CommIT (Communication & Information Technology) Journal 13(1), 17–23, 2019

A Comparison of Machine Learning Algorithms in Manufacturing Production Process

Rosalina
Program Studi Teknik Informatika, Fakultas Komputer, Universitas Presiden
Jawa Barat 17550, Indonesia
Email: rosalina@president.ac.id

***Abstract*—This research aims to improve the productivity and reliability of incoming orders in the manufacturing process. The unclassified data attributes of the incoming order can affect the order plan which will impact to the low productivity and reliability in the manufacturing process. In order to overcome the problem, machine learning algorithms are implemented to analyze the data and expected to help the manufacturing process in deciding the incoming order arrangement process. Four machine learning algorithms are implemented (Decision Tree, Nave Bayes, Support Vector Machine, and Neural Network). These machine learning algorithms are compared by their algorithm performance to the manufacturing process problem. The result of the research shows that machine learning algorithms can improve the productivity and reliability rate in production area up to 41.09% compared to the previous rate without any dataset arrangement before. The accuracy of this prediction test achieves 97%.**

***Index Terms*—Machine Learning Algorithm, Manufacturing, Reliability, Productivity**

I. Introduction

THE growth in market demand and customers' needs upon higher product and process quality and efficiency leads the companies to come up with new and innovative ways to enhance their production. When the manufacturing process is getting more complex, the information is an important part of the management of product and process quality [1, 2]. Tracing and managing products quality lead to large amounts of information and data that have to be handled. As there is much information, it can mean an impendence to the process quality [3]. The manufacturers have to know what and which information and data need to be brought together, how the products can be ordered, and how it can be utilized with maximal effect.

Many researchers have evaluated the performance of machine learning in manufacturing industry [4–8]. Applying machine learning in manufacturing industry aims to reduce product defect, raise the speed of production process, boost transition times, and cut down labor costs. Machine learning is also found to provide the promising potential to improve quality control optimization in manufacturing process [9]. Implementing machine learning algorithms provides information from existing datasets to predict future behavior of the system [10]. That information may support the company in its decision making.

Currently, the manufacturing industry is in the phase of cloud manufacturing [11]. It takes advantages of Internet of Things (IoT) and data science by equipping themselves with emerging infrastructure such as sensors and communications capabilities. It can reduce operational downtime, increase the level of automation the quality of the product, and increase the response time of the customer demands [12–14].

This research is taken in a multinational company which produces clothes label of many top brands in the world. The distribution of the label is following some of the seasons in the countries. The sales rating of the label is affected by those seasons. A significant rating of sales occurs at the beginning of every single season that most people begin to change their style of clothes to adapt to the changing weather. This kind of time will become the standard of determining the process in this company as a high season. In the high season, the company has to analyze the process since there is much order.

There are four significant products in the manufacturing process. First, digital products, it is the label with paper material such as price label tag, brands paper tag, and stickers. Second, fabric label is a label with a tape material such as the label on a cloth. Third, Radio Frequency Identification (RFID) label is the label using base roll tape with the security chip feature. Last, the woven is a brand label by using yarn material. One of the products that have a high order is

Received: Nov. 30, 2018; received in revised form: Mar. 21, 2019; accepted: Apr. 05, 2019; available online: Apr. 30, 2019.

the fabric label product since the technology used is quite simple and faster.

Since the fabric label process has the limitation of capacity, the process of order arrangement needs to be controlled based on the level of the complex order and the quantity of the incoming order when the order level is in the high season time. When the customers' demands keep increasing, the order will be placed in the subcontractor of the company to fulfill the sales level demand. It means that the company will use the third party to precede the incoming order.

When the incoming order quantity is over the production capacity, it will make the reliability time increase. Then, when the complexity of the incoming order is high, it will affect the level of productivity. These kinds of problems appear because there are some gaps due to a different standard to precede the order between in-house process and subcontractor process. This condition makes the company need to decide which one from the order type that will proceed in-house and which one will proceed in a subcontractor to control the result of metrics performance.

