OPERATIONS AND SUPPLY CHAIN MANAGEMENT
Vol. 16, No. 2, 2023, pp. 293 - 310
ISSN 1979-3561 | EISSN 2759-9363

Predictive Analytics to Improve Inventory Performance: A Case Study of an FMCG Company

Patdono Suwignjo
Department of Industrial and Systems Engineering,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Lisda Panjaitan
Department of Industrial and Systems Engineering,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Ahmed Raecky Baihaqy
Department of Industrial and Systems Engineering,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Ahmad Rusdiansyah
Department of Industrial and Systems Engineering,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
Email: arusdian@ie.its.ac.id *(Corresponding Author)*

ABSTRACT

Predictive analytics is a methodology used to predict the outcome of future events with the use of historical data. Predictive analytics comes in very handy in various fields such as finance, manufacturing, healthcare, and even supply chain. Not only in those fields, but predictive analytics is also useful in managing inventory. However, we find that there is a lack of studies focusing on the implementation of predictive analytics to predict inventory status (overstock, understock) by considering inventory level and demand forecast. This study is inspired by a real-world problem at one of the largest FMCG companies in Indonesia. With so many product types to manage, this company often faces problems of understocked and overstocked inventory. This study attempts to solve that problem by employing big data and predictive analytics approaches. The gradient boosting model is used because it is an improvement of the decision tree model. The data that are used as predictors are inventory level, inventory week cover, historical sales, and demand forecast. The target variable for classification is inventory status which is divided into three classes, namely understock, normal, and overstock. Meanwhile, the target variable for the regression model is the amount of understock/overstock. The result of the classification model has an accuracy of 0.84 for category 1 products, 0.76 for category 2 products, and 0.74 for category 3 products. While the result of the regression model is an R^2 of 0.89 for category 1 products, 0.76 for category 2 products, and 0.74 for category 3 products. The data that comes from the prediction model are visualized in a dashboard. The visualization dashboard displays the data using heatmaps and line graphs, so the information can be used for further analysis.

Keywords: *big data, FMCG, gradient boosting, inventory, predictive analytics*

1. INTRODUCTION

This section presents the introduction to predictive analytics and the reason of carrying out the study followed by explanations of the case study.

1.1 Predictive Analytics

This paper presents a study in which predictive analytics is used to predict inventory status by considering the forecast of incoming demand. Predictive analytics is one of the emerging topics lately. Predictive analytics is a method to predict the future outcomes and performance by using a statistical method and modeling techniques (Halton, 2023). Predictive analytics requires historical data as the input (Cote, 2021) so that by using the historical data, the pattern can be discovered and used as the basis of making predictions or estimations (Halton, 2023). Several tools or methods that are usually exploited to carry out predictive analytics include decision trees, regression, neural networks, cluster models, and time series modeling (Halton, 2023). Sometimes, artificial intelligence (AI) is used to help perform predictive analytics (Flynn, 2023). As a method, predictive analytics has some benefits and drawbacks. For example, predictive analytics allows us to save much time in terms of decision-making and gaining insight. However, since predictive analytics requires historical data as input, data gathering might cost a lot since predictive analytics requires a large size of data to yield more accurate results (Flynn, 2023).

The implementation of predictive analytics is so wide it can be applied to almost all fields. Predictive analytics can be used in the civil engineering field to predict and prevent floods (Sun *et al*., 2023). It is also possible to implement it in the healthcare field to predict the impact of the pandemic on hospital operations (Amrami *et al*., 2021) and the volume of inpatients (Pai and Dissanayake, 2022). Wang *et al*. (2021) reported the application of predictive analytics in the manufacturing industry, especially to improve the precision of cycle time estimation. Finance becomes one of the fields with the most implementation of predictive analytics. For instance, it is employed to predict earnings, stock price, return, credit score (Broby, 2022), and debt collection rate (Sanchez *et al*., 2022).

Not only those mentioned fields, but predictive analytics is also used in the supply chain field. Bag *et al*. (2022) used it to predict supply chain visibility and resilience with consideration of extreme weather events as the disrupting effect. Zheng *et al*. (2023) also used predictive analytics to assess the flexibility of supply chain networks. The implementation of predictive analytics to forecast the demand of retail industries can be found in Falatouri *et al*. (2022). Lastly, it is quite interesting to report that predictive analytics can predict vendor incoterm, especially in the pharmaceutical supply chain (Detwal *et al*., 2023). It is also essential to report on the implementation of predictive analytics in inventory management as a subset of supply chain management. Andrade and Cunha (2023) implemented predictive analytics to solve inventory problems by considering demand forecasts. Namir *et al*. (2021) took the benefit of predictive analytics to help in decision-making such as buying or selling items with the goal of maximizing revenue. Predictive analytics implementation is also found in Theodorou *et al*. (2023) to predict inventory performance in terms of inventory cost. The topic of implementation of predictive analytics to predict and determine whether the inventory is overstocked or understocked once compared to the forecast of demand is indeed interesting to inspect. However, there is a lack of study established especially in the application of predictive analytics to predict the inventory status by comparing the inventory level against the demand forecast.

1.2 Case Study

In this study, we would like to extend the implementation of predictive analytics to predict the inventory status by considering two aspects to be compared, namely inventory level and demand forecast. This study is inspired by a real-world problem occurring in one of the largest fast-moving consumer goods (FMCG) companies in Indonesia.

As an industry that moves quickly and has low prices, the FMCG industry must be able to compete in the market. One of these forms of competition can be judged by the availability of products in the market. If the product of a brand is not available in the market, then there is a possibility that consumers will move to another brand that is always available in the market and the company that owns the brand will lose consumers. In running a business, companies usually have the main goal of making a profit. However, to make a profit, the company must ensure that the product is available in the market. If the product is not available on the market, then the consumer cannot buy the product and the company does not make a profit.

Companies must have an effective and efficient supply chain system to ensure the availability of products in the market, such as planning, material acquisition, production, storage, and product delivery to consumers. Supply chain, as a vital factor in increasing competitiveness, regulates the flow of information, products, and cash in the overall business process (Nurkhasanah, 2021). One of the factors in the supply chain that has an important role in ensuring product availability in the market is inventory management. Companies engaged in manufacturing usually have several types of inventories, namely raw materials, work in process, and finished goods. Inventory is one of the largest assets for the company, namely in the form of materials and products that are stored in large quantities. In the end, the inventory will be converted into cash when the product has been received by consumers. As an asset (Sharma, 2023), inventory should have high liquidity. This liquidity can be judged by the movement of inventory in the warehouse. If inventory takes a long time to come out of the warehouse, then the company will also find it difficult to get cash flow. Keeping too much inventory is also not a good thing since it is considered as one of seven wastes (McBride, 2003).

Product availability in the market can also be one of the measures in terms of service level. One of the factors that affect this service level is the availability of inventory in the warehouse. When a company strives to increase the service level, sometimes the company piles up as many products as possible in the warehouse. However, this condition can increase costs in product storage and product damage, perishable products especially have a high risk of damage and the potential to expire before they reach consumers. And when companies strive to improve cost efficiency with the minimum possible amount of inventory, there will be a possibility of backordering or even not being able to meet consumer demand. Sometimes consumers understand the company by allowing the company to meet the demand in later periods. However, if such circumstances occur frequently and are not handled properly, consumers will turn to other products and the company may lose consumers.

This FMCG company under this study has a wide range of products manufactured to fulfill customer's daily needs such as food and refreshment (savory, tea & beverages, ice cream), household care (household cleaning, fabric cleaning, fabric conditioner), and personal care (skincare, deodorant, haircare, skin cleansing liquid, oral care, soap bar, lotion)The main customers of this company are modern trade (minimarkets, supermarkets), general trade (stalls, wholesalers, stores), distributive trade (wholesalers), and e-commerce. With so many types of products that must be managed, this FMCG company often faces understock and overstock problems. From the data for 12 weeks, only 15% of household products have a normal amount of inventory. 85% of the products are in an abnormal condition, 32% understock, and 53% overstock. The condition of inventory every week from week 1 to week 12 of 2022 can be seen in **Figure 1** below.

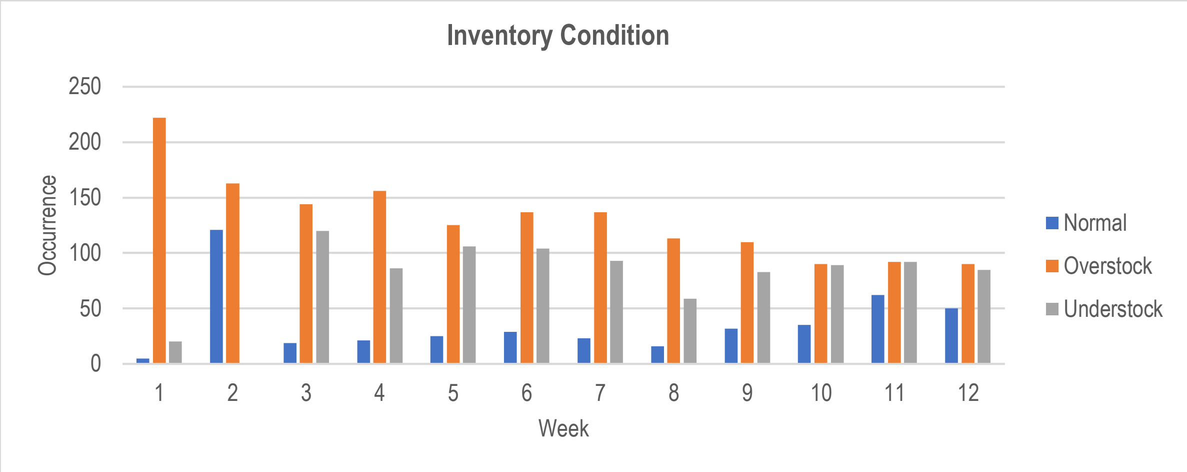


Figure 1 Finished goods inventory condition

From the chart above can be seen that understock and overstock problems occur almost every week. This problem should not be ignored because it can cause risks such as an increase in inventory costs (in case of overstock) and the potential for lost sales (in case of understock). Therefore, we conduct research to provide recommendations for solutions to these problems. The research was conducted with a predictive analytics approach, which uses the classification method to predict the occurrence of understock and overstock, especially in inventory finished goods. The method that is going to be used to predict inventory status is gradient boosting.

2. LITERATURE REVIEW

Inventory is crucial for a business and becomes a necessity for accommodating the imbalance between the inventory of goods in a location and the amount of consumption or sales in the corresponding location (Muckstadt & Sapra, 2010). By that statement, it can be inferred that the main function of inventory is to serve customers. Hence, having a sufficient inventory is necessary to fulfill the incoming customer's needs (Toomey, 2000). Owning a high level of inventory might be good for keeping the service level high. However, keeping too much inventory is unhealthy since it will inflate the total cost related to inventory. Hence it is necessary to keep in balance between the level of inventory, service level towards customers, and total inventory cost.

