Analysis of machine repair time prediction using machine learning at one of leading footwear manufacturers in Indonesia

Yulio Agefa Purmala, Sumarsono Sudarto
Department of Industrial Engineering, Universitas Mercu Buana, Jakarta, Indonesia

Article Info

Article history:

Received Aug 8, 2022
Revised Jan 31, 2023
Accepted Mar 10, 2023

Keywords:

Classification
Evaluation metrics
Machine breakdown
Machine learning
Repair time

ABSTRACT

Machine breakdowns in the production line mostly finish in more than 18 minutes, since the machine that needs repair more time is done on the production line, not in the machine warehouse. Historical machine breakdown data is digitally recorded through the Andon system, but it is still not being used adequately to aid decision-making. This research introduces an analysis of historical machine breakdown data to provide predictions of repair time intervals with a focus on finding the best algorithm accuracy. The research method uses machine learning techniques with a classification model. There are five algorithms used: logistic regression (LR), naive bayes (NB), k-nearest neighbor (KNN), support vector machine (SVM), and random forest (RF). The results of this study prove that historical machine breakdown data can be optimized to predict machine repair time intervals in the production line. The accuracy of LR algorithm is slightly better than the other algorithms. Based on the receiver operating characteristic–area under curve (ROC-AUC) performance evaluation metric, the quality value of the accuracy of LR model is satisfied with a percentage of 69% with a difference of 0.5% between the train and test data.

Corresponding Author:

Yulio Agefa Purmala
Department of Industrial Engineering, Universitas Mercu Buana
Jakarta, Indonesia
Email: yagefa@gmail.com

1. INTRODUCTION

Based on export data released by the *Direktorat Jenderal Bea dan Cukai Indonesia* (Indonesian Directorate General of Customs and Excise) taken from export declaration of goods documents in 2020, the value of Indonesia's exports decreased by 2.68% from the previous year, reflecting the impact of COVID-19. The decline in the value of Indonesia's exports in 2020 was caused by a 30.01% decline in commodity exports of oil and gas, as well as a 0.61% decline in non-oil and gas exports [1]. The role of Indonesia's oil and gas and non-oil exports has shifted. Non-oil and gas exports accounted for 94.94% of total exports in 2020, an increase of 1.97% from 2019. One of the non-oil and gas commodity sectors, namely the manufacturing industry sector, has the largest contribution to total exports. Its contribution will be 80.33% by 2020. One of the export commodities that experienced an increase was the footwear sector, especially sporting goods. Export value increased by 31.09% in 2020 compared to the previous year. Indonesia is among the top four countries with the highest number of footwear exports, still behind China, India, and Vietnam [2].

The footwear industry is a labor-intensive industry [2]. Indonesia has about 18,687 footwear business units with a workforce of 795,000 manpower, which are then followed by the required number of machines. This is to meet high demand for shoes, but still limited in technology and capital, so that Indonesia can compete with other exporting countries, especially China and Vietnam. If the focus is on the needs of the machine, then the function or type of machine used is also very diverse, adjusting to each process and model of footwear that must be done. This research will focus on the cutting, stitching, and assembly (CSA) process because most processes and stages of manufacturing and assembling components are carried out in that process. Based on data that has been collected in one of the leading footwear manufacturing industries in Tangerang, Indonesia, it takes about 10,633 machines in the CSA process, which consists of 106 types of machines and/or 375 series of machines that must be managed properly. With many machines that must be managed, the handling and control of machines must also be carried out very well. The effectiveness of several machines in supporting production activities to fulfill exports is closely related to the availability and reliability of machines. Technicians always try to ensure machine availability and reliability are maintained through regular maintenance activities [3], but due to the large number of machines used, machines sometimes break down before maintenance time is completed [4], [5].

In 2021, the number of machine breakdowns that occur in production lines is very high. Currently, all machine breakdowns that occur in the CSA process are well recorded using the Andon system. Besides being used as a tool to call technicians, Andon has also been used as a tool for storing historical information on machine breakdowns [6]. Five machines in particular had a high breakdown rate and made the biggest contribution to the overall breakdown, namely: 1N postbed stitching machine, 2N postbed stitching machine, computer stitching machine-Medium, skiving machine, and computer stitching machine-Large. In addition to availability, the reliability of the machine must also be controlled properly. The average time for repairing a machine breakdown from the five machines above is 27.6 minutes.

