Informed Decision Making on Prediction of Spare Parts Requirements Based-on Machine Learning Approach

Abdul Rahman Rizki Wijaya
Department of Industrial Engineering
Universitas Islam Indonesia
Jl. Kaliurang km 14.4, Sleman, Yogyakarta, Indonesia

Winda Nur Cahyo
Department of Industrial Engineering
Universitas Islam Indonesia
Jl. Kaliurang km 14.4, Sleman, Yogyakarta, Indonesia
winda.nurcahyo@uii.ac.id

Abstract

PT XYZ is one of a big company in Indonesia. In daily operations, PT XYZ performs many business processes, one of which is the procurement process. This process requires a prediction of the need for goods for the procurement process to be carried out because intermittent and lumpy characteristics. This is the background for conducting this research. This study aims to predict the level of demand for goods so that it can help the procurement department to make purchases of goods. The method used is K-Means for Clustering, Holt Winter for forecasting, and Association Rule to find out the combination of items that are often held. The clustering process produces 3 clusters, namely the first cluster which has members of 327 items with an average value of 68, the second cluster which has members of 18 items with an average value of 1281, and the third cluster which has members of 5 items with an average value of 2474. The resulting forecast covers the upcoming 3 months period with a value of 427 in January, 472 in February, and 476 in March. In addition, it is also known which items are likely to appear in that month using the probability method and the minimum value of the probability is 50%. In January there are 10 items that are possible to appear, 22 items in February, and 20 items in March. The rules generated from the association rule process are 24 rules. The rule is generated with a minimum support value of 10% and a minimum confidence of 50%. The integration between methods is also carried out to cover the existing deficiencies, so that more accurate results are obtained.

Keywords
Prediction, K-Means, Holt Winter, Association Rule, Integration, Machine Learning.

1. Introduction
Business processes are an important part of a company. Business processes are one of the determining factors for the smoothness, performance, and success of the company. The business process itself is a series of activities that are interconnected to achieve the desired results in supporting the running of a company (Harrington, 1991). A company carries out various kinds of business processes to support the running of the company. This process can be in the form of sales, production, procurement, or other processes. Each of these processes needs to be managed properly to improve the overall performance of the organization. One example of a business process carried out by the company is the process of procuring spare parts. This procurement is done to maintain the level of availability of these spare parts. Moreover, spare parts used for production machines. If the level of availability is not sufficient, it will affect the availability and reliability for machines in production. If this continues, it will have a negative impact on the company, such as decreased production productivity, longer machine breakdown times, until the production target is disrupted. Therefore, the need for spare parts needs to be predicted properly so that the procurement strategy can be made as good as possible. On the other hand, in predicting the need for machine spare parts, there are several difficulties to do so . According to Bacchetti and Saccani (2012), there are several aspects that cause the demand and management of spare parts to become a complicated problem, namely the high number of spare parts being managed and the existence of intermittent or lumpy demand. Intermittent requests are requests that occur at erratic intervals and with highly

variable amounts. Meanwhile, lumpy requests are requests with uneven timing and varying amounts required. Lumpy demand resulting in increased inventory investment requirements or longer response times than anticipated. On the other hand, this procurement process produces data which is a description of the transaction or process carried out. If this data is used properly, it will be able to provide valuable information for the company and support the company's management, such as the estimated level of spare part needs for a certain period. The data analysis approach for maintenance is one of a good approach to suggest the maintenance strategy (Cahyo et al., 2019)

Data is an asset for a company. The data generated by the company can be processed for certain purposes such as support in decision making. Decisions or policies that are supported by data will result in decisions that are right on target or effective. On the other hand, if the data held is not processed properly then this data will not provide any benefit. There is also the term "Drowning in Data but Starving for Insight" which means the company produces a lot of data but does not process it properly, so it does not provide any value. Therefore, the data generated is very important to be processed to produce a benefit for the company. In addition, the current trend of companies also shows the company's change to become a data-driven company or a data-informed company. Data-driven itself means that decisions made by a company are determined based on data. Meanwhile, data-informed means that company decisions are made based on data but are added with other factors such as experience, user research, or other important information.

Technology has become an important part of many aspects of life today. The existence of technology is expected to facilitate the work done by humans. In addition, technological developments are also increasingly rapid with new innovations being made. The development of this technology occurs in many aspects including data processing. The application of technology for companies can increase the productivity of a company, especially in data processing. This is because the bigger the company, the bigger the data will be. Thus, with the data processing technology it will be faster than manual processing and can minimize the error rate that occurs during the processing. Moreover, the industrial revolution 4.0 will result in the data produced by a company getting bigger so that tools are needed to process this data.

Machine learning is one of the technology choices for data processing. By using machine learning data processing for big data becomes easier. According to IBM, machine learning is an application branch of artificial intelligence that focuses on creating systems or algorithms that learn based on data continuously to improve accuracy. In machine learning applications, algorithms can be used to look for certain patterns or features in large data to make decisions or predictions from existing data. So, quite a lot of company data can be processed properly. In using machine learning, you can use the Python programming language because it provides and allows to run machine learning modules. Moreover, Python is a programming language that is popular in data science and is a programming language that is quite easy to understand.

