Contemporary Accounting Case Studies

Volume 3, Nomor 1, March 2024 Article 17

SALES FORECASTING USING MACHINE LEARNING WITH STRUCTURAL EQUATION MODELLING FRAMEWORK

Kris Sugiyanto
Master of Accounting Program, Faculty of Economics and Business, Universitas Indonesia
kris.sugiyanto@ui.ac.id

Setyo Hari Wijanto
Master of Accounting Program, Faculty of Economics and Business, Universitas Indonesia
setyohw@gmail.com

Contemporary Accounting Case Studies,
March 2024, Vol. 3, No. 1, pg. 313-337

SALES FORECASTING USING STRUCTURAL EQUATION MODELLING: CASE STUDY OF PT XYZ

Kris Sugiyanto*, Setyo Hari Wijanto
Master of Accounting Program, Faculty of Economics and Business, Universitas Indonesia

ABSTRACT

This research aims to analyze what variables significantly influence sales and make sales forecasts based on these variables. This research used the covariance-based Structural Equation Model framework with reflective indicator variables (CB-SEM) to develop and test the causal relationship of the demand and supply variables to performance that affect sales. Three latent variables proposed as constructs were performance, demand and supply. Meanwhile, eight variables proposed as indicator variables were sales, operating profit, inventory level, production output, sales promotion free goods, sales from distributors to stores, sales from stores to consumers and sales of similar cosmetic products. Sales forecast analysis was conducted by calculating the coefficient of determination (R-square), while the level of accuracy was determined by comparing the measured value and the predicted value. Performance variable was significantly influenced by demand, while supply had a positive effect on performance but not significantly. The indicator variables that are identified as predictors are inventory level, production output, promotion free goods and sales to stores. The polynomial regression had highest accuracy rate of 96.2%, followed by linear regression of 96%, while the lowest accuracy rate was 95.5% for random forest regression.

Keywords: *Multivariate Analysis, Sales Forecast, Structural Equation Model, Time-Series*

---

* Corresponding Author's Email: kris.sugiyanto@ui.ac.id

1. INTRODUCTION

Consumer demand is changing rapidly driven by the rapid development of information technology and consumer behavior patterns. Business entities are changing quickly to adjust their business models on production, distribution, and marketing lines to survive in the competitive business competition. The use of information technology in sales activities has a strategic role in reducing sales and distribution costs, creating product differentiation, reducing costs to retain suppliers and customers and creating barriers for competitors to enter the market (Poornima, 2019).

Changes in consumer behavior and changes in the business environment, including extraordinary events such as disease outbreaks, natural disasters, economic and industrial conditions, government regulations and business strategies from competitors, have become the focus of attention of business entities. The Covid-19 pandemic had a major impact on slowing economic growth because of restrictions on mobility and economic activity, which caused a decline in the performance of the manufacturing industry worldwide. The Covid-19 pandemic has also caused a shift towards direct sales to end consumers, so customer experience is very important to strengthen or maintain consumer loyalty to the brand of a product (Rashi et al., 2021).

The Prompt Manufacturing Index (PMI) data from 2016 to the first quarter of 2021 shows that the performance of the manufacturing industry in Indonesia experienced significant fluctuations during the Covid-19 pandemic. PMI measures the performance of the manufacturing industry based on four forming indicators, namely production volume, input goods order volume, finished goods volume and total workforce (Bank Indonesia, 2020).

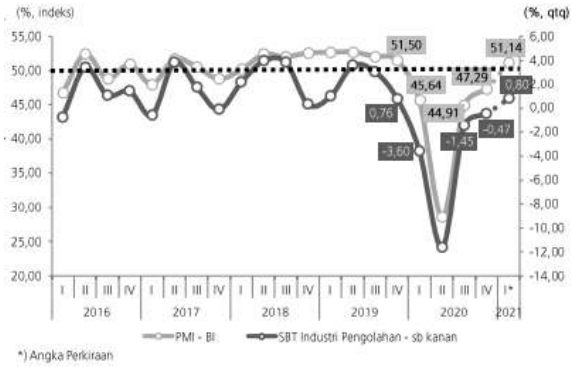

Figure 1 – Prompt Manufacturing Index (Bank Indonesia, 2020)

PT XYZ sold directly to a sole distributor covering all domestic sales, while export sales are determined by the overseas group of companies. PT XYZ also experienced up and down in financial performance with a declining trend, especially in sales and operating profit from the period 2017 to 2021. During the Covid-19 pandemic in 2020, PT XYZ experienced a decline in sales performance of IDR 922 billion and reported an operating loss amounting to IDR 114 billion, a decrease of IDR 305 billion when compared to 2019 as a result of the temporary suspension of production and sales activities, besides that production volume and inventories also decreased by 22%. PT XYZ's financial performance in 2017 to 2021 did not reach the set budget targets, both achieving the sales budget and the operating profit budget. In the last 5 years, there had been a significant decline in performance where sales fell by 32% while operating profit fell from IDR 225 billion to a loss of almost IDR 100 billion. Sales forecasting in PT XYZ was made by the Production and Sales Planning Department based on the average inventory balance for the previous 6 months and free goods sales promotion plans for certain products. The average discrepancy in actual sales achievement compared to the sales forecast using this method for the 2019 to 2021 period was above 25%.

The company determines goals, strategies and action plans to achieve the organisation's vision and mission. Economic resources and planning for managing these resources are needed to run a business in order to achieve business goals. Budget as a formal form of plans, goals and objectives of management that covers all operational aspects for a specified period of time. The budget provides targets for the financial aspects of work plans in an organization which also include alternative actions for making decisions and control applications for potential problems before work is carried out (Shim et al., 2005).

Sales forecasting is the process of using currently available sales data and past sales data to make predictions of future sales for planning purposes. A budget is a plan for achieving a target, while forecasting is a prediction of future conditions based on estimates of the dependent variable that influenced by a number of independent variables (Atrill and McLaney, 2021).

Sales forecasting methods that do not follow consumer demand trends cause at least two risks, inventory inefficiency and unavailability of inventory to meet consumer demand. These two risks cause losses and create derivative problems that affect the Company's performance. This case study aims to apply sales forecasts using Structural Equation Modelling (SEM) to measure the relationship between indicator variables and latent variables as a sales construct based on multivariate time series historical data. Sales prediction is performed using the KNIME

application with Linear Regression, Polynomial Regression and Random Forest Regression algorithms based on variables tested using the SEM method.

Multivariate statistics are increasingly popular because they are used to analyze complex data sets including dependent and independent variables simultaneously or in several stages. This statistical method is the development of univariate and bivariate statistical methods to analyze data with only one or two variables (Tabachnick and Fidell, 2013).