The export process will be disrupted if the production process has problems included because of machine breakdown time, so machines that are broken for more than 20 minutes must be worked in the machine warehouse. The next problem is that the technicians cannot predict it. In addition to the uneven distribution of technician expertise, this is also because they still tend to rely on experience and intuition in estimating repair time intervals. Historical data on machine breakdown is recorded digitally through the Andon system, but the data is still not properly utilized to assist decision making. The aim of this research is to analyze historical data on machine breakdown to provide predictions of time intervals for repairing machines with a focus on finding the best algorithm accuracy using a machine learning approach.

2. METHOD

This study uses the cross-industry standard process for data mining (CRISP-DM) model process for a universal data analysis approach [7], [8]. Several supervised machine learning classification method algorithms were chosen to get the best accuracy value [9], [10]. The algorithms are logistic regression (LR) [11], [12], naive bayes (NB) [13], [14], random forest (RF) [15], [16], k-nearest neighbor (KNN) [17], [18], and support vector machine (SVM) [19], [20]. Selected five classifications supervised machine learning based on each algorithm have on different dimension metrics there are parametric-simple for LR and NB, parametric-complex for SVM, non-parametric-simple for KNN, and non-parametic-complex for RF. Several dimensions of the variables involved in this study include reports of machine breakdown, such as repair time, response time, machine model or type, building location, type of breakdown, causes of breakdown, and repair solutions. Variable dimensions of assets, such as asset number, machine age, machine price, machine arrival date, and machine ownership status. Machine replacement, such as the location of the old and new buildings. Employee variables such as position and years of work experience. The Framework in this study describes how the concept of data utilization using machine learning methods [21], [22]. An overview of the framework can be seen in Figure 1, how data sources get and which variable is used, next go to how data are transformed with several data modeling, performance analysis, and finding the best model to get the output of the prediction accuracy.

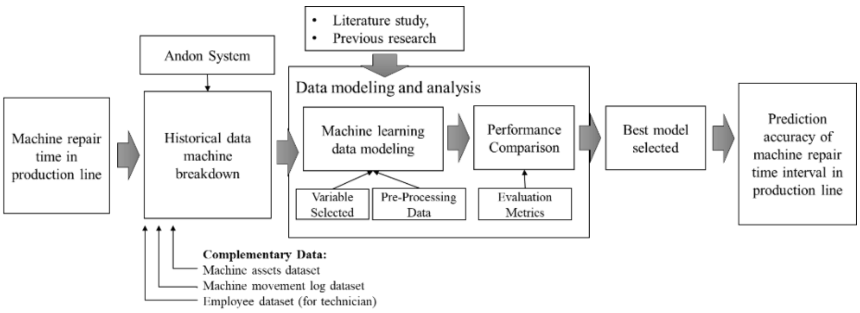

Figure 1. Research framework

There are two evaluation metric methods that will be used: confusion metrics [23] and receiver operating characteristic - area under curve (ROC-AUC) [24]. Confusion metrics is an evaluation model to find the values of accuracy, precision, recall, and F1 score by looking at the probability value between actual value and predicted value at one threshold. Meanwhile, ROC-AUC evaluates all possible performances at all thresholds that are under the ROC curve area. At the end of this study, ROC-AUC will be used as the main evaluation [25], while the evaluation of confusion metrics will be used as a supporting evaluation [26]. To be more effective and time efficient all machine learning methods are processed by platform orange data mining and do some setup parameters in some algorithms. LR set up on a ridge with a coefficient score is 11, RF setup with 10 number of trees, SVM setup with cost is 10 and regression loss epsilon is 0.10, and KNN setup with 10 number of neighbor and metric euclidean.

In addition to using a statistical approach, another way of determining variables in addition to using a statistical approach is through a domain expertise approach, with direct brainstorming to the engineering department leaders starting from the supervisory level to the manager level, through focus group discussions (FGD) [27]. After several independent variables have been determined, a correlation test is carried out with the dependent variable, that is, repair time interval, using the chi-squared test [28]. Prior to modeling, data will be divided into two with a ratio of 70:30, 70% as training data and 30% as test data. The flow chart of this research is shown in Figure 2.