There have been so many studies regarding inventory management problems. Most of them are conducted to aim for several objectives such as cost minimization, profit maximization, service level improvement, and planning accuracy. To achieve those goals, the studies did experiments on several aspects such as inventory or ordering policy and strategy for perishable items (Gioia *et al*., 2023; Guo *et al*., 2018; Gong *et al*., 2022), retail products (Rios and Vera, 2023; Sridhar *et al*., 2021), and spare parts or OEM items (Rinaldi *et al*., 2023; Zhang *et al*., 2019). The other aspects that are being changed include inventory level (Wolbert, 2013; Ohmori and Yoshimoto, 2020), inventory allocation (Bretthauer *et al*., 2010), product pricing (Rios and Vera, 2023), and supplier selection (Wibowo *et al*., 2021). Several methodologies are introduced to solve this kind of inventory management problem. For example, Dong *et al*. (2023) introduced a model predictive control (MPC) to solve the change in the forecast of demand so that total cost would be minimized. Zietsman & van Vuuren (2023) also exploited a decision support system (DSS) framework to solve their problem in inventory planning. Guo *et al*. (2018) utilized a risk management framework and used conditional value at risk (CVaR) in their study. Some other more popular approaches to solve inventory problems are optimization through various methods (Gong *et al*., 2022; Zhang *et al*., 2019; Dong *et al*., 2011), simulation (Sridhar *et al*., 2021; Rinaldi *et al*., 2023), and combination of simulation and optimization or Sim-Opt (Gioia *et al*., 2023; Wibowo *et al*., 2021).

Data mining refers to a process performed to reveal patterns and interesting insight into a large-sized set of data. Data mining allows us to gain a story behind data in a relatively quicker duration. According to Larose and Larose (2014), some functions of data mining include description, estimation, prediction, classification, clustering, and association. In its application, data mining requires cross-disciplines including statistics, machine learning, database system and warehouse management, and information retrieval (Han *et al*., 2011). Machine learning (ML) refers to a discipline of artificial intelligence that provides machines with the ability to automatically learn the data or past experiences which identifying patterns to construct a prediction with little to no human intervention (Kanade, 2022). There are several types of machine learning including supervised ML, unsupervised ML, semi-supervised ML, and reinforcement learning. ML could be implemented in various fields such as the healthcare industry, finance sector, retail sector, social media platform, etc.

Machine learning is also utilized in solving inventory problems. In this case, machine learning is primarily used to perform prediction or forecast. Lolli *et al*. (2017) exploited decision tree and Random Forest to perform inventory classification for electric resistance products. Seyedan *et al*. (2023) used deep learning to perform demand prediction and inventory management for retail items with the goal of cost minimization and service level improvement. Inventory management problems are also solved by Mahfouz *et al*. (2023) by using Federated Learning (FL). Shirisha *et al*. (2022) also used machine learning, especially predictive analytics, to predict the demand for agricultural products so that the inventory could be managed in such a way that the demand could be fulfilled. Reinforcement learning is also one of the methods in machine learning. This method is employed by Ahmadi *et al*. (2022) to determine the optimum ordering quantity by considering the remaining shelf life of a perishable pharmaceutical goods and Demizu *et al*. (2023) to calculate the optimal inventory level to be kept for new products with short life cycle so that the business is still profitable with maintain high service level. Yudhistyra *et al*. (2020) and Aamer *et al*. (2021). conducted a literature review regarding big data studies and ML implementation on logistics and supply chain.

Predictive analytics is an analysis that predicts future results using data from the past. The data in the past was processed using statistical modeling, data mining techniques, and machine learning. Businesses usually use these predictive analytics to identify future risks or opportunities (IBM, 2022). There are various types of models that can be used in making predictive analytics, such as linear models (linear regression, linear discriminants), logical approaches (decision trees), probabilistic approaches (naïve bayes), complex models (support vector machines) (Torgo, 2017). One of the predictive analytics methods that can be used is gradient boosting. Gradient boosting is a machine learning approach that serves to create an accurate predictor by combining weak and less accurate rules. Gradient boosting is a machine learning technique for regression and classification problems, resulting in a prediction model in the form of a collection of weak prediction models (usually decision trees). In this method, the model is built gradually by studying the prediction errors in the previous model.

It is interesting to report that gradient boosting is also applied to some research in the field of inventory management. One report by Andrade and Cunha (2023) in which an extension of gradient boosting, namely Extreme Gradient Boosting (XGBoost) was employed to solve the inventory problem by addressing the demand forecast in the retail. Another exploitation of the XGBoost method is reported by Namir *et al*. (2021) with the consideration of linking production to demand. Their research aimed to help stakeholders in making decisions (buy, sell, or hold) as well as the quantity to buy or sell to maximize revenue. Theodorou *et al*. (2023) also utilized gradient boosting in their research to predict their inventory performance measured by inventory cost. To enrich the utilization of gradient boosting in the inventory management field, we conducted this research to reach that goal. This research aims to predict the inventory level compared to upcoming demand to determine if the predicted inventory level will be understocked or overstocked. To conduct this research, the model of Gradient Boosting will be exploited as part of the predictive analytics method.

3. RESEARCH METHODOLOGY

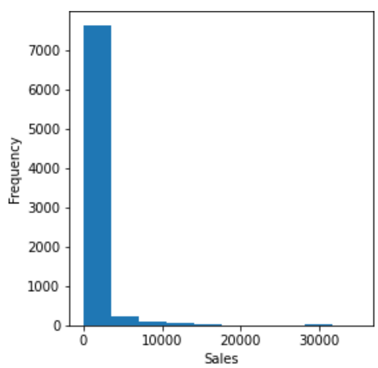
In this section, we will present the methodology and procedure used in this study. This section will emphasize data collection and processing, and also the gradient boosting model development.

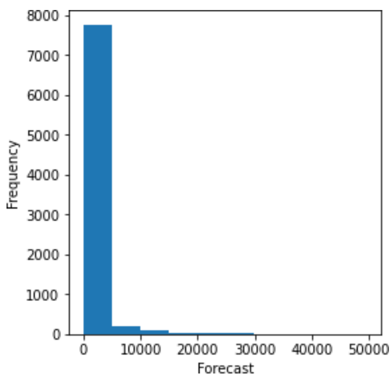
3.1 Data Collection and Processing

The first step to perform this research is collecting data. The data collected consists of inventory level, forecasted demand, actual sales based on historical data, inventory week coverage, inventory status, and total understock or overstock. Based on this data, we set the data of inventory level, forecasted demand, actual sales, and inventory week coverage as predictors. On the other hand, we set inventory status as well as total understock or overstock as target data.

Once the data is collected, it is necessary to perform data cleaning to ensure that the data is noise-free hence it is ready to be processed for further analysis. Data cleaning is performed by detecting the missing value and taking action upon it by either deleting the data or filling the missing value.

After we ensure that our data is cleaned, then the next step is to perform data exploration. In this step, we will perform further observation and exploration of our dataset to understand better about our data. **Figure 2** below shows the distribution of the actual data which are assigned as the predictor (sales, demand forecast, inventory level, week coverage (max IPM)).





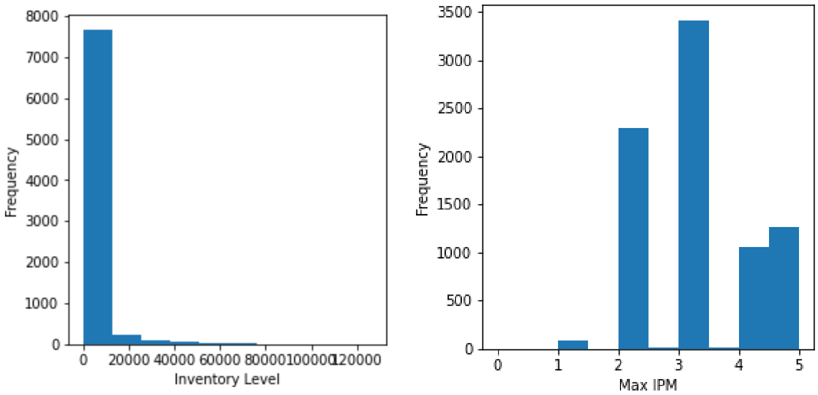


Figure 2 Distribution of actual data

From **Figure 2** above, we can learn that 3 out of 4 data are right skewed which are sales, forecasted demand, and inventory level. Hence, data transformation is required so that all data are presented on the same scale. This type of data transformation is also called data scaling. This activity is intended to avoid anomalies in data and improve data consistency.

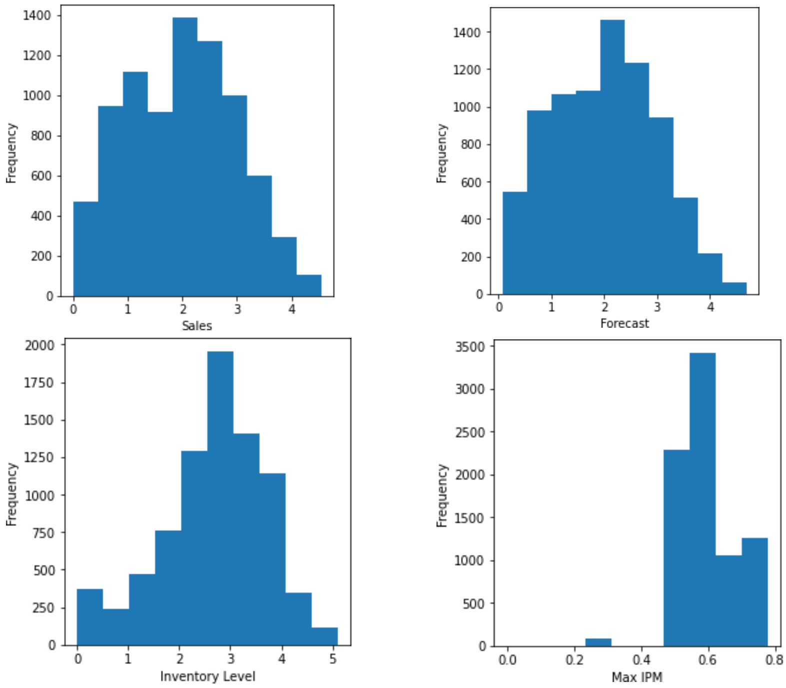


Figure 3 Distribution of transformed fata

Figure 3 above shows the predictor data distribution after being treated by data transformation. After transformation, we can see that our data is already normally distributed. Hence, the data is ready to be processed in the next step namely data conversion. Data conversion is a step in which numerical data will be transformed into categorical data.

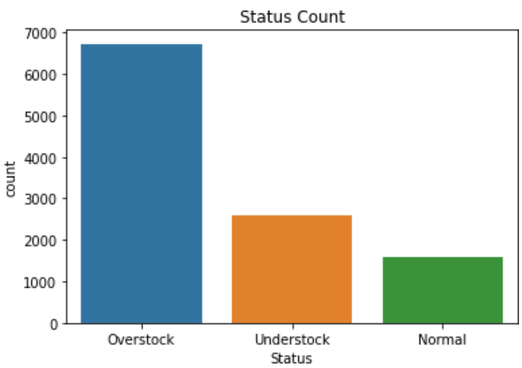


Figure 4 Frequency of data

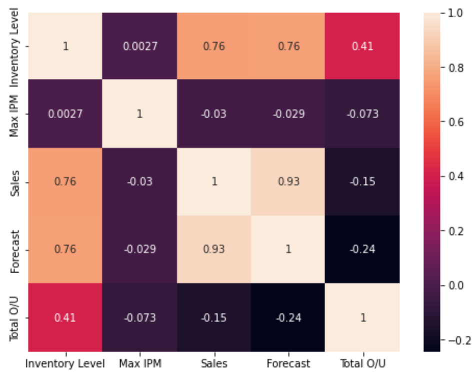


Figure 5 Correlation of data

The following data exploration to perform is to inspect the frequency of the inventory status. Here, in this case, the inventory status is divided into 3 distinct categories namely normal, understock, and overstock. The frequency distribution of each category is shown in **Figure 4**. From **Figure 4**, we can learn that the amount of data for each class is unbalanced where overstock status dominates the other. An unbalanced amount of data on a classification model can result in bias in predicting data against the majority class. A heatmap visualization is also presented by **Figure 5** to present the correlation value of each column which is expressed in numerical data type. The Sales and Forecast columns have a very good correlation, which is 0.93. In addition, Sales and Inventory Level as well as Forecast and Inventory Level also have a good correlation, which is 0.76.