The purpose of this research is to measure the relationship between indicator variables and latent variables, structural equation model between latent variables and make sales forecasts using the KNIME machine-learning application based on the results of analysis and evaluation of the indicator variables that affect sales through a series of tests with the SEM framework. Several multivariate techniques such as multiple regression and multiple analysis of variance can solve problems that are difficult to overcome with univariate analysis, namely single variable analysis and bivariate analysis. Other multivariate techniques such as Polynomial and Random Forest Regression are designed to deal with multivariate problems to identify the underlying structure of a set of variables or discriminant analysis to differentiate groups based on a set of variables (Hair et al., 2018).

2. LITERATURE REVIEW

2.1. COMPANY PERFORMANCE

The traditional approach is a method for measuring company performance that focuses on achieving targets from the financial aspect, namely sales and profit growth. These two performance measures are the main factors to determine the success of an organization in a certain period because they are the main requirement for an organization to be able to grow and maintain its survival (Baumol, 1960). However, in its development, these two performance benchmarks are no longer sufficient to measure the success of an organization, so other performance benchmarks are needed which are generally adapted to the vision and mission or the reason for the organization's existence in a business environment.

Sales occur when demand for goods and services meets supply of goods and services. Decisions about products, the amount of goods or services produced by the market, are determined from negotiations by buyers and sellers of these goods or services in a free market economic system. Buyers submit requests to sellers regarding what products or services must be made, how many quantities to provide, what specifications are desired, when the goods or services must be available and the price they are willing to pay for these goods and services. The seller then supplys product according to the wishes of the buyer at

the price proposed by the seller so that in the end an agreement is reached from the negotiations carried out by the seller and the buyer. Sales are the total amount of money received by a business entity during a certain period by selling goods and services (Nickels et al., 2019).

Companies make efforts to achieve sales growth and maximize profits simultaneously or jointly, sometimes even companies have to sacrifice sales growth to obtain minimal profits with the aim of maintaining survival or conversely sacrificing profits to pursue sales growth in order to compete with its competitors. Sales in the short term is a function of the number of goods sold and the price of these goods. The number of goods sold will increase if the price of goods is lowered, sometimes even beyond the expected minimum profit point, or conversely, sales will decrease if the price is increased to obtain the required minimum profit. Meanwhile, in the long term, sales are a function of the number of goods sold, the price of goods and other variables that are attempted by the Company to realize these sales, such as promotional and marketing activities. Efforts to achieve sales growth are limited by constraints on obtaining the minimum profit that can be received by the company to meet Demands of shareholders and other stakeholders and provide funds to maintain its survival (Baumol, 1960).

2.2. DEMAND

Demand for an item or service is influenced by various distribution channels used by sellers at various levels to distribute their products from the distributor level to the final consumer so that they are ready for consumption. Distribution channels are defined as a set of interconnected organisations in relation to efforts to provide a product of goods and services so that it can be used or consumed by consumers as end users. All types of distribution channels have an important role in the company's success and influence all other marketing decisions including the process of making products, distribution, sales to service activities to customers (Kotler and Keller, 2021).

2.3. SUPPLY

Supply is the number of products that producers are willing to sell at different prices for a certain period of time. The quantity supplied will increase as the price increases because sellers can make more profit at a higher price. The ability to predict the number of requests or a product and ensure sufficient supply to meet demand is very influential on sales. Inventory will increase if production volume exceeds sales volume and vice versa will decrease if production volume is less than sales volume. Therefore it is very important to maintain a balance between production levels, inventory quantities and sales volumes that reflect consumer demand. Products that are made but do not follow consumer demand patterns cause many problems in Supply chain, including unavailability of supplies during

periods of peak demand and conversely there can be significant oversupply during periods of low demand. These problems create cost inefficiencies and unnecessarily reduce company resources in Supply chain. Managing supply and demand through proper sales and production planning can significantly improve sales performance when applied to products with predictable demand variables (Chopra, 2018).

PT XYZ manages its supply chain mostly from domestic companies except for some raw materials that must be purchased from abroad to meet the quality standards required by the Company group. In addition, the Company also directly imports several raw materials from India and China because these goods are difficult to obtain from local suppliers. Raw material purchases are based on production plans derived from sales forecast requirements, where ongoing reviews are carried out periodically within three months based on the average lead time delivery from suppliers.

2.4. STRUCTURAL EQUATION MODELING

SEM consists of a set of variables that define constructs and then the constructs are related to each other. Latent variables are variables that cannot be directly measured or observed and form a construct. The latent variable is measured based on a set of other variables that can be measured by a series of tests. While other variables are variables that can be observed directly and are used to define latent variables. SEM is a statistical method that aims to test the hypothesis of a theory by using structural equation model analysis in the form of an interdependent relationship between latent variables measured by several indicator variables to provide answers to a phenomenon that arises (Schumacker and Lomax, 2010).

The SEM method used in this research is based on general variance analysis starting with calculating the covariance matrix or often referred to as covariance-based SEM. The SEM method was used in this study with the consideration that the structural equation model using the path model estimation technique is considered the most efficient for analyzing a series of multiple regression equations of various latent variables and indicator variables, where the structural model and measurement model are estimated simultaneously or together. This research was conducted based on supporting theory to identify factors that are relevant to sales forecasting to then be tested empirically on the research object, so that the SEM method is considered a good choice to use in this research.

2.5. FORECASTING

The ever-changing conditions of the organizational environment cause uncertainties that can significantly impact an organization. Every organization needs to make adjustments in response to these changes to survive, develop and achieve its goals. Forecasting is useful for anticipating future changes that affect business decisions and

activities. The approaches to forecasting are qualitative methods based on opinions, past experience and also the best guess that can be made by an expert in the field and quantitative methods using a series of data to create a forecasting model (Slack et al., 2013).

2.6. TIME-SERIES ANALYSIS

This analysis is used for research that use dependent variable data with long time period, at least 50 time periods or more where the time sequence is the main independent variable based on historical event data (Warner, 1998). Time-series analysis in this study is used to predict seasonal or cyclical sales movement patterns that are usually attached to sales of cosmetic products as the object of this research. Sales of cosmetic products in Indonesia are usually influenced by certain events such as religious holidays and year-end holidays where at these times sales tend to increase, whereas during the new school year sales tend to decrease. The purpose of this analysis is to identify patterns in a sequence of numbers over time that are interconnected over a period of time. As in any regression analysis, scores are decomposed into several potential elements.

The first potential element is a random process to identify patterns, i.e. trends over time where a linear pattern occurs if the mean is continuously increasing or decreasing over time, a quadratic pattern occurs when the mean first increases and then decreases over time or vice versa, or also other more complicated patterns. The second potential element is the remaining effect of the previous score while the third potential element is the remaining effect of the previous random process. These patterns are not mutually exclusive, two or all three can be intertwined in the random process that occurs (Tabachnick and Fidell, 2013).