PT XYZ is one of the companies in Indonesia. In operating their business, a lot of data will be generated from their business processes such as sales, procurement, and production processes. These data must be processed to provide knowledge or information for the management and as a supporter in policy making. One of the data generated is about the procurement of goods carried out by the company. PT XYZ carries out many procurement processes for many goods, but in this study, it will only cover one type of goods, namely one of the spare parts owned by PT XYZ. This research will focus on processing data on material procurement for 2010-2015 from PT XYZ to find valuable information that can be used by management in making decisions. The information to be extracted is related to the need for spare parts using PT XYZ procurement data. This research is expected to provide supporting information for decision making in the procurement process so that decision making can be easier and more precise.

2. Literature Review

The method used in the review on literature in this paper is Systematic Literature Mapping (SLM) as proposed in (Cahyo, 2021a) and (Cahyo, 2021b). Briefly, SLM combined Systematic Literature Review and K-Chart. In the process of finding the related literature, a search term is applied on Scopus indexed database ranging from 2018 to October 2022. The applied search term is “spare part requirement”. This search results 20 documents. However, only 14 documents are used in the SLM because the six excluded documents are not related to the issue of spare part requirement. The screenshot of the result in Scopus is shown in Figure 1 and the result of the SLM is shown in Figure 2.

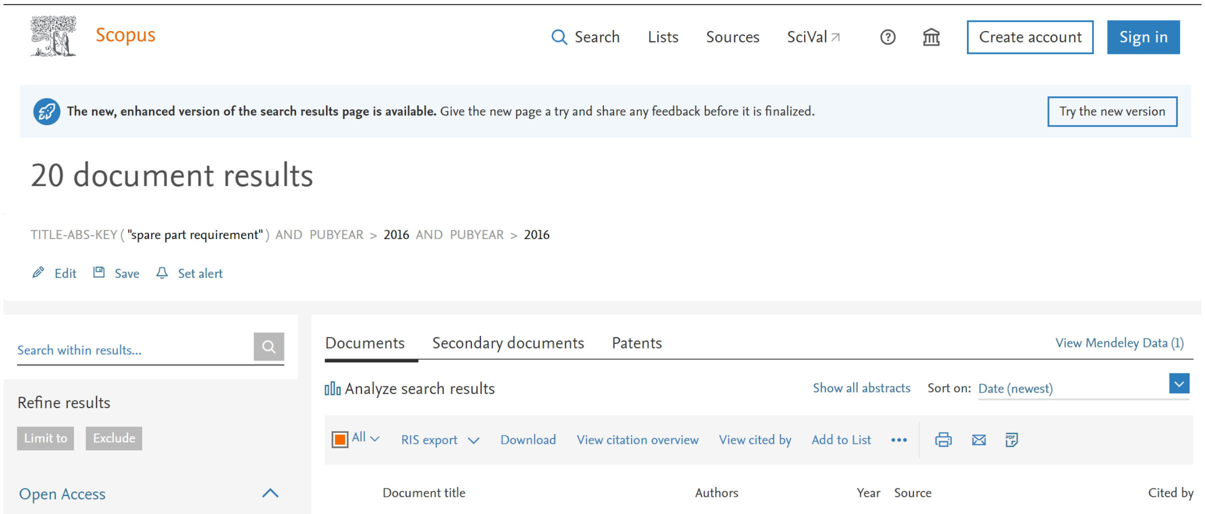

Figure 1. Result in Scopus

The result of the SLM in Figure 2 consists of several layer and its boxes. There are numbers in the boxes that show the related literature in the area shown in the boxes. The number in the boxes is associated with the list of literature in Table 1. It is indicated from the result of the SLM that most of research on spare part requirement are focusing on the manufacturing organisation as in Zhang, Deng, Miao, Liu, and Shao (2022), Fatemeh Faghidian, Khashei, and Khalilzadeh (2021), Ermawanto and Kurniati (2021), and (Angelina, Atmaji, & Santosa, 2020). Other areas of this research are in Transportation System, Military, Service Industry, Mining Industry, and Maintenance system. In the Method layer and Result Layer, it is also indicated that most of the result in the area are using Mathematical modelling and the result is dominated by Forecast/Prediction optimisation. By looking on Figure 1, it can be suggested that there are several research opportunities. One of the opportunities is applying machine learning approach to support informed decision making in the spare part requirement in oil and gas industries. This opportunity is raised because there are no supporting literature in the boxes related to the oil & gas industry in the system layer, machine learning approach in the method layer, and informed decision making in the result layer.

Table 1. The list of authors

No	Authors	No	Authors	No	Authors
[1]	Ermawanto and Kurniati (2021)	[6]	Q. Wang, Jia, Cheng, Weng, and Wang (2019),	[11]	Sun, Hao, Su, and Ren (2018)
[2]	Zhao, Pei, Hou, and Tian (2019)	[7]	Moharana, Sarmah, and Rathore (2019),	[12]	Fatemeh Faghidian et al. (2021),
[3]	Song, Wu, Kang, and Zhang (2021)	[8]	Angelina et al. (2020)	[13]	Nouri Qarahasanlou, Barabadi, Ataei, and Einian (2019)
[4]	Zhang et al. (2022)	[9]	N. Wang et al. (2019)	[14]	J. Wang, Pan, Wang, and Wei (2018)
[5]	Yang et al. (2022)	[10]	Yanluo (2021),		