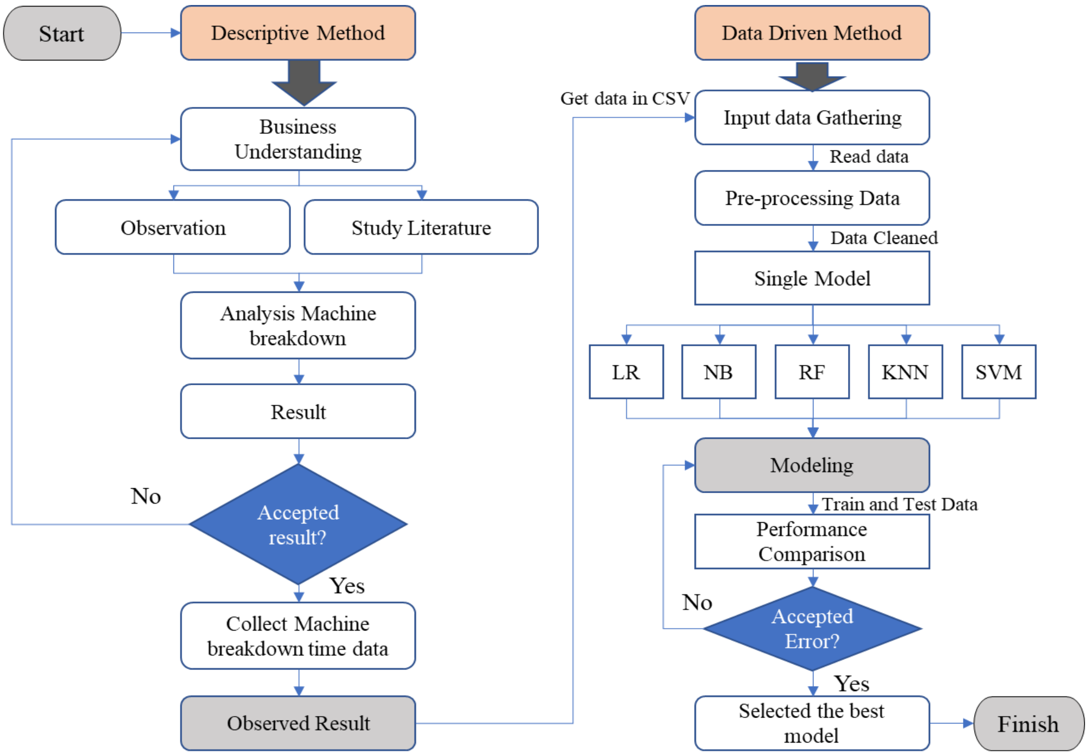

Figure 2. Research flow chart

3. RESULTS AND DISCUSSION

Several data sets are already recorded digitally, but there are also those that use worksheets in data collection and require a data transformation process from manual to digital with a computerized system. The following is the flow of data used in this research, as shown in Figure 3. Machine breakdown in Andon, spare parts transactions, and equipment assets data are recorded digitally and stored directly on the server, work order forms need to transform from paper to computerize before store on the server. The dataset in this research is exported directly from the internal server and cloud server. The data used for modeling analysis is in the (.csv) format to make analysis easier because this format can be used for many systems and is compact and straightforward [29]. Data that has been collected is then carried out a thorough depiction of data for each variable, as well as checking whether there is empty or missing data contained in each data set.