2.7. MULTIVARIATE ANALYSIS

Multivariate analysis refers to all statistical techniques that simultaneously process multiple measurements on individuals or objects under research. Any simultaneous analysis of more than two variables can be considered a multivariate analysis. Multivariate techniques are an extension of univariate analysis, i.e. single variable analysis and bivariate analysis such as cross-classification, correlation, analysis of variance and simple regression used to analyze two variables. For example simple regression with one predictor variable is extended in multivariate techniques to include multiple predictor variables. A single dependent variable in analysis of variance is expanded to include multiple dependent variables in multivariate analysis of variance (Hair et al., 2018).

2.8. BUDGET

A budget is a quantitative description of a plan for a specific period of time including planned sales volumes and revenues, total resources, costs and expenses, assets, liabilities

and cash flows. The budget as a plan consists of the amount of economic resources to be allocated and used for activities within a certain period of time and is measured in monetary units. The implementing party must act earnestly to realize the plan, while budgeting is a process for determining resource requirements which include the amount of capital, the amount of material and the number of people needed to carry out activities to achieve goals within a certain period of time (Erasmus, 2015).

2.9. PREVIOUS RESEARCH

Several previous studies related to this research: (Susetiyana, 2009) conducted research on customer loyalty analysis in the airfreight forwarder industry using the Structural Equation Modeling (SEM) method by proposing six constructs that were hypothesized to have an influence on customer loyalty. These constructs are perceived value, perceived quality, customer satisfaction, trust, customer complaint and image. (Boyapati & Mummidi, 2020) conducted research on sales forecasting using data from a company in the food industry, namely Support Vector Regression, Gradient Boosting Regression, Simple Linear Regression, and Random Forest Regression to identify independent variables that affect sales and make sales predictions. (Paduloh & Ustari, 2022) conducted research to minimize forecasting errors by analyzing which demand forecasting model is most suitable for demand conditions based on historical data for masterbatch product demand. The method used in this study is the time series model, which consists of the season naive method, holt exponential smoothing, triple exponential smoothing, and autoregressive integrated moving average (ARIMA).

3. RESEARCH METHODS

This research used a quantitative approach based on a certain number of samples that were analyzed to test the proposed hypothesis. The object of this research was PT XYZ, a company that produced and markets cosmetic products through a single distributor who had been working together for a long time. The data sources and information used were data for the period January 2019 to December 31, 2021.

Population and Sample

Population in this research were sales transactions and related data that affected sales during PT XYZ's establishment. Sampling was carried out based on a quota determined previously by the author because it was considered to reflect the latest conditions, namely monthly data from 2019 to 2021 with a total of 1,836 samples for each variable. A sample size of 100-150 data is considered adequate for SEM models with up to five latent variables (constructs) and each construct is explained by three or more indicators (Santoso, 2018).

Data Collection

Author obtained the required data and information from internal sources to be processed and used as indicator variables in creating structural equation models with the aim of testing hypotheses on latent variable constructs, performance, demand and supply. This research used 8 monthly data sets with a total of 1,836 data each as indicator variables that were thought to affect latent variables. The data used were sales data to distributors (K1), operating profit (K2), sales from distributors to stores (PT1), sales from stores to end consumers (PT2), sales of similar products in Indonesia (PT3), inventory levels (PW1), production output (PW2) and direct sales promotion of free goods to distributors (PW3).

Analysis Method

This research used a descriptive analysis method by developing SEM model to provide an overview of the conditions, characteristics and relationships of each latent variable construct studied. The descriptive model is a model intended to describe a concept or formation of variables and constructs that used in measurement tests (Waluyo, 2011). SEM analysis in this research was carried out to identify models that could explain the relationship between indicator variables and latent variable constructs based on theoretical studies so as to obtain an acceptable path diagram model. This research also used predictive analysis methods to analyse sales forecasts with multivariate methods from several regression algorithms to process historical data sets that had been tested and accepted as independent variables to predict future sales.

SEM Model

Latent variable constructs in this structural equation model were one endogenous variable Performance and two exogenous variables Demand and Supply. There were also eight endogenous variables as indicator variables that can be measured, K1, K2, PT1, PT2, PT3, PW1, PW2 and PW3 which were seen as indicators influenced by the same concept and underlying it latent variables. While the exogenous residual variables in this model were e1, e2, e3, e4, e5, e6, e7, e8 and e9.

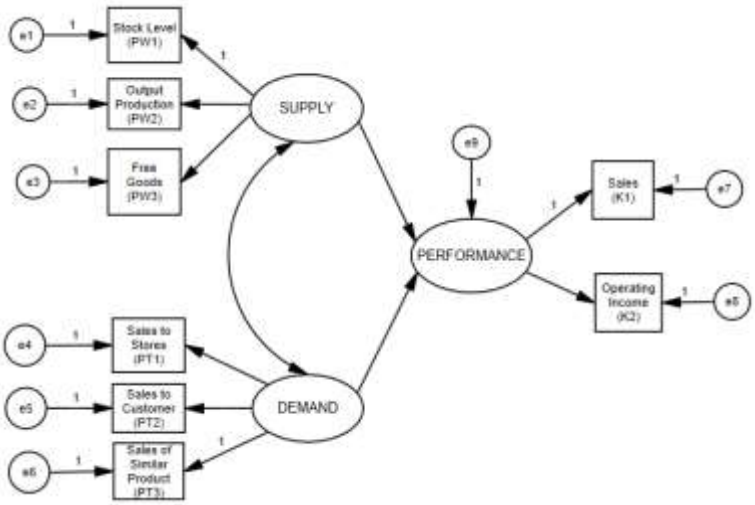

Figure 2 – SEM Model

Performance

The performance variable in this research was the achievement of organizational success measured using a traditional approach that prioritizes financial aspects. Indicator variables to measure performance were sales data to distributors (K1) and operating profit (K2).

Demand

Demand variables in this research were pulled factors which were sales in the distribution chain channel at each level used to market their products so that they can be received by end consumers and are ready for consumption (Kotler & Keller, 2021). Indicator variables to measure Demand were sales from distributors to stores (PT1), sales from stores to consumers (PT2) and sales of similar cosmetic products in the Indonesian market (PT3).

Supply

Supply variable in this research was a driving factor controlled by PT XYZ as a producer who produced and markets its products to buyers following consumer demand. Supply of goods was mainly influenced by the company's policy based on production needs derived from sales forecasts, besides that there was also an exceptional policy for a small number of material items that were difficult to obtain so that more finished goods stock was required. The indicator variables to measure this variable were inventory levels (PW1), finished goods production (PW2) and free goods sales promotion to distributors (PW3).

