

# PREDICTIVE ANALYTICS APPLICATION TO IMPROVE OVERALL EQUIPMENT EFFECTIVENESS (OEE) USING MACHINE LEARNING

Rikip Ginanjar  
Faculty of Computer Science  
President University  
Bekasi, West Java  
rikipginanjar@president.ac.id

Calvary Karamoy  
Faculty of Computer Science  
President University  
Bekasi, West Java  
chroniclechevalier@gmail.com

*Abstract—The fourth industrial revolution, or also known as Industry 4.0, enables information technology to be implemented within the industrial sector. Many opportunities to improve aspects within the manufacturing process could be done within the transition. One of the possible improvements could be done on the overall equipment effectiveness (OEE). OEE is used to calculate machine performance using variables that are related to the machine itself. However, OEE could only provide the current machine effectiveness without giving additional information for future performance. There is an opportunity to develop such a system or application that could create forecast analysis of the machine performance. With that system or application, it is possible to create a future baseline to benchmark as a target for future performance. Thus, improving the OEE by creating an accurate target for the process. Therefore, it is possible to create a predictive analytics application on a common platform used in the industrial sector to improve OEE.*

*Keywords— Industry 4.0, Predictive Analytics, Overall Equipment Effectiveness (OEE), Machine Performance, Forecast Analysis, Smart Manufacturing, Industrial IoT, Performance Benchmarking*

## I. INTRODUCTION

The industrial revolution, also known as Industry 4.0, is the latest industrial trend that applies the internet and other advanced technologies to their industrial process [2]. The information technology fields have become relevant to the industrial process in this revolution. The implementation of information technology could happen because of the existence of the Internet of Things (IoT). Nowadays, machines are expected to have the capability of calculating the process outcome, predicting their maintenance period, and providing reasons when something is not

supposed to happen. The implementation of these machines is the concept of creating a smart factory. A smart factory is a manufacturing factory that utilizes smart machines to do their industrial process. Industrial processes within this smart factory are envisioned to improve the manufacturing process. As information technology is now used as the foundation of the current industrial technology, there is an opportunity to increase the manufacturing process by implementing information technology to improve machine efficiency, also known as the Overall Equipment Effectiveness (OEE).

Information technology has now been implemented as an industrial process to create a smart factory that utilizes advanced machines or systems to improve their productivity. There is an opportunity to implement an information technology field to improve machine efficiency which adopts the concept of the smart factory. To improve machine effectiveness, also known as equipment effectiveness, an analytics application that can provide machine performance can be developed. The analytics application should also implement machine learning to provide predictive analytics of machine performance that can be used as a baseline for future production.

### 1. Limitations

The application made in this project has several limitations due to its application for every manufacturing area can be slightly different. The first limitation is that the prediction of the predictive analytics features might not be accurate in some cases. This limitation emerges

due to variables related to machine performance is not the same for every manufacturing process. As an example, a certain manufacturing process might have an operator who operates the machine, while other manufacturing process does not have an operator who operates the machine since the machine is capable of doing the production process on its own. Thus, some variables related to operator performance will not be included in the prediction analysis variables. In addition, the application will not include any variables other than set variables for the prediction process. In other words, whether there are variables for operator performance, the application will not take it as an input for the machine learning process.

## II. LITERATURE REVIEW

The machine performance is calculated using several tools or algorithms. The machine performance by itself is calculated using overall equipment effectiveness (OEE) calculation, while the prediction is done using regression methods.

### 2.1 Overall Equipment Effectiveness

Overall Equipment Effectiveness (OEE) is a calculation method to measure the effectiveness of machines or tools within the production processes [7]. There are three main components of OEE: availability, performance, and quality. Availability is the ratio of the available time for the machine or equipment to be used and the actual available time for the machine or equipment is being used. Performance is a relation between the planned or ideal machine or equipment productivity within a set of time with the actual machine or equipment productivity. Quality in OEE calculation is referring to the percentage of good quality products compared to the total of the products within a set of production processes. Thus, OEE can be formulated as the following:

$$OEE = A \times P \times Q$$

where  $OEE$  is the overall equipment effectiveness in percentage,  $A$  is the availability of the machine in percentage,  $P$  is the performance of the machine in percentage, and  $Q$  is the quality of the machine production in percentage.

### 2.2 Machine Learning

Machine learning is a subset of artificial intelligence that focuses on creating systems or algorithms that always learn from a dataset and increase its accuracy over time without certain programming. Machine learning is a computer's capability to learn without having to explicitly program the logic to the computer itself [6]. Machine learning can be classified into four types: supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is a machine learning algorithm that uses labeled datasets, where the input and output are identified [4]. Unlike supervised learning that has some sort of supervision for its input and output variables, unsupervised learning works on the assumption that there is no supervision. There semi-supervised learning combines supervised learning with unsupervised learning, where there are huge quantities of input data but not all of the data are labeled. On the other hand, reinforcement learning uses models to oversee the situation and determine appropriate action that would result in the best outcome.

