International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 3, June 2022, pp. 2839~2846
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i3.pp2839-2846

Simulation for predictive maintenance using weighted training algorithms in machine learning

Chanintorn Jittawiriyanukoon[1], Vilasinee Srisarkun[2]
[1]Graduate School of Business and Advanced Technology Management, Assumption University, Bangkok, Thailand
[2]Martin de Tours School of Management and Economics, Assumption University, Bangkok, Thailand

Article Info

Article history:

Received Jun 1, 2021
Revised Sep 29, 2021
Accepted Oct 11, 2021

Keywords:

Artificial intelligence
Intelligent automation
Internet of things
Machine learning
Mean time between failures
Predictive maintenance
Training algorithm

ABSTRACT

In the production, the efficient employment of machines is realized as a source of industry competition and strategic planning. In the manufacturing industries, data silos are harvested, which is needful to be monitored and deployed as an operational tool, which will associate with a right decision-making for minimizing maintenance cost. However, it is complex to prioritize and decide between several results. This article utilizes a synthetic data from a factory, mines the data to filter for an insight and performs machine learning (ML) tool in artificial intelligence (AI) to strategize a decision support and schedule a plan for maintenance. Data includes machinery, category, machinery, usage statistics, acquisition, owner's unit, location, classification, and downtime. An open-source ML software tool is used to replace the short of maintenance planning and schedule. Upon data mining three promising training algorithms for the insightful data are employed as a result their accuracy figures are obtained. Then the accuracy as weighted factors to forecast the priority in maintenance schedule is proposed. The analysis helps monitor the anticipation of new machines in order to minimize mean time between failures (MTBF), promote the continuous manufacturing and achieve production's safety.

Corresponding Author:

Chanintorn Jittawiriyanukoon
Graduate School of Business and Advanced Technology Management, Assumption University
Bangkok, Thailand
Email: pct2526@yahoo.com

1. INTRODUCTION

Currently, with the invention of industry 4.0, intelligent automation (IA), artificial intelligence (AI), machine learning (ML), and predictive maintenance (PM) methodologies have been widely required in various industries for managing the machinery health status in the production line [1]. The digital transformation towards internet of things (IoT), information and computerized control, and computer networks are making it possible to record substantial amounts of operational data from various types of machinery. This approach can help detect fault alarms to improve productivity and sustain their stable health statuses. Thus predictive maintenance (PM) is not trivial for retaining a smart industry [2]. AI approach has become a favorable mechanism and attracted researchers to analyze industrial data silos in few years [3]. Industry 4.0 technology plays a critical role to develop industrial automation systems and harvest considerable data automatically from industrial equipment. Upon data collection, data mining helps sort out the noise and the insight then ML's algorithms help analyze further. Nevertheless, it is complex to take proper ML approaches for specific amount of trained data, data class, and machines to extend ML in manufacturing systems [4]. Adoption of incorrect PM plan will waste time and money as well as procrastinate maintenance schedule. Hence, this paper's objectives are to propose how to decide a comprehensive solution towards ML applications based on existing algorithms and help engineers to opt right ML techniques for a feasible PM plan.

The PM plan per se can cope with machine's attrition as it has been written for hidden hazards, pollution, malfunctions, and injuries in the manufacturing environment. These massive amounts of data captured for analysis encompasses an insight which raises the productivity dynamically and provides a decision support for the entire organization, especially machines' health monitoring and knowledge-based scheduled maintenance [5]. It is currently feasible to accumulate huge volume of operational data created from numerous equipment in order to be mined in developing an IA system [6]. The whole datasets can also improve higher efficient techniques for the intelligent maintenance system, as discussed in [7].

ML builds algorithms which train the data collected and make predictions [8]. Data mining studies on the fundamental analysis of collected data for the extraction of insightful data. This ranges from vast databases using ML algorithms to all system that makes decisions. It focuses on a classification of a massive volume of information which will be valuable in order to make the correct decisions [9]. In this paper, ML algorithms to construct data models that will foresee class labels for the decision-making in maintenance activities are utilized. Upon knowledge-based schedule, the decisions that make it simple to comprehend PM are envisioned. The synthetic data are executed for the training procedure of the health of the equipment and the timing evaluation of the maintenance process. The decisions for scheduled maintenance do not absolutely invent from the dataset as the person-in-charge designates the decision rules, frequently leading to incorrect conclusions at maintenance activities. Not to mention, investment is costly and the equipment's monitoring using IoT is compulsory [10].

