Evaluation of Container Forecasting Methods for Analyzing Port Container Terminal Performance Using Agent-Based Simulation

Ryan Setiawan[1], Riyanarto Sarno[2], Abd. Charis Fauzan[3]
Department of Informatics
Faculty of Information Technology, Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
[1]ryan.stwn93@gmail.com, [2]riyanarto@if.its.ac.id, [3]fauzancharis@gmail.com

Abstract

Forecasting is a method to evaluate the performance prediction results of a future company, and in this paper 4 unvariate forecasting models are used for the number of container throughput volumes in a container terminal, where the four models are Moving Average (MA), Simple Exponential Smoothing (ES), Addictive and Multiplicative. And the method proposed for this paper is to prove the value of the prediction accuracy of the container throughput by applying data per day and will be comparing to obtain the measurement accuracy value of prediction result using Mean Square Error (MSE) and Root Mean Square Error (RMSE). It will get the smallest accuracy value as best prediction result from generating whole model for forecasting container throughput which will be implemented into Anylogic simulation as evaluation of event logs performance based on forecast result so that from result of anylogic forecast it will get performance time and cost in the form of event logs for the future time.

Keywords—Forecasting, Anylogic, Business Process, schedulue, container,

I. INTRODUCTION

Containerization is one of the transportation systems to improve transactions in a city or country. Containers at the terminal will pass through many ships on a journey and also require many processes at the time of departure to destination. Where the economic investment will greatly affect the number of container transactions in and out in the city container terminal [1] [2]. In this case is required forecasting which is a method of forecasting the occurrence of an organization in the company for the future based on time series by involving the calculation of data in the past recorded in an event logs so that it can model into the mathematical form for some time to come, event logs used are an information system based on the performance of an organization in the company at this time, where the data can contain the number of container and activities of the company from beginning to end of an organization process business is completed followed by the time of each activity to the cost required when the work done at that time.

So this forecasting model will be used to simulate the performance of the container terminal company of an event logs so as to evaluate the current activity is running based on event logs data. Based on several different forecasting methods that can be searched for the smallest error value of forecasting result, it will be used to evaluate the performance system of the organization based on the amount of container of the company's activity, sojourn time (including execution time + waiting time) from the Standard Operating (SOP).

There have been many researches on forecasting methods based on time series to predict some cases, such as W. Peng et al [3], which can be forecasting container throughput volumes in Taiwan using six univariate models including classical decomposition model, the trigonometric model, the regression model with seasonal dummy variables, the gray model, the hybrid gray model, and the SARIMA model. By applying monthly data and compiling prediction results based on mean absolute error, mean absolute percent error and root mean squared error, it is found that classical decomposition model is the best method to predict container throughput with seasonal variations. Then C. Chou et al [4], by linking empirical evidence and international trade container volume and economic growth in Taiwan. And using modified regression model to forecasting the volumes of Taiwan's import containers and the result is accuracy value generated from regression model shows high prediction value.

And the research then uses Agent-Based simulation which apply the range of modeling agent behavior in stock market, supply chains, and consumer [5]. The ABMS model thus understands factor factors that may be responsible for the business impact of simulation results using computers that support decision making and also on research C. M. Macal et al [6] discusses some illustrative applications in the use of toolkits for the development of Agent-Based methods. In this paper we use the application of Anylogic simulation in the experiment for Agent-Based Simulation (ABS) in the simulation to be able to monitor the business process and organizational form work in it, which in this study also case study taken to be simulated is port container terminal in Surabaya. Using some forecasting methods such as Moving Average (MA), Exponential Smoothing, Additive and Multiplicative to implement ABM in

Anylogic simulation, then we calculate the accuracy of prediction using Mean Square Error (MSE) and RMSE (Root Mean Square Error) methods. And from that result will yield value of forecasting result and minimum error value then the output of simulation is variables container amount for next month. So the data that will be used is the volume ports container terminal from January to March 2016 to generate event time and cost logs from the forecasting simulation for April 2016.