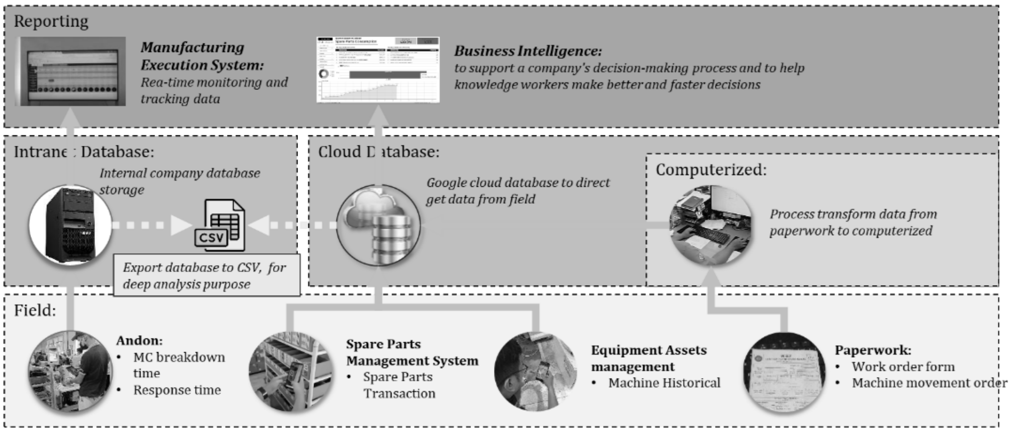

Figure 3. Diagram of data flow

3.1. Data preparation

Data that has been thoroughly prepared, inspected, and described is still very messy and needs to be preprocessed so that it is ready for use in machine learning modeling. Based on the initial description of each data set as shown in Figure 4, where there is missing data, which will then be processed to handle the missing data. Missing data that is still less than 5% does not need to be removed; it only needs to be imputationed by replacing it with the average or median value for numeric data types, and it is replaced based on the frequency with which it appears for categorical data types [30]. The variable target in this research is the repair time interval. Whether the required repair time is less than 18 minutes or not, the determination of the 18 minutes repair time standard is based on the machine breakdown history during 2021, where the average repair time is 24 minutes and the median time is 18 minutes, as shown in Figure 5.

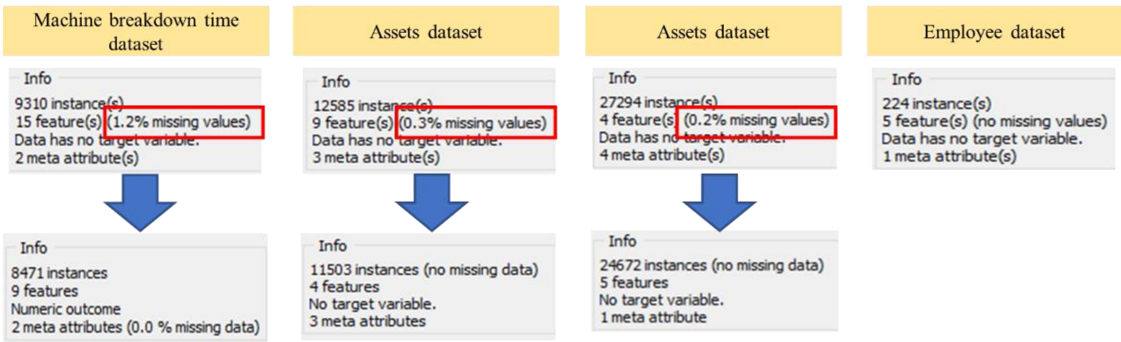

Figure 4. Pre-processing data results

Figure 5. Machine breakdown time variable in statistics

Based on the detailed analysis shown in Figure 6, the distribution of machine breakdown time is a positive skewed distribution. The median value is more appropriate to use for the distribution of positively skewed and negatively skewed data [31]. This is what underlies the determination of the machine repair time in this study, which is 18 minutes based on the median result of the breakdown time data.

