Forecasting Solid Alum Sales for Knowledge Management

Fransiska Prihatini Sihotang and Ermatita

December 23, 2021

International Conference on Global Optimization and Its Applications 2021 (ICoGOIA 2021), Bogor, December 23, 2021

Forecasting Solid Alum Sales for Knowledge Management

Fransiska Prihatini Sihotang[1*], Ermatita[2]

[1] *Faculty of Engineering, Sriwijaya University, Palembang, Indonesia*
[1] *Faculty of Computer Science and Engineering, MDP University, Palembang, Indonesia*
[2] *Computer Sience Faculty, Sriwijaya University, Palembang, Indonesia*

**Corresponding author email: fransiskaps@mdp.ac.id*

Abstract (200-300 kata)

Predicting or forcasting the number of sales is important for many manufactured companies. It will greatly affect the procurement of production of raw materials. Forecasting data is important in manufacturing companies to match demand and production capacity. This study tries to analyze the data sales of solid alum using the exponential smoothing method to obtain sales forecasts for the coming period. PT ABC is one of the companies that produce solid alum in Indonesia. Like any manufacturing company, this company also considers data to be important in making decisions. They confirm that data from previous sales are used to support the next decision relate to supplies. This research used 11 months of data sales. This study aims to produce an ARIMA prediction model and compare the prediction results with other methods, that is moving average and exponential smoothing. The comparison shows that the moving average model is more accurate than the other two models. The model produced in this study is ARIMA (2,3,1) model, which is an equation that can be used in sales forecasting. The results of these calculations can be used as knowledge for the production department to help estimate the stock of raw materials in the following periods.

Keywords: Forecasting, ARIMA, Solid Alum, Knowledge Management, Data Mining

1. Introduction

In simple terms, Knowledge Management (KM) is defined as doing what is necessary to get the most out of the source of knowledge. If managed properly, KM can have a broad impact on the organization in general (Becerra-Fernandes & Sabherwal, 2010). KM is considered an important thing, and it is a necessary factor for the competitive growth of the organization (Valmohammadi, 2010). And it has played an important role in the company (Kebede, 2010).

Data mining is one of the most important steps in knowledge discovery (Silwattananusarn, 2012). Data mining is the process of examining large amounts of historical data for new and meaningful patterns for future use. Data mining is used to discover insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data (Zaki & Meira, 2014). In short, data mining guarantees in helping all types of organizations to expose buried trends in historical data (Shastri et al., 2018), such as forecasting demand in food manufacturing companies (Fattah et al., 2018), predicting stocks (Azevedo et al., 2012) (Morariu & Borangiu, 2018), manufacturing process control (Vazan et al., 2017), even in other fields, such as predicting BCG vaccination (Shastri et al., 2018), and rainfall predicting (Mishra et al., 2017).

One of the uses of data mining is to make predictions or forecasts, and the most widely used forecasting method is exponential smoothing. For example in studies that estimate the BCG vaccine using 35 years of historical data (Shastri et al., 2018), the literature study on short-term prediction (Azevedo et al., 2012), predicting rainfall (Mishra et al., 2017), and to improving storage management in manufacturing companies (Luque et al., 2017).

Forecasting data is important in manufacturing companies to match demand and production capacity. Which of course will be related to company values. This study tries to analyze solid alum sales data using the exponential smoothing method to obtain sales forecasts for the coming period. The results of these calculations can be used as knowledge for the production department to help estimate the stock of raw materials in the following periods.

2. Literature Review

2.1 Knowledge Management in Manufacturing Company

Knowledge management (KM) is simply defined as doing what is necessary to get the most out of the source of knowledge, one source of knowledge (database). If managed properly, knowledge management will have a broad impact on the organization in general [1]. Knowledge management is an important and necessary factor for the competitive growth of an organization [2]. Knowledge management has played an important role in various companies [3]. In manufacturing companies, production and sales data are important knowledge and must be processed properly. The data is collected, stored, and then the pattern can be seen to be used as a guide in decision making.