II. RELATED WORk

In this paper will require some discussion of research that supports forecasting methods to predict data by using pre-existing methods with different cases. This is necessary to support the methods used in this study.

Some studies use forecasting Moving Average and gray prediction as one of its methods C. Hung et. al [7] as the monthly time series prediction in NNGC. Furthermore, the two methods of forecasting are combined with approaches support vector regression (SVR) from field data of computational intelligence, which later on the hybrid SVR, Moving Average, and gray prediction are divided from the criteria of MAPE, SMAPE, and RMSE. At the end of the study [8], the SVR hybrid integrity was used Bagging technique from both prediction methods with the same algorithm to generate training sets randomly to archive a better performance data mining. Then in another study that belongs to F. Nhita et al [9] is the same as using Moving Average but added with soft computing as algorithm that is used in general for forecasting especially focused on research that is to support rainfall forecasting. To get maximum result then use 5 types of Moving Average used in data processing. Where the results of rainfall forecasting are used for forecast crops, corns, and potatoes planting calendars, rainfall itself uses four methods of hybrid algorithms in soft computing such as ANFIS, Evolving Fuzzy, Fuzzy-Grammatical Evolution, and Artificial Neural Network (ANN-Nested) genetic algorithms. The later on forecasting results will be accumulated using MAPE (Mean Absolute Percent Error), from the results of the percent error measurement then found that ANFIS has the lowest error 0.1065. then for Planting calendar forecasting used the best prediction method of Modified Moving Average and ANFIS with an accuracy value of 91.67% for crops, corns, and potatoes.

Subsequent research used the method of forecasting Simple Exponential Smoothing (SES) and Simple Seasonal Exponential Smoothing to solve problems such as creating monthly reports with data at this time and for predicting using time series model. The analysis was performed to approximate the number of patient expected from 4 types of Trauma Oro-Maxillo-facial based on May-December 2015 data, and from the forecast results show the value of the analysis and prediction with 95% confidence intervals.

This is necessary to support the methods used in this study. S. Fan et al [10], proposed the semi-parametric additive model for the short-term load forecasting to estimate the relationships between the request and the driver variable. The additional point for forecast is to predict the half-hourly electricity demand for up to the next few days for power systems at the Australian National Electricity Market which will be implemented to the system operators for planning the generators. Using the nonlinear model additive method and nonparametric terms combined with the regression framework, which can see the complex non-liner value of the relationship between electricity demand and the drivers. Forecast results are demonstrated to a remarkably good performs model for both the previous data and real-time on-site implementation data. Furthermore, forecast of season time series Multiplicative Model N. Dongxiao et al [11] which forecasting maximum load data per month which will be in fitting error by GARCH model in modify so that prediction accuracy value is improved.

III. THE PROPOSED METHOD

This proposed method has using 4 forecasting methods as a comparison of the number of containers that will be predicted in the future through the existing data at this time. which will be the fourth forecast method is accumulated by measuring the accuracy of the predicted error value and the result of the lowest error value that will be used as a reference generate logs from forecasting to be simulated to obtain performance time and cost, the following will be explained in fig. 1 generated scheme of the proposed method to be worked out in this paper.

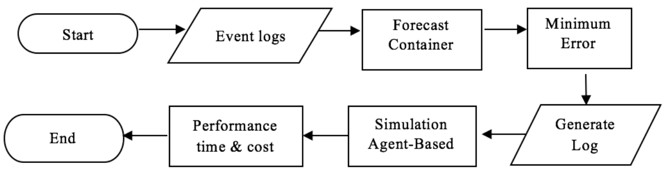

figure 1. Generate scheme for proposed method

Event logs are container throughput data with complete work activity from Terminal Ports container throughput container Surabaya followed by time and cost of every activity in January, February, and March in 2016. In this research will be forecasting with four methods to predicts the number of container throughput for the future.

