Enabling Predictive Maintenance Using Machine Learning in Industrial Machines With Sensor Data

Adelina Zian Andriani
Department of Industrial and Systems Engineering
Sepuluh Nopember Institute of Technology
Surabaya, 60111, Indonesia
adelina.19024@mhs.its.ac.id

Nani Kurniati and Budi Santosa
Department of Industrial and Systems Engineering
Sepuluh Nopember Institute of Technology
Surabaya 60111 Indonesia
nanikur@ie.its.ac.id, budi_s@ie.its.ac.id

Abstract

In line with the advancement of Industry 4.0 which provides opportunities for the utilization of sensors and Machine Learning (ML) technology, make Predictive Maintenance (PdM) practices much easier. Regarding implementing PdM with ML, manufacturers need to provide data that supports the machine learning process. However, the majority of data is unlabeled and still requires manual labeling to support the learning process, which is risky, costly, and labor-intensive. Therefore, the current research uses the integration of Active Learning (AL) and Semi-Supervised Learning (SSL) to solve labeling problems and support PdM models with a better level of generalization. First, unlabeled multi-sensor data stored on the main server database and slight labeled data becomes the research sample. Second, the AL scheme selects the most valuable unlabelled samples, to label and add to the training data set. Third, the SSL scheme to optimize the data usage, using the remaining samples to be labeled. Finally, based on the augmented training data set, the fault diagnostic model is trained to support the failure class prediction. Regarding the selection of the ML algorithm, the result of trained Random Forest Classification (RFC) could predict a fault diagnostic model of approximately 99,85%.

Keywords
Predictive Maintenance (PdM), Active Learning, Semi-Supervised Learning, and Fault Diagnostic

1. Introduction

The availability of production machines and equipment is one thing that needs to be considered to support the continuity of manufacturing process. According to Ran et al., (2019) the shutdown and unplanned downtime will disrupt the company's core processes and have the potential to cause losses and loss of reputation which is difficult to quantify. According to this, equipment maintenance practices are one of the policies that need to be prioritized in manufacturing. CKM Lee et al., (2017) and Susto et al., (2013) explain that machine maintenance strategies can be categorized into three types, that is Corrective Maintenance (CM), Preventive Maintenance (PM), and Predictive Maintenance (PdM). The technology revolutions and development of system complexity lead to the transition of machine maintenance management from CM to PM to PdM. In this case, PdM becomes a trade-off between the two strategies by optimizing the frequency of maintenance efforts and maintenance costs. Montero Jimenez et al., (2020) and Ran et al., (2019) state that PdM allows the lowest possible maintenance frequency to prevent unplanned CM without incurring excessive costs for doing too much PM.
Chuang et al., (2019), Lande et al., (2019), and Sakib & Wuest, (2018) define PdM as a maintenance activity carried out by detecting the changes on physical condition of machines, that aimed for maintenance activities and maximizing tool life without increases failure risk, which seeks to understand that maintenance must be taken to delay the failure. In contrast to other strategic approaches, Jimenez-Cortadi et al., (2020) explained that CM more focused on machine

maintenance efforts after the engine was damaged (which possibly to cause excessive costs) and PM tries to determine preventive actions, such as periodically planned maintenance, to prevent the occurrence of breakdowns and unplanned maintenance (which possibly to do an excess or lack of maintenance schedule), while PdM works to achieve pre-failure interventions appropriately. PdM will predict future failure events and plan the right maintenance solution at the right time. So, by knowing which equipment that requires maintenance, maintenance practices can be better planned. According to Namuduri et al., (2020) and Ran et al., (2019), in the end, PdM activities will provide benefits such as reduced downtime, increased availability and reliability of equipment, reduced losses due to damaged equipment, reduced warranty costs, and reduced costs operation.