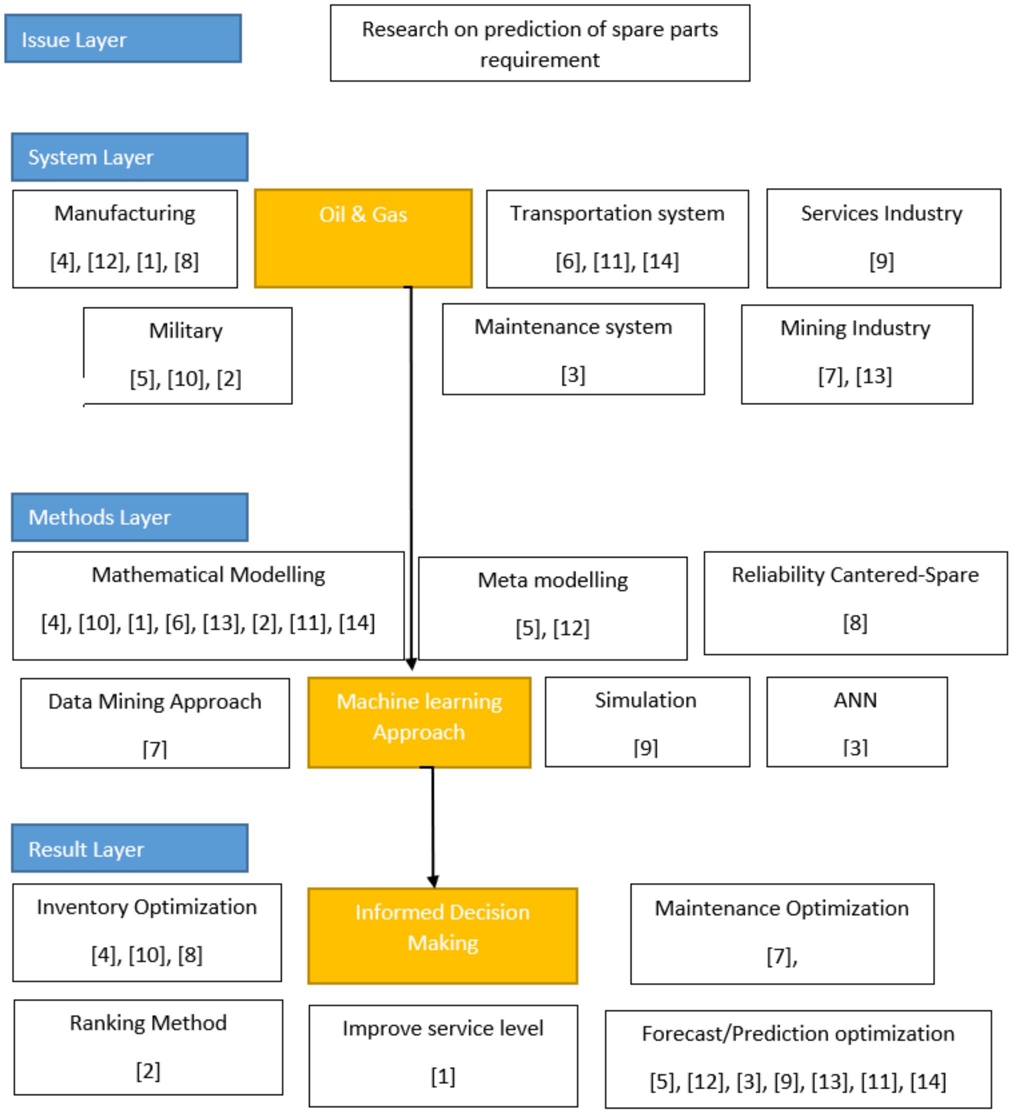

Figure 2. The result of Systematic Literature Mapping

3. Research Methods

3.1 Data Description

The data that used in the research is spare part purchasing transaction of PT XYZ from 2010-2015. This transaction consists of 5061 transaction items. The data also have 8 attribute that showed in Figure 3. Not all the attributes will be used in the data processing. Attribute Short Text, Item PO, OUn, and Crcy will be deleted because it's not needed.

```
Material No.                int64
Short Text                 object
Purch.Doc.                  int64
Item PO                     int64
PO. Date           datetime64[ns]
PO Quantity                 int64
OUn                        object
Crcy                       object
dtype: object
Dimensi:  (5091, 8)
```

Figure 3. Attribute of the data

Total of 394 material is manage in this transaction. This material categorized based on its dimension, however the categorization that is needed is based on moving speed and purchasing volume of the material. Another attribute like PO. Date will be broken down into day, month, and year.

3.2. Data Analysis

Data processing will be divided into 3 parts. First part is clustering of the materials. This part is to categorize materials based on moving speed and purchasing volume of the materials. Second part is demand forecasting of the materials using Holt Winters Triple Exponential Smoothing methods. In this part we want to know demand level and material that have high probability to appear. Last part is association rule or find out about combination of the materials that often to appear. These 3 parts will be analyzed using Python machine learning module. Clustering will use sklearn module, forecasting will use statsmodel, and association rule will use mlxtend module.

Every step of the data processing has different pre-processing. For clustering, the data need to be aggregated based on material code to know about total demand of each material. Beside of the total demand, procurement frequency also need to be calculated. For forecasting, the data will be aggregated based on transaction months. For association rule, purchasing document will be broken down so we know what material that included in each purchasing document.

4. Result and Discussion

4.1 Clustering

Before the clustering process can be carried out, it is necessary to know the level of Multicollinearity. This calculation is done by calculating the value of the Variance Inflation Factor (VIF). If the VIF value is less than 10 then the level of multicollinearity can still be tolerated, or the clustering process can be carried out. Based on the calculation results, it is known that the VIF value for the PO Quantity and Freq attribute is 3.16 for both attributes or in other words the level of multicollinearity is still tolerated, and clustering can be done.