PM based on advanced ML is possible to vigorously determine when the equipment needs to be regularly maintained. PM figures out intricate sign of faults which are impractical for persons. Activities of PM are set for the predictive condition of the equipment in order to schedule closely as the maintenance is executed. This method will reduce the cost and boost the productivity because of proactive on-demand maintenance jobs [11].

The PM with the manipulation of ML technologies and data mining rules on dataset based on the statuses of the equipment, leads to optimum decisions for the strategic planning of maintenance and upsurges the competitiveness of the industries and the effectiveness of the manufacturing [12]. ML application gives some benefits which include less equipment malfunction, maintenance cost saving, equipment spare-part life extension, operation safety, inventory cost reduction, increased productivity, and an increase in revenue growth. They somehow connect to maintenance activities [13]. Additionally, malfunction detection is crucial components of PM as it is essential in industries to perceive faults at early stage [14]. Methodologies for maintenance procedures can be classified:

a) Corrective maintenance (CM): it is non-scheduled maintenance but it is the straightforward among maintenance approaches which is done only when the machine fails. As a consequence, it may develop high downtime, a risk of secondary malfunctions and perhaps produce a number of flawed products in the manufacturing process.
b) Preventive maintenance (PRM): also known as routine-based maintenance. It defines routine maintenance based on anticipated schedule in order to avoid faults. It may induce to redundant maintenance which raises unnecessary operating costs.
c) Knowledge-based maintenance (KM): it is based on persistent machines or their health or equipment monitoring which can be collected if needed. The maintenance procedures can focus on the situation after one or more conditions of attrition of the process are detected. Note that it cannot be scheduled in advance.
d) Predictive maintenance (PM): it is based on statistics and forecast. Schedules are taken into account if needed. It depends on the time-series equipment monitoring like knowledge-based maintenance. It uses training algorithms and prediction applications to decide when maintenance process is needful, therefore the maintenance can be arranged. Moreover, it allows faults detection at initial stage based on the chronological dataset by executing ML tools and integrating variables (such as mean squared error (MSE) and correlation coefficient), statistical smoothing approaches, and engineering estimation techniques into consideration. It requires strategic planning of maintenance to lower faults rate, extend machine life-cycle and minimize costs. Maintenance types and their reliability are depicted in Figure 1.

PM performance represents the outstanding strategy on which many researches focus. The succeeding focuses on how the article is structured. The following section 2 describes training algorithms constructed in ML tool. Thirdly, the proposed method employed in this research and their supportive results for decision-making is discussed in section 3. In the end, conclusion and recommendations for future study are provided.

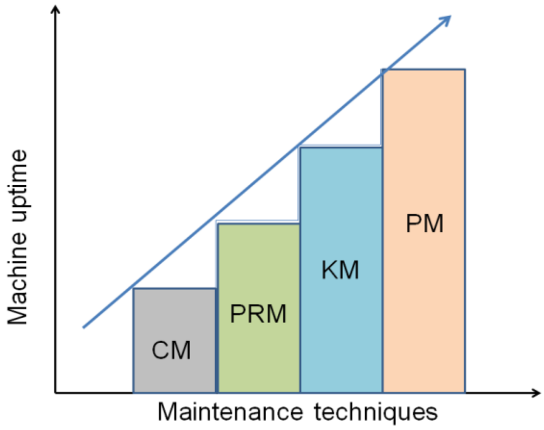

Figure 1. Maintenance models

2. TRAINING ALGORITHMS

Before the ML is carried on, first the historical data has to be mined into insights [15]. Data mining is a data preparation employed by industrial companies to transform a raw dataset into valuable information [16]. By using an application to eye for patterns in large dataset, companies can discover more about their equipment health to develop more effective maintenance strategy, increase productivity and reduce costs [17]. It is based on operational data collection, warehousing, and processing power. ML helps use insights to make healthier maintenance decisions. ML algorithms [18] discover patterns in insightful data, and create mathematical models upon these discoveries. Then the models to predict on future trend are used as references. For example, the ML model will predict how likely a machine needs a specific maintenance based on their past health [19]. The ML is employed to forecast whether defected products are trash and whether or not defected product examples are correctly labeled. Next, the ML is used to classify from these examples to predict how possible new finished products are defected or not. This technique of training-learning from dataset which has been labeled with the actual result is referred to the supervised ML [20].

