Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 3118~3125
e-ISSN: 2548-6861

Predictive Analytics of Food Retail Seasonal Trends with Advanced Forecasting Modeling

Anita Sindar Sinaga [1]*, Dameria Esterlina Br Jabat [2], Amalia Rossa [3]*, Dini Auliah [4]***
* Information Technology, STMIK Pelita Nusantara Medan
** Digital Business, STMIK Pelita Nusantara Medan
haito_ita@yahoo.com [1], sijabatdame@gmail.com [2], amalia11@gmail.com [3], aulia21@gmail.com [4]

Article Info

Article history:

Received 2026-06-02
Revised 2026-06-17
Accepted 2026-06-22

Keyword:

Sales Forecasting,
Food Retail,
Predictive Analytics,
Historical Analysis,
Forecast.
.

ABSTRACT

Food sales in food retailers generally increase on certain days. Three food categories served as data sources in this study: staple foods, ready-to-eat foods, and dairy products. Predictive analysis of seasonal trends in food retailers shows that macroeconomic factors, seasonal patterns, and religious holiday indicators play a significant role in shaping sales. Staples is the highest-revenue category, while frozen foods has the lowest volume of the three. Each highlighted sector, including dairy, is expected to experience a measurable increase in turnover over the coming period. All models exhibit varying accuracy in predicting 2026 sales compared to actual 2025 sales, evaluated using MAPE, RMSE, and MAE for key products. Moving Average and LSTM tend to be conservative, while ETS and ARIMA are more optimistic but remain limited by limited data. Random Forest also struggles to capture complex relationships. Prophet stands out for its ability to incorporate exogenous variables and handle seasonality, although caution is needed when interpreting future values. The MAPE values ranged from 1.89% to 5.34%, indicating excellent predictive accuracy, as MAPE values below 10% are generally considered high accuracy. The low RMSE and MAE values also indicate a relatively small difference between the 2026 prediction and the actual 2025 values.

I. INTRODUCTION

According to the July 2023 United States Department of Agriculture (USDA) report, "Indonesia: Food Retail," Indonesia's food retail sales reached US$100 million throughout 2022. The food retail industry faces highly dynamic demand, influenced by seasonal patterns such as holidays, national holidays, and lifestyles. Consumer tastes are shifting, and staple foods are no longer limited to rice. Consumers are comfortable purchasing essentials at retail outlets. They prioritize convenience, hygienic food, and ready-to-eat products. Some food items experience high demand around religious holidays. Unsatisfactory customer service, such as slow responses to complaints, can directly damage a retail store's reputation. Consumer demand trends are analyzed to ensure optimal service during peak demand while maintaining price stability. Prophet's ability to capture seasonality, trends, and holidays aligns perfectly with the dynamic nature of food demand. Integration of anomaly detection techniques can be explored to identify unexpected demand fluctuations, potentially caused by external factors or emerging trends. In this Predictive Analytics of Food Retail Seasonal Trends with Forecasting Modeling study, consumer purchasing patterns are identified according to seasonal periods and lifestyles so that they can be used to optimize inventory management, reduce the risk of overstocking or understocking, and improve product distribution efficiency [1].

The visual characteristics of the Moving Average (MA) are very simple and focus on smoothing short-term trends. Exponential Smoothing (ETS/Holt-Winters) is able to capture trend factors and seasonal patterns. Forecasting handles various types of data, from large-scale quantitative historical data (numbers/statistics) to qualitative data (expert opinions and subjective trends). Moving Average (MA) calculates the average of data values over a specific time

period sequentially [2]. Exponential Smoothing (ETS) works by giving greater weight to recent sales data and exponentially decreasing weight to older data.

To generate accurate seasonal trend predictions for food retail from inconsistent historical data, algorithm-based interpolation or imputation techniques are used to logically fill in the gaps in sales records. To capture complex and non-linear seasonal patterns, Machine Learning algorithms such as Random Forest are used to process multiple external variables (such as holidays) without the risk of overfitting, or can be combined with Hybrid approaches [3]. The courage to test the lower limits of these algorithms capabilities provides new theoretical insights into the behavior of machine learning and deep learning models in the face of retail data scarcity [4]. Predictive analytical research on seasonal trends in food retail using forecasting modeling is the availability of a predictive model that is able to accurately identify consumer demand patterns according to seasonal periods, so that it can be used to optimize inventory management, reduce the risk of overstocking or understocking, and improve product distribution efficiency [5]. The main objective is to conduct predictive analysis of seasonal trends in food retail in Indonesia using various forecasting models.

II. METHOD

The entire process begins with Data Preparation to clean and align the raw data format for consistency, followed by Data Exploration (EDA) to explore growth patterns and seasonal spikes in food types [6]. Through Data Splitting, the dataset is divided into training and testing sections before entering the Forecasting Model stage [7].