1. Model MA (Moving Average)

This is one of the time series models that we use to forecasting the number of containers per day, where in this model the dataset of container quantities is currently divided into 2 parts: training data and data testing. In the data training used as a reference container amount for the forecast while the data testing as a comparison of forecast data results using Equation 1 we can get the data forecasting container throughput.

$$\text{MA} = \frac{Y_t + Y_t\text{-}1 + \ldots + Y_t\text{-}(n\text{-}1)}{n} \quad (1)$$

The Moving Average model is a simplest forecasting model using Yt is the value of a data based on the current time, and Ft is the actual forecasting result value with the time of adding t+1 and n is the amount of training data.

2. The simple exponential smoothing model (SES)

Simple exponential smoothing can also be used as a forecasting method based on time series because it is almost equal to the moving average but done by adding the calculation using the constant value as the smoothing value ranges from 0 to 1.

$$St = \alpha * Yt + (1 - \alpha) * Ft\text{-}1 \quad (2)$$

Where:

St = forecast for period t.
Yt + (1-α) = The actual value of time series
Ft-1 = forecast at time t-1 (previous time)
α = constant alignment between 0 and 1

This SES method of constants is related to the accuracy of the smoothing of the forecast result, where the value gives a major relation with the current and the value 0 is related to the previous value so that the forecast error can be slightly reduced.

3. Addictive and Multiplicative Models

As described in [10], for forecasting addictive and multiplicative methods the predicted results are determined on a seasonally significant weekly basis for the case in this paper, and this model also influences the trend in time. So it can be done by Equation 3 for the additive model can be described time series model of the overall value of trend, seasonal effect, cyclical effect and irregual effect.

$$Y_t = Trend + \text{Cycle} + \text{SF} + \text{IF} \quad (3)$$

And for Multiplicative model can be described with Equation 4 which is.

$$Y_t = Trend \times \text{Cycle} \times \text{SF} \times \text{IF} \quad (4)$$

From the Equations 3 and 4 the variables values can be explained that from time series model based on t, SF is the value of the seasonal component to t, and also IF the value of irregular component on t. thus the four models of this time series will be calculated the number of forecast throughput containers for the future [12] [13].

4. Defined Trace based on Event Logs

Based on the data of event logs of Container Terminal Surabaya, it is found that there are 12 traces of all activities using the Gaussian Distribution through Equation 5 that are running from January 2016 to March 2016.

$$f_{(x)} = ae\text{-}\frac{(x\text{-}b)^2}{2c^2} \quad (5)$$

Where a = mean value of the number of frequency container and e = standard deviation from the previous container amount of data that is in January until march 2016 so that the data obtained trace number will be used to determine the performance of container amount of forecast result so it can be simulated to get sojourn time, waiting time, execution time and cost for the future.

5. Measurement of Forecasting Accuracy

The accuracy of forecasting will be measured using the function and Mean Square Error (MAE), and Root Mean Square Error (RMSE) [14] [15], where in this case we are both criteria of measuring accuracy for the four forecasting models as the method of evaluation of the forecasting technique used by measuring the accuracy predicted results of a model using Equation 6 and 7 to get the value of accuracy prediction error.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (X_t \text{-} f_t)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\chi_t \text{-} f_t)^2}{n}} \quad (7)$$

Where Xt and Ft are actual and predicted values of time series in period i for whole n. this obviously gives the impact measurement accuracy has a positive value and the smaller the value obtained on each measurement method accuracy is calculated the better the performance of the forecasting method.

IV. Experiment and Discussion

4.1. Comparison of Forecasting Methods

The volume of container throughput of data Even Logs obtained from January 2016 to March 2016 is calculated into 4 methods of forecast. And we display into the Table 1. of forecast results in Surabaya Container Terminal based on the value of the comparison of the number of container forecasting based on time series.

Table 1. Performance of Various methods of forecasting container throughput

Forecasting Method	Accuracy Measure	
	MSE	RMSE
Moving Average	254134.1	504.11
Exponential Smoothing	406618.9	637.66
Additive	440274.70	663.53
Multiplicative	969337.55	1470.63

The comparison of accurate results between the four forecasting methods using data from TPS container throughput is seen as Moving Average method as a forecasting method which has the lowest accuracy measurement accumulation value of both performance measure, although Multiplicative has low accuracy value with RMSE measurement method but for measurement of error percentage MSE method for forecasting Multiplicative cannot be used as reference of next evaluation simulation.