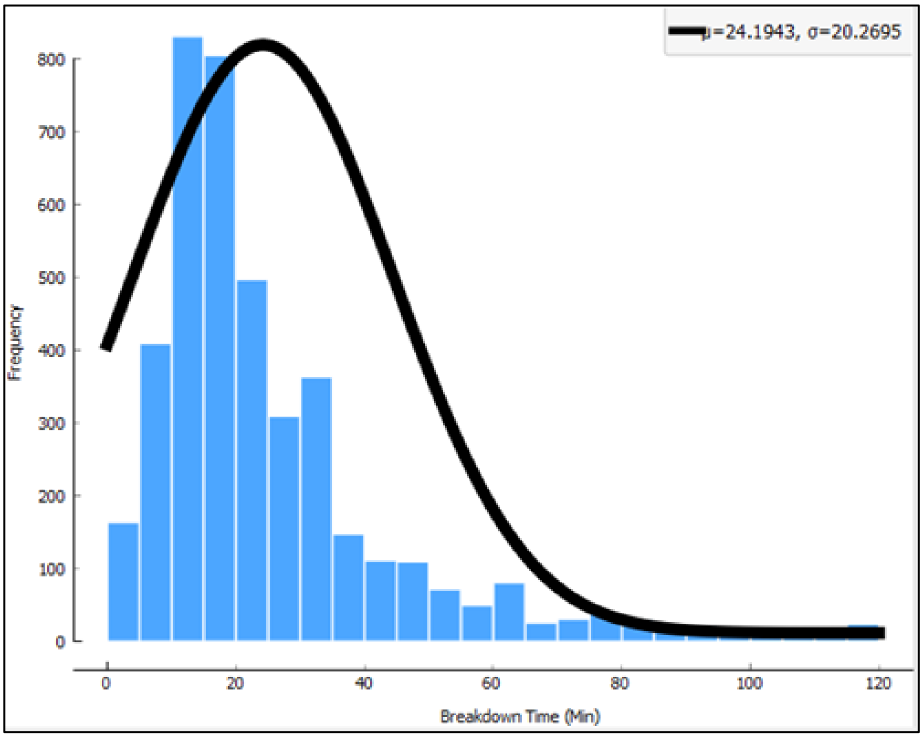

Figure 6. Machine breakdown time has a positive skewed distribution

3.2. Data modeling and analysis

The data set is divided into two parts for modeling evaluation: training data and testing data. The distribution of training and test data from the total data set is 70:30. The entire data utilized for modeling comes from 4,163. Therefore, when employing a 70:30 ratio, 2,915 training data and 1,248 test data will be employed. The cross-validation approach will be used to confirm the model's performance for several evaluations. In this research, the modeling will be evaluated using 5-fold cross validation [32].

As seen in Table 1, the results of train data modeling using AUC evaluation, for LR=0.688, NB=0.677, RF=0.644, KNN=0.638 and SVM=0.530. Results of test data modeling results using AUC evaluation, for LR=0.693, NB=0.670, RF=0.673, KNN=0.626 and SVM=0.531.

Table 1. Training and test data score comparison

Algorithm	AUC		Accuracy		Precision		Recall		F1	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
LR	0.688	0.693	0.646	0.643	0.647	0.644	0.646	0.643	0.646	0.643
NB	0.677	0.670	0.632	0.617	0.632	0.617	0.632	0.617	0.631	0.617
RF	0.644	0.673	0.607	0.631	0.607	0.631	0.607	0.631	0.607	0.631
KNN	0.638	0.626	0.600	0.696	0.603	0.596	0.600	0.596	0.597	0.596
SVM	0.530	0.531	0.515	0.505	0.515	0.506	0.515	0.505	0.514	0.479

Predictions made by test data are the result of learning from training data. Machine learning algorithms will perform modeling based on their respective formulations and techniques and then determine the probability of whether to enter the classification under 18 minutes or above 18 minutes for repair time by considering the variables that have been determined in this study. Table 1 compares the outcomes of the modeling evaluation between the training and test data. From this comparison, the difference between training and test data is not considerably different, hence LR has reasonable modeling accuracy. The results of the test data modeling show that the LR, NB, RF, and KNN algorithms are in the satisfactory category, while the SVM algorithm is in the unsatisfactory category. The results of the confusion metrics between the actual and predicted data can be seen in Figure 7.