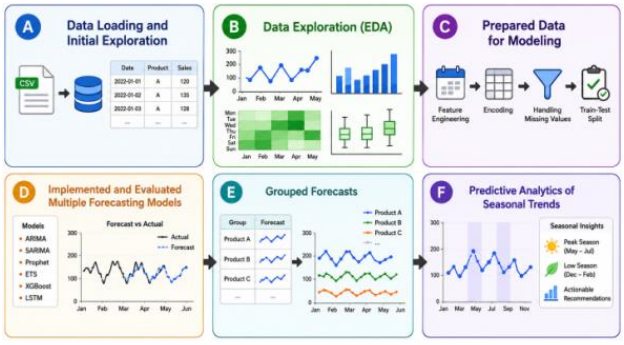

Figure 1. Scope of Research

The 2026 prediction execution is carried out in the Model Application and Forecasting stage using a combination of six models with diverse approaches, ranging from Linear Regression as the basis, a conservative Moving Average (MA), Exponential Smoothing (ETS) that optimistically captures linear trends, and ARIMA [8]. The projection results are then grouped into specific market segments through Categorized Forecasting to identify potential food subsectors supported by Correlation Analysis to measure the influence of external variables such as holidays or lifestyles [9].

A. Data Loading and Initial Exploration

Data source from Bapanas Macro Sectoral & Indonesian FMCG Statistics (2018-2025) also from the integration of Bapanas & One Data Indonesia's macro-commodity strategic data. Generating descriptive statistics for staple food categories such as rice, cooking oil, sugar, and instant noodles remains a key category with consistent demand. Frozen food categories, such as chicken nuggets, sausages, and meatballs, are particularly popular due to their convenience. Dairy products such as UHT milk, yogurt, and cheese show strong demand growth. External factors predict retail trends in Indonesia, as the market is highly dynamic, including significant cultural and religious factors and lifestyle changes [10]. Temporal Aggregation was used to convert transaction frequency data into consistent time intervals [11]. Data was also grouped by product category (Staples, Frozen Food, Dairy) to see monthly performance trends over the period. 2018–2025.

B. Data Exploration (EDA)

Exploratory Data Analysis (EDA) is essential for understanding dataset characteristics, identifying patterns, detecting anomalies, and verifying assumptions [12]. It examines summary statistics, unique values, and sales distributions. This initial data exploration helps confirm data types, identify sales ranges and distributions, inform model selection and feature engineering steps, and examine how sales vary across potential product categories [13]. It examines unique values and counts for categorical columns. Average sales per category are displayed.

C. Prepared Data for Modeling

This includes collecting and integrating detailed historical sales data along with relevant exogenous variables such as holiday calendars, conducting in-depth exploratory data analysis (EDA) to understand seasonal trends and patterns, developing and evaluating advanced time series models (such as Exponential Smoothing, LSTM, ARIMA, and Random Forest), generating sales forecasts for future periods with confidence intervals, and finally, providing actionable insights and recommendations for business strategies based on the refined forecasts and identified seasonal pattern. The crucial initial process for correcting or removing incorrect, corrupted, duplicate, or incomplete data involves a series of actions taken to ensure the data is ready for use and of good quality. This includes addressing missing values and outliers [14]. Data modeling creates a blueprint for how data will be stored, accessed, and used within the system. This involves identifying key data objects, their attributes, and their relationships [15]. This process ensures that data is structured logically and efficiently, supporting current and future business needs. For the 2026 sales predictions, various forecasting models were applied a Linear Regression baseline, a Moving Average (MA) for conservative two-point averages; Exponential Smoothing (ETS) for optimistic, linear growth extrapolations, ARIMA all of which yielded

cautious projections heavily constrained by the limited historical data [16].

D. Implemented and Evaluated Forecasting Models

Various forecasting models were applied to predict 2026 sales. Moving Average (MA) provides a conservative estimate by averaging the last two data points, while Exponential Smoothing (ETS) often shows a more optimistic extrapolation, capturing a linear growth trend. ARIMA was applied for time series forecasting, with parameters chosen to be minimal due to data limitations. The Holt-Winters time series model is a Holt variant for linear trends [17]. MAPE measures accuracy as a percentage, which is easy to interpret. Lower values indicate higher accuracy. RMSE measures the magnitude of the average error. It gives greater weight to larger errors, and its units are the same as the predicted variable. Performance metrics on the test set are used MAE, RMSE, MAPE. Performance metrics on the entire dataset use MAPE. MAPE Value Limits Based on the classic evaluation criteria from Lewis (1982) which is used internationally in supply chain management and statistics, a model is said to be successful based on the percentage of error [18]. The Limits of MAE, MSE, and RMSE Values of these three matrices do not have fixed limit numbers (such as having to be below 1 or below 10).

E. Grouped Forecasts

Grouped forecasts, also known as hierarchical forecasting, predict time series data at various levels of aggregation [19]. To group data by product category, apply a forecasting model based on the monthly sales categories of Staples total sales, Frozen Food total sales, and Dairy Products for 2018–2025. The monthly sales chart grouped by category for 2018 –2025 shows the monthly sales trends for Dairy Products, Frozen Food, and Staples, making it easy to see an overview of performance over a given period. Aggregate forecasts for 2026 are grouped by category to highlight differences across sectors. The results are then averaged across multiple models to provide a balanced overall projection [20].