3.2 Gradient Boosting Model Development

In this section, we will present the process of developing our prediction model.

3.2.1 Model Definition and Algorithms

The algorithm model used in the creation of classification and regression models is Gradient Boosting. This model is created using several decision trees, where each decision tree built learns from mistakes in the previous decision tree. Train dataset, which is 80% of the dataset, is used to build the model. The model is built using a library on the learn skicit, namely sklearn.ensemble. In the classification model, the data used as predictors are Sales, Forecast, Max IPM (Inventory Week Cover), and Inventory Level data. While the target variable is Inventory Status data (understock, normal, overstock). **Figure 6** below shows the algorithm of the classification model that is going to be developed.

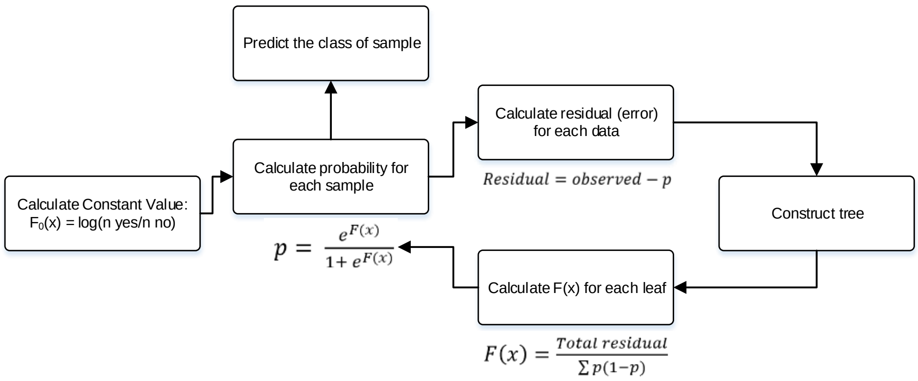


Figure 6 Classification model algorithm

In regression models, the predictor data used is the same as the predictor data in the classification model. However, the target variables used are different, namely the number of deficiencies when understocking and the amount of excess on the overstock. **Figure 7** below shows the algorithm of the regression model that is going to be developed.

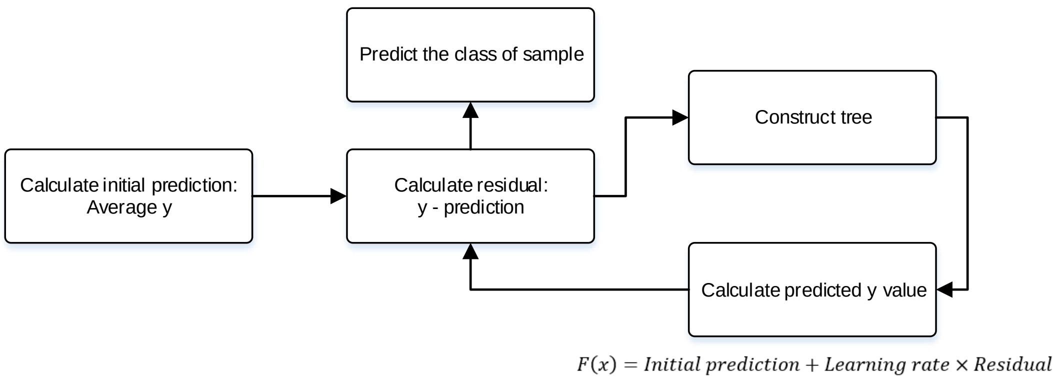


Figure 7 Regression model algorithm

3.2.2 Classification Model Development

As mentioned previously, in the classification model, the predictors include sales, forecast, max IPM, and inventory level while the target variable is inventory status (understock, normal, overstock). The classification model is created by developing 5 different decision trees with a maximum depth of 3 for each decision tree. The data sample used to develop these decision trees is 50 data. **Figure 8** below shows the decision trees developed for the classification model.

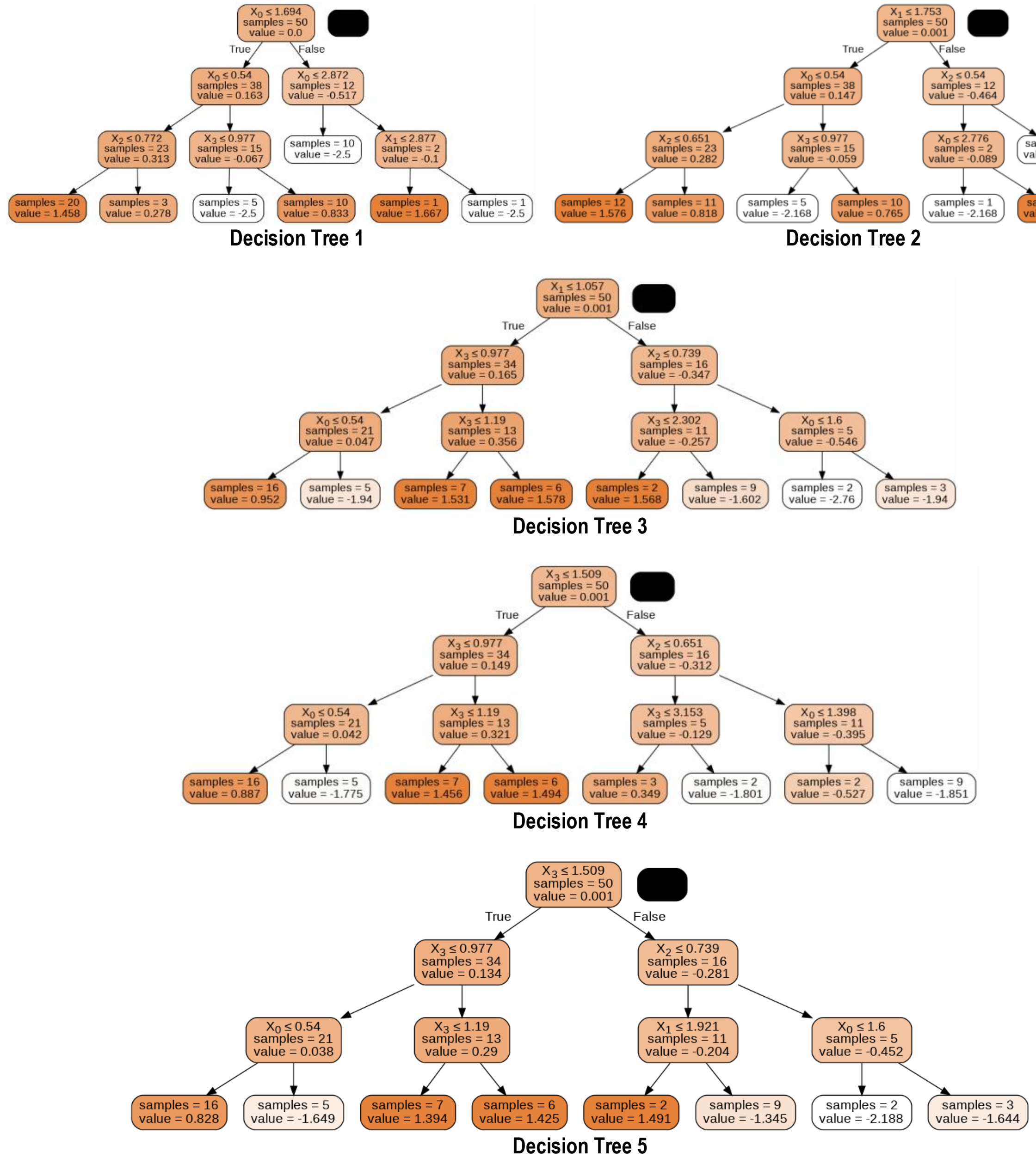
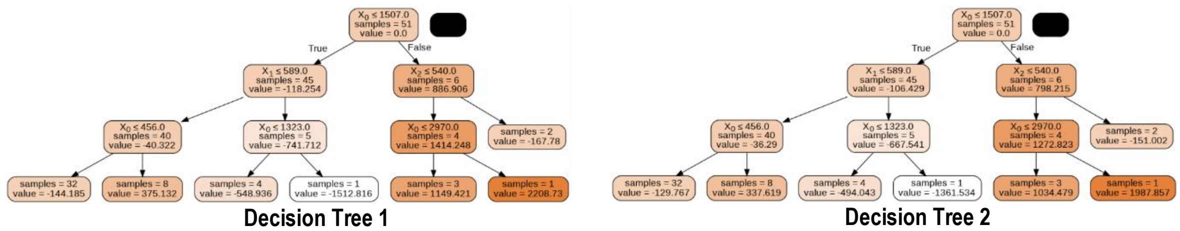


Figure 8 Decision trees of classification model

Once the decision trees are developed, the testing data will be inserted into each decision tree and will be calculated its result. The result will be tested against the actual data to measure the validity of the decision tree.

3.2.3 Regression Model Development

The following model to develop is the regression model. As mentioned before, the predictor of the model is the same as the classification model. However, the difference lies on the target variable where the regression model used the total of understock or overstock as the target variable instead of inventory status. The regression model is created by developing 5 different decision trees with a maximum depth of 3 for each decision tree. The data sample used to develop these decision trees is 50 data. **Figure 9** below shows the decision trees developed for the classification model.