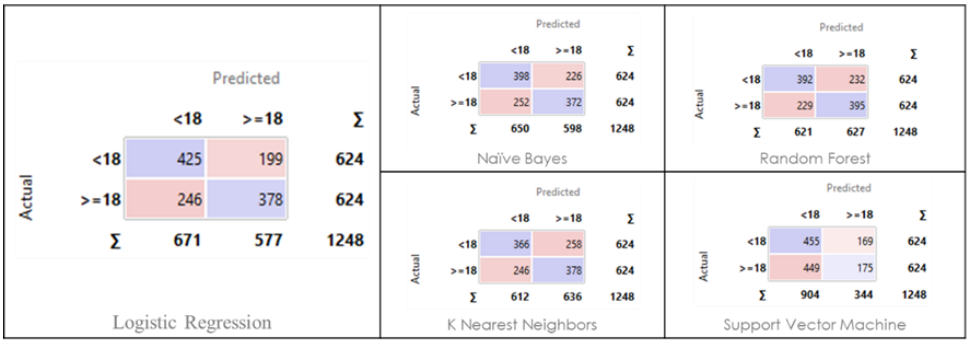

Figure 7. Confusion metrics of test data

3.3. Model performance evaluation and quality model result

This research seeks to determine the degree of modeling accuracy, using both training and test data. The gap between training and test data is not a serious issue if the difference is not too large, regardless of whether it is overfitting or underfitting. In accordance with the outcomes of machine learning modeling using five classification algorithms, LR is an algorithm that fits the variables in this study with the difference between the training data and the test data being -0.005, meaning that test data is better at 0.005 than training data, or at a percentage of 0.5%, according to Figure 8.

Algorithm	AUC Train	AUC Test	AUC Gap	
Logistic Regression	0.688	0.693	-0.005	69% = Satisfactory
Naïve Bayes	0.677	0.670	+0.007	67% = Satisfactory
Random Forest	0.644	0.673	-0.029	66% = Satisfactory
K Nearest Neighbors	0.638	0.626	+0.012	63% = Satisfactory
Support Vector Machine	0.530	0.531	-0.001	53% = Un-satisfactory

Figure 8. The AUC of the evaluation metrics gap between training and test data

Based on the conclusion of data evaluation on training data and test data, for LR, NB, RF, and KNN algorithms in this study, the modeling quality is in the satisfactory category because the AUC value is around 0.6 to 0.7 [24]. As for the SVM algorithm in this study, the quality of the modeling is in the unsatisfactory category because the AUC value is below 0.6, but the modeling is still accepted because the AUC value is still above 0.5. LR algorithms are the best result since the dependent variable is binary and data structure is simple.

4. CONCLUSION

This research wants to know the level of accuracy of the modeling that has been done both on the training data and test data. The difference between training data and test data is not a significant problem if the difference between the two is not too great, be it overfitting or underfitting. The results of this study prove that historical machine breakdown data can be optimized to predict machine repair time intervals in the production line within under 18 minutes. The accuracy of LR algorithm is slightly better than the other algorithms. Based on ROC-AUC performance evaluation metric, the quality value of the accuracy of LR model is satisfied with a percentage of 69% with a difference of 0.5% between the train and test data. In further research, the variables used can be enriched, so that the percentage of the results of the analysis of the resulting model will be even better. In addition, it can be developed to the implementation stage and integrated into existing maintenance systems to provide real-time predictions.

ACKNOWLEDGEMENTS

The article is part of my study thesis report for the industrial engineering master's degree program at Universitas Mercu Buana in Jakarta, Indonesia. I am grateful to Dr. Sumarsono Sudarto, my supervisor, for mentoring and supporting me during this project.

REFERENCES

[1] K. F. Arifah and J. Kim, "The importance of agricultural export performance on the economic growth of Indonesia: the impact of the COVID-19 pandemic," *Sustainability*, vol. 14, no. 24, p. 16534, Dec. 2022, doi: 10.3390/su142416534.

[2] M.-L. Tseng, T.-D. Bui, M. K. Lim, and S. Lewi, "A cause and effect model for digital sustainable supply chain Competitiveness under Uncertainties: enhancing digital platform," *Sustainability*, vol. 13, no. 18, p. 10150, Sep. 2021, doi: 10.3390/su131810150.

[3] V. L. Harmon *et al.*, "Reliability metrics and their management implications for open pond algae cultivation," *Algal Research*, vol. 55, p. 102249, May 2021, doi: 10.1016/j.algal.2021.102249.

[4] Y. Wang, C. Deng, J. Wu, Y. Wang, and Y. Xiong, "A corrective maintenance scheme for engineering equipment," *Engineering Failure Analysis*, vol. 36, pp. 269–283, Jan. 2014, doi: 10.1016/j.engfailanal.2013.10.006.