4.2. Generated Trace Using Forecast Data

So from the TPS case and the results from Table 1 based on the measurement value of MSE and RMSE concluded that Moving Average is selected as data for forecasting container throughput which will be incorporated into the trace to simulate to ABS model in AnyLogic and it will be showing in the Table 2. to be calculated performance data container throughput in the next month.

Tabel 2. Percentage Frequency Container Throughput

No	Trace	Frequency Total	StDev
Trace 1	Quarantine, Dry, Green Line	12953	511.23
Trace 2	Quarantine, Dry, Red Line	578	27.465
Trace 3	Quarantine, Reefer, Green Line	2721	418.66
Trace 4	Quarantine, Reefer, Red Line	96	18.357
Trace 5	Quarantine, Uncontainer, Green Line	14	8.0829
Trace 6	Quarantine, Uncontainer, Red Line	0	0
Trace 7	Dry, Green Line	50428	415.30
Trace 8	Dry, Red Line	1804	164.96
Trace 9	Reefer, Green Line	2463	679.97
Trace 10	Reefer, Red Line	106	52.548
Trace 11	Uncontainer, Green Line	174	12.49
Trace 12	Uncontainer, Red Line	10	2.51

From this data, we can know the amount of container throughput in the next month so that it can be simulated on Anylogic with Agent-Based Simulation, along with message from the activity, sojourn time (Execution time + Waiting time), and cost of any activity in Anylogic simulation. In Table 3. we will present the performance evaluation results from Agent-Based simulation based on sojourn time from every activity that run on AnyLogic simulation from January to April 2016 (from initial data until forecast data), and sojourn time value is shown in Average time from January to April 2016.

Table 3. Evaluation Performance Trace Sojourn Time

Trace	Total Average (Second)	
	Performance January – March 2016	Performance April 2016 (MA)
Quarantine, Dry, Green Line	25354.77	27099.64
Quarantine, Dry, Red Line	27235.92	27659.01
Quarantine, Reefer, Green Line	17422.05	18986.21
Quarantine, Reefer, Red Line	9633.59	10358.85
Quarantine, Uncontainer, Green Line	6870.52	6961.03
Quarantine, Uncontainer, Red Line	0	0
Dry, Green Line	23159.05	23958.53
Dry, Red Line	25375	25702.10
Reefer, Green Line	16190.18	16299.52
Reefer, Red Line	7183.27	739.462
Uncontainer, Green Line	25081.52	29684.47
Uncontainer, Red Line	3646.22	3354.84

Based on Table 3. There is average sojourn time per each activity in a trace divided into 12 traces, and as previously stated that in sojourn time there is execution time and waiting time for each activity based on completed time. To be able to know the execution time it is assumed that the lowest sojourn time of an activity is the value of the value, so the result is in Table 4. where the value of execution time is obtained from time calculation value per trace.

Table 4. Evaluation Performance Trace Execution Time

Trace	Total Average (Second)	
	Performance January – March 2016	Performance April 2016 (MA)
Quarantine, Dry, Green Line	12629.88	12184.64

Quarantine, Dry, Red Line	13108.09	11993.54
Quarantine, Reefer, Green Line	9581.94	9523
Quarantine, Reefer, Red Line	6154.11	6278.82
Quarantine, Uncontainer, Green Line	2036.74	2197.07
Quarantine, Uncontainer, Red Line	0	0
Dry, Green Line	12455.58	12185.34
Dry, Red Line	13885.41	14137.23
Reefer, Green Line	10143	10319.75
Reefer, Red Line	4483.76	4816.30
Uncontainer, Green Line	12385.34	12543.23
Uncontainer, Red Line	2163.91	2296.34

Furthermore, after we get execution time and sojourn time then we can get the total average performance waiting time per trace in January to March to get the total average performance waiting time with this Equation 8 then deal for the next month.

$$WT = ST - ET \tag{8}$$

Where WT (waiting time) is the value we looking for, from the sojourn time value per every minus activity by the lowest sojourn time of each activity or we can say the execution time value is continuously. And from that Equation obtained performance waiting time from the performance of a container container throughput terminal Surabaya from the data Table 3, 4, 5 then next data we can generate to get the performance data ST, ET, and WT for data forecast of data on month to April.