Measurement Model

The measurement model test aims to ensure that the constructs used meet valid and reliable criteria by creating a Confirmatory Factor Analysis (CFA) model for each

acceptable exogenous and endogenous construct. The CFA model is acceptable if it has a good validity and reliability model data match. The test of validity is indicated by the t value and the loading factor must be above the critical value of 1.96 and the loading factor is greater than 0.5. The test results for variables that do not meet the valid criteria cannot be included in further testing. The factor load for each indicator on its latent variable is presented in the form of a relationship depicted in the path diagram obtained by running the IBM SPSS AMOS Version 24 application. Reliability testing was done by calculating Construct Reliability (CR) and Variance Extracted (VE) from each observed variable (Hair et al., 2018).

The formulas for calculating CR and VE are:

$$CR = \frac{(\Sigma \text{ Std. Loading})^2}{(\Sigma \text{ Std. Loading})^2 + \Sigma\, \varepsilon\, j}$$

$$VE = \frac{(\Sigma \text{ Std. Loading})^2}{\Sigma \text{ Std. Loading}^2 + \Sigma\, \varepsilon\, j}$$

If the calculation result of CR is greater than 0.70 and VE is greater than 0.50, it can be said that the reliability is good (Wijanto, 2008).

Goodness of Fit Test

Furthermore, testing the results of the CFA confirmatory analysis was conducted by referring to the fit model criteria in the Goodness of Fit index. This statistical test was performed using the maximum likelihood estimation procedure to produce an estimated covariance matrix that will be compared with the sample covariance matrix to determine the suitability of the proposed model hypothesis.

Structural Model

This test was done by combining CFA models of exogenous and endogenous constructs that have been accepted into one overall model to be evaluated and analyzed to see the overall model fit and evaluation of the structural model so that an acceptable overall model is obtained. This analysis was carried out by looking at the coefficients of the structural equation which show a certain level of significance. Structural model analysis aims to test the hypotheses proposed in the research with a significance level of 0.05, the t value of the structural equation must be greater than or equal to 1.96 or greater than 2 (Wijanto, 2008).

The structural equation of the hypothesis proposed in this research is as follows:

H_1 : Supply has a positive and significant effect on Performance.
H_2 : Demand has a positive and significant effect on Performance.

Sales forecasts were made after the results of testing the measurement model and structural equation model were obtained. If the hypotheses proposed in this research matched the test results, the latent variables and indicators that were accepted are then used to make sales forecasts. Sales forecast predictions were made using the KNIME application to run algorithms as sales predictors. Sales forecasts were made using three regression algorithms namely linear, polynomial and random forest to predict sales of eight PT XYZ product brands. Test the prediction results of the regression model by calculating the Coefficient of Determination (R-square), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) between the measured and predicted values to assess the performance of the three regression algorithms.

Sales forecast accuracy criteria were measured using the R-square value with the following theoretical reference:

- If the R-square value is smaller than 0.3, the prediction results are considered unrelated or very weak.
- If the R-square value is greater than 0.3 and smaller than 0.5, the prediction result is considered weak.
- If the R-square value is greater than 0.5 and smaller than 0.7, the prediction result is considered moderate.
- If the R-square value exceeds 0.7, the prediction result is considered strong (Moore et al., 2011).

4. ORGANIZATION PROFILE

The object of this research was PT XYZ, a manufacturing company that produces and markets cosmetic goods in the form of hair care, skin care and makeup, fragrances and others. PT XYZ order materials from qualified suppliers based on production needs, the materials are received only if they meet quality standard tests and then store them in the material warehouse. PT XYZ processes raw materials into finished goods through raw material preparation, mixing, filling up or primary packing, packing of product packaging to the carton box or secondary packing and store the finished goods before dispatch to the external distributor who distribute goods to agents, stores and end-consumers. PT XYZ was established in the framework of Foreign Investment Law No. 1 of 1967 jo.Law No. 11 of 1970 in Jakarta. The Company began commercial production in 1971 with a factory location in Bekasi Industrial Estate, West Java. The Company's head office is located at Jl. Jend. Sudirman, Central Jakarta. The Group's products are marketed domestically and internationally, including to the United Arab Emirates, Malaysia, Japan, Thailand, Philippines, India, Singapore, Korea, Vietnam, China, Hong Kong and Taiwan. PT XYZ has a people-oriented Corporate Mission with the concept of "Healthy, Clean and Beautiful", confidently challenging boundaries to bring inspiration and memory to the lives of its customers.

5. RESULT AND DISCUSSION

The analysis of SEM model with maximum likelihood estimation (MLE) method consisted of two testing steps, measurement model testing and structural model testing. Before testing the structural equation model proposed in the hypothesis, the author first conducted data cleansing using the IBM SPSS Statistics Version 25 application to correct errors and identified data deficiencies to make the data set complete and accurate. This process included identifying outliers and dealing with them, testing the normality of the data and dealing with data that was not normally distributed.

Missing Value

There were 66 data of variable K1 with zero sales value, the data was excluded from the test sample because it was not relevant to the purpose of this research which was to make sales forecasts. After the data was excluded, the amount of data that qualified for further processing becomes 1,770 data. In addition, there were 1,155 data on operating profit variables based on brand group K2 with negative values and these data were replaced with zero values.

Outlier Test

Outlier identification was done using a boxplot graph that shows five statistical measures, namely minimum, first quartile, median, third quartile, and maximum to display the distribution of scale variables and determine outliers. Based on this test, 47 data were found in the outlier category and the data were then removed with the aim that the analysis of the SEM test results would not be biased. After data cleansing of outliers, 1,723 data were obtained to be used in making SEM models.

Normality Test

Normality test for the data was conducted using the Kolmogorov-Smirnov method with IBM SPSS Statistics Version 25. The test results showed that all eight brand categories for all product categories reached a significance value above 0.5 or it can be concluded that all data used was normally distributed.

Measurement Model Test

Factor loadings for each indicator on its latent variable were presented in the form of relationships depicted in the path diagram obtained using IBM SPSS AMOS Version 24 application. The measurement model fit test was carried out for each construct separately through an evaluation of the validity and reliability of the constructs used in this research to meet the valid and reliable criteria.

Table 1. SEM model regression result

			Estimate	S.E.	C.R.	P	Label
PERFORMANCE	<---	SUPPLY	0.142	0.118	1.203	0.229	par_8
PERFORMANCE	<---	DEMAND	0.112	0.046	2.442	0.015	par_9
PW1	<---	SUPPLY	1				
PW2	<---	SUPPLY	0.512	0.012	42.149	***	par_2
PT3	<---	DEMAND	1				
PT2	<---	DEMAND	0.003	0.004	0.593	0.553	par_3
PT1	<---	DEMAND	0.212	0.037	5.656	***	par_4
PW3	<---	SUPPLY	0.028	0.001	42.562	***	par_6
K1	<---	PERFORMANCE	1				
K2	<---	PERFORMANCE	0.076	0.005	14.55	***	par_7

Table 2. Loading factor of indicator variables

			Estimate
PERFORMANCE	<---	SUPPLY	0.312
PERFORMANCE	<---	DEMAND	0.665
PW1	<---	SUPPLY	0.798
PW2	<---	SUPPLY	0.688
PT3	<---	DEMAND	0.134
PT2	<---	DEMAND	0.014
PT1	<---	DEMAND	0.894
PW3	<---	SUPPLY	0.842
K1	<---	PERFORMANCE	0.959
K2	<---	PERFORMANCE	0.336

Table 1 and 2 are the result of the SEM regression model which showed the relationship between latent variables, demand and supply to the latent variable performance and the relationship between indicator variables to their respective latent variables.