[5] V. Polotski, J.-P. Kenne, and A. Gharbi, "Optimal production and corrective maintenance in a failure-prone manufacturing system under variable demand," *Flexible Services and Manufacturing Journal*, vol. 31, no. 4, pp. 894–925, Feb. 2019, doi: 10.1007/s10696-019-09337-8.

[6] K. Demeter, D. Losonci, and J. Nagy, "Road to digital manufacturing - a longitudinal case-based analysis," *Journal of Manufacturing Technology Management*, vol. 32, no. 3, pp. 820–839, Jun. 2020, doi: 10.1108/jmtm-06-2019-0226.

[7] F. Martinez-Plumed *et al.*, "CRSP-DM twenty years later: from data mining processes to data science trajectories," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/tkde.2019.2962680.

[8] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017, doi: 10.1109/access.2017.2756872.

[9] T. Witkowski, P. Antczak, and A. Antczak, "Machine learning-based classification in manufacturing system," in *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, Sep. 2011, pp. 580–585, doi: 10.1109/idaacs.2011.6072833.

[10] E. A. Adje, V. R. Houndji, and M. Dossou, "Features analysis of internet traffic classification using interpretable machine learning models," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 3, pp. 1175–1183, Sep. 2022, doi: 10.11591/ijai.v11.i3.pp1175-1183.

[11] B. Pavlyshenko, "Machine learning, linear and Bayesian models for logistic regression in failure detection problems," Dec. 2016, doi: 10.1109/bigdata.2016.7840828.

[12] R. D. Joshi and C. K. Dhakal, "Predicting type 2 diabetes using logistic regression and machine learning approaches," *International Journal of Environmental Research and Public Health*, vol. 18, no. 14, p. 7346, Jul. 2021, doi: 10.3390/ijerph18147346.

[13] M. Koutina and K. L. Kermanidis, "Predicting postgraduate students' performance using machine learning techniques," in *IFIP Advances in Information and Communication Technology*, Springer Berlin Heidelberg, 2011, pp. 159–168.

[14] H. Zhang, N. Cheng, Y. Zhang, and Z. Li, "Label flipping attacks against Naive Bayes on spam filtering systems," *Applied Intelligence*, vol. 51, no. 7, pp. 4503–4514, Jul. 2021, doi: 10.1007/s10489-020-02086-4.

[15] D. Wu, C. Jennings, J. Terpenny, R. X. Gao, and S. Kumara, "A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests," *Journal of Manufacturing Science and Engineering*, vol. 139, no. 7, Apr. 2017, doi: 10.1115/1.4036350.

[16] T. C. McCandless and S. E. Haupt, "The super-turbine wind power conversion paradox: using machine learning to reduce errors caused by Jensen's inequality," *Wind Energy Science*, vol. 4, no. 2, pp. 343–353, Jun. 2019, doi: 10.5194/wes-4-343-2019.

[17] E. Ruiz, D. Ferreño, M. Cuartas, A. López, V. Arroyo, and F. Gutiérrez-Solana, "Machine learning algorithms for the prediction of the strength of steel rods: an example of data-driven manufacturing in steelmaking," *International Journal of Computer Integrated Manufacturing*, vol. 33, no. 9, pp. 880–894, Sep. 2020, doi: 10.1080/0951192x.2020.1803505.

[18] A. M. Musolf, E. R. Holzinger, J. D. Malley, and J. E. Bailey-Wilson, "What makes a good prediction? feature importance and beginning to open the black box of machine learning in genetics," *Human Genetics*, vol. 141, no. 9, pp. 1515–1528, Dec. 2021, doi: 10.1007/s00439-021-02402-z.

[19] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction," *Expert Systems with Applications*, vol. 83, pp. 405–417, Oct. 2017, doi: 10.1016/j.eswa.2017.04.006.

[20] H. R. Baghaee, D. Mlakic, S. Nikolovski, and T. Dragicevic, "Support vector machine-based islanding and grid fault detection in active distribution networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 3, pp. 2385–2403, Sep. 2020, doi: 10.1109/jestpe.2019.2916621.