2.2 Mining a Knowledge

Data mining is the process of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data. Based on facts, data mining is part of a larger knowledge discovery process, which includes pre-processing tasks such as data extraction, data cleaning, data fusion, data reduction, and feature construction, as well as post-processing steps such as pattern and model interpretation, hypothesis confirmation and generation, and so on. This knowledge discovery and data mining process tend to be highly iterative and interactive (Zaki & Meira, 2014).

2.3 Moving Average Method

Moving Average is an old forecast method that gives good results when the demand pattern is the rather flat and horizontal type. Parameter m in this method specifies the number of recent requests to use in generating the average. Each of the history demands is given equal weight in computing the average than the average becomes the forecast for the future months (Thomopoulos, 2016). The formula can be expressed as follows.

$$Ft = \frac{\Sigma_{i=1}^{n} A_{t-1}}{m} \tag{1}$$

Ft is the new forecast and At-1 is the previous period's actual demand.

2.4 Exponential Smoothing Method

Exponential smoothing methods in classical time series obtain forecasts as the weighted moving average of all past observations where the assigned weights decrease exponentially (Gardner, 2006). Calculations using exponential smoothing using the following equation.

$$Ft = Ft - 1 + \alpha\,(At\text{-}1 - Ft\text{-}1) \tag{2}$$

Ft is the new forecast, Ft-1 is the previous period forecast, α is the smoothing constant, and At-1 is the previous period's actual demand.

2.5 Autoregressive integrated moving average

Autoregressive Integrative Moving Average (ARIMA) models are a form of the Box-Jenkins model. ARIMA is an inventory modeling approach that can be used to calculate the probability of future values. The advantage of ARIMA modeling compared to simple forecasting and smoothing methods is that it is more flexible in adjusting data. However, identifying and fitting a model may be time-consuming, and ARIMA modeling is not easily automated (Kolker, 2016).

ARIMA combines elements in the autoregressive and moving average models. All data in the ARIMA analysis are assumed to be stationary. If the data is not stationary, it must be adjusted to correct for its instability using differencing. The resulting model is said to be an integrated (differenced) model. This is the source of the "I" in the ARIMA model.

The general model that represents a non-stationary time series is the ARIMA (p, d, q) model. Where p, d, q are the orders for the autoregressive process, differentiation, and moving average. The ARIMA model (p, d, q) can be written in a general form as follows (Wei, 2006).

$$\boldsymbol{\left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right)(1 - B)^d \tilde{z}_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t} \tag{3}$$

2.6 Mean Absolute Deviation (MAD)

MAD is a method for using a forecasting method using absolute errors. MAD measures the accuracy of forecasts by calculating the average of the estimated errors (the absolute value of each error). MAD is useful when measuring forecast errors in the same units as the original series. The value of MAD can be calculated using the following formula.

$$MAD = \frac{\Sigma \text{ the absolute value of the forecast error}}{n} \tag{4}$$

2.7 Mean Absolute Percentage Error (MAPE)

MAPE is calculated using the absolute error in each period divided by the actual observed value for that period. Then, the mean absolute percentage error is calculated. This approach is useful when the size or size of the forecast variable is important in evaluating the accuracy of the forecast. MAPE indicates how much error in prediction is compared to real values. The value of MAPE can be calculated using the following formula.

$$MAPE = \frac{\Sigma\frac{|e_i|}{X_i}\times 100\%}{n} = \frac{\Sigma\frac{|X_i - F_i|}{X_i}\times 100\%}{n} \tag{5}$$

2.8 Tracking Signal

Tracking Signal is a method to validate forecast results. Tracking signals are a measure of how well a forecast predicts actual values. The tracking signal is obtained by dividing the running sum of forecast error by MAD. Tracking Signal values can be calculated using the following formula.

$$Tracking\ signal = \frac{RSFE}{MAD} \tag{6}$$

3. Materials and Methods

3.1. Materials

The object for this study is PT ABC, one of manufacturing company that produce Solid Alum in Indonesia. The data needs of this study are in the form of actual data obtained from solid aum sales data at PT ABC. This data will be used to calculate sales for the following month so that it can be used as a consideration in the supply of raw materials. Chemical sales data obtained from PT ABC, starting from January 2019 to October 2020, can be seen in Table 1. The data is processed using Microsoft Excel, Minitab, and R.