The method used for clustering is K-Means. This method is used because the number of clusters generated will be determined at the beginning as desired and the cluster members are only joined in an exclusive 1 cluster. The clustering process will use a module from scikit-learn or also often called sklearn. Sklearn itself is a machine learning module in Python that allows to perform various processes such as classification, regression, and clustering.

The use of the K-Means method requires determining the number of clusters at the beginning or before the cluster formation process. The number of clusters produced is determined to be 3 clusters. This is because clusters related to the number of procurements are expected to be divided into 3 clusters. In addition, by using the elbow method, 3 clusters are the optimal number of clusters for the existing data. The conclusion is because the value shown after the

3rd point in the elbow method tends to be flat, so 3 clusters is the optimal value. The elbow method process is carried out using the sklearn module. Therefore, the module must be imported before it can be used. The result of the clustering is shown in Figure 4.

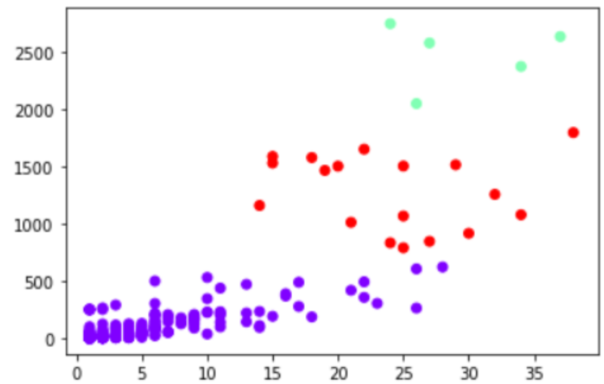

Figure 4. The result of the Clustering

The first cluster formed had 327 members. In the scatter diagram, the first cluster is indicated by a purple dot. Based on the cluster formed, it can be seen about the characteristics of the members who are joined. This process can be done using the dataframe.describe() syntax in Python. The results regarding the characteristics of the first cluster are shown in Figure 5.

	PO Quantity	Freq	Clusters
count	327.000000	327.000000	327.0
mean	67.538226	4.067278	0.0
std	107.242247	4.416408	0.0
min	1.000000	1.000000	0.0
25%	6.000000	1.000000	0.0
50%	20.000000	3.000000	0.0
75%	79.500000	5.000000	0.0
max	621.000000	28.000000	0.0

Figure 5. The characteristics of the first cluster

The second cluster formed has 18 members. In the scatter diagram, the second cluster is indicated by a red dot. Based on the cluster formed, it can be seen about the characteristics of the members who are joined. This process can be done using the dataframe.describe() syntax in Python. The results related to the characteristics of the second cluster are shown in the Figure 6.

	PO Quantity	Freq	Clusters
count	18.000000	18.000000	18.0
mean	1281.166667	24.055556	2.0
std	322.533035	6.795087	0.0
min	789.000000	14.000000	2.0
25%	1024.750000	19.250000	2.0
50%	1359.500000	24.500000	2.0
75%	1524.250000	28.500000	2.0
max	1794.000000	38.000000	2.0

Figure 6. The characteristics of the second cluster

The third cluster formed has 5 members. In the scatter diagram, the third cluster is indicated by the cyan colored dot. Based on the cluster formed, it can be seen about the characteristics of the members who are joined. This process can be done using the dataframe.describe() syntax in Python. The results related to the characteristics of the third cluster are shown in the Figure 7.

	PO Quantity	Freq	Clusters
count	5.000000	5.00000	5.0
mean	2474.000000	29.60000	1.0
std	274.411188	5.59464	0.0
min	2047.000000	24.00000	1.0
25%	2371.000000	26.00000	1.0
50%	2576.000000	27.00000	1.0
75%	2632.000000	34.00000	1.0
max	2744.000000	37.00000	1.0

Figure 7. The characteristics of the third cluster

4.2 Forecasting

The transformation process needs to be carried out so that the data held can be used for the forecasting process using the Holt Winter method. Existing data needs to be grouped by transaction month so that the data is in the form of an aggregate each month. This grouping is also done because the forecast will use the month period. The results of the grouping show that the data consists of 48 months from 2012 to 2015.

The forecasting method used is Holt Winter Seasonal with Triple Exponential Smoothing. In this forecast, it only focuses on clusters 2 and 3 because the materials incorporated in these clusters are slow moving and fast moving or in other words, procurement is carried out periodically. Cluster 1 is not a part of forecasting with this method because most of the combined materials have non-moving properties or procurement is not done regularly. In addition, the data for 2010 and 2011 were not used in forecasting because the distribution of the data is quite random so that it will complicate and affect the final forecasting result. As an initial description of the data distribution, it can be seen using the statsmodel module, so a graph will be generated as shown below. If observed carefully, the distribution of the data shows that there is a seasonal nature at the beginning of each year. The result of the forecasting is shown in Figure 8.

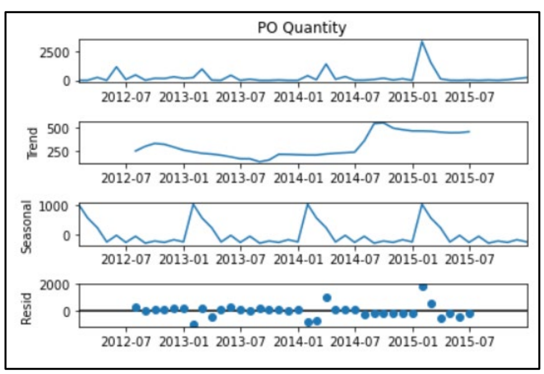

Figure 8. The result of the forecasting

The comparison of forecasted data and actual data is shown in Figure 9 bellow. The blue line shows the actual data and the orange line shows the forecast results. With this graph can be seen the comparison and variance of the actual data and forecasting results. In the forecast, there are variations between the actual value and the forecast or what is often referred to as forecasting error. The Figure 10 shows the error value of the forecasting results by using the mean absolute error.