Jimenez et al., (2020) stated that the main idea of PdM is the management and the use of historical data to detect the trends or behavior to predict the equipment failure itself. According to Susanto & Kurniati, (2020) the success of implementing condition-based maintenance depends on the manufacturing ability to manage historical data information that comes from monitoring equipment conditions. In line with technological advances in Industry 4.0, monitoring equipment condition can be easier to do now. Santosa & Umam, (2018) also states that the availability of data in data mining (regarding to machine learning process) is an absolute thing, which comes from an object being observed, and in a data set itself represent the objects and their related features or attributes. Azar & Naderkhani, (2020) explain that the presence of sensors, networks, and integrated systems makes access to the acquisition, condition monitoring, and data storage easier. In this case, the sensors that can record equipment conditions in real-time, possibly to produce a very huge data (Big Data), which can support the implementation of PdM.

Coleman et al., (2017) showed that with the development of Augmented Intelligent and Augmented Behavior which developed event processing tools, data processing platforms, and analytic algorithms (i.e Machine Learning (ML)), were able to support PdM problem-solving although in complex systems. Orrù et al., (2020) state that advances computational allow the application of ML algorithms to find correlations and identify patterns in complex (e.g multi-sensor) data and in large volumes (e.g Big Data), by using trained data to provide a reliable model. Saravanan & Sujatha, (2018) and Solhmirzaei et al., (2020) explain that ML is an AI technique related to the study, design, and development of algorithms, which can learn from data and make predictions on that data without making explicit programming. Adhikari et al., (2018), Bekar et al., (2020), and Saravanan & Sujatha, (2018) state that based on the way ML carries out the learning strategies, ML techniques are categorized into supervised learning, unsupervised learning, and semi-supervised. learning. Comprehensive research in PdM is commonly developed using various ML algorithm techniques, according to the problem cases condition.

The latest research on ML applications in PdM activities was carried out by Chen et al., (2019) for fault diagnostic activity on gearbox element objects that derived from vibration signals data, with Random Forest (RF) alghorithm for classifier tool. This research provides insight in the form of an active learning (AL) scheme with a pool-based sampling scenario and an uncertainty sampling as a query strategy. The research proposed according to the limitations of labeled data for learning process and the opportunity to utilize the available unlabeled data. The AL scheme was proposed to facilitate the labeling process which is actually costly and requires a lot of manpower, in order to determine which sample data should be selected for labeling to support machine learning process. In this study, the RF algorithm was trained using training data to identify the gearbox element failure modes, and the learning was repeated until the desired model performance achieved. Based on the research, the application of this concept guarantees the improvement of fault diagnostic accuracy through the established fault pattern classification.

Another approach can use the integration of semi-supervised learning (SSL) and active learning (AL) for labeling problems and optimization of data usage. Ellefsen et al., (2019) and Youn & Kim (2012) explain that several previous PdM studies in the SSL technique can improve accuracy for fault diagnostic and prognostic problems. According to Le et al., (2016), the SSL approach is theoretically unquestionable and has been implemented successfully. However, Leng et al., (2013) stated that if the data set label for the machine learning process is too small to be the initial base for extracting the intrinsic elements of an unlabeled data set, it tends to produce a weak classifier model and cause errors in the training dataset. Thus, further research combines the SSL method integrated with AL to solve the labeling problem and support model generalization. Han et al., (2016), Huang et al., (2014), and Ren et al., (2020) explain that AL carries out a labeling process based on the lowest level of confidence to select the most informative and representative sample that has the potential to increase the performance model from the set of unlabelled data to be labeled by expertise. On the other hand, SSL at each iteration selects the sample with the highest level of confidence, to then added and automatically by machine anotators to predict labels without human involvement. However, if the initial model is weak, many of the predicted labels are incorrect, and caused the training data set to be biased. Therefore, a combination of AL and SSL methods can be used to form better prediction models, because AL was able to prepare the best training dataset to used by SSL in predicting unlabeled data.