Table 1: Sales of Solid Alum Actual Data

Period	Sales	Period	Sales
Jan-19	1.390.350	Dec-19	1.224.450
Feb-19	1.081.950	Jan-20	1.235.750
Mar-19	1.141.900	Feb-20	1.129.000
Apr-19	1.087.700	Mar-20	1.668.050
May-19	1.295.800	Apr-20	1.402.750
Jun-19	845.500	May-20	831.050
Jul-19	1.223.350	Jun-20	1.108.950
Aug-19	1.069.200	Jul-20	1.066.950
Sep-19	1.191.200	Aug-20	923.900
Oct-19	1.107.500	Sep-20	1.217.800
Nov-19	861.250	Oct-20	812.850

3.2. Methods

This research begins by identifying the problem by conducting interviews with the production supervisor at PT ABC. Then it was found that sales forecasts were needed to facilitate forecasting future production capacity. The next step is to conduct a literature study regarding the forecasting method to be used: moving average, exponential smoothing, and ARIMA. In determining the ARIMA model, 3 main stages are carried out: identification, estimation, and verification. The stages of this research can be seen in Figure 1.

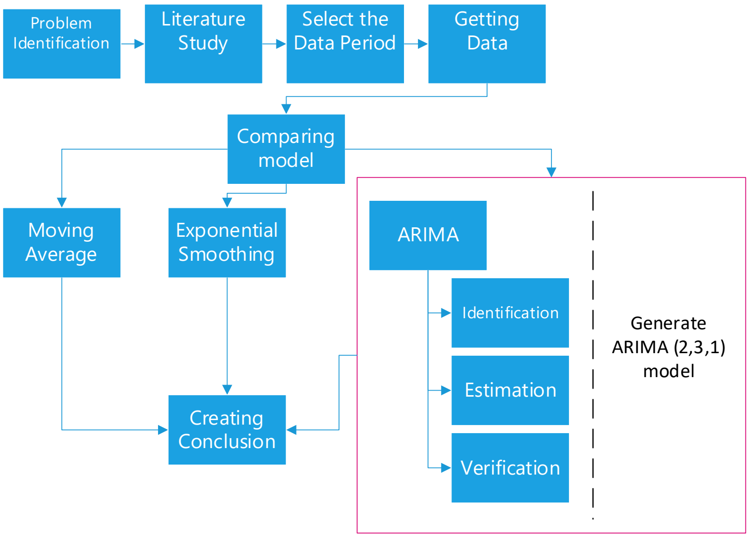

Figure 1: Research Stages

The output of this study is a comparison of forecasting results and an ARIMA model (2,3,1).

4. Results and Discussion

4.1. Solid Alum Sales Forecasting using Moving Average

Forecasting calculations using the moving average method using 4-month historical data can be seen in Table 2.

Table 2: Forecast Using Moving Average

Period	Sales (kg)	Forecast
Jan-19	1.390.350	-
Feb-19	1.081.950	-
Mar-19	1.141.900	-
Apr-19	1.087.700	-
May-19	1.295.800	1.175.475
Jun-19	845.500	1.151.838
Jul-19	1.223.350	1.092.725
Aug-19	1.069.200	1.113.088
Sep-19	1.191.200	1.108.463
Oct-19	1.107.500	1.082.313
Nov-19	861.250	1.147.813
Dec-19	1.224.450	1.057.288
Jan-20	1.235.750	1.096.100
Feb-20	1.129.000	1.107.238
Mar-20	1.668.050	1.112.613
Apr-20	1.402.750	1.314.313
May-20	831.050	1.358.888
Jun-20	1.108.950	1.257.713
Jul-20	1.066.950	1.252.700
Aug-20	923.900	1.102.425
Sep-20	1.217.800	982.713
Oct-20	812.850	1.079.400
Nov-20		1.005.375

The results of forecasting solid alum sales using the moving average method are 1.005.375 kilograms, with the MAD value of 195.035 and the MAPE value of 19.