This condition is affecting the reliability of production lead time. It will affect the customers' dependency to put their order in this company next time. Another impact is the increment of the customers' possibility to claim the lateness of their order. It means it will consume additional cost. Another impact is in the productivity of the process. When productivity is low, it shows that the operator results in the low output while they consume higher working hours. It will also consume additional cost for extra hours. When this condition continuously occurs, it will affect company performance.

Therefore, it is important to find a proper technique to improve the productivity and reliability rate in this manufacturing process. This research aims to predict the effect of incoming order localized to the productivity and reliability of lead time order by using machine learning algorithm. It will be applied in the manufacturing process. In this research, four machine learning algorithms will be implemented (Decision Tree, Naive Bayes, Support Vector Machine, and Neural Network). These machine learning algorithms will be compared by their algorithm impact on the manufacturing process problem to know which machine learning algorithms will give the best performance to this manufacturing process database.

II. Literature Review

A. Manufacturing Process

The globalization of markets, growth of expectations, and full competition in all spheres of the relation between the customers and the suppliers are the factors extorting the quality in the strategy of the enterprise [15]. The quality in the competitive industry means the necessity of fulfillment for customers' requirements who occupy the leading position on the market. Meanwhile, the quality of the production process means a reliable and sustainable process. Therefore, the common objective of manufacturing enterprises is to increase overall production productivity and reliability.

In other words, they look for output maximization of their current resources by reducing the wastes in equipment and process reliability. Equipment and process reliability jointly create reliable production. The system of reliability assessment and prediction has become an increasingly important aspect of the process in different stages. It is important to develop efficient reliability assessment techniques for complicated systems with several methods and different failure mechanisms to ensure adequate performance under extreme and uncertain demand [15]. Reliability requirement for production process ensures the sustainability of the whole enterprise.

B. Reliability

In this research, reliability is the measurement of how often the products are shipped on time to the promised date given to the customer. Reliability of an incoming order lead time is measured by a formula which will be presented in percentage rate (0%–100%). The bigger the percentage is, the better the reliability is performed.

To ensure that the reliability rates are fair for the company and customers, there are business rules that should be accepted by both sides. Here are the business rules:

- Final promised date: ship date confirmed to customers up to 24 hours after incoming order receipt;
- All orders must be shipped in a sequence of promised date;
- The promised date must be confirmed to the customer within 24 hours of placing an order;
- Include all inter-company shipments;
- Include all order lines: both on-site and off-site production (including non-platform orders), as well as sampling orders received by customer service;
- Do not include order sampling production.

The formula to measure the reliability rate percentage is shown in Eq. (1).

$$\text{Reliability (\%)} = \frac{n}{N} \times 100. \quad (1)$$

It shows n as the number of order lines shipped/before the promised date and N as the total number of order lines shipped. For example, the printed label product in one week is recorded. Then, the number of order shipped or before the promised date is 90 order lines. Meanwhile, the total number for that order in the week is 95 order lines. The reliability rate calculation is shown as follows:

$$\text{Reliability (\%)} = \frac{90}{95} \times 100 = 94.7\%. \quad (2)$$

To see whether the reliability rate is performed well or not, the standard is created. For the reliability rate, the standard is 96%. It means that the order shipped on time should be less than the total order in that week.

C. Productivity

Utilization Equipment Effectiveness (UEE) is an appropriate measure for manufacturing organizations. It has been used broadly in the manufacturing industry to monitor and control the performance (time losses) of an equipment or work station within a production system. The UEE allows the manufacturing industry to quantify and to assign all the time losses that affect equipment in the production to three standard categories. Being standard and widely acknowledged, UEE has been a powerful tool for production systems performance benchmarking and characterization, and as the starting point for several analysis techniques for continuous improvement and research. Despite this widespread and relevance, the use of UEE presents limitations [16].

UEE measures how efficiently the equipment is utilized when the schedule is operated. The equation used is a comparison of the outstanding units and the total target per order line. Equation (3) shows the formula to measure UEE.

$$\text{UEE (\%)} = \frac{\text{Total good units produced}}{\text{Total target units}} \times 10. \quad (3)$$

The target units are different for each order lines. It depends on the product type which is the units of the product such as meters, pieces, clicks, and others. In this research, the unit of the product is measured in meter, which the global standard is 1.200 meters per hour.