F. Predictive Analytics of Seasonal Trends

This predictive analytics integration aims to optimize inventory management by reducing the risk of overstocking and understocking, improve distribution efficiency by improving the product distribution chain based on consumer purchasing patterns, and maintain price stability by ensuring optimal service during peak demand periods without compromising price stability. SARIMA (Seasonal Autoregressive Integrated Moving Average) is a time series forecasting method that predicts future values based on historical data, specifically taking into account trends and recurring (seasonal) patterns. This method is an extension of the conventional ARIMA. SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) is a statistical method for forecasting time series data. This method works by combining past trends (AR and MA), data stabilization processes (I), recurring/seasonal patterns (S), and the influence of external variables (X). In this study, SARIMAX is used for time series with clear trends and seasonal components [18].

III. RESULTS AND DISCUSSION

A. Data Preparation for Model Selection

Data source from BAPANAS Macro Sectoral & Indonesian FMCG Statistics. The data used are historical monthly sales figures (2018–2025) for major categories such as Staples (Rice, Cooking Oil), Dairy, and Frozen Food. Food types are grouped into 3 categories, namely, Dairy Products, Frozen Food, and Staples. Detecting outliers using the IQR Method by applying the Interquartile Range (IQR) method to detect outliers in numeric columns identified from df_integrated. This method is robust to skewed distributions. Visualizing outliers for numerical columns Sales. The Middle of the Range (IQR) of Sales Distribution shows that 50% of the sales data falls within a fairly narrow range, between 0.3 and 1.0. The median, shown within the box, indicates the middle sales value is around 0.45, which is consistent with the highest-frequency peak we saw in the previous histogram.

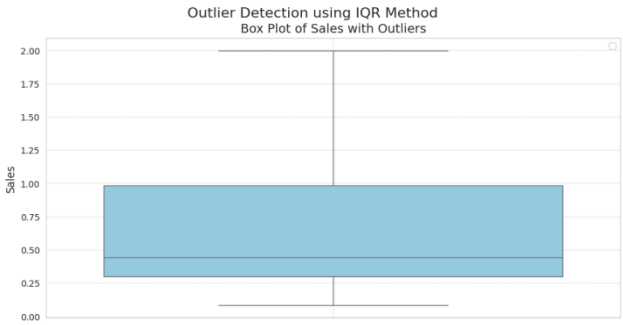

Figure 2. Outlier Detection using IQR Method

Unusual data points that may affect the model. The outlier detection process in this research provided insight that cultural and religious factors (such as Eid al-Fitr and Christmas) often create seasonal spikes that appear to be outliers but are actually recurring patterns. By separating anomalies from trends and seasonality through time decomposition, researchers can ensure that models like ETS and ARIMA remain stable and accurate (MAPE <1%) even when working with very limited historical data. This validation ensures that the data is logically structured to support business decisions such as inventory management and distribution efficiency.

To facilitate the selection of an appropriate time series model, the dataset is first restructured. This format is crucial for applying time series analysis techniques. This step helps inform decision-making for subsequent modeling by examining summary statistics, unique values, and sales distributions. Numeric data refers to data that can be expressed as numbers. Unique values and counts for categorical columns. Before categorical data can be used as

input to a machine learning model, it must first be converted into numeric data.

This series of processes, especially the prepared data for modeling section, is a crucial step in producing a logically and efficiently structured data blueprint to support 2026 sales predictions. This process of converting categorical data into numeric representation is known as encoding. Nine product names were obtained rice, instant noodles, cooking oil, chicken eggs, chicken nuggets, sausage, processed meatballs, uht milk, yogurt, and cheese. Unique values are Rupiah, Tons, and Packs. Average Sales per Category that is Staples (333.196057 sales), Dairy (161.617024 sales), and Frozen Food (105.018278 sales). This initial data exploration helps confirm data types, identify sales ranges and distributions, and see how sales vary across product categories.

B. Exploratory Data Analysis (EDA) Phase

The Sales Distribution graph is part of the Exploratory Data Analysis (EDA) phase. The Kernel Density Estimate (KDE) curve (the curved blue line) is used to visualize the frequency and density distribution of sales data. Outlier and Anomaly Detection using The Exploratory Data Analysis (EDA) stage is used specifically to detect anomalies and outliers before the forecasting stage is carried out using the Kernel Density Estimate (KDE) curve on the sales distribution to visualize data frequencies and identify unusual data points that can affect model performance. EDA also serves to verify assumptions and ensure the sales distribution range is appropriate before entering the feature engineering stage.

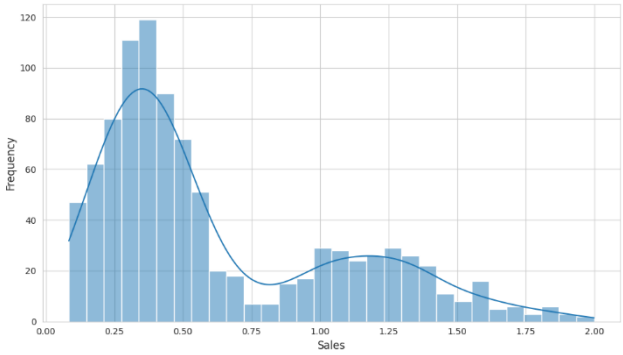

Figure 3. the Sales Distribution 2018 – 2025

This graph is most likely used to provide an overview of the historical sales data profile before further modeling is carried out. The X-axis shows the sales value (in trillions of Rupiah or millions of tons/packs), while the Y-axis shows the frequency of occurrence of that value. This analysis helps understand the characteristics of the dataset, detect anomalies or outliers, and determine the most appropriate forecasting model for predicting sales trends in the food retail sector.