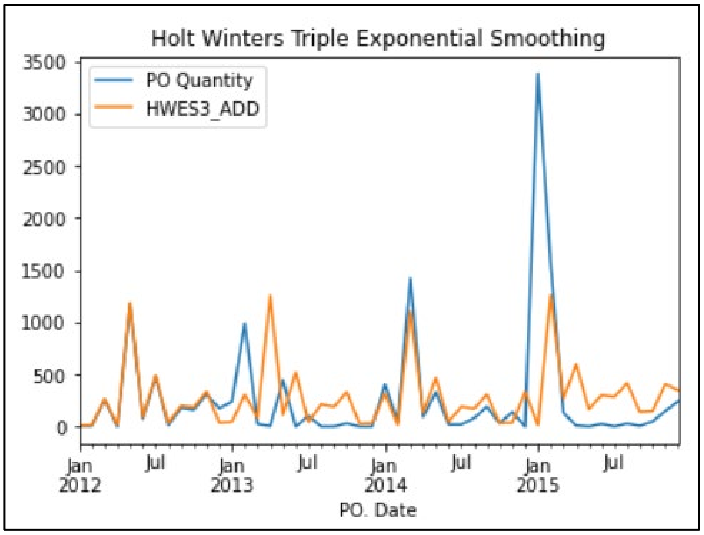

Figure 9. The comparison between actual and the result of forecasting

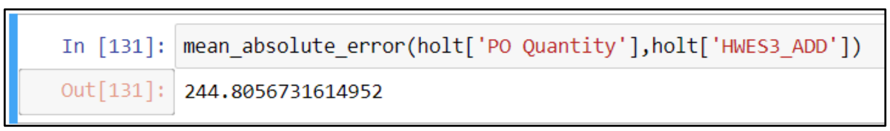

Figure 10. The error values of the forecasting results by using the mean absolute error.

Forecasting related to the number of needs has been done using the Holt Winter method. The results of forecasting have not been able to describe the material needed. Therefore, further analysis can be done to predict what materials are needed each month. This analysis can be done by using the probability of the material appearing each month. Using Python makes it possible to look for such opportunities. The process carried out by this syntax is to find out the probability of occurrence of a material every month during 2010 to 2015. Then after that, the probability of the emergence of material for the following year can be known.

The results for the material predictions in January are shown in Table 2. Probabilities for a material to appear in January are indicated by the Sum column. This column shows the number of times a material appears in each year. The minimum value specified for taking this Sum is 3 or 50%. Thus, the material that is likely to appear or be needed with a 50% chance in January is 10 materials. The results for the material predictions in February are shown in 3able 3. Probabilities for a material to appear in January are indicated by the Sum column. This column shows the number of times a material appears in each year. The minimum value specified for taking this Sum is 3 or 50%. Thus, the material that is likely to appear or be needed with a 50% chance in February is 22 materials (Table 3).

Table 2. The results for the material predictions in January

Material Code	2010	2011	2012	2013	2014	2015	Sum
39	1	0	0	1	1	0	3
45	1	0	0	1	1	0	3
46	1	0	0	1	1	0	3
47	1	1	0	1	1	0	4
67	0	1	0	0	1	1	3
71	0	1	0	0	1	1	3
74	0	1	0	0	1	1	3
82	1	1	0	1	1	1	5
83	1	1	1	1	1	1	6
152	0	1	0	1	0	1	3

Table 3. The results for the material predictions in February

Material Code	2010	2011	2012	2013	2014	2015	Sum
47	0	1	0	1	1	1	4
49	1	1	0	1	0	1	4
50	1	1	0	1	0	0	3
52	1	1	0	1	0	0	3
64	1	1	0	1	0	0	3
67	1	1	0	1	0	0	3
69	1	1	0	1	0	0	3
74	1	1	0	1	0	1	4
78	0	1	0	1	1	0	3
79	1	1	0	0	0	1	3
82	1	1	0	1	1	0	4
83	1	1	1	1	0	0	4
85	1	1	0	1	1	0	4
86	1	1	0	1	0	1	4
112	1	1	0	1	0	0	3
194	0	1	0	0	1	1	3
197	0	1	0	1	0	1	3
199	0	1	0	1	0	1	3
212	0	1	0	0	1	1	3
323	1	1	0	0	0	1	3
328	1	1	0	1	0	1	4
329	0	1	0	1	0	1	3

The results for the material predictions in March are shown in Table 4. Probabilities for a material to appear in January are indicated by the Sum column. This column shows the number of times a material appears in each year. The minimum value specified for taking this Sum is 3 or 50%. Thus, the material that is likely to appear or be needed with a 50% chance in March is 20 materials.