4.2. Solid Alum Sales Forecasting using Exponential Smoothing

The actual data on solid alum sales in Table 1 are calculated using the exponential smoothing formula using an alpha value of 0.1 to 0.9. Then we get the sales forecasting results for November 2020 as shown in Table 3.

Table 3: Forecast for Each Alpha

Alpha	Forecast (kilograms)	MAD	MAPE
0.1	1.131.379,5	198.007	0.195
0.2	1.065.351,508	187.254	0.181
0.3	1.023.350,907	187.591	0.180
0.4	990.062,8437	1.167.979	0.181
0.5	962.656,4612	186.248	0.179
0.6	937.629,7823	193.770	0.186
0.7	911.852,2068	203.111	0.194
0.8	883.245,3238	1.167.979	0.203
0.9	850.551,0639	1.167.979	0.215

Based on Table 3, it is known that the smallest MAD and MAPE value is obtained at an alpha value of 0.5. Then the alpha value of 0.5 is used for forecasting. The simulation for calculating the MAD and MAPE value is obtained from Table 4.

Table 4: MAD And MAPE Calculations Use Alpha 0.5

Period	Sales (kg)	Forecast (kg)	Sales-forecast (kg)	Sales-forecast/forecast
Feb-19	1.081.950	1.390.350	308.400	0.2850
Mar-19	1.141.900	1.236.150	94.250	0.0825
Apr-19	1.087.700	1.189.025	101.325	0.0932
May-19	1.295.800	1.138.363	157.438	0.1215
Jun-19	845.500	1.217.081	371.581	0.4395
Jul-19	1.223.350	1.031.291	192.059	0.1570
Aug-19	1.069.200	1.127.320	58.120	0.0544
Sep-19	1.191.200	1.098.260	92.940	0.0780
Oct-19	1.107.500	1.144.730	37.230	0.0336
Nov-19	861.250	1.126.115	264.865	0.3075
Dec-19	1.224.450	993.683	230.767	0.1885
Jan-20	1.235.750	1.109.066	126.684	0.1025
Feb-20	1.129.000	1.172.408	43.408	0.0384
Mar-20	1.668.050	1.150.704	517.346	0.3102
Apr-20	1.402.750	1.409.377	6.627	0.0047
May-20	831.050	1.406.064	575.014	0.6919
Jun-20	1.108.950	1.118.557	9.607	0.0087
Jul-20	1.066.950	1.113.753	46.803	0.0439
Aug-20	923.900	1.090.352	166.452	0.1802
Sep-20	1.217.800	1.007.126	210.674	0.1730
Oct-20	812.850	1.112.463	299.613	0.3686
MAD Value of Alpha 0.5 (average)			186.248	
MAPE Value of Alpha 0.5 (average)				0.1792

4.3. Solid Alum Sales Forecasting using ARIMA

Based on the Box–Jenkins Approach, this study will be carried out in three parts: identification, estimation, and verification. Based on data processing experiments using the Minitab application, the Arima model to be used for forecasting is Arima (2,3,1).

1. Identification

The model shown in Figure 2 is based on the demand for solid alum in PT ABC from January 2019 until November 2020.