### 2.3 Regression Algorithm

Regression algorithm is one of the machine learning algorithms. Supervised machine learning is a type of learning which extracts structures or patterns from labeled datasets. As supervised learning relies on labeling its datasets, supervised learning can be classified into two types: linear regression and non-linear regression. Linear regression is a statistical tool that is used to find the relation between one or more variables with another variable, where those variables can be defined as the independent variables and the dependent variable [3].

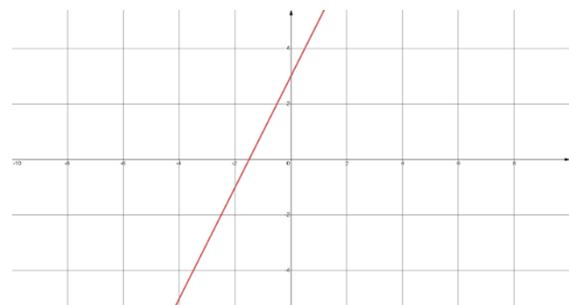


Figure 1  $y = 3x + 2$  Graph

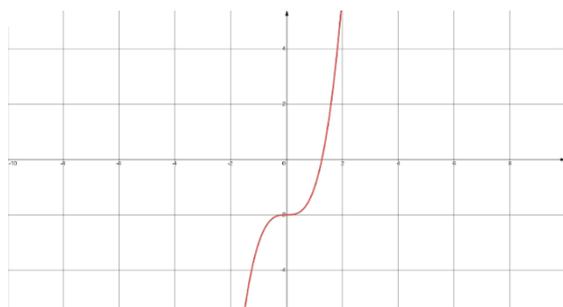


Figure 2  $y = x^3 - 2$  Graph

Figure 1 represents the linear regression functions, while figure 2 represents the non-linear regression functions. The linear functions can be defined as functions with the concept where the input and output are constants. Thus, the linear regression functions are always in the form of the first degree of a polynomial. In a two-dimensional space, linear functions are forming straight lines.

### 2.3.1 Multi-Linear Regression

Multi-linear regression is a statistical analysis method used to determine the correlation between two or more independent variables with the dependent variable [10]. This method is also known as multivariate linear regression. The main difference between multi-linear regression and standard linear regression is how the method handles the independent variables. The standard or univariate linear regression only involves one independent variable, while multi-linear regression utilizes two or more independent variables within its statistical approach. It can be defined that multi-linear regression is a further expansion from the standard linear regression.

### 2.3.2 Multi-Polynomial Regression

Polynomial regression is one of regression statistical analysis methods to find the relationship between independent variables and dependent variables where the dependent variable is labeled as an nth degree polynomial of independent variable [11]. Thus, polynomial regression can be defined as a quadratic function, while also being a cubic function. The degree of the polynomial function will determine the function and the overall result.

### 2.3.3 Random Forest Regression

Random forest is an algorithm used to classify large amounts of data by merging trees by conduction training on the owned sample data [12]. In other words, random forest is the

combination of trees of the best decision tree which are then combined into one model. The aforementioned trees came from the decision tree which represents the structure of its model that resembles a tree. The random forest algorithm can also be used as a regression by expanding the tree completely so that each leaf has exactly one value. A prediction can then be gained by returning the individual response variable from the distribution of the leaves.

## III. EXPERIMENTAL RESULT

The experimental result describes the result of the testing scenarios that were used to evaluate the application. The result of the testing scenarios can be seen in table 1.

Table 1 Testing Scenarios

No	Scenario	Expected Result	Evaluation
1	Start the web application	The web application is loaded	As expected
2	Register user account	The user account is registered	As expected
3	Login with their user account	The user account is authenticated	As expected
4	Open the Dashboard page	The data from the server is fetched and the Dashboard components are loaded.	As expected
5	Open the Trends page	The data from the server is fetched and the Trends component are loaded	As expected,

6	Upload new SSC file	The data from the client is sent and the processed data is returned	As expected
7	Change Selected Date	The data of the Dashboard component is updated	As expected
8	Start the Telegram bot	The Telegram bot is started	As expected
9	Login to the Telegram bot	The user account is authenticated	As expected
10	Check the OEE status	The Telegram bot send the OEE status message	As expected
11	Change the broadcast setting	The Telegram bot send the OEE status message based on the user's selected setting	As expected
12	Change the broadcast time	The Telegram bot send the OEE status message based on the user's selected setting	As expected
13	Logout from the Telegram bot	The Telegram bot logs out the user	As expected

14	Login with incorrect username and/or password	The user is not logged in.	As expected
15	Register with already registered username	The user is not registered.	As expected
16	Login with incorrect username and/or password to the Telegram Bot	The user is not logged into the Telegram bot.	As expected

As for the prediction, the function is using a dataset that has an average availability of 96,74%, average performance of 95.84%, average quality of 97.29%, an average OEE of 90.28%, and an evaluation score of 2.7572. The detailed summary of the datasets can be seen in table 2.