Exploratory Data Analysis (EDA) is used to understand the characteristics of integrated datasets, identify hidden patterns, and detect anomalies before forecasting. Trend, Seasonal, and Fluctuation Analysis dissects historical data into key components to isolate seasonal effects. Through time series plots, consistent upward or downward movements from 2018 to 2025 are identified. Products such as Rice (Staples), Chicken Nuggets (Frozen Food), and UHT Milk (Dairy Products) show a steady positive growth trend during this period. Seasonal patterns capture recurring spikes in demand during specific periods, such as religious holidays (Eid al-Fitr and Christmas) and national holidays. Analytical models break down the data into trend, seasonal, and residual (irregular fluctuations) to identify external variables such as lifestyle changes or the influence of retail discounts that influence consumer behavior. Average yearly sales per category 2018-2025 that is Dairy Products as much as 20.202128, Frozen Food 13.127285, Staples 41.649507 (in trillions of Rupiah or millions of tons/packs).

C. Sales Trends for Model Selection

Modeling occurs in two main phases: a training phase using 80% historical data, where models like Moving Average (MA) method, Forecasting with Exponential Smoothing (ETS), including Holt-Winters, ARIMA learn to recognize sales patterns from a bimodal distribution (low peaks around 0.4 and high peaks around 1.2) and relate them to variables like seasonality or holidays; and a testing phase using 20% of the data that has never been seen before, serving as a generalization test to ensure the model can predict both low sales periods and the spikes in the second peak.

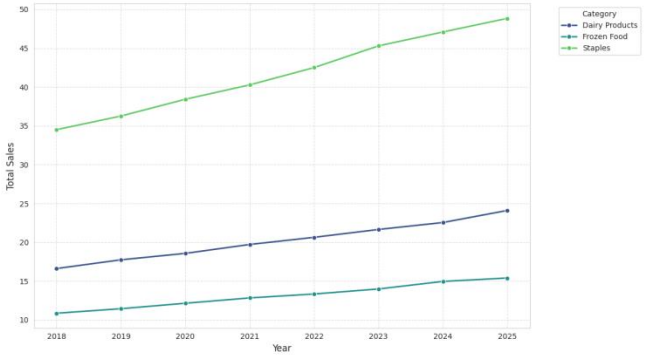

Figure 4. Total Sales by Category 2018 – 2025

This split ensures that sales projections for future periods are based on the logic of stable patterns, rather than simply rigidly following historical data. The bimodal sales distribution with two frequency peaks influenced the choice of forecasting model because it required an algorithm capable of handling non-normal data and multiple behaviors. Random Forest was chosen because of its nonparametric nature, its ability to divide the data according to different conditions e.g., weekdays vs. holiday seasons, and its ability to reduce the influence of outliers due to rightward skewness.

ETS and ARIMA were used to capture seasonal patterns and trends, with ETS highlighting growth momentum during

high sales peaks, while ARIMA managed autocorrelation in fluctuations. Moving averages were added to dampen volatility due to frequency dips in the 0.75–0.90 range, allowing the model to focus more on the main pattern rather than noise. Thus, the combination of ensemble learning (Random Forest) and seasonal decomposition (ETS/ARIMA) proved to be a more appropriate strategy than a single linear model in mapping the two modes of retail sales. Before creating a forecasting model, aggregate based on the objective column and time series. There are two ways to categorize forecasts a bottom-up approach (per category) and a top-down approach (aggregation). Temporal aggregation converts the frequency of transaction or event data into consistent time intervals in the Total Monthly Sales for January - June 2018 Table.

TABLE I
TOTAL MONTHLY SALES JANUARY - JUNI 2018

Category	Date	Sales	Unit Data
2018-01-01	Dairy Products	1.392756	Billion IDR
2018-01-01	Frozen Food	0.918761	Billion Packs
2018-01-01	Staples	3.010717	Billion IDR
2018-02-01	Dairy Products	1.366197	Billion IDR
2018-02-01	Frozen Food	0.872612	Billion Packs
2018-02-01	Staples	2.765318	Billion IDR
2018-03-01	Dairy Products	1.274694	Billion IDR
2018-03-01	Frozen Food	0.821788	Billion Packs
2018-03-01	Staples	2.792395	Billion IDR
2018-04-01	Dairy Products	1.397828	Billion IDR
2018-04-01	Frozen Food	1.000646	Billion Packs
2018-04-01	Staples	3.053599	Billion IDR
2018-05-01	Dairy Products	1.558931	Billion IDR
2018-05-01	Frozen Food	1.050250	Billion Packs
2018-05-01	Staples	3.161221	Billion IDR
2018-06-01	Dairy Products	1.766092	Billion IDR
2018-06-01	Frozen Food	1.016042	Billion Packs

The percentage difference between top-down and bottom-up Forecasts metric is used to measure how consistent predictions made at the aggregate level (the total category) are compared to the total of individual-level predictions (each product). A small difference indicates that the model is stable across different levels of the data hierarchy. The Bottom-Up Category Forecasts projection results for 2026 show that the Dairy Products (16.53) and Staples (16.14) categories are predicted to dominate sales volume, far surpassing the Frozen Food category (5.48). These projections use the Prophet model's aggregation method at the product level, meaning the prediction for each category is obtained by summing the individual forecasts from the various products below it to more accurately capture the details of seasonal trends.