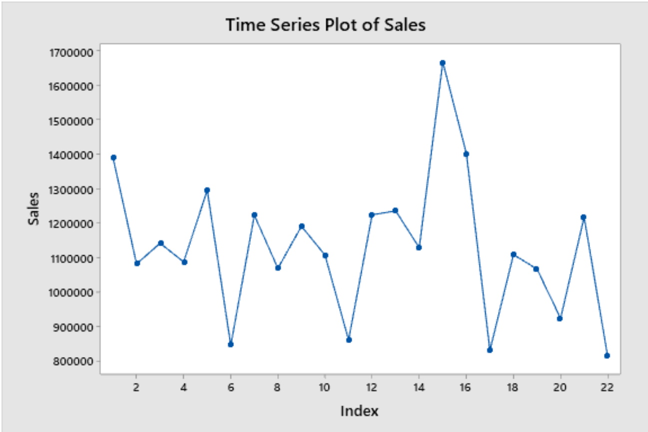

Figure 2: Time Series Plot of Solid Alum Sales

Then, it starts with the initial preprocessing of the data to make it stationary and then choosing possible values of p and q which of course can be adjusted as model fitting progresses. ACF and PACF correlograms can be seen in Figure 3 and Figure 4.

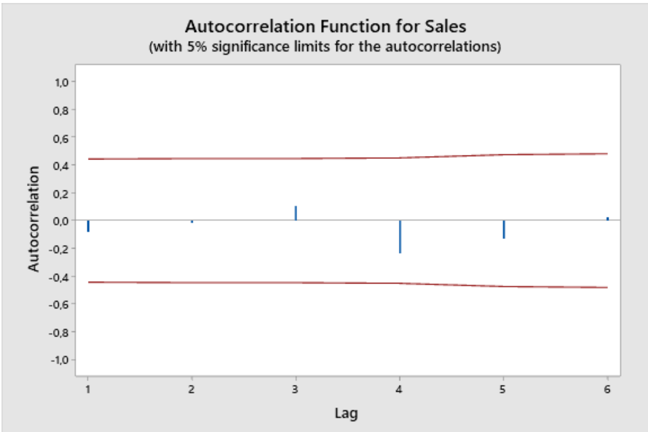

Figure 3: Autocorrelation Function of Solid Alum Sales

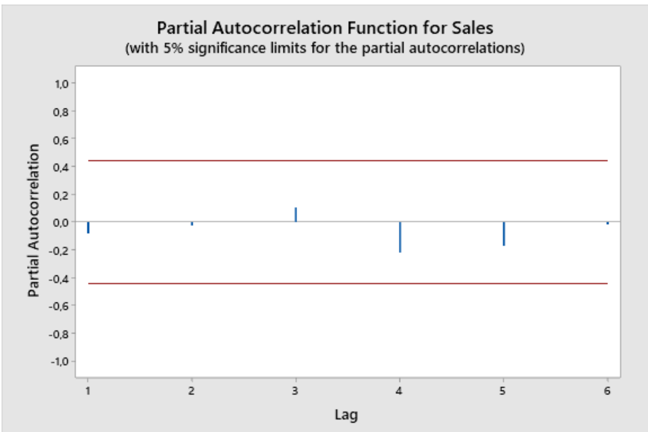

Figure 4: Partial Autocorrelation Function of Solid Alum Sales

In Figures 3 and 4 it is known that there is no cut-off, and by using the R application as shown in Figure 5, based on The Augmented Dickey-Fuller stationary test, the p-value is 0.6347. Thus it can be concluded that the data is not stationary because the p-value is greater than alpha if alpha is 0.05. Then we need differences to make data stationary. In this study, three differences were needed to make the data stationary, as seen in Figure 6.