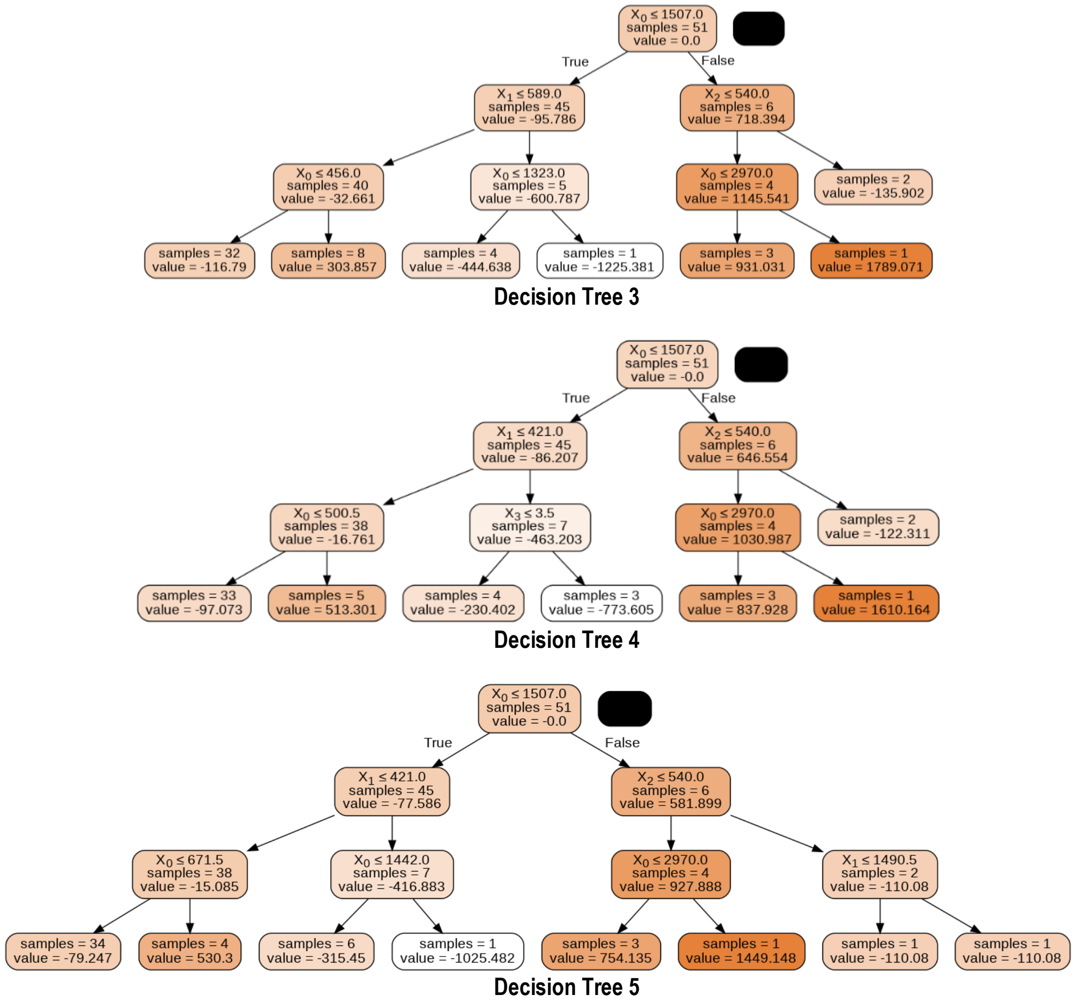


Figure 9 Decision trees of regression model

Once the decision trees are developed, the testing data will be inserted into each decision tree and will be calculated its result. The result will be tested against the actual data to measure the validity of the decision tree.

4. RESULTS AND DISCUSSION

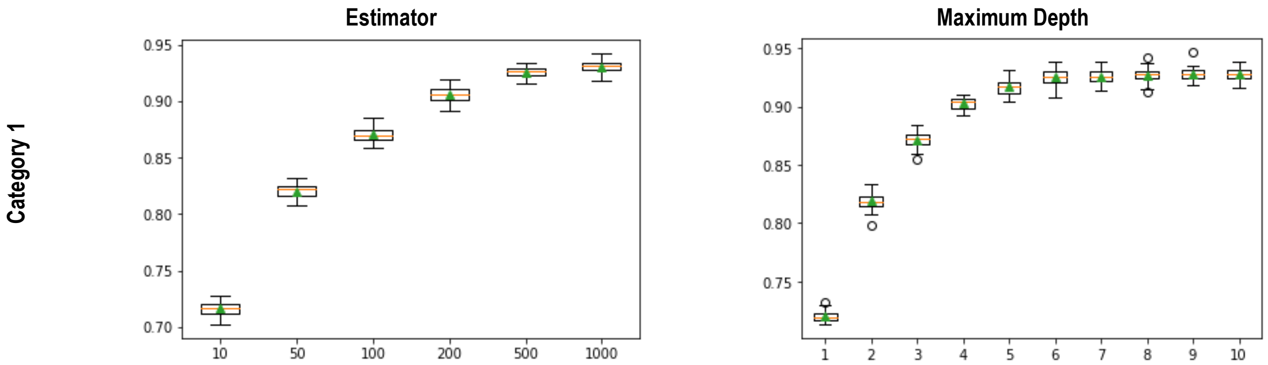
In this part, we will present, analyze, and discuss the results of our research. This section will be more focused on the selection of the best model by doing some comparison, model accuracy measurement, and scenario development and testing.

4.1 Selection of The Best Model

There are several parameters that can be changed to form a model with better accuracy, such as the maximum number of decision trees built, the maximum depth of the decision tree, and others. Before implementing the model, experiments will be conducted by changing these parameters. The experiment conducted is to change the number of decision trees (or called n_estimators) and the maximum depth of the decision tree (or called max_depth). Here is a plot box that shows the accuracy of each model with the number of estimators and the number of depths altered. Both classification and regression models will be tested on data from 3 different product categories of the company.

4.1.1 Classification Model Selection

Figure 10 below shows the boxplot which illustrates the accuracy of classification model in which the number of decision trees (or numerators) and maximum depth are modified.



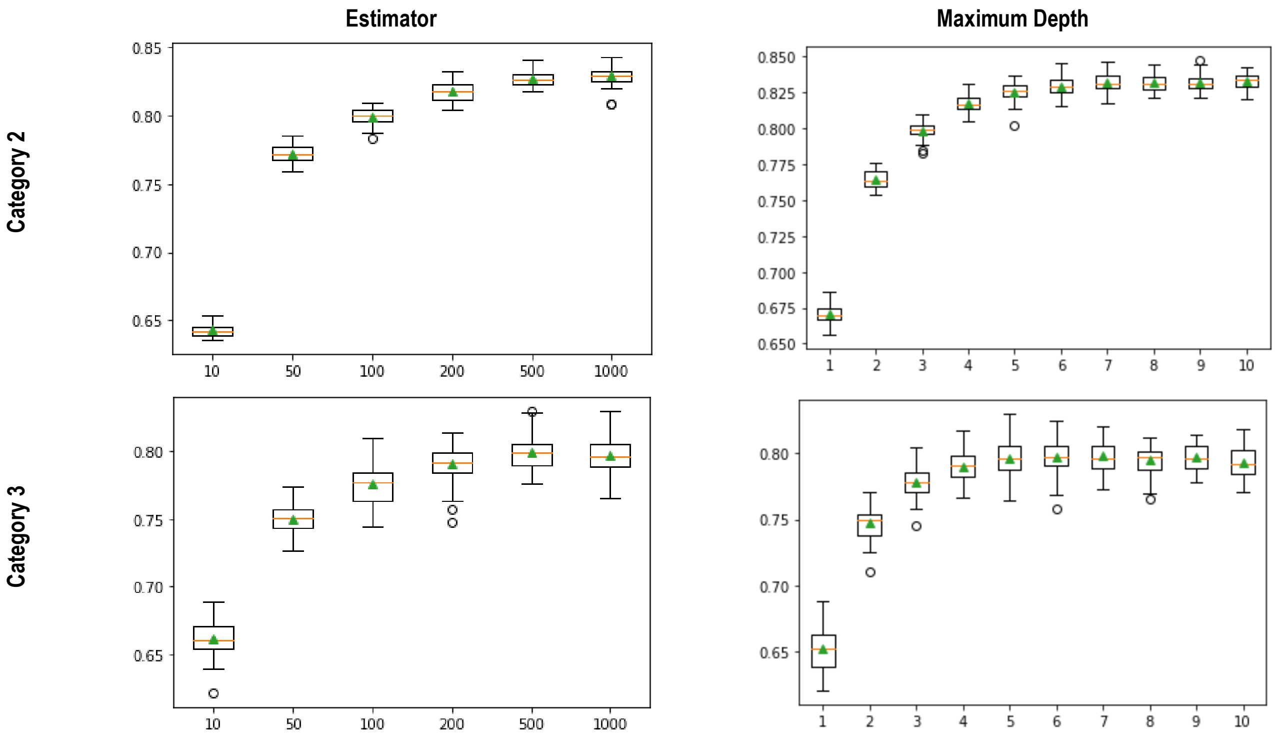


Figure 10 Evaluation of classification model accuracy based on (a) estimators and (b) maximum depth

From **Figure 10**, it is implied that the larger the estimators and the deeper the decision tree is, the more accurate the classification model is. Then, we tried to run the model for each parameter as many as 50 times and we would obtain an average value of accuracy shown by **Table 1** and **2** for estimators and maximum depth consecutively.

Table 1 Classification model accuracy for each product category based on estimators

n_estimator		10	50	100	200	500	1000
Accuracy	Category 1	0.7165	0.8203	0.8703	0.9059	0.9253	0.9306
	Category 2	0.6427	0.7720	0.7993	0.8177	0.8268	0.8290
	Category 3	0.6616	0.7501	0.7759	0.7906	0.7994	0.7967

Table 2 Classification Model Accuracy for Each product Category Based on Maximum Depth

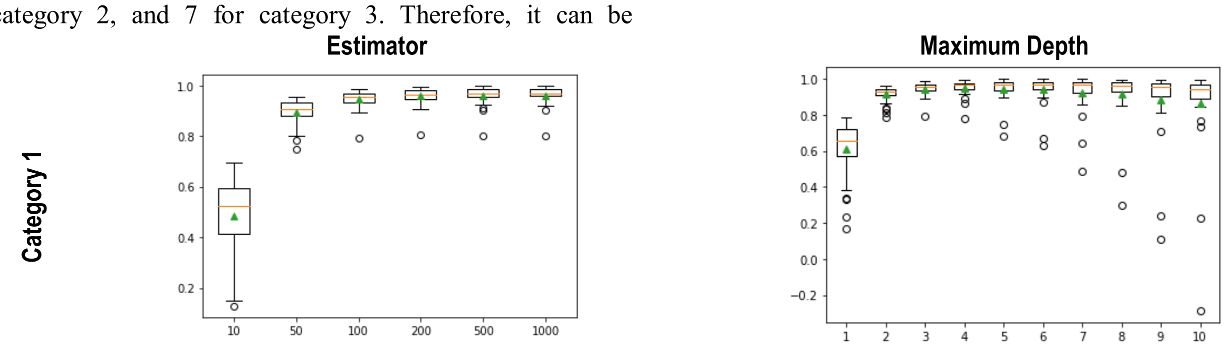
max_depth		1	2	3	4	5	6	7	8	9	10
Accuracy	Category 1	0.720	0.819	0.871	0.903	0.917	0.925	0.926	0.927	0.928	0.927
	Category 2	0.671	0.764	0.799	0.817	0.826	0.829	0.832	0.832	0.832	0.833
	Category 3	0.652	0.747	0.778	0.789	0.796	0.797	0.798	0.795	0.797	0.793

Based on **Table 1**, we tried several different numbers for estimators which are 10, 50, 100, 200, 500, and 1,000. From the result, 1,000 estimators suit the best for data of product category 1 and 2 while for product category 3, 500 estimators work the best.

From **Table 2**, we also tried several different maximum depths of the decision tree starting from 1 to 10. It can be inferred that the maximum depth of the decision tree that yields the highest accuracy is 9 for category 1, 10 for category 2, and 7 for category 3. Therefore, it can be concluded that each product category data will have their own optimal accuracy conditions.