Combine the target column with grouping so the model can learn the historical patterns of each entity specifically in the Dairy Product Category Sales Projection Table in 2026. For more powerful and actionable business insights, combine the sales figures per individual unit, total sales in billions of IDR, and total sales in billions of packs. Combine the predictions from each tree to produce the final output. In numerical prediction (regression) tasks such as sales forecasting, the aggregation process is usually done by averaging the predictions from all trees. If each tree provides a prediction. Combine the target column with grouping so the model can learn the historical patterns of each entity specifically in the Dairy Product Category Sales Projection Table in 2026. For more powerful and actionable business insights, combine the sales figures per individual unit, total sales in billions of IDR, and total sales in billions of packs. Combine the predictions from each tree to produce the final output. In numerical prediction (regression) tasks such as sales forecasting, the aggregation process is usually done by averaging the predictions from all trees.

This smoothing process aims to reduce the error of each tree, reduce variance, and produce more stable and accurate predictions compared to using only one Decision Tree. Forecasting with Moving Average (MA) is used as a conservative approach in this implementation, the model uses a window of 2, which means the 2026 prediction is calculated based on the average of 2018 sales data and 2025 estimates.

TABLE II
SALES FORECAST COMPARISON (MA VS. ETS) WITH FORECAST GAP

Product	Category	Year	MA Forecast	ETS Forecast	Forecast Gap
Rice	Staples	2026	11.420	12.053	0.633
Rice	Staples	2027	11.530	12.473	0.943
Instant Noodles	Staples	2026	14.765	15.000	0.235
Instant Noodles	Staples	2027	14.808	15.155	0.347
Cooking Oil	Staples	2026	3.140	3.325	0.185
Cooking Oil	Staples	2027	3.170	3.450	0.280
Chicken Eggs	Staples	2026	5.835	6.173	0.338
Chicken Eggs	Staples	2027	5.892	6.398	0.505
Chicken Nuggets	Frozen Food	2026	3.900	4.555	0.655
Chicken Nuggets	Frozen Food	2027	4.010	4.990	0.980
Sausage	Frozen Food	2026	5.030	5.960	0.930
Sausage	Frozen Food	2027	5.185	6.580	1.395
Processed Meatballs	Frozen Food	2026	2.680	3.020	0.340
Processed Meatballs	Frozen Food	2027	2.740	3.245	0.505
UHT Milk	Dairy Products	2026	11.775	13.453	1.678
UHT Milk	Dairy Products	2027	12.062	14.568	2.506
Yogurt	Dairy Products	2026	4.475	5.533	1.058

Yogurt	Dairy Products	2027	4.662	6.233	1.571
Cheese	Dairy Products	2026	1.490	1.580	0.090
Cheese	Dairy Products	2027	1.505	1.640	0.135

The ETS method produces higher prediction values than MA for all products, which indicates that ETS is better able to capture the upward sales trend pattern than the MA method. The results of the total projection per category in 2026 show that staple foods dominate with a value of 39.55 in various combined units, followed by dairy products at 20.57 and frozen foods at 21.44. Product predictions using the Moving Average (MA) method show variations in units used according to product type, such as rice with a value of 13.66 billion IDR, instant noodles with 17.15 billion packages, chicken nuggets worth 4.3 trillion IDR, and UHT milk with 14.11 trillion IDR. This projection provides a detailed overview of the contribution of each category and product to total sales, while serving as an important basis for analyzing consumption trends and future business strategies.

Forecasting model evaluation 2026 Forecasts and 2025 Actuals) as follows all models showed varying degrees of accuracy when forecasting 2026 sales against 2025 actuals, with Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) being calculated for Staples, Frozen Food and Dairy Products. Moving Average (MA) with generally provided conservative forecasts (Staples 11.42 Billion USD) and was simple to interpret but might underestimate accelerating trends. Exponential Smoothing (ETS) as follows Often provided slightly more optimistic extrapolations (Staples 12.05 Billion) by accounting for linear trends, making it a pragmatic choice for limited data. LSTM as follows Showed conservative forecasts (Staples 11.87 Billion) and struggled with the limited data points, often leading to warnings during execution and potentially less reliable predictions. ARIMA also faced challenges with limited data, leading to forecasts that were sometimes close to ETS (Staples 12.08 Billion but with inherent instability due to data scarcity. Random Forest (RF) with generated forecasts (Staples 12.34 Billion) that also exhibited limitations due to sparse data, as it struggled to build robust models for complex relationships. Prophet with demonstrated for key products (Staples 16.14 Billion, Frozen Food 5.48 Trillion, Dairy Products 16.53 Trillion) including exogenous variables, showing its potential for handling seasonality and regressors, but requiring careful interpretation of future exogenous values.