III. Research Method

Direct observation in the manufacturing process is conducted to gather initial data and determine which focused area is in the research. The researcher continues by collecting information by interviewing and having a discussion with the related people. To understand the whole process that the data are taken, the researcher creates the flow process of the observation area. Figure 1 shows the flow process diagram of processing an order.

The incoming order process diagram shows that there are 11 main processes. The process starts in the commercial department. The customer service will follow up with the incoming order process. The customer will be directly handled by customer service admin to get the information of incoming order placed.

Then, the information of the incoming order will be proceeded by the planning team in the production area. It will be arranged in some production sub-area such as the filmmaker that will create the film and others. After the film is done, the filmmaker operator will send the film to the rubber plate maker who will process the film into the rubber plate. Then, the kitting order operator will check and combine all the information in order instruction paper with the material needed. These materials are sent to the printing area. The production is started by the printing operator who prints the requested order quantity.

The printing result will become the Work in Progress (WIP) of printing that will be taken by cut and fold operator. The label which is in the form of the roll will be cut into the pieces of the label based on the requested label length in the design. These labels will be scaled and checked by Quality Control (QC) and sorted before it will be preserved in the huge oven. This process is to make sure the label is in good qualify.

The final process in production is packing. All finished goods are packed based on the customers' code. Then, the finished goods are sent to the warehouse to proceed to delivery to the customers' address.

In this research, data are collected. It is divided into two phases during the data collection (data gathering and observation) and calculation (data tabulation and calculation). The first phase of data collection and calculation is as the guidance in choosing the topic. The data gathered in the first phase will proceed into data calculation that will be needed to complete this research. The next phase is data tabulation and calculation to analyze the dataset. This phase consists of the data quality. The quality of each attribute will be checked to prevent the missed data of converting data format (data are converted into the suitable format of the application), discretizing the continuous data to prevent the thousand continuous data, and setting class chosen from the attribute list.

The process starts by building the model using machine learning algorithm. Once the model is done and resulted in high accuracy, the top attributes are picked from the rank results. The top attributes are

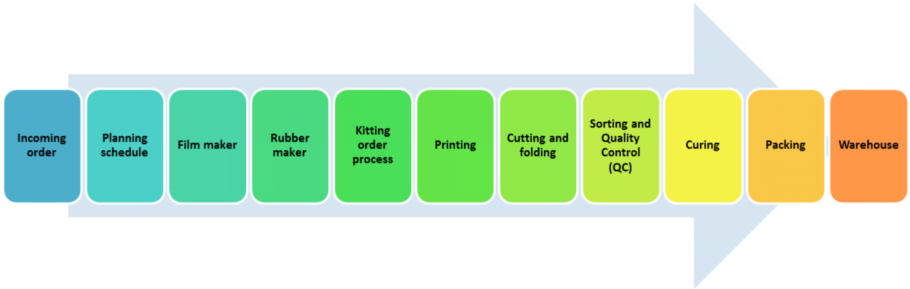

Fig. 1. Incoming order process flow.

the attributes that contribute to the productivity and reliability rate the most. From these top attributes, the researcher will find other data and try to arrange the data manually based on the attribute priority. The chosen data are the next quarter production data in the 2^{nd} quarter in 2014. The data consist of 1.000 datasets since the data are only for testing or called as the predicted dataset. Meanwhile, 3.000 datasets are used as a training dataset to create the model test.

To create the high accuracy of the dataset, the researcher divides the training dataset into two categories: 66% of the training datasets are used as a sub-training dataset, and the 34% is used as a sub-training dataset. The data that have been arranged will be predicted by using the machine learning algorithm again. Once it finishes, the result will be compared and checked with the actual result of the dataset to see the accuracy of the prediction.

Figure 2 shows the data analysis process. The collected data with the product attributes are classified using machine learning algorithms. The results help the decision maker to decide the order placed. With the well-planned incoming order, it will impact the reliability and the productivity of that manufacturing process.