Table 4. The results for the material predictions in March

Material Code	2010	2011	2012	2013	2014	2015	Sum
32	1	0	1	0	1	0	3
36	1	1	0	1	1	0	4
39	1	1	1	1	1	1	6
45	1	0	0	1	1	0	3
46	1	0	0	1	1	1	4
47	1	0	1	1	1	1	5
51	1	0	1	0	1	1	4
52	1	1	0	1	1	1	5
69	1	0	0	0	1	1	3
71	1	0	0	0	1	1	3
78	1	1	0	0	1	0	3
82	1	1	0	0	1	1	4
83	1	0	1	0	1	1	4
86	1	1	0	0	1	1	4
114	1	1	1	0	1	0	4
196	0	0	0	1	1	1	3
197	0	1	0	1	1	1	4
326	1	1	0	1	0	0	3
328	1	1	1	0	1	1	5
334	1	1	0	0	1	0	3

4.3 Association Rule

Before performing an AR analysis of the data, it is necessary to adjust or transformations of the data. This is so that the form of the data can be used for further processing. After adjustments are made, it can be seen what material is contained in a transaction. In addition, it is also known that the number of Purch.Doc. or 269 transactions.

The AR process is carried out using an a priori algorithm. The process will also use the mlxtend module in Python because it provides an a priori algorithm. The parameters used to find AR rules are the support and confidence values of the material combination. The minimum support value used is 10% and the minimum confidence is 50%. Altogether 24 AR rules were formed with a minimum support of 10% and a minimum confidence of 50% as shown in the Table 5.

Table 5. The results of Association Rule

Antecedents	Consequents	Support	Confidence	Lift
{'78'}	{'82'}	0.1301	0.7778	3.0322
{'85'}	{'82'}	0.1115	0.7692	2.9989
{'85'}	{'83'}	0.1115	0.7692	2.9560
{'51'}	{'328'}	0.1004	0.7297	3.5690
{'86'}	{'83'}	0.1115	0.6977	2.6811
{'66'}	{'82'}	0.1004	0.6923	2.6990
{'45'}	{'46'}	0.1264	0.6415	3.0815
{'49'}	{'83'}	0.1041	0.6364	2.4455
{'46'}	{'47'}	0.1301	0.6250	2.5473
{'46'}	{'45'}	0.1264	0.6071	3.0815
{'45'}	{'47'}	0.1190	0.6038	2.4608
{'46'}	{'39'}	0.1152	0.5536	2.5239
{'82'}	{'83'}	0.1413	0.5507	2.1164
{'83'}	{'82'}	0.1413	0.5429	2.1164
{'69'}	{'46'}	0.1004	0.5400	2.5939
{'69'}	{'82'}	0.1004	0.5400	2.1052
{'47'}	{'46'}	0.1301	0.5303	2.5473
{'45'}	{'82'}	0.1041	0.5283	2.0596
{'328'}	{'47'}	0.1078	0.5273	2.1490
{'39'}	{'46'}	0.1152	0.5254	2.5239
{'45'}	{'39'}	0.1004	0.5094	2.3227
{'39'}	{'47'}	0.1115	0.5085	2.0724
{'82'}	{'78'}	0.1301	0.5072	3.0322
{'47'}	{'83'}	0.1227	0.5000	1.9214

4.4. Integration

The results of the three methods used can be integrated to achieve the desired goals. In this study, one of the objectives to be achieved is to get an overview of future material needs. This can be started by looking at the forecasting results. Forecasting will describe the amount of material needed in the aggregate. Figure 11 shows the result of forecasting material needs, where horizontal axis is the periods (month) and the vertical axis is material need (Q).

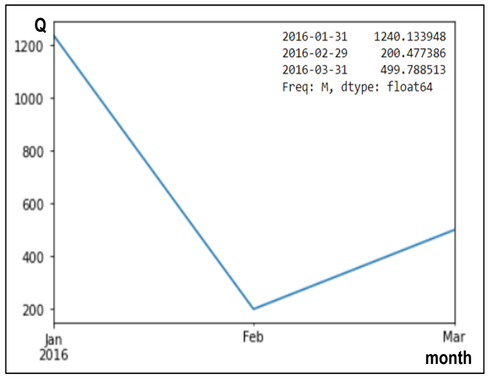

Figure 11. The result of forecasting material needs

However, these results have not been able to provide an overview of what materials are needed in the relevant month. Therefore, the use of opportunities can be used to make up for these deficiencies. The minimum value of the determined opportunity is 50% because this value will be able to provide an overview of the tendency of the emergence of a material.

If we observe these results carefully, we will find a probability value equal to 50%. This value has not been able to provide an overview of the trend of occurrence of a material because this value is in the middle. To be able to cover this deficiency can include the results of clustering processing. Clustering will be able to show the characteristics of a material. For example, if a material has a 50% chance to appear and join the third cluster, then that material has a weighting factor because the third cluster is the cluster with the highest average or can be said to be high demand and fast moving. The second cluster is also a weighting factor. However, for the first cluster, it becomes a which is a weakening factor because the characteristics of this cluster are low demand and non-moving. Therefore, the materials belonging to the second and third clusters are recommended to be procured, but the first cluster is not recommended to be procured. The relationship between the clustering and the association rule in January, February and March are shown in Table 6, Table 7, and Table 8 accordingly.