[21] A. F. Vermeulen, *Industrial machine learning*, First Edit. Berkeley, California: Apress Media, 2020.

[22] S. Fahle, C. Prinz, and B. Kuhlenkötter, "Systematic review on machine learning ML methods for manufacturing processes-identifying artificial intelligence AI methods for field application," *Procedia CIRP*, vol. 93, pp. 413–418, 2020, doi: 10.1016/j.procir.2020.04.109.

[23] A. Emmanuel, *Building machine learning powered applications going from idea to product*, First Edit. Sebastopol, California: O'Reilly Media, Incorpotated, 2020.

[24] O. P. Trifonova, P. G. Lokhov, and A. I. Archakov, "Metabolic profiling of human blood," *Biochemistry Moscow Supplement Series B: Biomedical Chemistry*, vol. 7, no. 3, pp. 179–186, Jul. 2013, doi: 10.1134/s1990750813030128.

[25] C. X. Ling, J. Huang, and H. Zhang, "AUC: a better measure than accuracy in comparing learning algorithms," in *Advances in Artificial Intelligence*, Springer Berlin Heidelberg, 2003, pp. 329–341.

[26] H. Hamdani, H. R. Hatta, N. Puspitasari, A. Septiarini, and H. Henderi, "Dengue classification method using support vector machines and cross-validation techniques," *IAES International Journal of Artificial Intelligence (IJ-A)*, vol. 11, no. 3, pp. 1119–1129, Sep. 2022, doi: 10.11591/ijai.v11.i3.pp1119-1129.

[27] T. O.Nyumba, K. Wilson, C. J. Derrick, and N. Mukherjee, "The use of focus group discussion methodology: insights from two decades of application in conservation," *Methods in Ecology and Evolution*, vol. 9, no. 1, pp. 20–32, Jan. 2018, doi: 10.1111/2041-210x.12860.

[28] D. Shi, C. DiStefano, H. L. McDaniel, and Z. Jiang, "Examining chi-square test statistics under conditions of large model size and ordinal data," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 25, no. 6, pp. 924–945, Mar. 2018, doi: 10.1080/10705511.2018.1449653.
[29] Y. A. Purmala, "Footwear machine breakdown time dataset 2020 to 2021," *Github*. https://github.com/yap23/Footware-machine-breakdown-time/blob/main/4. Thesis - Machine Breakdown (Orange) Final.csv (accessed Jan. 31, 2023).
[30] Y. Dong and C.-Y. J. Peng, "Principled missing data methods for researchers," *SpringerPlus*, vol. 2, no. 1, May 2013, doi: 10.1186/2193-1801-2-222.
[31] M. K. Cain, Z. Zhang, and K.-H. Yuan, "Univariate and multivariate skewness and kurtosis for measuring nonnormality: prevalence, influence and estimation," *Behavior Research Methods*, vol. 49, no. 5, pp. 1716–1735, Oct. 2017, doi: 10.3758/s13428-016-0814-1.
[32] Z. Xiong, Y. Cui, Z. Liu, Y. Zhao, M. Hu, and J. Hu, "Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation," *Computational Materials Science*, vol. 171, p. 109203, Jan. 2020, doi: 10.1016/j.commatsci.2019.109203.

BIOGRAPHIES OF AUTHORS

Yulio Agefa Purmala holds a Bachelor of Engineering in Electrical Engineering from Insititut Teknologi Indonesia. He is currently pursuing his master's degree in Industrial Engineering at Universitas Mercu Buana (UMB). His current research interest included data mining, machine learning, automation, and manufacturing system. He can be contacted at email: yagefa@gmail.com or 55319110023@student.mercubuana.ac.id.

Sumarsono Sudarto received the D.Eng. degree (Doctor of Engineering) in system cybernetic from Hiroshima University, Hiroshima, Japan, in 2016 and the M. Eng degree in Industrial Engineering from Universitas Indonesia, Jakarta, Indonesia, in 2011. He is an assistant Professor of Industrial Engineering in Universitas Mercu Buana (UMB) since 2018. In addition, He is Founder and CEO of Elite Tutor Indonesia, one of institution to help dyslexic children. He can be contacted at email: sumarsono@mercubuana.ac.id.