IV. Results and Discussion

By conducting 66% of datasets as the training data, and 34% of datasets for the validation test, the data mean the accuracy of the result is coming out after training and data validation. Using Decision Tree, the test instance classifies 3.115 instances (98.4201%) correctly and 50 instances (1.5799%) incorrectly. In Naive Bayes, it classifies 1.035 instances (96.1896%) correctly and 41 instances (3.8104%) incorrectly.

For Support Vector Machine, the correctly classified instances are 1.057 (98.2342%), and incorrectly classified instances are 19 (1.7658%). Next, in Neural Network, 3.105 instances (98.2043%) are correctly classified, and 980 instances (1.8957%) are incorrectly classified.

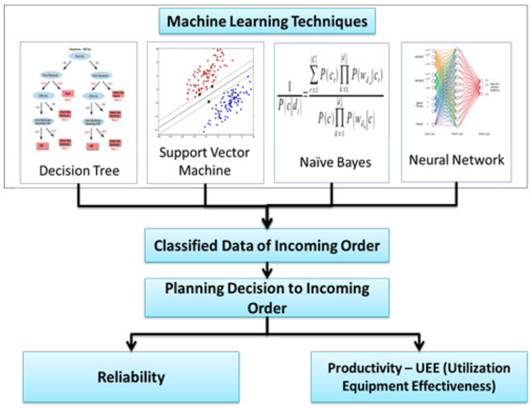

Fig. 2. Data analysis process flow.

To see the contribution of each machine learning technique through the data experiment test, the researcher compares the accuracy of results in these machine learning algorithms. Those are counted in UEE and reliability test. The purpose of comparing these results is to see which machine learning algorithms contribute the most during the experiment. The detail of the comparison is shown in Table I. Moreover, those comparison results are ranked to see the highest accuracy of machine learning algorithm performance in this experiment. The rank of the machine learning algorithm is sequenced from the highest as shown in Table II.

The data comparison in Table 2 shows that Decision Tree results in the highest accuracy with 98.116%. The second algorithm is Support Vector Machine with 95.1673%. Then, Neural Network is the third algorithm about 94.3128%. The Naive Bayes is in the last rank with 90.1952%.

To achieve the objective of this research, the researcher uses the attribute evaluator to get the sequence of priority of the used attributes. Since Decision Tree has the highest accuracy rate, it is chosen as the model of attribute evaluator ranking filter.

TABLE I
MACHINE LEARNING ALGORITHM RESULT COMPARISON.

Algorithms	Correctly Classified		Incorrectly Classified		Testing Time (s)
	# Instances	%	# Instances	%	
Decision Tree	1 059.00	98.4201	17.00	1.5799	0.07
Naive Bayes	1 035.00	96.1896	41.00	3.8104	0.01
Support Vector Machine	1 057.00	98.2342	19.00	1.7658	46.20
Neural Network	3 105.00	98.1043	60.00	1.8957	4 339.66

TABLE II
THE ACCURACY OF MACHINE LEARNING ALGORITHM RANK.

Algorithms	Accuracy (%)
Decision Tree	98.1160
Naive Bayes	95.1673
Support Vector Machine	94.3128
Neural Network	90.1952

TABLE III
TOP FIVE ATTRIBUTES.

Attributes	Ranked
Order Quantity	0.2635899
Label Length (mm)	0.0419638
Percentage Allowance	0.0402605
Customer Code	0.0400268
RBO	0.0307018

TABLE IV
PREDICTION OF ACCURACY.

	After Manual Arrangement			
	Prediction		Actual	
	Achieve	Not	Achieve	Not
Number	776	24	800	0
Accuracy	97.00%	3.00%	-	-

From 23 attributes that have been selected by the attribute evaluator, the rank filter shows the most contributing attribute to the actual results of UEE and reliability. Based on that, the researcher chooses the top five best attributes from the result to arrange the testing data. Table III shows the results.

Order quantity is the most contributing attribute in the database. It has 0.2635899 points. It is followed by label length (mm) with 0.0419638 points, the percentage of allowance with 0.0400268 points, customers' code with 0.0400268 points, and Retail Brand Order (RBO) with 0.0307018 points. These top attributes are used by the researcher to arrange the testing data.