TABLE III
FORECASTING MODEL EVALUATION

Evaluation Model	Staples	Frozen Food	Dairy Products
MA MAPE	1.89%	5.34%	4.66%
MA RMSE	0.38	0.38	0.38
MA MAE	0.34	0.34	0.34
ETS MAPE	3.55%	10.56%	8.93%
ETS RMSE	0.73	0.71	0.70
ETS MAE	0.67	0.65	0.62
LSTM MAPE	6.12%	17.78%	14.32%
LSTM RMSE	1.38	1.37	1.30
LSTM MAE	1.07	1.02	1.01
ARIMA MAPE	4.16%	10.31%	9.67%
ARIMA RMSE	0.8	0.76	0.54
ARIMA MAE	0.98	0.77	0.69
RF MAPE	0.67%	0.75%	0.62%
RF RMSE	0.63	0.62	0.60
RF MAE	0,54	0.58	0.53

Due to the limited historical data of only 3 years of complete data. The performance evaluation Matrix (MAPE, RMSE, and MAE) table for the 2026 Prediction compared to the 2025 Actual shows a comparison of the accuracy levels of several forecasting methods, namely Moving Average (MA), Exponential Smoothing (ETS), Long Short-Term Memory (LSTM), ARIMA, and Random Forest (RF) on staples, Frozen Food, and Dairy Products commodities. Based on the evaluation results, the Moving Average (MA) method provided the best performance with the lowest error value for all commodities, indicated by MAPE values of 1.89% for staples, 5.34% for Frozen Food, and 4.66% for Dairy Products, as well as relatively small RMSE and MAE values. MAPE value < 10% is very good with very accurate forecasting model capability. MAPE value 10% - 20% is considered good and operationally acceptable. Conversely, the LSTM method produced the highest error value, especially for Frozen Food with MAPE 17.78%, indicating that the model is less than optimal in capturing historical data patterns in this dataset. The ETS and ARIMA methods provided a medium level of accuracy, but still had larger errors than MA. Meanwhile, Random Forest demonstrated fairly stable performance, but the MAPE value was still higher than MA. Overall, the evaluation results indicate that the Moving Average method is the most consistent and effective model for predicting 2026 demand based on actual data patterns from 2018. From this table, MA (Moving Average) provides the best performance for all commodities because it has the lowest error value compared to other models.

Comparison of ARIMA and ETS model accuracy based on visualized metrics, highlighting which model performs better for which products and under which conditions. Filtered comparison for MA and ETS models. MAPE 1.89%–5.34% indicates a small level of prediction error, so the Moving Average model can be considered successful in making predictions because the difference between the predicted 2026 and actual 2025 values is relatively low.

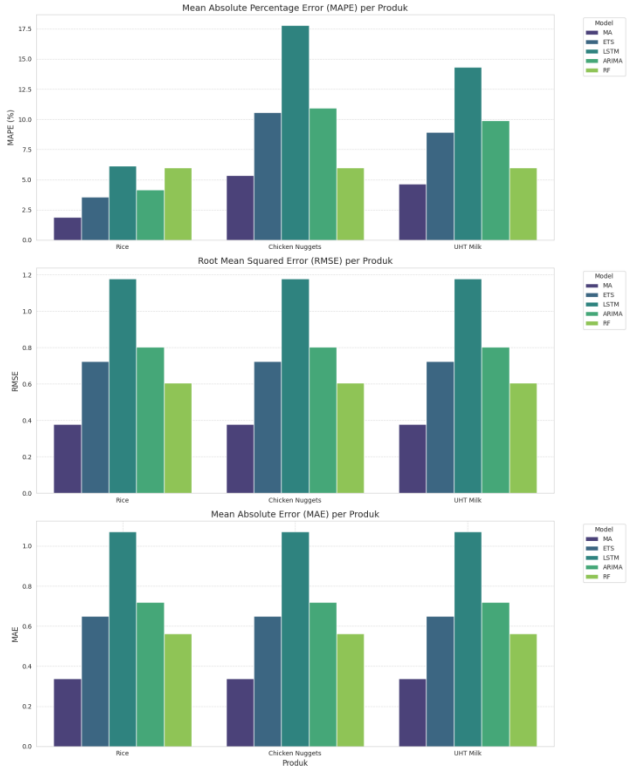

Figure 5. Comparison of Model Evaluation Metrics

TABEL IV
METRIC TYPE MODEL

Product	Metric Model	Value	Model Name	Metric Type
Rice	MA MAPE	1.890034	MA	MAPE
Chicken Nuggets	MA MAPE	5.339806	MA	MAPE
UHT Milk	MA MAPE	4.655870	MA	MAPE

Simple models like Linear Regression, ETS, Moving Average these models will continue to extrapolate the observed trend or average from the 2023-2025 data. Holt's Linear Trend (part of ETS) will likely remain the most pragmatic choice due to its ability to capture and extend existing linear trends, albeit with decreasing confidence as the projections become further removed from historical data. Complex models (LSTM, ARIMA, Random Forest) without the addition of significant and more granular historical data, these models will likely continue to struggle to produce stable and accurate predictions for 2027. Their performance will remain conservative or even unstable due to the lack of strong patterns to learn from such limited data. LSTM 'retracing warnings' and ARIMA's sensitivity to parameters will become increasingly relevant with further extrapolation.