Validity Test.

Table 3. Estimated loading factor

Indicators Variable	Loading Factor	
	Supply	Demand
Stock Level (PW1)	0.798	
Output Production (PW2)	0.688	
Free Goods (PW3)	0.842	
S1 (PT1)		0.894
S2 (PT2)		0.014

S3 (PT3)		0.134

Table 3 shows the results of the estimated loading factor values of all exogenous variables on their endogenous variables. The results of data processing with IBM SPSS AMOS Version 24 application showed that the exogenous indicator variable K2 had no significant effect on the latent variable Performance because it had a value smaller than 0.5. The exogenous indicator variables PT2 and PT3 also had no significant effect on the latent variable Demand. While all other exogenous indicator variables had a positive and significant effect on their respective endogenous latent variables as indicated by the estimated loading factor value greater than 0.5. The variable relationships that did not meet the valid criteria cannot be included in further testing.

Validity testing aims to see how appropriate the variables used in the research were indicated by the standardized loading factor value with a standard measurement greater than 0.5. Supply variable was a confirmatory analysis which was measured using observable variables as indicators, namely inventory levels (PW1), production results (PW2) and sales promotion of free goods to distributors (PW3). Demand variable was a confirmatory analysis which was measured using observable variables as indicators, namely sales from distributors to stores (PT1), sales from stores to consumers (PT2) and sales of similar products in the Indonesian market (PT3). Table 4 shows the results of the loading factor calculation for each observed variable on its latent variable. Based on these calculations, supply variable were measured by indicators PW1, PW2 and PW3 which had a loading factor of more than 0.5. Demand variable were only measured by the PT1 indicator which had a loading factor of more than 0.5, while PT2 and PT3, although they had a positive influence on demand, were not significant so they cannot be used as indicators to measure demand variable.

Reliability Test

Table 4. Realibility test result

Variable	Penawaran			Permintaan		
Indicator	Loading	Loading2	Error	Loading	Loading2	Error
Stock Level (PW1)	0.798	0.637	0.202			
Output Production (PW2)	0.688	0.473	0.312			
Free Goods (PW3)	0.842	0.709	0.158			
Sales 1 (PT1)				0.894	0.799	0.106
Sales 2 (PT2)				0.014	0.000	0.986
Sales 3 (PT3)				0.134	0.018	0.866
Sum of Std. Loading	2.328			1.042		
Sum of Std. Loading 2		1.819			0.817	

Sum of error		0.672	1.958
Construct Reliability	0.9	0.7	
Variance Extract	0.7		0.5

The reliability test aims to test the consistency of the indicator variable data that affects the latent variable. Reliability testing was completed by calculating the CR and VE of each observed variable. Table 4 shows the results of the calculation of CR and VE for the two latent variables, namely supply obtained results of 0.9 and 0.7, while demand latent variable obtained results of 0.7 and 0.5. Based on the results obtained, it can be concluded that the two latent variables can be explained by their indicator variables because they had met the measurement model criteria. If the calculation result of CR is greater than 0.70 and VE is greater than 0.50, it can be said that the model criteria are good.

Goodness of Fit Test

Table 5. Goodness of fit result

Index Goodness of Fit	Expected Standard	Result	Criteria
Chi-Square	≤ 16.919	10.394	Good Fit
Probabilitas	≥ 0.05	0.320	Good Fit
GFI	≥ 0.90	0.998	Good Fit
AGFI	≥ 0.90	0.994	Good Fit
NCP	Expected small	1.394	Moderate Fit
CFI	≥ 0.90	1	Good Fit
RMSEA	≤ 0.08	0.009	Good Fit

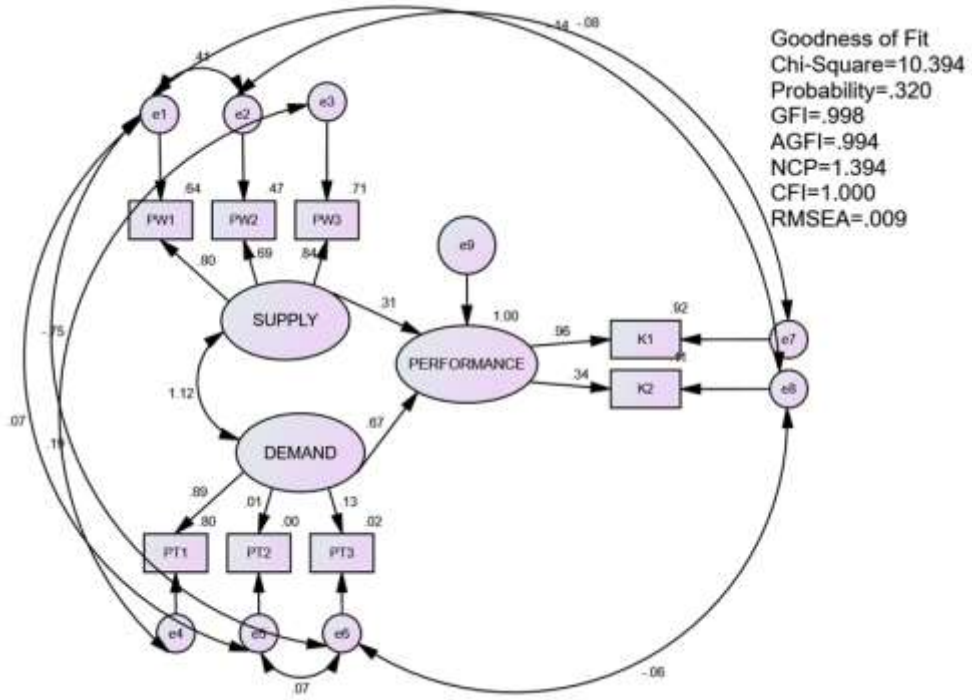

Figure 3. Structural model result

Structural model analysis in SEM was also carried out by testing the overall model fit as seen based on the Goodness-of-fit Index (GFI) statistical indicator from the IBM SPSS AMOS Version 24 output. The estimation results for all Goodness of Fit criteria had met the expected value standards as can be seen in Table 5, so that from the results of the analysis of the reliability of the output for testing the overall model it can be concluded that the model was a good fit.