Table IV shows the accuracy result of final prediction of UEE and reliability. The data prediction has 776 datasets as "Achieve" and 24 datasets as "Not". Meanwhile, the actual UEE and reliability have 800 datasets as "Achieve" and 0 datasets as "Not". Then, the actual data has 0 datasets as "Not" since the manual data arrangement is conducted through the data from the top five attributes. The accuracy of that final prediction is 97%.

TABLE V
UEE AND RELIABILITY PREDICTION.

UEE and Reliability Improvement Rate	
Before	550
After	776
Percentage	41.09%

The final data prediction of UEE and reliability are compared with the actual data before the arrangement is conducted. The data of UEE and reliability are improved by 41.09% as shown in Table V.

As Decision Tree results in the highest accuracy percentage with 98.1160% of accuracy, it is important to analyze the characteristic of this machine learning algorithm. It is to see the possible reason for this accuracy result in this research. The characteristic of Decision Tree can be represented by seeing its advantages. It compares the competing alternatives even without complete information in terms of risk and probable value. The Expected Value (EV) term combines relative investment costs, anticipated payoffs, and uncertainties into a single numerical value. The EV reveals the overall merits of competing alternatives. It means that Decision Tree has four main benefits [17]:

- Decision trees can handle both nominal and numerical attributes.
- Decision trees representation is rich enough to represent any discrete-value classifier.
- Decision trees are capable of handling datasets that may have errors.
- Decision trees are capable of handling datasets that may have missing values.
- Decision trees are considered to be a nonparametric method. This means that decision trees have no assumptions about the space distribution and the classifier structure.

Another resource shows that the superiority of Deci-

TABLE VI
THE COMPARIOSN OF MACHINE LEARNING ALGORITHMS (FOUR POINTS REPRESENT THE BEST).

Performance	Decision Tree	Neural Network	Naive Bayes	Support Vector Machine
General	2	3	1	4
Speed of learning with respect to number of attributes and the number of instances	3	1	4	1
Speed of classification	4	4	4	4
Tolerance to missing values	3	1	3	2
Tolerance to irrelevant attributes	3	1	2	4
Tolerance to redundant attributes	2	2	1	3
Tolerance to highly interdependent attributes (e.g. parity problems)	2	3	1	3
Dealing with discrete/binary/-continuous attributes	4	2	1	1
Tolerance to noise	2	2	3	2
Dealing with danger of over fitting	2	1	3	2
Attempts for incremental learning	2	3	4	2
Explanation ability/transparency of knowledge/classifications	4	1	4	1
Model parameter handling	3	1	4	1
Total performance	36	25	35	30

sion Tree compares to other common machine learning algorithms. These machine learning algorithms are ranked in every single performance category that commonly appears in machine learning problems. The performance category includes the general performance, speed of learning, speed of classification, tolerance to the missing values, tolerance to irrelevant attributes, tolerance to the redundant attribute, tolerance to highly interdependent attributes, dealing with the attribute format type, tolerance of noise, dealing with danger of overfitting, attempt for incremental learning, model transparency, and the model parameter handling. Each category is given a point in per machine learning algorithm. As a result, each machine learning algorithm has its total point and show the comparison of the effectiveness of each one. The comparison result is shown on Table VI.

V. CONCLUSION

The UEE and reliability are improved by 41.09% with 97% of data prediction accuracy. Then, five attributes contribute the most to the productivity and reliability in this training dataset. These attributes are ranked after experimenting with machine learning algorithms through the application with information gain attribute rank. The Decision Tree successfully has the highest accuracy of dataset prediction with 98.1160%.

As an additional conclusion, machine learning algorithms are successfully applied in the manufacturing process. This proves that information technology learning tools can be applied in the manufacturing process as a problem solution. As the additional objective of this research is expected to find out the root causing the problem, the researcher concludes that one of the causes in the lack of UEE and reliability result is the lack of order quantity arrangement in the process.

This research can be expanded by combining manufacturing improvement techniques with machine learning algorithms. Thus, a study to improve the productivity and reliability by combining learning tools will be very beneficial for the company and the knowledge for people. The machine learning techniques can predict datasets. Then, this prediction can be used as the planning arrangement or it can be called as the earlier improvement before the real process is conducted. Moreover, for the future research manufacturing improvement techniques such as Single Minutes Exchange of Dies, Total Preventive Maintenance, 5S techniques can be used as the main improvement in the process.