D. Forecasting Model

This qualitative data explains the driving factors behind the growth trends seen in the source chart, where the Staples, Frozen Food, and Dairy Products categories are all projected to increase in 2026 compared to their historical averages. Rice products in the Staples category are expected to experience a Very High Surge, driven by household stockpiles and social assistance distribution. For example, Chicken Nuggets in the Frozen Food category are predicted to experience a High Surge due to their practicality as a pre-dawn meal and breaking-the-fast meal for children, while UHT milk in the Dairy Products category experiences a Moderate Surge as a popular component in Eid al-Fitr gift packages or hampers. This explanation provides in-depth context for the increased sales volumes shown in the 2026 forecast for each of these categories.

TABLE V
QUALITATIVE SEASONAL FACTOR DEMAND

Foods	Surge	Explanation
Rice	Very High Surge	Social Aid & Household Stock
Chicken Nuggets	High Surge	Practical Choice for Suhoor & Iftar for Children
UHT Milk	Moderate Surge	Parcel / Hampers Components for Eid

Rice experiences a Very High Surge during Ramadan and Eid. While the MA and ETS models show continued growth, and LSTM is more conservative, none of these models explicitly account for a seasonal spike within the year 2026. Their forecasts represent annual totals/averages. To capture this surge, we would need monthly or quarterly data to observe the within-year patterns. Chicken Nuggets (High Surge) have a High Surge due to being a practical choice during these periods. The forecasting models extrapolate the overall upward trend, but without sub-annual data, they cannot predict the specific peaks associated with holidays. UHT Milk sees a Moderate Surge as a component for parcels/hampers. Again, the annual forecasts reflect a general upward trajectory, but the magnitude of a holiday-driven surge within 2026 remains unquantified by these models.

The main objective is to conduct predictive analysis of seasonal trends in food retail in Indonesia using various forecasting models. This involves collecting and integrating detailed historical sales data along with relevant exogenous variables such as holiday calendars, conducting in-depth exploratory data analysis (EDA) to understand seasonal trends and patterns, developing and evaluating advanced time series models (such as SARIMAX, Exponential Smoothing, LSTM, ARIMA, and Random Forest), generating sales forecasts for future periods (2026 and 2027) with confidence intervals, and finally, providing actionable insights and recommendations for business strategy based on the refined forecasts and identified seasonal patterns.

The 2026 sales projection of 70.65 is a direct numerical summation of various product categories, but this figure should be interpreted with caution as it is derived from mixed units, ranging from Billions and Billions of Packs to Million Tons and Trillions of Rupiah. Without conversion to a consistent currency or physical unit, this aggregate figure does not represent a tangible financial or physical quantity,

so it is recommended to aggregate by individual units or convert all data to a single currency for more accurate business insights. While the previous box plot analysis confirmed that the distribution of individual sales data falls within the normal range with no outliers, the inconsistency of the units within this 70.65 total makes it merely a raw numerical aggregation that requires extreme caution when used for strategic analysis.

When applied to retail stores with highly impulsive customers or extreme sales variability (many outliers), this model may require parameter adjustments to avoid prediction deviations. The model's experience with zero data periods (such as the 2021) anomaly seen in the previous chart) suggests that it has a robust framework for recovering from data disruptions. This characteristic is useful for retail stores in locations likely to experience similar operational disruptions, but for stores with stable operations without data gaps, the model may need to be optimized to be more sensitive to changes in microtrends.

This model is highly applicable as a methodological framework to other retail stores in Indonesia because it takes into account local seasonal factors and data stability. For optimal results, the model must be retrained using specific data from the new store to adapt to its customers unique shopping behavior.

IV. CONCLUSION

The business objective of these grouped forecasts is to help retailers identify food subsectors with growth potential, optimize inventory management to avoid overstocking or understocking, and improve distribution efficiency to address seasonal spikes such as Eid al-Fitr or Christmas. Staples (20600.71) is the category with the highest sales volume. Products in this category include staples such as rice, instant noodles, cooking oil, and chicken eggs, which have consistent demand in the Indonesian market. Dairy Products (5936.84) show strong demand growth, including products such as UHT milk, yogurt, and cheese. Frozen Food (16250.43) has the lowest projection among the three but remains popular due to its convenience for consumers, including chicken nuggets, sausages, and processed meatballs. The Moving Average (MA) in this context provides a conservative approach by averaging sales data from the last two points (2024 and the 2025 estimate), so that the projection results focus more on smoothing short-term trends rather than pursuing aggressive linear growth. To make more accurate and relevant predictions for 2027 and the future Implementing SARIMAX Model with Exogenous Variables e.g., holiday calendar, promotional data, economic indicators will be crucial to accurately capture the impact of external factors and intra-year seasonal patterns.