4.1.2 Regression Model Selection

Figure 11 below shows the boxplot which illustrates the accuracy of classification model in which the number of decision trees (or numerators) and maximum depth are modified.



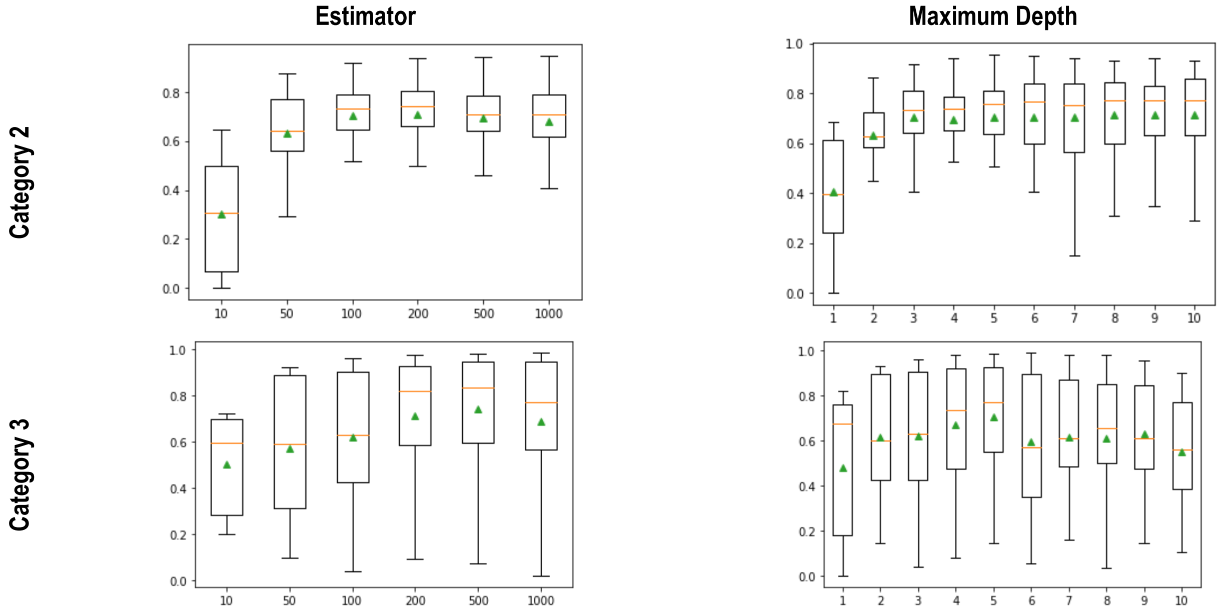


Figure 11 Evaluation of regression model accuracy based on (a) estimators and (b) maximum depth

From **Figure 11** above, we can learn that the more estimators implemented and the deeper the decision tree is, the higher accuracy of the model will be. However, at some point, there is an accuracy decline detected which is caused by overfitting. Then, we tried to run the model for each parameter as many as 50 times and we would obtain an average value of accuracy shown by **Table 3** and **4** for estimators and maximum depth consecutively for the regression model.

Table 3 Regression model accuracy for each product category based on estimators

n_estimator		10	50	100	200	500	1,000
R^2 Score	Category 1	0.4828	0.8930	0.9443	0.9574	0.9607	0.9606
	Category 2	0.3001	0.6314	0.7033	0.7098	0.6953	0.6818
	Category 3	0.5051	0.5707	0.6191	0.7117	0.7415	0.6918

Table 4 Regression model accuracy for each product category based on maximum depth

max_depth		1	2	3	4	5	6	7	8	9	10
R^2 score	Category 1	0.608	0.915	0.945	0.951	0.943	0.937	0.926	0.908	0.885	0.858
	Category 2	0.408	0.632	0.707	0.697	0.706	0.705	0.704	0.713	0.713	0.716
	Category 3	0.483	0.614	0.621	0.671	0.706	0.597	0.614	0.613	0.629	0.549

Based on **Table 3**, we tried several different numbers for estimators which are 10, 50, 100, 200, 500, and 1,000. From the result, product category 1 and 3 will have the highest accuracy with 500 estimators and 200 estimators for product category 2.

From **Table 4**, we also tried several different maximum depths of the decision tree starting from 1 to 10. It can be inferred that the maximum depth of the decision tree that yields the highest accuracy is 6 for category 1, 10 for category 2, and 5 for category 3. Therefore, it can be concluded that each product category data will have its own optimal accuracy conditions like the classification model previously.

4.2 Measurement of Model Accuracy

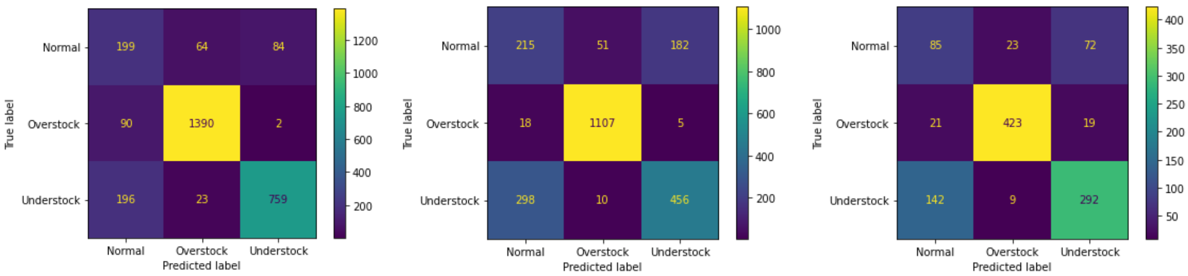
Measurement of the model accuracy is conducted to find out whether the model that has been created can predict the target label well. A model is able to predict well if it has an accuracy of more than 70%. To measure the accuracy of the model, we divided our dataset into a train dataset and a test data set. **Table 5** below shows the detail of how we divide our dataset of each product category into train dataset and test dataset.

Table 5 Data*set* allocation for Each Product Category

Product Category	Train Dataset	Test Dataset	Total Dataset
Category 1	8,096	2,807	10,903
Category 2	11,957	2,342	14,299
Category 3	3,983	1,086	5,068

4.2.1 Classification Model Accuracy

The selected model from the previous section later will be tested for its accuracy in this section. The model is selected based on the performance measured by accuracy. For instance, the model which is used for product 1 will have 1,000 estimators and 9 maximum depth of the decision tree, and so on for the remainder. As for the classification model, the accuracy will be measured using a confusion matrix. **Figure 12** shows the confusion matrix of each category.



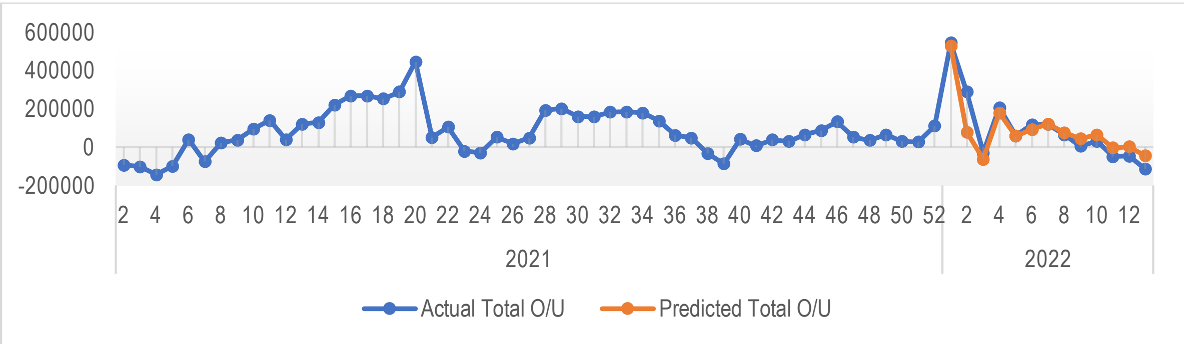
(a) Category 1 (b) Category 2 (c) Category 3

Figure 12 Confusion matrix of classification model of each category

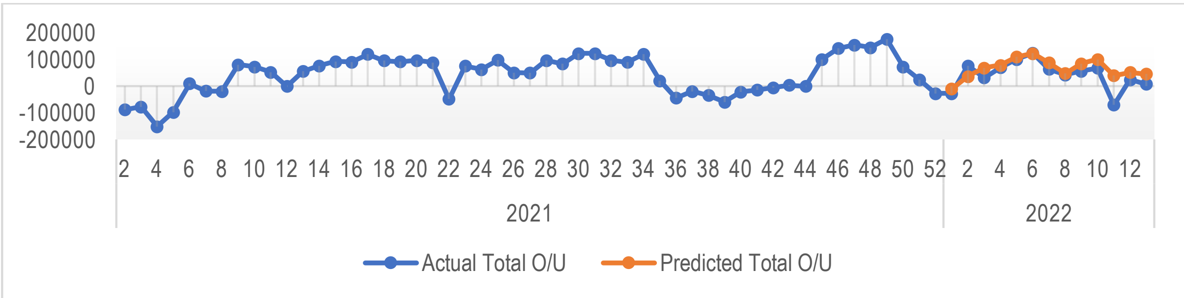
The accuracy of the model is measured by counting the amount of data in which the prediction match with the true label. For example, the number of correct predictions for product 1 is 2,348 data (199 normal, 1390 overstock, and 759 understock). Hence, the model could predict 2,348 outcomes correctly out of 2,807 data. Therefore, the accuracy of category 1 is 0.84. By doing the same thing for the remainder of the category, we obtained that the accuracy of the classification model for category 2 and 3 are 0.76 and 0.74 respectively.

4.2.2 Regression Model Accuracy

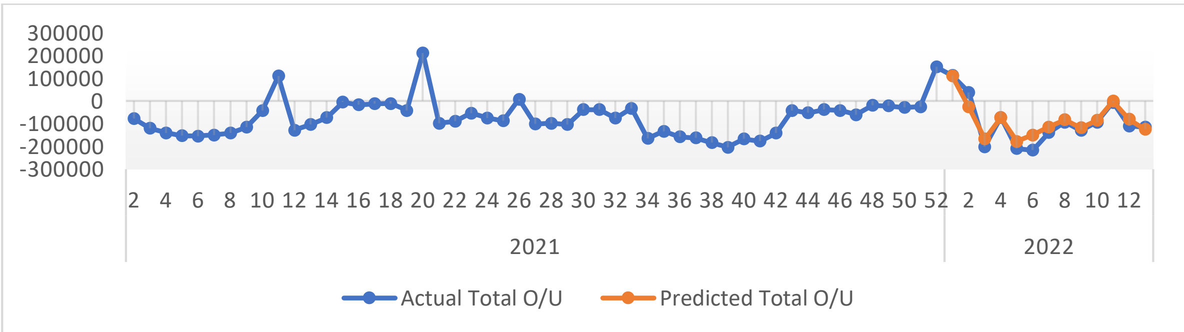
For the regression model, the accuracy is tested by comparing the predicted value against the actual value of overstock or understock visualized in a line chart. Later, the model accuracy of each product category will be measured using R-Square (R^2). **Figure 13** below shows the prediction produced by the regression model based on historical data.