ACKNOWLEDGEMENT

This research is funded by the research institute and community service on junior research grant programs at President University.

REFERENCES

[1] V. Albino, P. Pontrandolfo, and B. Scozzi, "Analysis of information flows to enhance the coordination of production processes," *International Journal of Production Economics*, vol. 75, no. 1-2, pp. 7–19, 2002.

[2] B. J. Hicks, S. J. Culley, and C. A. McMahon, “A study of issues relating to information management across engineering smes,” *International Journal of Information Management*, vol. 26, no. 4, pp. 267–289, 2006.

[3] M. H. Jansen-Vullers, C. A. van Dorp, and A. J. Beulens, “Managing traceability information in manufacture,” *International Journal of Information Management*, vol. 23, no. 5, pp. 395–413, 2003.

[4] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, “Deep learning for smart manufacturing: Methods and applications,” *Journal of Manufacturing Systems*, vol. 48, pp. 144–156, 2018.

[5] S. Shah, S. Reddy, A. Sardeshmukh, B. Gautham, G. Shroff, and A. Srinivasan, “Application of machine learning techniques for inverse prediction in manufacturing process chains,” in *Proceedings of the* 3^{rd} *World Congress on Integrated Computational Materials Engineering (ICME 2015)*. Colorado Springs, Colorado: Springer, May 31–June 4 2015, pp. 261–268.

[6] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, “Machine learning in manufacturing: Advantages, challenges, and applications,” *Production & Manufacturing Research*, vol. 4, no. 1, pp. 23–45, 2016.

[7] T. Wuest, C. Irgens, and K. D. Thoben, “An approach to monitoring quality in manufacturing using supervised machine learning on product state data,” *Journal of Intelligent Manufacturing*, vol. 25, no. 5, pp. 1167–1180, 2014.

[8] H. Li, “An approach to improve flexible manufacturing systems with machine learning algorithms,” in *IECON 2016 -* 42^{nd} *Annual Conference of the IEEE Industrial Electronics Society*. Florence, Italy: IEEE, Oct. 23–26 2016, pp. 54–59.

[9] C. Apté, R. Sasisekharan, V. Seshadri, and S. M. Weiss, “Case studies in high-dimensional classification,” *Applied Intelligence*, vol. 4, no. 3, pp. 269–281, 1994.

[10] E. Alpaydin, *Introduction to machine learning*, 2^{nd} ed. Cambridge, Massachusetts: The MIT Press, 2010.

[11] Y. T. Lee, S. Kumaraguru, S. Jain, S. Robinson, M. Helu, Q. Y. Hatim, S. Rachuri, D. Dornfeld, C. J. Saldana, and S. Kumara, “A classification scheme for smart manufacturing systems’ performance metrics,” *Smart and Sustainable Manufacturing Systems*, vol. 1, no. 1, p. 52, 2017.

[12] P. Lalanda, D. Morand, and S. Chollet, “Autonomic mediation middleware for smart manufacturing,” *IEEE Internet Computing*, vol. 21, no. 1, pp. 32–39, 2017.

[13] F. Tao and Q. Qi, “New IT driven service-oriented smart manufacturing: Framework and characteristics,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, no. 99, pp. 1–11, 2017.

[14] Y. Ye, T. Hu, C. Zhang, and W. Luo, “Design and development of a CNC machining process knowledge base using cloud technology,” *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3413–3425, 2018.

[15] M. Kostina, T. Karaulova, J. Sahno, and M. Maleki, “Reliability estimation for manufacturing processes,” *Journal of Achievements in Materials and Manufacturing Engineering*, vol. 51, no. 1, pp. 7–13, 2012.

[16] V. Cesarotti, A. Giuiusa, and V. Introna, “Using overall equipment effectiveness for manufacturing system design,” in *Operations Management*. IntechOpen, 2013.

[17] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, “Supervised machine learning: A review of classification techniques,” *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, pp. 3–24, 2007.