REFERENCES

[1] N. S. Poli and A. S. Sikder, “Predictive Analysis of Sales Using the Apriori Algorithm : A Comprehensive Study on Sales Forecasting and Business Strategies in the Retail Industry,” no. 1, pp. 1–15, 2023.

[2] S. Demand and R. Mohammed, “Moving Average Smoothing for Gregory-Newton Interpolation: A Novel Approach for Short-Term Demand Forecasting,” *IFAC Pap.*, vol. 55, no. 10, pp. 749–754, 2022, doi: 10.1016/j.ifacol.2022.09.499.

[3] S. Usoro, P. Orekha, and D. A. Galadima, “Reducing food waste through predictive analytics and cold chain monitoring : A U . S . Retail Review,” 2026.

[4] C. Martins, J. P. Silva, and C. Martins, “Predictive Analytics for Food Retail: Improving Seasonal Sales Forecasts through Advanced Models,” 2025.

[5] J. Lin, “Application of machine learning in predicting consumer behavior and precision marketing,” pp. 1–12, 2025, doi: 10.1371/journal.pone.0321854.

[6] V. Gustine, S. Tendean, and J. Tjen, “An Optimized Demand Response-ARMA Model for Inventory Management Under Intermittent Product Demand,” vol. 8, no. 2, 2025, doi: 10.32877/bt.v8i2.3326.

[7] H. Khoshvaght, R. Ramyad, A. Razmjou, and M. Khiadani, “A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction,” *J. Environ. Chem. Eng.*, vol. 13, no. 6, p. 119675, 2025, doi: 10.1016/j.jece.2025.119675.

[8] M. Mujiyanto, S. Nurindahsari, and R. N. Izza, “Telematika Comparison of Linear Regression , ARIMA , Simple Exponential Smoothing , hybrid ARIMA-LSTM , and EWMA in Forecasting Commodity Prices,” vol. 17, no. 2, pp. 112–127, 2024.

[9] S. M. Damayanti, “Correlation Analaysis Of Market Performance And Bankruptcy Risk (Case Of Healthcare Subsector Companies Listed In Indonesian Stock Exchange Period,” no. August, pp. 7–9, 2019.

[10] T. Ericko, M. D. Lauro, A. Winatar, and T. Handhayani, “An Analysis And Forecasting The Foodstuffs Prices In Surabaya Traditional Market Using LSTM,” vol. 10, no. 2, pp. 34–42, 2024, doi: 10.24014/coreit.v10i2.27855.

[11] M. Obaidat, H. A. Almomani, R. Mallouhy, A. Almotari, and O. T. A. L. Meanazel, “A Hybrid Machine Learning Framework for Daily Dairy Demand Forecasting : Integrating SARIMAX and XGBoost for Seasonal Production Optimization,” *IEEE Access*, vol. 13, no. September 2025, pp. 162668–162680, 2026, doi: 10.1109/ACCESS.2025.3610316.

[12] D. R. Brillinger, H. K. Preisler, A. A. Ager, and J. G. Kie, “An exploratory data analysis (EDA) of the paths of moving animals,” vol. 122, pp. 43–63, 2004, doi: 10.1016/j.jspi.2003.06.016.

[13] S. A. Usman, M. Bhattacharjee, A. A. Alsukhailah, and A. D. Shahzad, “Identifying the Best-Selling Product using Machine Learning Algorithms Identifying the Best-Selling Product Using Machine Learning Algorithms,” pp. 0–15, 2025, doi: 10.20944/preprints202502.1356.v1.

[14] T. Hidayat, R. Mahardiko, and A. M. Rosyad, “The Analysis Of Data Preparation To Validate ModeL,” vol. 6, no. 2, pp. 23–34.

[15] R. Paper and K. Kandananond, “A Comparison of Various Forecasting Methods for Autocorrelated Time Series Regular Paper,” vol. 4, pp. 1–6, 2012.

[16] A. Hassan, M. Alam, and A. Faeique, “Forecasting Monthly Inflation in Bangladesh : A Seasonal Autoregressive Moving Average (SARIMA) Approach,” vol. 2, no. 2, pp. 25–43, 2023, doi: 10.1991/jefa.v7i2.a61.

[17] S. Muna, U. Airlangga, and E. Java, “Application Of The Holt-Winters Exponential Smoothing,” vol. 10, no. July, pp. 53–60, 2021, doi: 10.20473/jbk.v10i1.2021.53.

[18] S. Guritno, “A Comparative Study Of Forecasting Models For Trend And Seasonal Time Series : Does Complex Model Always Yield Better Forecast Than Simple Models ?,” vol. 7, no. 1, pp. 22–30, 2005.

[19] K. Misiurek and T. Olkuski, “Review of Methods and Models for Forecasting Electricity Consumption,” 2025.

[20] A. D. Maharani and N. K. Mawardi, “Effects of Staple Food Consumption Pattern on Achieving Food Security in Girisekar Village ,” pp. 334–340.