Table 6. The relationship between clustering and association rule in January

January			
Material Code	**Probability**	**Cluster**	**AR Material**
39	50.0%	2	46,47
45	50.0%	1	46,47,82,39
46	50.0%	1	47,45,39
47	66.7%	1	46,83
67	50.0%	1	-
71	50.0%	0	-
74	50.0%	1	-
82	83.3%	2	83,78
83	100.0%	1	82
152	50.0%	0	-

With the application of the factors previously described, in January, materials 71 and 152 are not recommended to be procured because these materials have a 50% chance and are in the first cluster which is a debilitating factor. With the application of the factors previously described, in February for materials 50, 52, 79, 112, 194, 197, 199, 212, and 329 it is not recommended to procure because these materials have a 50% chance and are in the first cluster which is a weakening factor. In addition, material 49, which is a material that belongs to the first cluster or has non-moving and low demand properties, will also be eliminated.

Table 7. The relationship between clustering and association rule in February

February			
Material Code	**Probability**	**Cluster**	**AR Material**
47	66.7%	1	46,83
49	66.7%	0	83
50	50.0%	0	-
52	50.0%	0	-
64	50.0%	1	-
67	50.0%	1	-
69	50.0%	2	46,82
74	66.7%	1	-
78	50.0%	1	82
79	50.0%	0	-
82	66.7%	2	83,78
83	66.7%	1	82
85	66.7%	1	82,83
86	66.7%	1	83
112	50.0%	0	-
194	50.0%	0	-
197	50.0%	0	-
199	50.0%	0	-
212	50.0%	0	-
323	50.0%	2	-
328	66.7%	1	47
329	50.0%	0	-

With the application of the factors previously explained, in March for materials 71, 196, and 334 it is not recommended to procure because these materials have a 50% chance and are in the first cluster which is a weakening factor. In addition, materials 51, 52, 114, and 197 which are materials belonging to the first cluster or have non-moving and low-demand properties will also be eliminated.

Finally, AR analysis can be added to find out which material combinations occur frequently. This is useful in saving logistics costs or ordering costs because discounts for certain order quantities may apply. For example, material 39 can be ordered together with material 46 and 47 with a level of confidence that can be seen in AR table.

The research conducted in this report is intended to predict the need for spare parts for PT XYZ. Therefore, three methods are used, namely clustering, forecasting, and association rules. With the use of these three methods, it can be seen the number of materials needs and materials that are likely to arise as well as the supporting factors. However, in the data processing carried out there is potential for other developments to find out more knowledge from the existing data.

The development that can be done is related to the results of the forecasting method. The forecasting process only uses one method without any comparison with other methods. This can be the development of this research to find out the most appropriate forecasting method for the available data and reduce the error value of forecasting results. In addition, the addition of the data used can be done to improve the accuracy of forecasting because machine learning requires a large amount of training data for the learning process. The results of this forecasting have shortcomings in providing an overview of material requirements at the individual level. Therefore, research to find out this can be done.

Table 8. The relationship between clustering and association rule in March

March			
Material Code	**Probability**	**Cluster**	**AR Material**
32	50.0%	2	-
36	66.7%	1	-
39	100.0%	2	46,47
45	50.0%	1	46,47,82,39
46	66.7%	1	47,45,39
47	83.3%	1	46,83
51	66.7%	0	328
52	83.3%	0	-
69	50.0%	2	46,82
71	50.0%	0	-
78	50.0%	1	82
82	66.7%	2	83,78
83	66.7%	1	82
86	66.7%	1	83
114	66.7%	0	-
196	50.0%	0	-
197	66.7%	0	-
326	50.0%	1	-
328	83.3%	1	47
334	50.0%	0	-

5. Conclusion

Maintaining inventory levels of the required spare parts is important in running the company's operations. If this inventory level is not maintained and suffers from a shortage, it will result in the cessation of the operation process because machine repair cannot be carried out. One way that can be done to maintain this level of inventory is to predict the need for existing spare parts. On the other hand, this prediction process has its own challenges, such as the lumpy and intermittent nature of spare parts requirements. Both properties will complicate the prediction process because the data on the level of demand becomes very volatile and can be erratic. These problems can be overcome by the results of research conducted. The results of this research consist of three parts, namely clustering, forecasting, association rules, and integration of the three methods.

The results of the clustering show that the cluster formed consists of 3 groups with the division of low demand – non-moving, middle demand – slow moving, and high demand – fast moving. This division is based on the average value of each cluster or on low demand and non-moving has an average value of 67.54 and 4, middle demand and slow moving have an average of 1281.17 and 24, and high demand and fast moving have an average the average is 2474 and 30. In addition, the low demand cluster has 327 material members, middle demand has 18 material members, and high demand has 5 material members.

The results of the forecasting show procurement need for January of 1241, February of 201, and March of 500. In addition to the prediction of the number of needs, predictions are also made on what materials are likely to appear with a minimum probability value of 50%. In January materials 39, 45, 46, 47, 67, 74, 82, and 83 are likely to appear. In February materials 47, 64, 67, 69, 74, 78, 82, 83, 85, 86, 323, and 328 are likely to appear. In March materials 32, 36, 39, 45, 46, 47, 69, 78, 82, 83, 86, 326, and 328 are likely to appear. However, because the material is in the middle value, it is necessary to carry out additional analysis to determine the trend of occurrence.

The association rules or material combinations obtained from the AR analysis are 24 combinations. This combination has a minimum support value of 10% and a minimum confidence of 50%. The existence of this result can provide benefits such as savings on shipping costs and get a discount on a certain number of purchases.