2.1. Decision tree algorithm

Decision tree (DT) training is one of the effective training algorithms because of its interesting characteristics such as easiness, directness, no required factors, and can execute mixed-type of dataset [21]. In DT training, a decision is prompted from a set of labeled training examples denoted by an attribute and a class. Due to the spatial search, DT is typically a selfish, top-down and iterative process starting with the whole training dataset and a null tree. An attribute which best splits the training dataset is selected as the partitioning attribute for the root and the training dataset is then divided into separate subsets corresponding the values of the breaking attribute. For each disjoint subset, the algorithm continues iteratively until all examples in a subset join to the same class. This algorithm has a time complexity, $O(p^2 \bullet k)$, where p is the number of attributes and k is the size of the training example. As big data is common nowadays, the attention in developing DT algorithms is quickly growing. The reason is that DT functions appropriately on big data, besides its interesting characteristics. It is apparent that DT outperforms naive Bayes on big data [22]. DT algorithm adopts a searching (heuristic), which processes the purity of the subsequent subsets upon the division of attributes. Information gain (G), defined as a standard search:

$$G(D, A) = E(D) - \sum_a \frac{|D_a|}{|D|} E(D_a), \quad (1)$$

where D is a set of training examples, A is an attribute, a is a value within A, D_a is a disjoint subset of D containing the examples with $A = a$, and $E(D)$, Entropy of D, is described as

$$E(D) = -\sum_{n=1}^{|K|} P_D(k_n) log P_D(k_n), \quad (2)$$

where $P_D(k_n)$ is calculated by the proportion of examples joining k_n in D, and |K| is class size. Tree grows as a partitioning process of training data iterates and D is the training data corresponding to the present child node. So, $P_D(k_n)$ is exactly $P(k_n|x_p)$ on the complete training data, where x_p is the attributes set following the path from present child node to root node.

2.2. Random forest algorithm

Random forest (RF) is associating a series of tree classifiers provided that it grows each tree for a random selection of the subset class then merges these results for the final sorting result [23]. It presents high accuracy, indulges outlier, handles noise very well and is not overfitting. It is found to be one of the most popular methods data mining and information to the industrial area. The RF is a classifier containing a collection of tree structure {T (x, R_k)} in which {R_k} is randomly independent distributed vector and each tree votes for a random selection of the subset class x. Upon k recursive process, classifiers sequence {t_1(x), t_2(x),…,t_k(x)} is obtained. RF uses these results for classification models then the final selected result is computed by majority vote where the decision is:

$$T(x) = \arg\max_{c} \sum_{i=1}^{j} I(t_i(x) = c) \tag{3}$$

2.3. Logistic regression algorithm

Problems concerning classifications of output into classes are extensive in computational PM [24]. For instance, the classification of manufacturing variables [e.g., ambient temperature (Ta), sewage activity (Sw), machine health (Mh)] as leading to improvement of an observed industry is one common problem in predictive methodology. Classification model helps challenge to reply ad hoc questions like whether Ta of 25C, Sw of 0.98, and Mh of "STABLE" would lead to stabilized situation, or what the probability of the continuous health growth of this production would be under condition as such. Among the particular benefits of such classifiers in the industry is that they can be used as vital tools in maintenance [25]. Logistic regression (LR) has been used broadly for classification and approximation of the probability of machine health under a dataset of conditions. Some studies use LR to develop a nonlinear equation for estimating the probability of revenue growth as well. LR represents a statistical model where P_{LR}, the probability of the outcome can be defined as a set of variable x and their associated coefficient b:

$$\ln\left(\frac{P_{LR}}{1-P_{LR}}\right) = b_0 + \sum_{i=1}^{k} b_i x_i \tag{4}$$

From (4), we can derive a nonlinear of probability of occurrence (P_{LR}(x)) as a function of variable x as shown in (5). The variables, x, are independent and they (x_1,x_2,…,x_k) can either represent regressed variables such as Ta, Sw, or Mh.

$$P_{LR}(x) = \frac{1}{1+e^{-(b_0+\sum_{i=1}^{k} b_i x_i)}} \tag{5}$$

In conclusion, these above mentioned training algorithms, namely DT, RF and LR, are considered to analyze an improvement of PM in the next section.