Category 1



Category 2



Category 3

Figure 13 Prediction result using regression model for each category

Figure 13 shows the results of predictions for 2022 based on models that have been created using data from 2021. On the chart, the prediction results (line graph orange) follow the shape or pattern of the actual data (blue line graph). The R^2 for the regression model of category 1, category 2, and category 3 are 0.89, 0.76, and 0.74 consecutively. These values are considered pretty good considering that our threshold of declaring a model is good is having an accuracy rate higher than 70% or 0.7.

4.3 Identify Understock and Overstock Scenarios

The scenario of understock and overstock is obtained from the results of the classification model prediction. The following is a mapping of understock and overstock occurrence and the data range for each column designed for product category 1 shown by **Table 6**.

Table 6 Ranges for each data for category 1

Categorization	Sales	Forecast	Inventory Level	Week Cover
Very low	0-11	0-19	0-17	
Low	12-159	20-200	18-347	1-2
Medium	160-2,000	201-2,000	348-6,474	3-4
High	>2,000	>2,000	>6,474	>5

For the range, we also established in a similar way for category 2 and 3. However, the value of the range might differ from shown here which is category 1. Next is scenario development. The scenarios are developed by combining the four conditions (i.e., very low, low, medium, high) of those four aspects (i.e., sales, forecast, inventory level, week coverage). By performing this combination, we end up with 256 conditions to be checked if the overstock or understock inventory takes place. **Figure 14** below shows the outcome of combining the condition of predictors where the outcome is the inventory status of overstock or understock. **Figure 14** shows the outcome for category 1 specifically.

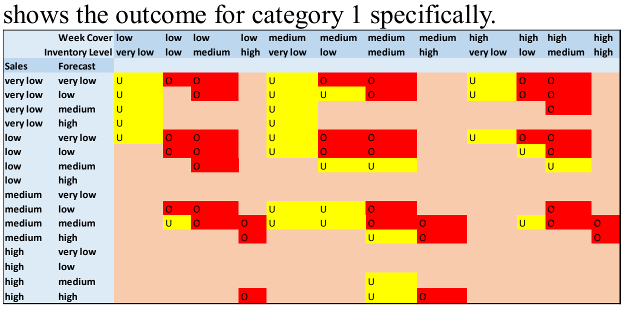


Figure 14 Summary of inventory status scenarios of category 1

From **Figure 14** above, the letter U signifies Understock, and the letter O signifies Overstock. From the picture can be seen the areas that show the occurrence of understock and overstock. Most of the understock occurrences are in inventory-level areas with very low and low conditions. Overstock occurrences are in sales areas with very low to medium conditions, while inventory levels are in low to high conditions. We perform a similar procedure and generate a similar figure like **Figure 14** above dedicated for category 2 and 3. We found that, in general, the result for category 2 is somewhat like category 2 where only overstock and understock conditions were generated. However, we found interesting findings on the result of category 3 where there is a normal status generated instead of only overstock and understock status. Also, most of the understock occurrences are in inventory-level areas with very low and low conditions. Overstock occurrences are in sales areas with very low to medium conditions, while inventory levels are in low to high conditions.

To have more detailed information upon what happened from **Figure 14**, **Table 7** is presented. **Table 7** below will show a summary of the scenario along with the probability of the occurrence of that scenario designed for category 1.

Table 7 Summary of inventory status scenarios of category 1

Sales	Forecast	Week Cover	Inventory Level	Status	Probability
Low	Very Low	High	Very Low	Understock	1
Medium	Low	Medium	Very Low	Understock	1
Very Low	High	Low	Very Low	Understock	1
Very Low	High	Medium	Very Low	Understock	1
Very Low	Low	High	Very Low	Understock	1
Very Low	Low	Low	Very Low	Understock	0.952380952
Medium	Medium	Medium	Low	Understock	0.941176471
Medium	Medium	High	Low	Understock	0.928571429
Very Low	Medium	Low	Very Low	Understock	0.923076923
Very Low	Low	Medium	Very Low	Understock	0.916666667
Very Low	Medium	Medium	Very Low	Understock	0.913043478
Medium	Medium	Medium	Very Low	Understock	0.9
Low	Very Low	Medium	Very Low	Understock	0.875
Low	Medium	Medium	Low	Understock	0.857142857
Low	Very Low	High	Medium	Overstock	1
Low	Very Low	Low	Medium	Overstock	1
Low	Very Low	Medium	Medium	Overstock	1
Medium	Medium	Low	High	Overstock	1
Very Low	Low	Low	Medium	Overstock	1
Very Low	Low	Medium	Medium	Overstock	1
Very Low	Very Low	High	Medium	Overstock	1
Very Low	Very Low	Low	Medium	Overstock	1
Very Low	Very Low	Medium	Medium	Overstock	1
Low	Low	Low	Medium	Overstock	0.977777778
Low	Very Low	Low	Low	Overstock	0.96875
Very Low	Very Low	Low	Low	Overstock	0.963636364
Medium	Low	Low	Medium	Overstock	0.962962963
Medium	Medium	Medium	High	Overstock	0.952380952
Low	Very Low	Medium	Low	Overstock	0.898305085
Very Low	Very Low	Medium	Low	Overstock	0.851351351
Low	Low	Medium	Medium	Overstock	0.807407407

From **Table 7** above, there is a combination with the highest probability. In this case, we will showcase the occurrence whose probability is higher than 0.8. There is a probability of the 30 highest combinations, which are 14 combinations for understock occurrences and 16 combinations for overstock occurrences. In addition, there are several combinations with probability 1. This explains that in combination, the model predicts the same occurrence. For example, from 20 data with sales conditions: low; forecast: very low; week cover: high; and inventory level: low, the model predicts all these data as understock, and the results are validated correctly according to the actual data.

There is also a probability of less than 1, such as the example of 0.807 in overstock conditions. This explains that out of 20 data with sales conditions: low; forecast: low; week cover: medium; and inventory level: medium, the model predicts 16 data as overstock and validated according to the actual data. Meanwhile, the other 4 data are correctly predicted as understock or overstock or predicted as overstock but do not match the actual data.

Table 8 Summary of inventory status appearance on all categories

Category	Understock		Overstock		Normal	
	App.	*Hi Prob.*	*App.*	*Hi Prob.*	*App.*	*Hi Prob.*
Category 1	Yes	1.00	Yes	1.00	-	-
Category 2	Yes	0.77	Yes	1.00	-	-
Category 3	Yes	0.89	Yes	1.00	Yes	0.4

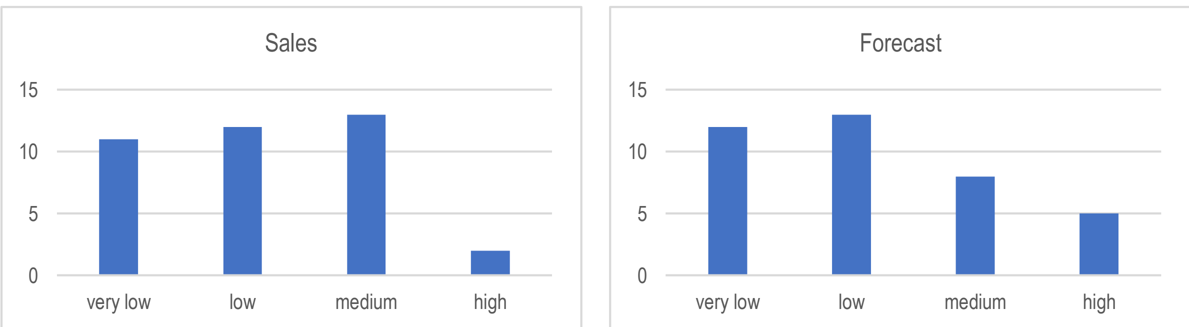
App.: appearance
Hi Prob.: highest probability

We also generate a similar table for category 2 and 3. In summary, as for category 2, there is a probability of 22 highest combinations where 8 of them are for understock, and 14 are for overstock. Since the highest probability of overstock is 1, therefore in this case the overstock will have a higher probability to happen. Meanwhile, the highest probability of understock for category 2 is only 0.77. By this value, it infers that there is still room for a normal or overstock event to take place. Next is an explanation for category 3. The highest probability of understock, overstock, and normal status occur are 0.89, 1, and 0.4 respectively. It is important to note that the probability of the normal status appearing is in the range of 0.2 to 0.4. Hence, we can conclude that its probability is relatively small. The summary of this information is presented in **Table 8**.

The graph below will show the distribution of each category of data in each inventory state for category 1. In **Figure 15** below, understock occurrences mostly occur when sales are in a very low to medium condition. In terms of forecasts, understock occurrences mostly occur when they are medium and low. From inventory week covers, understock occurrences mostly occur when they are in a medium state. And in terms of inventory level, understock occurrences mostly occur when they are very low and low.

We replicate the same procedure for category 2 and 3. As for category 2, the understock condition mostly happens when both sales and demand forecast are low or medium, inventory week cover is medium or high, and inventory level is very low. And for category 3, the understock condition appears when sales and forecast are very low to high (all conditions), inventory week cover is medium or high, and inventory level is very low. The summary of these facts is shown in **Table 9** below.

Figure 15 Distribution of possible understock scenarios of category 1



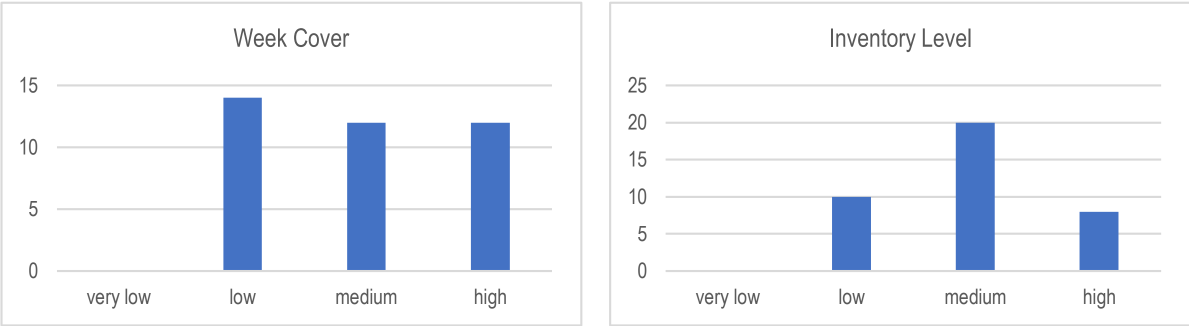


Figure 16 Distribution of possible overstock scenarios of category 1

In **Figure 16** above, overstock occurrences mostly occur when sales are in a very low to medium condition. In terms of forecasts, overstock occurrences mostly occur when they are very low and low. From inventory week cover, overstock occurrences evenly occur when in a low to high state. And in terms of inventory level, overstock occurrences mostly occur when in a medium condition. We perform a similar analysis of overstock condition for category 2 and 3. For category 2, overstock happens when sales are very low to medium, the demand forecast is very low or low, and both inventory week cover and inventory level are low to high. For category 3, overstock takes place when sales are low or medium, demand forecast is very low to medium, inventory week cover is low to high, and lastly inventory level is low to medium. It was also mentioned that the interesting finding emerges from category 3 where normal status appeared. Normal status takes place on category 3 when sales are very low or high, forecast is low to medium, inventory week coverage is medium to high, and this also applies to inventory level. The summary of this information is presented in **Table 10** for overstock and **Table 11** for normal status respectively.