Integration between the three methods can be done to overcome the existing shortcomings. The results of the forecast show the level of spare parts demand for the next 3 months, but the results are still aggregate or a combination of all existing materials. Therefore, to predict at the individual level, the probability method can be used. The results of clustering are added to determine the characteristics of the existing material. Thus, the priority level can be determined and the material which has a 50% chance can be identified. With clustering integration, it is known that there are several materials that are not recommended to be procured because they are considered unlikely to occur. Rules formed from AR analysis can be added to determine the accompanying materials or combinations of materials that tend to occur. In other words, the results of the AR analysis are also a weighting factor for a material that will appear or tend to occur in a transaction.

References

Angelina, C. F., Atmaji, F. T. D., & Santosa, B., *Spare part requirement and inventory policy for Rovema's 1 machine using Reliability Centered Spare (RCS) and Min-Max stock methods.* Paper presented at the IOP Conference Series: Materials Science and Engineering, 2020.

Bacchetti, A., & Saccani, N., Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega, 40*(6), 722-737, 2012. doi:https://doi.org/10.1016/j.omega.2011.06.008

Cahyo, W. N., *Finding Novelty of Research with Systematic Literature Mapping (SLM).* Paper presented at the Journal of Physics: Conference Series, 2021a.

Cahyo, W. N., Novelty Does Matter: SLM is a Convenient Way Finding a Research Originality. *Journal of Industrial Engineering Management, Special Edition*, 2021b.

Cahyo, W. N., Prawahandaru, H., Swasono, B. A., Raben, R. S. I., Sutartono, R. T., & Immawan, T., Data-Based Maintenance Strategy Analysis using Operational Excellence Approach in Engineering Asset Management. *International Journal of Integrated Engineering, 11*(5), 2019.

Ermawanto, D., & Kurniati, N., *Examining inventory management for spare-part warehouse.* Paper presented at the Proceedings of the International Conference on Industrial Engineering and Operations Management, 2021.

Fatemeh Faghidian, S., Khashei, M., & Khalilzadeh, M., Improving intermittent demand forecasting based on data structure. *Journal of Engineering Research (Kuwait), 9*(1), 188-199, 2021. doi:10.36909/JER.V9I1.8667

Harrington, H. J., Improving business processes. *The TQM Magazine, 3*(1). 1999. doi:10.1108/eb059514

Moharana, U. C., Sarmah, S. P., & Rathore, P. K., Application of data mining for spare parts information in maintenance schedule: a case study. *Journal of Manufacturing Technology Management, 30*(7), 1055-1072, 2019. doi:10.1108/JMTM-09-2018-0303

Nouri Qarahasanlou, A., Barabadi, A., Ataei, M., & Einian, V., Spare part requirement prediction under different maintenance strategies. *International Journal of Mining, Reclamation and Environment, 33*(3), 169-182, 2019. doi:10.1080/17480930.2017.1373883

Song, W., Wu, J., Kang, J., & Zhang, J., Research on maintenance spare parts requirement prediction based on LSTM recurrent neural network. *Open Physics, 19*(1), 618-627, 2021. doi:10.1515/phys-2021-0072

Sun, Y., Hao, X., Su, Z., & Ren, H., An Ordering Decision-Making Approach on Spare Parts for Civil Aircraft Based on a One-Sample Prediction. *IEEE Access, 6*, 27790-27795, 2018. doi:10.1109/ACCESS.2018.2818404

Wang, J., Pan, X., Wang, L., & Wei, W., Method of spare parts prediction models evaluation based on grey comprehensive correlation degree and association rules mining: A case study in aviation. *Mathematical Problems in Engineering, 2018*. doi:10.1155/2018/2643405

Wang, N., Xia, G., He, Z., Su, L., Zhang, Y., & Wang, X., *Research on spare parts requirement quantity forecasting for communication equipment.* Paper presented at the Proceedings - 2019 International Conference on Communications, Information System, and Computer Engineering, CISCE, 2019.

Wang, Q., Jia, X., Cheng, Z., Weng, Q., & Wang, S., *Study on spare parts requirement modelling of multi-component system under stochastic shock condition.* Paper presented at the IOP Conference Series: Materials Science and Engineering, 2019.

Yang, Y., Liu, W., Zeng, T., Guo, L., Qin, Y., & Wang, X., An Improved Stacking Model for Equipment Spare Parts Demand Forecasting Based on Scenario Analysis. *Scientific Programming, 2022*. doi:10.1155/2022/5415702

Yanluo, G., *Prediction model of aeronautics equipment spare parts considering spare part sharing.* Paper presented at the IOP Conference Series: Materials Science and Engineering, 2021.

Zhang, L., Deng, Q., Miao, B., Liu, X., & Shao, H., Parallel service mode of production and inventory for spare part inventory optimization. *Knowledge-Based Systems, 241*, 2022. doi:10.1016/j.knosys.2022.108282

Zhao, Q., Pei, Y., Hou, P., & Tian, C., Ranking method for the importance of aircraft damage spare parts. *Vol. 459. Lecture Notes in Electrical Engineering,* pp. 1457-1468. 2019.

Biographies

Abdul Rahman Rizki Wijaya holds his degree in Industrial Engineering from The Department of Industrial Engineering Universitas Islam Indonesia. He was also actively supported the laboratory of Enterprise Resource Planning in the Faculty of Industrial Technology Universitas Islam Indonesia during his study in this university.

Winda Nur Cahyo is an Associate Professor and the head of Master Program in Industrial Engineering, Faculty of Industrial Technology Universitas Islam Indonesia. He holds a PhD from the University of Wollongong in Australia. His main research interest is in the area of Engineering Asset Management. There are several national and international trainings and projects in this area are conducted under his direction. Winda is also one of the declarator of Indonesia Asset Management Community.