3. PROPOSED METHOD AND EXPERIMENTAL ANALYSIS

In the investigation, the training method to improve the PM of the manufacturing system is met with machine learning. The employment of ML algorithms (DT, RF and LR) succeeds the method of a new process for the predictive analysis of the maintenance activities of the production. The approaches are given by the software tool of ML. The tool only requires distinct types of dataset like for example csv, arff, and tab files, which provide a collected data of machines as the malfunction appeared in the line, from monitoring system of each individual machine, the total malfunctions that the machine experienced during a year, the periodicity of the malfunctions of all equipment or the ad hoc repair at the malfunctions section and the numerical distinction of the machines age, which will be taken into the account of the research. ML is a subset of AI and is described as any algorithms or software that has the ability of data training. ML supports in solving many complication such as in image processing, computer vision, robotics, big data, and speech detection [26]. Additionally, ML approach is designed to develop knowledge out of insightful data upon data mining. PM procedure and technologies to improve PM such as smart network, sensors, AI, cloud computing systems are presented by [27]. This collected information is extensively utilized for early fault detection, malfunction identification, machine health evaluation and a prediction of the future status of the equipment. Technologies that drive PM are sensor networks, deep learning, and intelligent automation. Smart sensor networks (built-in or external) are to detect machine environment information. The networks gather data and transfer by using wireless technology. Technologies integration consents data accumulation and management via IoT, deep learning technology helps process data, and data analytics, where intelligent automation performs virtualization, displays service on the platform via apps, and signals to support the operation.

3.1. Proposed estimation

As for the ML-based classifier, the training of the insightful data upon data mining and employing the three training approaches (DT, RF and LR) discussed earlier are prompt. The probabilities for receiver operating characteristic (ROC) curve, area under ROC (AUC), and accuracy are taken into account as a weight required to be related with each classifier in the targeted dataset. Using training and targeted dataset associated, the accuracy of this weighted network is obtained and results the forecast model to anticipate maintenance activities. To quantify the performance characteristics of each classifier, the AUC and ROC curves are computed for training. These values for the 3 classifiers are developed. Upon normalization of these weights, the weights for the resulting estimator are designated. If supplementary information is needed by adjustment, the weight corresponded with this estimator is called as the adjustment weight. An estimation weight is used as the design weight has been calibrated for compensation or supplementary data. The measures of the three training ML algorithms, that is, DT's, RF's, and LR's algorithms are calculated and the estimators, which are, the normalized weighted estimators, for forecasting the PM. Figure 2 lists the proposed PM's algorithm. Note that in the weighted cases, the estimators performed better than considering only single but unstable solution. It is motivating to see that, in all three algorithms, the variance of the weighted estimator is relatively small as they develop comparatively close precision values. These effects towards the PM will be discussed in the next section.

Proposed PM Algorithm

Requirement:
Insightful data matrix $[D]_{xy}$ with *x* records and *y* attributes
/** Training dataset
Targeted dataset $[T]_{ab}$ with *a* records and *b* attributes
/** Prediction dataset with *c* classes

Initialize:
ALG(1) = DT's algorithm
ALG(2) = RF's algorithm
ALG(3) = LR's algorithm
A_{N1}, A_{N2}, A_{N3} = Accuracy upon normalization
A_1 = Accuracy for DT's algorithm
A_2 = Accuracy for RF's algorithm
A_3 = Accuracy for LR's algorithm

for *k* = 1 to 3 do
  $A_k = 0$; $A_{Nk} = 0$;
end for

Execute :
for *m* = 1 to 3 do
  Run ALG(*m*) with $[D]_{xy} \rightarrow A_m$
end for
Normalize all A → all A_N

For *i* = 1 to c do
Perform prediction analysis with $[T]_{ab} \rightarrow T_i$
For *j* = 1 to 3 do
$S_i = \sum_{q=0}^{j} A_{Nq} T_i$
end for
end for
Return $\arg\max_{i \in S}(S_1, S_2, S_3, \dots, S_i)$/** Solution for this computation