Table 2 Datasets Per Date

Date	Availability	Performance	Quality	OEE
12/13/2021	95.94%	94.99%	95.44%	86.98%
12/14/2021	96.41%	96.24%	97.18%	90.16%
12/15/2021	96.09%	94.24%	97.75%	88.52%
12/16/2021	97.46%	95.97%	98.89%	92.50%
12/17/2021	97.57%	96.95%	97.25%	91.99%
12/18/2021	96.99%	96.63%	97.12%	91.03%

Based on the data of the datasets which has a dataset evaluation score of 2.7572, the regression method that has the closest evaluation score is the multi-linear regression method which has a 2.722 evaluation score. On the other hand, multi-polynomial regression has a score of 2.5883 and random forest regression has a score of 2.7035. The detailed result summary can be seen in table 3.

Table 3 Prediction Result Per Date

Date	Availa bility	Perfor mance	Qua lity	OE E
12/13/ 2021	95.94%	94.99%	95.4 4%	86.9 8%
12/14/ 2021	96.41%	96.24%	97.1 8%	90.1 6%
12/15/ 2021	96.09%	94.24%	97.7 5%	88.5 2%
12/16/ 2021	97.46%	95.97%	98.8 9%	92.5 0%
12/17/ 2021	97.57%	96.95%	97.2 5%	91.9 9%
12/18/ 2021	96.99%	96.63%	97.1 2%	91.0 3%

#### IV. DISCUSSION

This section elaborates the decision of using multi-linear regression, multi-polynomial regression, and random forest regression as the machine learning model for the OEE prediction. There are several reasons why the aforementioned regressions are used for the application. Based on the researches which used OEE predictive analytics as their topic, several models were being used for the machine learning prediction. The most commonly used model is support vector regression (SVR). Based on the researches result, SVR was proven to be effective for handling OEE prediction. Hence, the regression algorithms were then decided to be used for the application. However, although SVR was proven to be effective, the algorithm itself is still not accurate for some cases. Therefore, different kinds of regression models are selected to give alternate predictions which can provide different results compared to SVR prediction.

Amongst all the regression algorithms, multi-linear regression, multi-polynomial regression, and random forest regression were selected due to their nature. Multi-linear regression can give a linear or straight prediction. Multi-polynomial regression can give a very curvy prediction. Random forest regression can give a slightly random prediction. By comparing the three different algorithms, the best prediction result can be obtained for most cases.

#### V. CONCLUSION

The OEE predictive analytics application was meant to create an OEE monitoring medium for people within the manufacturing area. The purpose of this project is to develop the OEE predictive analytics application in order to give an overview of the machine performances. Based on the data provided in the application, the user whether they are from the production side of view or planner side of the view is able to determine what is the necessary action to do within the manufacturing process. Thus, the machine performances can be fully utilized and the output result can be controlled while maintaining the best quality product.

The application is working as expected but there is still room for improvement. For starters, the data gained from the SSC can be expanded to get more variables that are related to machine performance. The accuracy of the prediction could also be furtherly increased by having another type of model implemented on the application. For future works, the application could also have multi-logarithmic regression as the intermediary of both linear and polynomial functions. In addition, the application could also implement other types of regression such as logistic regression or artificial network regression to improve the accuracy of the prediction for some other cases.

#### VI. ACKNOWLEDGMENT

The author would like to thank: Mr. Rikip Ginanjar as my thesis advisor that has helped me and has given me ideas for my final project, all the lecturers that have shared their valuable knowledge and wisdom during my study, and all my friends that have given me their support along the way.

References

- [1] Ahuja, I., & Jaimal, K. (2008). Total productive maintenance: Literature review and directions. *International Journal of Quality & Reliability Management*, 25(7):709-756. doi:10.1108/02656710810890890
- [2] André, J. C. (2019). Industry 4.0: Definition and the Acceleration of Innovations. *Industry 4.0*, 1-65. doi:10.1002/9781119644668.ch1
- [3] Gujarati, D. N. (2018). *Linear Regression: A Mathematical Introduction*. SAGE Publications, Inc.
- [4] Murphy, K. P. (2021). *Machine Learning, Second Edition: A Probabilistic Perspective*. United States: MIT Press.
- [5] Okpala, C., & Anozie, S. (2018). Overall Equipment Effectiveness and the Six Big Losses in Total Productive Maintenance.
- [6] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *Computer Games I*, 206-226. doi:10.1007/978-1-4613-8716-9\_15
- [7] Stamatis, D. H. (2010). *The OEE Primer: Understanding Overall Equipment Effectiveness, Reliability, and Maintainability*. Productivity Press.
- [8] Anusha, C., Umasankar, V. (2020). Performance prediction through OEE-Model. *International Journal of Industrial Engineering and Management*, 93–104. doi: 10.24867/ijiem-2020-2-256 34-f659cc1d0f10
- [9] Sudarsanam, S.K. (2017). A Neuro Fuzzy Model for OEE Prediction. *International Conference on Business Analytics and Intelligence 2017*, Bangalore, India.
- [10] Freedman, D. (2005). *Statistical Models: Theory and Practice*. Cambridge University Press.
- [11] Ostertagová, E. (2012). Modelling using polynomial regression. *Procedia Engineering*, 48, 500–506. doi: 10.1016/j.proeng.2012.09.545
- [12] Tin Kam Ho. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. <https://doi.org/10.1109/34.709601>