Structural Model Test

Table 6. Hypothesis test result

Hypothesis	Path			C.R.	P	Result
H_1	SUPPLY	-->	PERFORMANCE	1,203	0,229	Rejected
H_2	DEMAND	-->	PERFORMANCE	2,442	0,015	Accepted

This analysis was conducted to test the hypotheses proposed in the research with level of significance as the calculation results for a significance level of 0.05, the t value of the structural equation must be greater than or equal to 1.96 or for practical purposes greater than 2.

Hypothesis testing showed the accepted hypothesis was H_2, demand had a positive and significant effect on performance, while H_1 was rejected because although supply had a positive effect on performance, it was not significant with a Critical Ratio (C.R) value of

1.203 and 2.442 respectively while the criteria for an accepted C.R value must be greater than 1.96.

Sales Forecast Analysis

Table 7. Multivariate coefficient correlation of SEM variables

	K2	K1	PW3	PT1	PT2	PT3	PW2	PW1
K2	1							
K1	0.322	1.000						
PW3	0.299	0.854	1					
PT1	0.305	0.870	0.663	1				
PT2	0.005	0.014	0.013	0.013	1			
PT3	-0.009	0.131	0.127	0.12	0.072	1		
PW2	0.244	0.680	0.579	0.689	0.011	0.103	1	
PW1	0.202	0.809	0.672	0.8	0.052	0.232	0.729	1

The multivariate correlation coefficient value of each indicator variable on the sales of PT XYZ (K1) in Table 7 obtained the results of three indicator variables, PW1, PW2 and PW3 had positive and significant correlation coefficient with value above 0.5, while only one demand indicator variable, PT1 had a positive and significant correlation coefficient with a value above 0.5. Thus, there were four independent variables that can be used as predictors to make a sales forecast for PT XYZ (K1), variables PT1, PW1, PW2 and PW3. The total sample data used to make sales forecasts was 1723 data and was accepted as an indicator variable that had a significant effect on sales. Sales forecasting was divided into eight brand categories namely GB, JA, PU, PX, SP, TC, PX-DEO and MD because each product category with each of the same brands had similar sales patterns. Indicator variable data was divided into two groups, training data and prediction data with a portion of 70% and 30% to make sales forecasts.

Sales forecasting result used three regression algorithms, linear regression, polynomial regression and random forest regression are as follow:

Table 8. Sales forecast result

	Linear Regression	Polynomial Regression	Random Forest Regression
R-square	0.960	0.962	0.955
Accuracy	Strong	Strong	Strong

6. CONCLUSION AND RECOMMENDATION

Based on the results of the tests that have been carried out, it can be concluded that sales indicator variable K1 can explain the performance latent variable while the operating income indicator variable K2 was considered unable to explain the performance latent variable.

Indicator variables of stock level PW1, output production PW2 and promotion of free goods PW3 can explain supply latent variables and indicator variable sales to stores PT1 can explain demand latent variable. Meanwhile, the indicator variables sales to consumers PT2 and sales of similar product PT3 were considered unable to explain the latent demand variable.

Sales forecasts for three regression algorithms, i.e. linear regression, polynomial regression and random forest regression, all showed strong R-square values. The highest R- square value for sales forecasts using the polynomial regression algorithm and the lowest used the random forest regression algorithm. All sales forecast methods produced in this research provide accuracy results of more than 90%, which were higher than the current method used by PT XYZ which was around 75%.

PT XYZ can improve budget accuracy by making sales forecasts using multivariate analysis based on the SEM method. The SEM method allows companies to analyze the causal relationship between predictor variables and proposed sales based on supporting theories of performance, demand and supply as latent variables so that companies can obtain valid and reliable input variables to be used as predictors in making sales forecasts. PT XYZ can use sales forecasts as a reference in making strategies and sales promotions so that it is expected that business risks due to oversupply and inventory vacancies will be lower.

REFERENCES

Atrill, P., & McLaney, E. (2021). *Management Accounting for Decision Makers*. Pearson Education Limited.

Baumol, W. J. (1960). Business Behavior, Value and Growth. *Economica - The London School of Economics and Political Science and The Suntory and Toyota International Centres for Economics and Related Disciplines*, *27*(107), 274.

Boyapati, S. N., & Mummidi, R. (2020). *Predicting sales using Machine Learning Techniques* (Issue May). Blekinge Institute of Technology, 371 79 Karlskrona, Sweden.

Chopra, S. (2018). *Supply Chain Management: Strategy, Planning, and Operation* (Vol. 51, Issue 170). Pearson Education, Inc.

Erasmus. (2015). *Global Innovative Leadership Module: Budget and Financial Planning*.

Hair, J., Black, W., Babin, B., & Anderson, R. (2018). *Multivariate Data Analysis.pdf: Vol. 8 edition*. Cengage Learning EMEA.

Indonesia, B. (2020). *Prompt Manufacturing Index Bank Indonesia (PMI-BI): Vol. IV*.

Kotler, P., & Keller, K. (2021). *Marketing Management*. Pearson Education, Inc.

Moore, D. S., Notz, W. I., & Fligner, M. A. (2011). *The Basic Practice of Statistics* (6th editio). W. H. Freeman.

Nickels, W. G., McHugh, J. M., & McHugh, S. M. (2019). *Understanding Business* (12th editi). McGraw-Hill Education.

Paduloh, P., & Ustari, A. (2022). Analysis and Comparing Forecasting Result Using Time Series Method to Predict Sales Demand on Covid-19 Pandemic Era. *Journal of Engineering and Management in Industrial System - Universitas Brawijaya*, *10*(1), 36–48.

Poornima, G. (2019). Information Technology in Marketing and Its Role in Business. *Journal of Emerging Technologies and Innovative Research - Ahmedabad*, *6*(3), 86–94.

Rashi, P., Ankur Singh Bist, Ari Asmawati, Mukti Budiarto, & Yustika Prihastiwi, W. (2021). Influence of Post Covid Change In Consumer Behaviour of Millennials On Advertising Techniques and Practices. *Aptisi Transactions on Technopreneurship (ATT) - Pandawan Incorporation*, *3*(2), 85–92.

Santoso, S. (2018). *Konsep Dasar Dan Aplikasi SEM Dengan Amos 22*. Gramedia.

Schumacker, R. E., & Lomax, R. G. (2010). *A Beginner's Guide to Structural Equation Modeling*. Routledge Taylor & Francis Group.

Shim, J. K., Siegel, J. G., & Shim, A. I. (2005). *Budgeting Basics & Beyond*. John Wiley & Sons, Inc.

Slack, N., Brandon-Jones, A., & Johnston, R. (2013). Operations Management. In *Operations Management II*. Pearson Education Limited.