Figure 2. Proposed PM algorithm

3.2. Simulation results

The classification accuracy can be resolved by allocating the minimum distance from a coordinator (*x*,*y*) on the ROC curve to the upper-left space of the same as the precision of each classifier depends on the insights. The performance of the three classifiers can be refereed by executing them at the obtained optimum. It is found that the LR classifier is placed first regarding its precision. The LR-based classifier gets higher precision compared to other two methods. However, the research investigates the moderation of these three approaches by taking the weighted value as a proportion in terms of their ranks in order to predict the PM. The simulations are performed by using the ML tool such as Weka. Using a synthetic data with 96 training substances (observed class is machine usage frequency) which are insights mined from collected data, while using 6-machine-statuses for prediction models is simulated. Results from individual algorithm are compared to the proposed algorithm which is described in Figure 2 as shown in Table 1. Machines which are rarely used (R) are observed and are considered as higher priority in PM. As shown in the table, the proposed PM based on accuracy results from each algorithm is computed for a guided forecast in maintenance. By this a high accurate LR algorithm alone is not appreciated and other lower accurate algorithms are ignored, but all algorithms which are playing significantly in their own different judgments upon their trainings are taken into account. Although results based on proposed method seem to be different figures compared to those from three algorithms but these figures are from the bottom line of taking up the accuracy values. The PM supports maintenance people that they should keep a close look (high priority) on M1 equipment for their maintenance activities.

Table 1. Simulation results for the first insight

Classifier	Precision (%)	AUC (%)	Accuracy (%)	PM (%,R)					
				M1	M2	M3	M4	M5	M6
DT	66.9	57	62.4	100	0	100	100	100	0
RF	64.1	48.7	74.2	89	76	85	87	66	68
LR	69	44.9	77.4	87	73	73	77	78	78
Proposed	-	-	-	**91.5**	52.8	85	87.2	80.3	51.8

Next another synthetic data with 1083 training substances (observed class is destructive) which are insights mined from collected data is simulated, while using 10-destructive-mahines for prediction models as a reference. Results from individual algorithm are compared to the proposed algorithm which is discussed in Figure 2 as shown in Table 2. Machines which are destructible (D) are considered as higher priority in PM. As displayed in the table, the proposed PM recommends that the maintenance team should prioritize M10 machine for the maintenance.

Table 2. Simulation results for the second insight

Classifier	Precision (%)	AUC (%)	Accuracy (%)	PM (%,D)									
				M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
DT	77.6	55	78	84	84	16	84	16	16	84	84	84	84
RF	80	70	84	56	79	76	76	53	90	74	79	91	90
LR	83.1	75	85.5	77	52	50	50	54	76	56	57	90	93
Proposed	-	-	-	72	71.2	48.1	69.5	41.7	61.8	70.9	73	88.4	**89.1**

The synthetic data with 187 training substances (observed class is extra maintenance cost) which are insights mined from collected data is used, while 3 machines for prediction models are observed. Results from individual algorithm are compared to the proposed algorithm which is explained in Figure 2 as shown in Table 3. The machines which involve with extra cost (C) in the PM purpose are specifically considered. As listed in the underneath table, the proposed PM recommends that the maintenance team should stress M1 machine for their first priority in maintenance.

Table 3. Simulation results for the third insight

Classifier	Precision (%)	AUC (%)	Accuracy (%)	PM (%, C)		
				M1	M2	M3
DT	64.2	61	65	38	0	38
RF	60	60	64	53	93	56
LR	65	56.3	68	39	31	35
Proposed	-	-	-	**43.2**	40.9	42