Table 5. Evaluation Performance Trace Waiting Time

Trace	Total Average (Second)	
	Performance January – March 2016	Performance April 2016 (MA)
Quarantine, Dry, Green Line	12724.89403	11989.21081
Quarantine, Dry, Red Line	14127.82105	13009.94466
Quarantine, Reefer, Green Line	7840.110735	7185.40611
Quarantine, Reefer, Red Line	3479.484162	3825.567337
Quarantine, Uncontainer, Green Line	4833.779	5128.359523
Quarantine, Uncontainer, Red Line	0	0
Dry, Green Line	10703.46202	10843.02224
Dry, Red Line	11489.58981	12047.34778
Reefer, Green Line	6047.115689	6578.477142
Reefer, Red Line	2699.518324	256.7476077
Uncontainer, Green Line	12696.1796	12429.0993
Uncontainer, Red Line	1482.318774	1574.307012

From table 3, 4, and 5 the average performance of April 2016 we got based from the event logs of AnyLogic to forecast simulation results, which the performance value is calculated with the normal random value that is influenced by the value of the standard deviation on the devoted activity in a traces.

As for the cost that will be shown in Table 6. That is the total average total cost of each activity in the traces, just as before the data that we know today is from January to March 2016 will produce data forecast results that is in April 2016.

Table 6. Evaluation Performance Trace Cost activity.

Trace	Total Average (USD)	
	Performance January – March 2016	Performance April 2016 (MA)
Quarantine, Dry, Green Line	59391.03	60473.47
Quarantine, Dry, Red Line	84800.73	87205.41
Quarantine, Reefer, Green Line	59410.76	60069.86
Quarantine, Reefer, Red Line	84820.46	89933.45
Quarantine, Uncontainer, Green Line	59445.58	60839.22
Quarantine, Uncontainer, Red Line	84855.27	82544.44

Dry, Green Line	57701.49	52316.61
Dry, Red Line	83111.19	81925.06
Reefer, Green Line	57721.22	54125.62
Reefer, Red Line	83130.92	81241.16
Uncontainer, Green Line	57756.03	58008.19
Uncontainer, Red Line	83165.73	83354.62

Based on Table 6. Evaluation of performance cost in January to March for each trace is not much different compared to performance cost data for April in 2016, it proves that there is no significant movement of marketing or installation of new equipment at Terminal port container Peti Kemas Surabaya at that times.

4.3. Discussion

From the experimental results why we need to find the value of performance time and cost are displayed based on the total average value in the Table 3, 4, and 5 this is because to be able to run the simulation in Anylogic we must generate event data logs at this time until it can be done forecast for performance time and cost in the next month in April 2016 based on the number of container throughput of each trace. And for the amount of container throughput based on existing traces we can do with the normal Gaussian by Equation 5 then we will from the average frequency value and standard deviation per day in Table 2 for the month of April 2016, so it is known the number of containers on the whole trace based on the number of containers that have been in forecast.

V. Conclusion and Future Works

From the results of the comparison of the four methods of forecasting which have been discussed in this paper are Moving Average, Simple Exponential, Addictive, and Multiplicative based on the number of container throughput per day in January, February, and March 2016 at Container Ports containers ports Surabaya, we were to compare the value of accuracy error prediction using MSE and RMSE for each model. So from the results of both methods comparison was obtained Moving Average model proved to get the value of accuracy of the smallest error prediction, so that the amount of container throughput of Moving Average forecasting used into Agent-Based Simulation using Anylogic. And the result of this experiment is to determine the current time and cost in order to generate new event logs for AnyLogic tools simulation and message from ABS simulation, and to determine the trace number of container throughput is done on Equation 5. so we can specify the container throughput is through the trace which is so no missing data from forecast results.