Susetiyana, H. (2009). *Analysis Loyalitas Pelanggan pada Industri Airfreight Forwarder dengan Menggunakan Structural Equation Modeling (SEM)*. Universitas Indonesia.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics*. Pearson Education, Inc. https://doi.org/10.1515/snde-2014-0102

Waluyo, M. (2011). *Panduan dan Aplikasi Structural Equation Modeling, Untuk Aplikasi Penelitian Industri, psikologi, Soial dan Manjemen*. Indeks.

Warner, R. M. (1998). *Spectral Analysis of Time-Series Data*. The Guilford Press.

Wijanto, S. H. (2008). *Structural equation modeling dengan Lisrel 8.8 : konsep dan tutorial*. Graha Ilmu.

APPENDIX

Appendix 1. Data for Structural Equation Model

Year	Quarter	Brand	K1	K2	PW1	PW2	PW3	PT1	PT2	PT3
2019	1	GB	214.896	32.020	407.139	137.947	11.596	345.464	105.064	2.391.888
		JA	3.493	1.816	10.090	6.057	-	2.665	19.979	179.356
		MD	14.336	2.433	26.721	48.705	356	1.369	1.741	66.424
		PU	29.392	6.501	63.643	22.022	891	57.105	47.901	1.460.938
		PX	181.301	-	435.542	116.019	8.591	197.190	159.413	8.766.520
		PX Deo	18.159	189	25.778	2.314	1.193	27.083	899	49.009
		SP	6.163	1.662	8.424	1.627	362	8.632	12.714	542.826
		TC	43.595	6.083	61.481	13.887	3.149	10.812	28.818	339.762
	2	GB	214.509	-	369.343	166.543	8.629	378.748	89.973	2.624.893
		JA	4.010	-	8.365	6.494	-	2.502	26.951	184.191
		MD	13.006	-	33.283	54.570	331	1.168	1.265	68.589
		PU	17.926	-	41.714	18.698	154	40.689	25.237	1.366.851
		PX	183.269	5.690	589.294	159.155	11.497	224.287	164.389	9.609.662
		PX Deo	19.592	-	18.391	2.472	1.480	28.915	29.736	50.557
		SP	4.677	83	5.976	1.830	206	9.152	11.062	621.931
		TC	25.659	1.285	57.999	15.559	1.547	19.155	14.933	374.362
	3	GB	202.510	13.854	433.455	182.821	7.361	344.234	80.050	2.458.990
		JA	2.356	-	3.047	3.629	110	1.577	8.746	139.513
		MD	14.540	-	31.788	59.165	199	1.345	14.386	70.245
		PU	28.473	8.245	46.297	21.282	629	44.195	51.599	1.106.065
		PX	200.061	499	575.459	176.480	10.890	224.307	221.285	9.935.644
		PX Deo	17.342	983	26.032	2.656	1.153	30.240	12.499	48.242
		SP	5.888	51	5.765	1.909	378	8.863	24.026	565.093
		TC	44.555	1.483	39.872	16.218	3.331	11.571	6.465	355.205
	4	GB	187.991	-	371.167	203.817	9.119	311.453	86.784	2.505.959
		JA	3.340	-	3.899	5.145	-	2.010	11.031	163.901
		MD	15.834	-	27.147	61.518	-	8.277	11.532	71.961
		PU	53.813	-	47.069	34.161	1.715	75.244	39.843	1.484.000
		PX	168.827	-	620.755	207.851	6.785	221.412	170.273	10.017.110
		PX Deo	11.318	142	30.090	3.310	341	28.238	6.635	49.514
		SP	3.906	27	6.844	1.895	177	9.400	18.358	577.024
		TC	47.661	186	56.007	16.336	1.718	18.613	29.494	352.078

2020	1	GB	168.172	19.111	346.079	229.598	7.801	296.132	99.401	2.494.807
		JA	2.057	2.321	7.171	6.613	1	2.145	29.086	173.438
		MD	11.649	4.914	28.419	65.667	181	1.320	1.434	64.033
		PU	33.503	8.041	49.409	41.616	27	72.317	66.649	1.560.959
		PX	112.629	18.441	655.258	222.651	3.811	198.705	234.820	9.984.894
		PX Deo	14.467	5.275	19.141	3.857	883	24.136	8.839	46.916
		SP	5.111	1.373	4.285	1.933	321	7.649	9.659	570.567
		TC	39.853	15.186	49.118	17.340	2.569	27.237	14.799	319.070
	2	GB	105.415	12.662	307.746	263.605	4.106	221.699	97.518	1.903.645
		JA	2.305	1.054	7.413	6.964	-	1.633	6.237	114.658
		MD	16.487	2.068	37.789	72.388	220	3.712	30.711	43.382
		PU	33.510	1.744	47.469	227.274	1.858	55.893	39.993	1.180.267
		PX	84.667	-	621.178	238.872	5.648	130.019	183.044	7.194.726
		PX Deo	13.888	186	19.809	5.003	845	26.740	8.864	37.756
		SP	2.859	777	5.047	2.025	105	6.522	8.104	422.668
		TC	32.050	3.639	40.535	19.102	2.566	9.759	20.352	210.734
	3	GB	127.123	60.831	219.490	293.202	8.850	250.951	103.273	2.025.521
		JA	2.869	2.684	4.023	8.077	-	2.160	4.419	118.349
		MD	9.553	3.066	28.455	76.236	110	1.447	13.956	45.187
		PU	39.403	16.794	26.694	228.948	4.450	68.749	52.044	1.256.152
		PX	119.404	13.280	492.584	258.586	14.659	144.354	249.489	7.902.753
		PX Deo	13.888	11.351	10.112	5.331	18	24.936	2.378	39.737
		SP	4.807	2.281	3.006	2.151	339	6.569	30.794	446.557
		TC	38.807	9.659	27.469	20.947	3.112	7.567	16.087	222.120
	4	GB	141.529	44.159	199.762	297.206	10.885	239.269	102.289	2.225.207
		JA	3.027	1.252	3.634	7.605	-	2.519	8.906	130.604
		MD	12.256	2.889	25.157	78.673	56	1.197	6.438	50.146
		PU	37.949	10.318	33.135	56.623	3.944	54.644	37.565	1.308.015
		PX	129.300	6.067	429.312	260.268	16.208	152.633	234.357	8.198.619
		PX Deo	16.665	10.980	10.244	5.131	923	27.517	677	42.504
		SP	3.480	1.107	3.338	2.092	231	7.425	34.343	540.632
		TC	50.271	5.744	27.391	19.642	6.239	8.605	10.448	238.662
2021	1	GB	139.213	-	198.236	302.727	10.789	223.804	98.689	2.070.561
		JA	4.614	13	1.953	7.359	241	1.735	2.409	127.139
		MD	11.856	111	24.382	79.649	173	7.566	10.886	48.425
		PU	34.653	521	41.293	60.297	3.009	53.691	66.348	1.290.334