4. CONCLUSION

In this article, the proposed approach for the PM is introduced for guided maintenance schedule and particularly the priority of the targeted machines which need immediate maintenance. After the mining of factory dataset, ML's training algorithms (DT, RF and LR) are utilized for building guidelines and predictive models. The article stresses out the significance of the PM of the equipment from insights upon the extraction of the synthetic dataset available from the industrial companies, which can be employed for the development of training models in ML. With the proposed method, the skill to opt for the instant schedule to the maintenance team is demonstrated. The proposed method adopts ML's freeware like Weka which affords needful information for the hasty, steady, effective, and authentic decision from the engineering viewpoint to dodge downtime of machines which will risk the productivity and minimize the cost of the regular maintenance as well. The predictive model starts, where it has been pinpointed as a vulnerable data that is likely to become the reason of malfunction of the whole regular machines. The knowledge management (KM) for the specific component of the equipment contributes the capability to schedule to instantaneous maintenance priority of the specific equipment, as its stability can affect the entire system. Additionally, the predictive result may not be pleasing any individual ML algorithms, since the weighted values upon normalization are used for the design of a different maintenance plan, targeting at the appropriate operation of the machine and cutting resources downtime/malfunction cost. The quality of the prediction model can grow with the upsurge in the enhancement of collected data. Finally, the proposed method can help create faster PM undertaking a decline in the total maintenance cost, by lowering overtime payment and the procurement of unnecessary spare parts of the machine. The management of the enterprise with KM for fast decision-making and effective decisions, can use a new maintenance plan of monitoring for the availability and reliability of the under maintenance equipment machine. The imminent research will focus to forecast faults, and machine predictive condition while the maintenance activity cannot fit on the usage plan of the equipment. The advancement of the proposed method can develop specific maintenance rules to automate maintenance activities for an industrial intelligent automation. Not to mention, in a company, the improvement of many procedures that generate complex data in the manufacturing can be taken into consideration, such as timetable of the engineers' shifts, real-time basis monitoring, machines' health tracking, and the safety of the environment.

REFERENCES

[1] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers and Industrial Engineering*, vol. 137, Nov. 2019, doi: 10.1016/j.cie.2019.106024.

[2] W. Zhang, D. Yang, and H. Wang, "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2213–2227, Sep. 2019, doi: 10.1109/JSYST.2019.2905565.

[3] J. Lee, J. Singh, M. Azamfar, and V. Pandhare, "Industrial AI and predictive analytics for smart manufacturing systems," in *Smart Manufacturing*, Elsevier, 2020, pp. 213–244.

[4] S. Chibani and F. X. Coudert, "Machine learning approaches for the prediction of materials properties," *APL Materials*, vol. 8, no. 8, Aug. 2020, doi: 10.1063/5.0018384.

[5] N. Mahantesh, P. Aditya, and U. Kumar, "Integrated machine health monitoring: a knowledge based approach," *International Journal of Systems Assurance Engineering and Management*, vol. 5, no. 3, pp. 371–382, Jul. 2014, doi: 10.1007/s13198-013-0178-1.

[6] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013, doi: 10.1109/TII.2013.2243743.

[7] E. M. Frazzon, T. H. Holtz, L. S. Silva, and M. C. Pires, "Simulation-based analysis of intelligent maintenance systems and spare parts supply chains integration," *Journal of Intelligent Systems*, vol. 28, no. 1, pp. 31–42, Jan. 2019, doi: 10.1515/jisys-2016-0303.

[8] A. Bousdekis, D. Apostolou, and G. Mentzas, "Predictive maintenance in the 4th industrial revolution: benefits, business opportunities, and managerial implications," *IEEE Engineering Management Review*, vol. 48, no. 1, pp. 57–62, Mar. 2020, doi: 10.1109/EMR.2019.2958037.

[9] R. S. Peres, A. Dionisio Rocha, P. Leitao, and J. Barata, "IDARTS – Towards intelligent data analysis and real-time supervision for industry 4.0," *Computers in Industry*, vol. 101, pp. 138–146, Oct. 2018, doi: 10.1016/j.compind.2018.07.004.

[10] M. A. Fabricio, F. H. Behrens, and D. Bianchini, "Monitoring of industrial electrical equipment using IoT," *IEEE Latin America Transactions*, vol. 18, no. 8, pp. 1425–1432, Aug. 2020, doi: 10.1109/TLA.2020.9111678.

[11] N. Karballaeezadeh *et al.*, "Intelligent road inspection with advanced machine learning; Hybrid prediction models for smart mobility and transportation maintenance systems," *Energies*, vol. 13, no. 7, Apr. 2020, doi: 10.3390/en13071718.

[12] N. Sakib and T. Wuest, "Challenges and opportunities of condition-based predictive maintenance: a review," *Procedia CIRP*, vol. 78, pp. 267–272, 2018, doi: 10.1016/j.procir.2018.08.318.