Besides Anylogic simulation will result in evaluation of performance time and cost as shown in Table 3 and 4, from the current time to the future. So the results of this forecasting paper evaluation suggest to calculate the optimization value so it can be obtained the most optimal trace based on time and cost, and can be calculated performance performance evaluation of forecast data results to get the appropriate time with the cost cost is not too high. This can be used as the next research method from the data of forecast simulation data at Terminal container port throughput Surabaya.

References

[1] V. Gosasang, W. Chandraprakaikul, S. Kiattisin, "An application of nueral networks for forecasting container throughput at bangkok port," Proceedings of the world congress on engineering, vol. 1, 2010.

[2] A. Jamal, R. Sarno, H. Ginardi, "Analyzing the benefit of implementing integrated DGPS and terminal operating system at yard terminal Surabaya," Journal of Theoretical and Applied Information Technology. vol. 94. no. 4. pp. 948-959. 2017.

[3] W. Y. Peng, C. W. Chu, "A comparison of univariate methods for forecasting container throughput volumes", vol. 50. Matthematical and computer medelling, vol. 50, pp. 1045-1057. 2009

[4] C. C. Chou, C. W. Chu, G. S. Liang, "A modified regression model for forecasting the volumes of Taiwan's import containers," Mathematical and Computer Modelling, vol. 47, pp. 797-807. 2008.

[5] A. Ningpramuda, R. Sarno, E. Suryani, A. C. Fauzan, "Dynamic simulation of electricity supply and demand for industry sector in East Java," Int. Conf. on Science in Infor. Technology (ICSITech). 2017.

[6] C. M. Macal, M. J. North, "Agent-based modeling and simulation," Proceedings of the 2009 winter simulation conf, vol. 9, pp. 86-98. 2009.

[7] C. Hung, X. Y. Huang, H. K. Lin, Y. H. Hou, "Integrated time series forecasting approaches using moving average, grey prediction, support vector regression and bagging for NNGC," IEEE. 2010.

[8] F. Nhita, D. Saepudin, D. Triantoro, Adiwijaya, U. N. Wisesty, "Implementation of moving average and soft computing alorithm to supportplanting seasonal calendar forecasting system on mobile device," International Conf. on Science in Inf. Technology ICSITech, pp. 114-118. 2016.

[9] C. G. Dascalu, G. Zegan, E. R. Cernei, R. B. Mavru, "Time series analysis in predicting the Oro-macillo-facial traumas," IEEE Int. conf. on E-Health and Bioengineering – EHB, pp. 19-21. 2015.

[10] S. Fan, R. J. Hyndman, "Short-term load forecasting using semi-parametric additive models," IEEE power and energy society general meeting, 2011.

[11] N. Dongxiao, Z. Yunyun, L. Jinpeng, "The application of time series seasonal multiplicative model and GARCH error amending model on forecasting the monthly peak load," International forum on computer science-technologu and applications, pp. 135-138. 2009.

[12] M. A. Yaqin, R. Sarno, A. C. Fauzan, " Scalability measurement of business process model using business process similarity and complexity," IEEE 4th Inter. Conference on Elec. Engineering, Computer Science and Informatics (EECSI), 2017.

[13] A. C. Fauzan, R. Sarno, M. A. Yaqin, A. Jamal, "Extracting common fragment based on behavioral similarity using transition adjacency relations for scalable business processes," IEEE 11th International Conference on Information & Communication Technology and System (ICTS), 2017.

[14] A. C. Fauzan, R. Sarno, M. A. Yaqin, "Petri net arithmetic models for scalable business processes," IEEE 3rd Inter. Conference on Science in Infor. Technology (ICSITech), 2017.

[15] A. C. Fauzan, R. Sarno, M. A. Yaqin, "Performance measurement based on coloured petri net simulation of scalable business process," IEEE 4th Inter. Conference on Elec. Engineering, Computer Science and Informatics (EECSI), 2017.