Table 9 Distribution summary of condition leading to understock

Category	Sales	Forecast	Week Cover	Inventory Level
Category 1	Very low - medium	Low - medium	Medium	Very low
Category 2	Low - medium	Low - medium	Medium - high	Very low
Category 3	Low - high	Low - high	Medium - high	Very low

Table 10 Distribution summary of condition leading to overstock

Category	Sales	Forecast	Week Cover	Inventory Level
Category 1	Very low - medium	Very low - low	Low - high	Medium
Category 2	Low - medium	Very low - low	Low - high	Low – high
Category 3	Low - medium	Very low - medium	Low - high	Low - medium

Table 11 Distribution summary of condition leading to normal

Category	Sales	Forecast	Week Cover	Inventory Level
Category 1	-	-	-	-
Category 2	-	-	-	-
Category 3	Very low / high	Low - medium	Medium - high	Medium - high

4.4 Dashboard Development

The model development dashboard was developed using Python and a dashboard for analysis of predicted results from models was developed using Microsoft Excel. This dashboard can be used for different data according to the user's wishes. Here is the display of the dashboard developed shown in **Figure 17**.

Figure 17 Display of prediction model dashboard

The dashboard will run the prediction model using the data entered the system. The result of the prediction that can be seen on the dashboard is the accuracy of the prediction model and the prediction result data. Users can enter the data they want to use in the prediction model. Users can also change the ratio of train data and test data. Users can also change the ratio of train data and test data. Next, the user can select the columns to be used as independent variables and dependent variables. There is a box to select parameters in the model, namely n_estimators (number of decision trees) and max_depth (maximum depth of the decision tree). Furthermore, there are several models that can be chosen. If each box already matches the user's needs, then click run to run the model. The table on the dashboard is a display of data that has been uploaded by the user into the system. There are records which are the total amount of data, train and test is the amount of train data and test data after splitting data. Next, there are the accuracy parameters of the model. There is a boxplot that shows the accuracy of each model. The system will choose the best model for predicting the data that has been uploaded. As in the image above, the best model is Gradient Boosting so the system will use that model to predict the data. In the last column, there is the accuracy of the model, confusion matrix, and prediction result data that can be downloaded by the user. Furthermore, the data can be entered into the dashboard in Microsoft Excel.

A dashboard that can be used to analyze the results of these predictions is created in Microsoft Excel by using Pivot Table. The predicted result data from the model built in Python is fed into Microsoft Excel and visualized using a pivot table. The following is a view of the visualization of the prediction results.

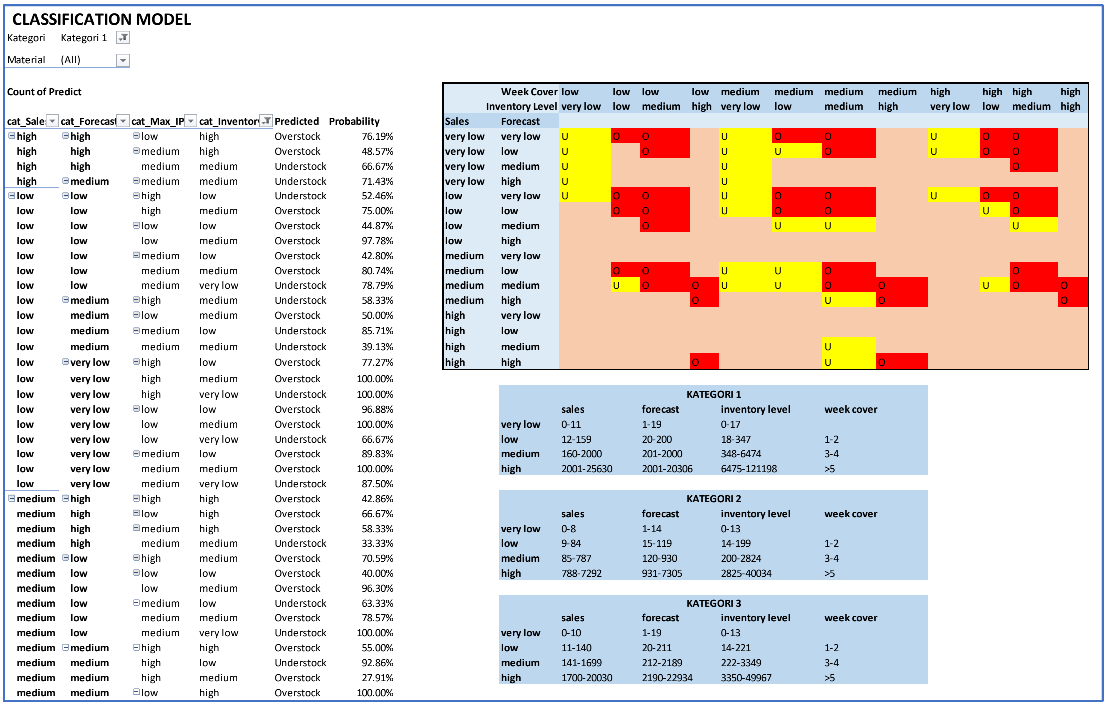
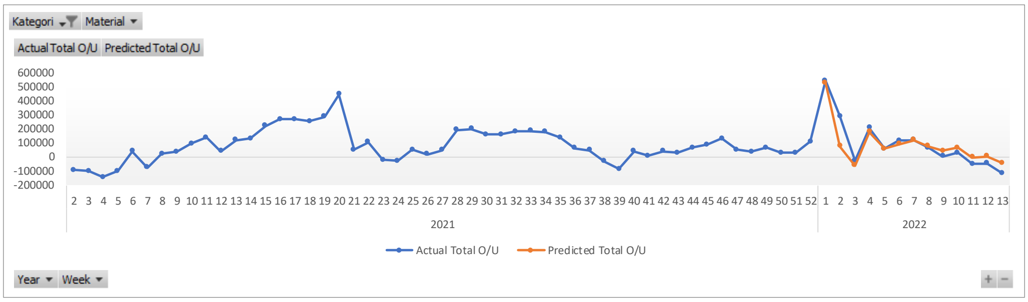


Figure 18 Display of prediction result dashboard using classification model

REGRESSION MODEL



Metrik	Kategori 1	Kategori 2	Kategori 3
R^2	0.8975	0,7609	0,7267
MAE	374.8685	262.9345	735.2275
MSE	3663878.466	1343980.097	13260989.09
RMSE	1914.126032	1159.301556	3641.5641

Figure 19 Display of prediction result dashboard using regression model

There are two types of information that can be retrieved from this dashboard. The regression data graph on the right can be used by the work level in the execution section to find out the predicted number of understocks or overstocks. So, this work level can regulate the number of product shipments. If it is predicted that an understock will occur, then the stakeholder can increase the number of shipments of products to the DC. However, if it is predicted that overstock will occur, then the stakeholder can reduce the number of product shipments to the DC. There is a heatmap which is the result of a classification model that can be used by work levels in the top management to find out product patterns that have understock or overstock occurrences. From the heatmap, the stakeholder can further analyze the factors causing the occurrence of the incident. For example, most understock occurrences occur when the week cover inventory on the product is low, so the management level can conduct research or analyze the week cover level on the product. The stakeholder analyzes whether the week cover on the product needs to be increased or whether there are other factors that cause the understock occurrences to occur.

5. CONCLUSION

This research developed a prediction model using Gradient Boosting. The model developed is used to predict the occurrence of inventory overstock or understock. We developed two types of models which are the classification model (used to predict the inventory status) and the regression model (used to predict the amount of overstock and understock). We assigned four variables as predictors namely sales, demand forecast, inventory coverage week, and inventory level. The model we developed appeared to satisfy our accuracy threshold of 70%. The classification models we developed have an accuracy of 0.84, 0.76, and 0.74 for product category 1, category 2, and category 3 respectively. Additionally, the accuracy of our regression models, measured by R^2, are 0.89, 0.76, and 0.74 for category 1, category 2, and category 3 consecutively. Since the accuracy rates are greater than 70% or 0.7, our model is robust enough to predict another dataset. We suggest extending this research by employing a larger data set, considering more product families, and disaggregating the data by observing products in the level of stock keeping unit (SKU). It is also recommended to consider special events such as seasonality and promotions which affect the behavior of the data.

REFERENCES

Aamer, A. M., Putu, L., Yani, E., Made, I., & Priyatna, A. (2021). Data Analytics in the Supply Chain Management: Review of Machine Learning Applications in Demand Forecasting. *Operations and Supply Chain Management*, *14*(1), pp. 1–13.

Ahmadi, E., Mosadegh, H., Maihami, R., Ghalehkhondabi, I., Sun, M., & Süer, G. A. (2022). Intelligent Inventory Management Approaches for Perishable Pharmaceutical Products in a Healthcare Supply Chain. *Computers and Operations Research*, 147, pp. 1–20. https://doi.org/10.1016/j.cor.2022.105968

Amrami, K., Domnick, R., Heinzen, E., Helfinstine, K., Jayakumar, A., Johnson, P., Knepper, C., Marcelletti, D., Mickelson, M., Rojas, R., st. George, M., Tande, A., Walvatne, K., Wutthisirisart, P., Pollock, B. D., Carter, R. E., Dowdy, S. C., Dunlay, S. M., Habermann, E. B., … Shah, N. D. (2021). Deployment of an Interdisciplinary Predictive Analytics Task Force to Inform Hospital Operational Decision-Making During the COVID-19 Pandemic. *Mayo Clinic Proceedings*, *96*(3), pp. 690–698. https://doi.org/10.1016/j.mayocp.2020.12.019

Andrade, L. A. C. G., & Cunha, C. B. (2023). Disaggregated Retail Forecasting: A Gradient Boosting Approach. *Applied Soft Computing*, *141*, pp. 1-16. https://doi.org/10.1016/j.asoc.2023.110283

Bag, S., Rahman, M. S., Srivastava, G., Chan, H. L., & Bryde, D. J. (2022). The Role of Big Data and Predictive Analytics in Developing a Resilient Supply Chain Network in the South African Mining Industry Against Extreme Weather Events. *International Journal of Production Economics*, *251*, pp. 1-15. https://doi.org/10.1016/j.ijpe.2022.108541

Bretthauer, K. M., Mahar, S., & Venakataramanan, M. A. (2010). Inventory and Distribution Strategies for Retail/E-Tail Organizations. *Computers and Industrial Engineering*, *58*(1), pp. 119–132. https://doi.org/10.1016/j.cie.2009.09.005

Broby, D. (2022). The Use of Predictive Analytics in Finance. *Journal of Finance and Data Science*, *8*, pp. 145–161. https://doi.org/10.1016/j.jfds.2022.05.003

Cote, C. (2021). *What is Predictive Analytics? 5 Examples*. Harvard Business School Online. <https://online.hbs.edu/blog/post/predictive-analytics>, accessed on 21 August 2023.