[13] J. Wan *et al.*, "A manufacturing big data solution for active preventive maintenance," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2039–2047, Aug. 2017, doi: 10.1109/TII.2017.2670505.

[14] M. Fernandes, A. Canito, J. M. Corchado, and G. Marreiros, "Fault detection mechanism of a predictive maintenance system based on autoregressive integrated moving average models," in *Advances in Intelligent Systems and Computing*, vol. 1003, Springer International Publishing, 2020, pp. 171–180.

[15] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of big data challenges and analytical methods," *Journal of Business Research*, vol. 70, pp. 263–286, Jan. 2017, doi: 10.1016/j.jbusres.2016.08.001.

[16] K. Sumiran, "An overview of data mining techniques and their application in industrial engineering," *Asian Journal of Applied Science and Technology*, vol. 2, no. 2, pp. 947–953, 2018.

[17] P. Gackowiec, "General overview of maintenance strategies – concepts and approaches," *Multidisciplinary Aspects of Production Engineering*, vol. 2, no. 1, pp. 126–139, Sep. 2019, doi: 10.2478/mape-2019-0013.

[18] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, Mar. 2021, doi: 10.1007/s42979-021-00592-x.

[19] J. H. Jeong, D. Y. Cheon, and S. H. Han, "A better maintenance strategy, a more sustainable Hanok: Towards Korean traditional public facilities," *Buildings*, vol. 9, no. 1, p. 11, Jan. 2019, doi: 10.3390/buildings9010011.

[20] Osisanwo F. Y., Akinsola J. E. T., Awodele O., Hinmikaiye J. O., Olakanmi O., and Akinjobi J., "Supervised machine learning algorithms: classification and comparison," *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128–138, Jun. 2017, doi: 10.14445/22312803/ijctt-v48p126.

[21] R. H. A. Al-Sagheer, "Popular decision tree algorithms of data mining techniques: a review," *International Journal of Computer Science and Mobile Computing*, vol. 6, no. 6, pp. 133–142, 2017.

[22] D. Berrar, "Bayes' theorem and naive bayes classifier," in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, Elsevier, 2018, pp. 403–412.

[23] Q. Ren, H. Cheng, and H. Han, "Research on machine learning framework based on random forest algorithm," in *AIP Conference Proceedings*, 2017, vol. 1820, doi: 10.1063/1.4977376.

[24] A. Borucka and M. Grzelak, "Application of logistic regression for production machinery effciency evaluation," *Applied Sciences (Switzerland)*, vol. 9, no. 22, p. 4770, Nov. 2019, doi: 10.3390/app9224770.

[25] Z. M. Çinar, A. A. Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability (Switzerland)*, vol. 12, no. 19, Oct. 2020, doi: 10.3390/su12198211.

[26] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," in *2018 5th International Conference on Industrial Engineering and Applications, ICIEA 2018*, Apr. 2018, pp. 355–361, doi: 10.1109/IEA.2018.8387124.

[27] Z. Z. Hu, S. Leng, J. R. Lin, S. W. Li, and Y. Q. Xiao, "Knowledge extraction and discovery based on BIM: a critical review and future directions," *Archives of Computational Methods in Engineering*, Apr. 2021, doi: 10.1007/s11831-021-09576-9.

BIOGRAPHIES OF AUTHORS

Chanintorn Jittawiriyanukoon received the B.Eng degree in electrical engineering from Kasetsart University, Bangkok, Thailand in 1984, M.Eng. and D.Eng. degrees in data communication from Osaka University, Osaka, Japan in 1987 and 1990 respectively. He is presently an assistant professor in Information Technology and also a Director of Master program in Information Technology Management at Graduate School of Business and Advanced Technology Management, Assumption University, Bangkok, Thailand. His research focuses include big data curation, computer vision, fast routing algorithm, adaptive rate control (ARC) algorithm, queuing network analysis and high-speed-network performance evaluation. He can be contacted at email: pct2526@yahoo.com.

Vilasinee Srisarkun holds a PhD in Information Technology from University of Wollongong, Australia. She is currently a lecturer at Martin de Tours School of Management and Economics, Assumption University, Thailand. Her researches include human-computer interaction, information management. She can be contacted at email: vicki911@gmail.com.