		PX	72.441	-	427.571	265.109	7.172	134.868	215.947	8.298.153
		PX Deo	13.672	-	21.180	5.161	902	25.354	3.089	42.777
		SP	4.256	111	3.404	2.095	392	8.967	24.986	492.002
		TC	33.609	6.115	42.005	19.394	3.409	15.787	39.056	225.868
	2	GB	126.573	-	201.528	250.370	7.000	216.931	110.644	2.332.382
		JA	3.212	-	3.092	7.684	15	1.873	32.149	137.337
		MD	11.477	-	26.031	74.152	165	993	9.141	52.939
		PU	39.389	13	43.352	52.302	1.386	57.160	56.439	1.636.561
		PX	99.443	805	426.710	247.626	8.654	151.829	206.630	9.331.189
		PX Deo	19.530	-	21.046	4.437	16	24.478	380	48.012
		SP	4.870	85	3.270	2.095	303	7.066	15.508	592.961
		TC	26.691	6.175	62.321	16.114	2.640	16.519	8.769	249.419
	3	GB	129.507	4.041	203.007	226.927	5.057	200.912	100.637	2.004.623
		JA	3.954	205	2.199	8.001	274	1.808	15.786	117.018
		MD	11.218	259	30.110	76.535	82	4.721	2.932	46.285
		PU	34.592	-	41.695	48.028	984	48.091	58.121	1.439.459
		PX	89.292	3.392	404.541	218.267	4.163	114.898	244.036	7.889.195
		PX Deo	10.527	1.298	14.690	4.154	4	17.181	14.112	42.048
		SP	3.513	413	4.760	2.112	203	6.801	19.039	479.930
		TC	32.213	6.292	51.902	14.408	1.441	14.609	17.328	202.968
	4	GB	158.711	4.419	228.623	227.042	5.014	262.853	75.537	2.414.955
		JA	2.219	155	2.737	4.883	8	1.529	6.492	92.978
		MD	11.530	210	28.107	83.473	224	1.273	781	56.642
		PU	44.525	3.875	46.862	46.483	2.663	66.134	52.777	1.487.873
		PX	115.883	8.323	340.199	225.509	3.332	153.261	230.727	9.067.790
		PX Deo	18.779	1.352	13.762	4.044	1.386	27.817	4.554	48.144
		SP	4.033	451	4.487	2.069	191	7.276	23.092	626.450
		TC	47.132	11.760	67.559	13.421	3.350	9.608	8.511	242.826

Appendix 2. Sales Forecast Result

Year	Quarter	Brand	K1	Prediction Linear Regression	Prediction Polynomial Regression	Prediction Random Forest Regression
2019	1	GB	68.785	65.736	66.807	60.579
		JA	3.102	3.236	3.337	3.090
		PU	20.749	22.245	21.785	18.503
		PX	51.446	46.977	45.952	34.227
		PX Deo	2.953	1.431	(1.662)	3.483
		SP	450	451	478	402
		TC	334	347	188	375
	2	GB	51.158	50.026	49.138	44.212
		JA	966	722	688	736
		MD	13.006	13.096	23.391	15.147
		PU	9.130	10.286	8.720	9.395
		PX	19.745	22.772	21.839	21.272
		PX Deo	10.767	10.452	11.897	9.759
		SP	1.095	1.198	1.273	1.039
	3	GB	99.651	88.906	93.480	82.981
		PU	5.060	5.211	4.966	4.953
		PX	58.972	63.675	62.817	58.536
		PX Deo	7.057	6.500	5.009	9.507
		SP	2.491	2.617	2.846	2.947
		TC	21.146	15.679	19.104	14.973
	4	GB	54.563	56.292	55.853	55.450
		JA	452	350	347	388
		MD	9.835	8.181	7.370	9.155
		PU	8.562	6.687	6.505	6.255
		PX	16.772	17.757	16.942	16.432
		PX Deo	6.541	5.629	5.118	5.687
		SP	488	479	499	391
		TC	18.347	6.042	4.510	8.576
2020	1	GB	54.568	52.813	53.325	51.798
		JA	792	1.239	1.412	1.429
		PU	11.807	12.493	11.307	12.148
		PX	33.465	38.778	37.488	31.183
		SP	633	892	832	933
		TC	33.660	25.121	42.265	29.849
		GB	31.312	29.653	29.350	32.056
	2	JA	252	266	181	324
		MD	4.651	4.252	2.878	5.287
		PU	626	573	239	902

		PX	16.389	13.629	12.261	16.983
		SP	1.739	1.622	1.566	1.976
		TC	30.676	31.285	32.359	38.712
		GB	27.019	29.335	29.655	28.210
	3	JA	534	725	690	636
		PU	8.677	8.223	7.394	8.182
		PX	27.787	28.672	27.965	30.709
		SP	229	239	228	252
		TC	853	684	851	1.097
		GB	43.683	45.037	44.693	42.316
	4	JA	749	773	787	737
		PU	4.404	4.015	4.287	3.893
		PX	49.465	53.972	53.898	57.161
		PX Deo	12.978	13.654	11.362	10.187
		SP	1.596	2.111	2.169	1.945
		TC	8.830	8.463	7.211	8.875
2021	1	GB	15.025	15.772	15.895	16.061
		JA	1.811	1.862	1.750	1.663
		MD	2.538	3.496	(466)	3.182
		PU	10.519	11.121	12.030	11.045
		PX	15.229	19.027	17.732	22.141
		PX Deo	4.988	4.836	3.038	4.314
		SP	1.640	1.944	2.033	2.341
	2	TC	210	229	303	328
		GB	26.470	28.679	28.971	28.291
		JA	542	758	744	803
		MD	5.611	4.838	3.755	8.106
		PU	10.847	8.160	7.796	7.917
		PX	12.416	14.539	13.071	16.208
		SP	665	653	612	626
	3	TC	962	1.255	1.116	1.364
		GB	30.863	25.911	25.773	25.865
		JA	1.662	2.146	2.566	1.627
		PU	7.507	5.864	5.961	5.876
		PX	37.684	24.673	23.279	34.369
		PX Deo	6.780	3.532	5.229	4.040
	4	SP	1.539	1.626	1.649	1.701
		TC	1.000	747	241	860
		GB	52.994	47.868	48.892	49.632
		JA	735	702	751	745
		MD	2.323	3.749	(565)	3.173
		PU	5.726	5.865	6.182	5.339
		PX	29.140	27.276	27.333	28.927
		PX Deo	6.168	4.719	4.238	4.715
		SP	542	523	453	571
		TC	1.611	1.706	1.915	1.461
Jumlah			**1.266.744**	**1.215.577**	**1.218.127**	**1.209.521**