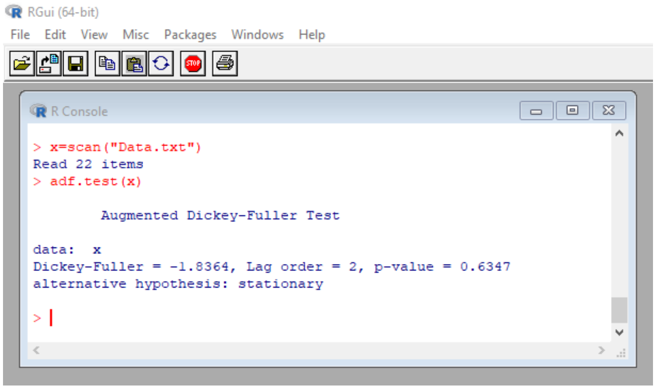

Figure 5: Augmented Dickey-Fuller stationarity test using the R application

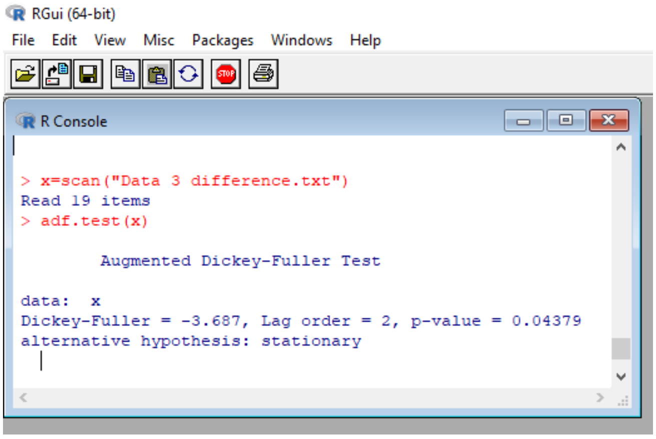

Figure 6: Augmented Dickey-Fuller stationarity test using the R application after 3 Difference

2. Estimation

Theoretically, the estimation process is carried out by including various models. However, referring to the parsimony principle, the simplest model will be used. In this case, the model with the smallest mean square error will be selected. In this case, there is 3 times the difference so that the ARIMA model to be used is ARIMA (p, 3, q).

3. Verification

The results of data processing using Minitab software are shown in Table 5.

Table 5: ARIMA Model Experiment Using Minitab

ARIMA Models	P. Value (L-Jung Box)	Mean Square Error
ARIMA (1,3,1)	0.003	1.51715E+11
ARIMA (1,3,2)	0.009	1.47949E+11
ARIMA (1,3,3)	-	-
ARIMA (2,3,1)	0.618	1.02264E+11
ARIMA (2,3,2)	0.037	1.89994E+11
ARIMA (2,3,3)	0.1	1.88039E+11
ARIMA (3,3,1)	-	-
ARIMA (3,3,2)	-	-
ARIMA (3,3,3)	-	-

To test the adequacy of the model, refer to (Wei, 2006) using the L-Jung Box Test, a good ARIMA model can be used if the P-Value of the Ljung-Box Test > Alpha (0.05).

Based on Table 5, the ARIMA (1,3,1), ARIMA (1,3,2), and ARIMA (2,3,2) models are rejected because the p-value (L-Jung Box) is smaller than 0.05. Meanwhile, other models cannot be predicted using Minitab. Then the model to be verified is two other models, that is ARIMA (2,3,1) and ARIMA (2,3,3). With the principle of parsimony, the model chosen is the one with the smallest Mean Square Error value, that is ARIMA (2,3,1). Parameters of the ARIMA model (2,3,1) can be seen in Table 6.

Table 6: Final Estimates of Parameter ARIMA (2,3,1)

Type	P-Value
AR 1	-0.936
AR 2	-0.567
MA 1	1.120

The general form of the equation for the ARIMA (2,3,1) model is as follows.

$$(1-\emptyset_1 B-\ \emptyset_2 B^2)(1-B)^3\tilde{z}_t=(1-\theta_1 B)a_t \tag{7}$$

By entering the parameters in the Table 6, the equation is as follows.

$$(1+0.936\,B+0.567\,B^2)(1-B)^3\tilde{z}_t=(1-1.120B)a_t \tag{8}$$

4.4. Comparing Forecasting Results

Forecasting results using the Moving Average Method for the next period is 1,005,375 kilograms, with the Exponential Smoothing method is 962,656 kilograms, and with ARIMA (2,3,1) model is 707,800 kilograms. The actual data is 1,051,150 kg. Comparison of forecasting results can be seen in Table 8.

Table 7: Tracking Signal Calculations

Method	Forecasting Result (kg)	Accuracy
Moving Average	1,005,375	95.65 %
Exponential Smoothing	962,656	91.6 %
ARIMA (2,3,1)	707,800	67,3 %

Looking at the accuracy comparison, the moving average model has a higher level of accuracy.

