pp. 40–50, 2025.