Demizu, T., Fukazawa, Y., & Morita, H. (2023). Inventory Management of New Products in Retailers Using Model-based Deep Reinforcement Learning. *Expert Systems with Applications*, *229*, pp. 1-11. https://doi.org/10.1016/j.eswa.2023.120256

Detwal, K. P., Soni, G., Jakhar, S. K., Shrivastava, D. K., Madaan, J., & Kayikci, Y. (2023). Machine Learning-based Technique for Predicting Vendor Incoterm (Contract) in Global Omnichannel Pharmaceutical Supply Chain. *Journal of Business Research*, *158*, pp. 1-12. https://doi.org/10.1016/j.jbusres.2023.113688

Dong, H., Zheng, H., & Li, Y. P. (2011). Model Predictive Control for Inventory Management in Supply Chain Planning. *Procedia Engineering*, *15*, pp. 1154–1159. https://doi.org/10.1016/j.proeng.2011.08.213

Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive Analytics for Demand Forecasting - A Comparison of SARIMA and LSTM in Retail SCM. *Procedia Computer Science*, *200*, pp. 993–1003. https://doi.org/10.1016/j.procs.2022.01.298

Flynn, S. (2023). *What is Predictive Analytics?*. Business News Daily. <https://www.businessnewsdaily.com/4938-predictive-analytics.html>, accessed on 21 August 2023.

Gioia, D. G., Felizardo, L. K., & Brandimarte, P. (2023). Simulation-based Inventory Management of Perishable Products via Linear Discrete Choice Models. *Computers and Operations Research*, *157*, pp. 1-12. https://doi.org/10.1016/j.cor.2023.106270

Gong, M., Lian, Z., & Xiao, H. (2022). Inventory Control Policy for Perishable Products Under a Buyback Contract and Brownian Demands. *International Journal of Production Economics*, *251*, pp. 1-12. https://doi.org/10.1016/j.ijpe.2022.108522

Guo, Y., Wood, J., Pan, W., & Meng, Q. (2018). Inventory Optimization of Airport Perishable Emergency Supplies with Replacement Strategy Facing Stochastic Occurrence Time by CVaR Approach. *International Journal of Disaster Risk Reduction*, *31*, pp. 170–183. https://doi.org/10.1016/j.ijdrr.2018.04.027

Halton, C. (2023). *Predictive Analytics: Definition, Model Types, and Uses.* Investopedia. <https://www.investopedia.com/terms/p/predictive-analytics.asp>, accessed on 21 August 2023.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques*, 3rd Edition.

IBM. (2022). *What is Predictive Analytics*? <https://www.ibm.com/analytics>, accessed on 21 August 2023.

Kanade, V. (2022). *What is Machine Learning? Definition, Types, Applications, and Trends for 2022*. <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/>, accessed on 13 July 2023

Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining.*

Lolli, F., Ishizaka, A., Gamberini, R., Balugani, E., & Rimini, B. (2017). Decision Trees for Supervised Multi-criteria Inventory Classification. *Procedia Manufacturing*, *11*, pp. 1871–1881. https://doi.org/10.1016/j.promfg.2017.07.326

Mahfouz, M. A., Mosaad, S. M., & Belal, M. A. (2023). Inventory Data Mining System Using Federated Machine Learning Strategy. J*ournal of Southwest Jiaotong University*, *58*(3), pp. 610-623. https://doi.org/10.35741/issn.0258-2724.58.3.51

McBride, D. (2003). *The 7 Wastes in Manufacturing*. EMS Consulting Group. <https://www.emsstrategies.com/dm090203article2.html#:~:text=The%20concept%20of%20the%20seven,Lean%20Manufacturing%20or%20Lean%20Thinking.>, accessed on 21 August 2023.

Muckstadt, J. A. & Sapra, A. (2010). *Principles of Inventory Management: When You Are Down to Four*, *Order More*. http://www.springer.com/series/3182

Namir, K., Labriji, H., & ben Lahmar, E. H. (2021). Decision Support Tool for Dynamic Inventory Management using Machine Learning, Time Series and Combinatorial Optimization. *Procedia Computer Science*, *198*, pp. 423–428. https://doi.org/10.1016/j.procs.2021.12.264

Nurkasanah, I. (2021). Reinforcement Learning Approach for Efficient Inventory Policy in Multi- Echelon Supply Chain Under Various Assumptions and Constraints. Journal of Information Systems Engineeringand Business Intelligence, *7*(2), pp. 138-148. https://doi.org/10.20473/jisebi.7.2.138-148

Ohmori, S., & Yoshimoto, K. (2020). A Robust Optimization for Multi-Period Lost Sales Inventory Control Problem. *Operations and Supply Chain Management*, *13*(4), pp. 375–381.

Pai, D. R., & Dissanayake, C. K. (2022). Predictive analytics methods for investigating inpatient volume in rural hospitals. *Healthcare Analytics*, *2*, pp. 1-7. https://doi.org/10.1016/j.health.2022.100113

Rinaldi, M., Fera, M., Macchiaroli, R., & Bottani, E. (2023). A New Procedure for Spare Parts Inventory Management in ETO Production: A Case Study. *Procedia Computer Science*, *217*, pp. 376–385. https://doi.org/10.1016/j.procs.2022.12.233

Rios, J. H., & Vera, J. R. (2023). Dynamic Pricing and Inventory Control for Multiple Products in a Retail Chain. *Computers and Industrial Engineering*, *177*, pp. 1-12. https://doi.org/10.1016/j.cie.2023.109065

Sánchez, C., Maldonado, S., & Vairetti, C. (2022). Improving Debt Collection via Contact Center Information: A Predictive Analytics Framework. *Decision Support Systems*, *159*, pp. 1-11. https://doi.org/10.1016/j.dss.2022.113812

Seyedan, M., Mafakheri, F., & Wang, C. (2023). Order-up-to-level Inventory Optimization Model Using Time-series Demand Forecasting with Ensemble Deep Learning. *Supply Chain Analytics*, *3*, pp. 1-13. https://doi.org/10.1016/j.sca.2023.100024

Sharma, D. (2023). *Is Inventory an Asset? (Plus Type of Inventory)*. Indeed. <https://www.indeed.com/career-advice/career-development/is-inventory-an-asset >, accessed on 21 August 2023.

Shirisha, G., Divyamani, M.R., Neha, A., Sravani, D., & Suresh, Y. (2022). Machine Learning Based Predictive Analytics for Agricultural Inventory Management System. *In International Research Journal of Modernization in Engineering Technology and Science*, *4*(7), pp. 2569 – 2575.

Sridhar, P., Vishnu, C. R., & Sridharan, R. (2021). Simulation of Inventory Management Systems in Retail Stores: A Case Study. *Materials Today: Proceedings*, *47*, pp. 5130–5134. https://doi.org/10.1016/j.matpr.2021.05.314

Sun, L., Xia, J., She, D., Guo, Q., Su, Y., & Wang, W. (2023). Integrated Intra-storm Predictive Analysis and Real-time Control for Urban Stormwater Storage to Reduce Flooding Risk in Cities. *Sustainable Cities and Society*, *92*, pp. 1-14. https://doi.org/10.1016/j.scs.2023.104506

Theodorou, E., Spiliotis, E., & Assimakopoulos, V. (2023). Optimizing Inventory Control through a Data-driven and Model-independent Framework. *EURO Journal on Transportation and Logistics*, *12*, pp. 1-14. https://doi.org/10.1016/j.ejtl.2022.100103

Toomey, J. W. (2000). *Inventory Management: Principles, Concepts and Techniques* (Vol. 12). Springer US. https://doi.org/10.1007/978-1-4615-4363-3

Torgo, L. (2017). *Data Mining with R Learning with Case Studies* Second Edition.

Wang, Y.-C., Chen, T., & Hsu, T. C. (2021). A Fuzzy Deep Predictive Analytics Approach for Enhancing Cycle Time Range Estimation Precision in Wafer Fabrication. *Decision Analytics Journal*, *1*, pp. 1-8. https://doi.org/10.1016/j.dajour.2021.100010

Wibowo, A. T., Arvitrida, N. I., & Widodo, E. (2021). Raw Material Order Allocation Problem Using Mixed Integer Linear Programming and Simulation. *Operations and Supply Chain Management*, *14*(4), pp. 444–455.

Wolbert, M. (2013). *Predictive Analytics for Inventory in a Sporting Goods Organization* [Master's Thesis, Massachusetts Institute of Technology].

Yudhistyra, W. I., Marta Risal, E., Raungratanaamporn, I., & Ratanavaraha, V. (2020). Exploring Big Data Research: A Review of Published Articles from 2010 to 2018 Related to Logistics and Supply Chains. *Operations and Supply Chain Management*, *13*(2), pp. 134–149.

Zhang, C., Gao, W., Yang, T., & Guo, S. (2019). Opportunistic Maintenance Strategy for Wind Turbines Considering Weather Conditions and Spare Parts Inventory Management. *Renewable Energy*, *133*, pp. 703–711. https://doi.org/10.1016/j.renene.2018.10.076

Zheng, J., Alzaman, C., & Diabat, A. (2023). Big Data Analytics in Flexible Supply Chain Networks. *Computers and Industrial Engineering*, *178*, pp. 1-16. https://doi.org/10.1016/j.cie.2023.109098

Zietsman, H. J., & van Vuuren, J. H. (2023). A Generic Decision Support Framework for Inventory Procurement Planning in Distribution Centres. *Computers and Industrial Engineering*, *177*, pp. 1-15. https://doi.org/10.1016/j.cie.2023.109057

Patdono Suwignjo is an associate professor at the Industrial and Systems Engineering Department, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. He received an Engineer degree (1983) in Mechanical Engineering from ITS, a Master's degree (1991) was from the University of New South Wales, Australia, and his Doctoral degree (1999) is from the University of Strathclyde, Glasgow, UK. For many years he has been a consultant and advisor in several state-owned companies in Performance Measurement Systems, Strategic Management, Business Process Management, Risk Management, and Feasibility Study. From 2011 to 2020, he was appointed for several positions in The Minister of Education and Culture, i.e., as Senior Advisor; Secretary for the Directorate General of Higher Education; Director-General of Organizational Management for Knowledge, Technology, and Higher Education; and Ad interim Director General of Vocational Education. In June 2020, he retired from the government position and returned to campus.

Lisda Panjaitan received her BEng in Industrial Engineering from Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia in 2022. She was a Supply Chain intern in Unilever Indonesia in 2022. Currently, she is a Demand Planning Officer in PT Enseval Putera Megatrading Tbk Jakarta.

Ahmed Raecky Baihaqy earned his bachelor’s degree in industrial engineering from Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia in 2019. He was active as a laboratory assistant at Quantitative Modelling and Industrial Policy Analysis (QMIPA) Laboratory of Industrial Engineering ITS also had been appointed as coordinator of laboratory assistant. Currently, he is pursuing a master’s degree in industrial engineering at ITS fully funded by Indonesia Endowment Fund for Education. Ahmed has more than 3 years of experience in logistics and supply chain management at one of biggest 3PL companies. His research interests include discrete-event simulation, inventory modelling, reliability modelling, as well as logistics and supply chain.

Ahmad Rusdiansyah received his B.Eng in Industrial Engineering from Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia in 1992, the MEng in Industrial Engineering from Dalhousie University, Halifax NS, Canada in 1999 and the DrEng in Industrial Engineering and Management from Tokyo Institute of Technology, Tokyo, Japan in 2005. He also holds CSCP and CLTD certificates of APICS. He is head of the Transportation and Distribution Logistics (TDLog) research group. His research interests include production planning and inventory control, and transportation and distribution systems modelling. Currently, he is starting to explore research in logistics analytics.