5. Conclussion

This study produces forecasting for the next period using 3 methods: moving average, exponential smoothing, and ARIMA. The three methods can be used to predict solid alum sales. The ARIMA model created from this study is ARIMA (2,3,1) in the form of mathematical equations. The comparison of the three methods shows that the moving average method has a higher level of accuracy.

References

Azevedo, J. M., Almeida, R., & Almeida, P. (2012). Using Data Mining with Time Series Data in Short-Term Stocks Prediction: A Literature Review. *International Journal of Intelligence Science*, *02*(04), 176–180. https://doi.org/10.4236/ijis.2012.224023

Becerra-Fernandes, I., & Sabherwal, R. (2010). *Knowledge Management System And Processes*.

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1–9. https://doi.org/10.1177/1847979018808673

Gardner, E. S. (2006). Exponential smoothing: The state of the art-Part II. *International Journal of Forecasting*, *22*(4), 637–666. https://doi.org/10.1016/j.ijforecast.2006.03.005

Kebede, G. (2010). Knowledge management: An information science perspective. *International Journal of Information Management*, *30*(5), 416–424. https://doi.org/10.1016/j.ijinfomgt.2010.02.004

Kolker, A. (2016). *Healthcare Management Engineering: What Does This Fancy Term Really Mean? The Use of Operations Management Methodology for Quantitative Decision-Making in Healthcare Settings* (Vol. 4, Issue 1). Springer New York Dordrecht Heidelberg London.

Luque, A., Aguayo, F., Lama, J. R., & González-Regalado, E. (2017). Enhanced manufacturing storage management using data mining prediction techniques. *Procedia Manufacturing*, *13*, 956–963. https://doi.org/10.1016/j.promfg.2017.09.166

Mishra, N., Soni, H. K., Sharma, S., & Upadhyay, A. K. (2017). A comprehensive survey of data mining techniques on time series data for rainfall prediction. *Journal of ICT Research and Applications*, *11*(2), 167–183. https://doi.org/10.5614/itbj.ict.res.appl.2017.11.2.4

Morariu, C., & Borangiu, T. (2018). Time series forecasting for dynamic scheduling of manufacturing processes. *2018 IEEE International Conference on Automation, Quality and Testing, Robotics, AQTR 2018 - THETA 21st Edition, Proceedings*, 1–6. https://doi.org/10.1109/AQTR.2018.8402748

Shastri, S., Sharma, A., Mansotra, V., Sharma, A., Bhadwal, A. S., & Kumari, M. (2018). A Study on Exponential Smoothing Method for Forecasting. *International Journal of Computer Sciences and Engineering*, *6*(4), 482–485. https://doi.org/10.26438/ijcse/v6i4.482485

Silwattananusarn, T. (2012). Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012. *International Journal of Data Mining & Knowledge Management Process*, *2*(5), 13–24. https://doi.org/10.5121/ijdkp.2012.2502

Thomopoulos, N. T. (2016). Elements of Manufacturing, Distribution and Logistics. In *Elements of Manufacturing, Distribution and Logistics*. https://doi.org/10.1007/978-3-319-26862-0

Valmohammadi, C. (2010). Identification and prioritization of critical success factors of knowledge management in Iranian SMEs: An experts view. *African Journal of Business Management*, *4*(6), 915–924.

Vazan, P., Janikova, D., Tanuska, P., Kebisek, M., & Cervenanska, Z. (2017). Using data mining methods for manufacturing process control. *IFAC-PapersOnLine*, *50*(1), 6178–6183. https://doi.org/10.1016/j.ifacol.2017.08.986

Wei, W. W. S. (2006). Time Series Analysis: Univariate and Multivariate Methods. In *Technometrics* (Vol. 33, Issue 1, p. 614). Greg Tobin. https://doi.org/10.2307/1269015

Zaki, M. J., & Meira, W. J. (2014). *Data Mining and Analysis*. Cambridge University Press.