This paper proposes the design development of AL and SSL integration in the context of PdM through the presence of Big Data. The structure of the paper is presented as follows. In Section 2, the related literature review is described.

The research method is described in Section 3. Case study and results of the evaluation approach in Section 4. Finally, conclusions are given in Section 5.

1.1 Objectives

The current research uses the integration of AL and SSL to solve the problem of labeling and optimization of data usage in support of the PdM model with a better generalization. Which is intended to obtain a predictive model for identifying future equipment failure classes to support the fault diagnostic process, which is applied to case studies in the cement processing manufacturing industry.

2. Literature Review

In the previous discussion, the semi-supervised learning (SSL) and Active Learning (AL) approaches are in a situation that demonstrated that a condition has less labeled data than unlabeled data. Both approaches try to use the information from the two data for the learning process to making predictions. The focus of these two approaches is to process the intrinsic elements in unlabeled data and to label the unlabeled data by considering the presence of labeled data. To reduce the workload of manual labeling, AL aims to find the most informative samples in the dataset that are not labeled at each iteration, then expertise will manually do labeling. Meanwhile, SSL aims to carry out the sample labeling process by the machine itself. Leng et al., (2013) describe the main characters of AL and SSL as follows:

1. AL in each iteration takes the most informative sample with the lowest level of confidence. The sample is then assigned to expertise for the labeling process. In this case, the human role is still involved during the training process. The selection of the most informative samples will improve model performance and can accelerate the convergence speed.
2. SSL on it each iteration selects the sample that has the highest level of confidence, to then adds and machine-predicted labels without human involvement. However, if the initial model very weak, many of the labels were miss predictable and would result in the training data being heavily misinformed.

Leng et al (2013) stated that combining the SSL and AL methods could minimize the manual labeling efforts. Some AL and SSL integration concepts can start the process using the SSL approach first, then in the final phase use the AL approach, or another research use the AL concept in the initial phase and SSL in the later phase. Research by Kontonatsios et al., (2017) on the case of text classification to support citation screening, the initial phase begins with labeling a small sample randomly, then the SSL method (label propagation) is used for automatic labeling process, and the AL method is used in the end phase to carry out labeling on the remaining unlabeled samples. In contrast, in research conducted by Fazakis et al., (2019) in the Plethora Experiment, in which the integration method is carried out by developing AL to annotate unlabelled samples by interactively asking expertise based on various query strategies, until the last phase is trying to automatically perform labeling of unlabeled data by exploiting previously learned knowledge using SSL method, finally the model is created in both cases with the aim of application on unknown test cases. Similar research was also carried out by Camargo et al., (2020) in the case of biological data classification, the initial phase began with the AL method to select informative samples for labeling by expertise, then this labeled data was combined with the unlabeled data in the SSL method mechanism. This study proposed the integration of SSL and AL, with the AL implementation methodology first and then continued by SSL. The integration of the two approaches developed in this study is similar with the research of Fazakis et al., (2019) and Camargo et al., (2020). The AL approach is integrated with SSL, where expertise only labels the most informative samples according to the AL algorithm, while the remaining samples are automatically labeled by machine annotators with the SSL algorithm.

3. Methods

The PdM implementation primarily seeks to assess the health of the equipment and find detailed information regarding current or future failures through historical data analysis. Barraza-Barraza et al., (2014) and Jardine et al., (2006) stated that the activities in PdM include fault diagnostics, that include detection, isolation, and identification of failures that occur, as well as fault prognostics to predict future failures such as future degradation or potential failures and estimating the Remaining Useful Life (RUL) of equipment. The previous research conducted by Chen et al., (2019) was aimed at the fault diagnostic process, with supporting methods using AL based on a pool-based sampling scenario and an uncertainty sampling query strategy. In this regard, the current research develops a similar concept but by adding the integration of the AL strategy with SSL for a more optimal labeling process and optimally utilizing available data, in support of the generalization level and performance of the classifier model.

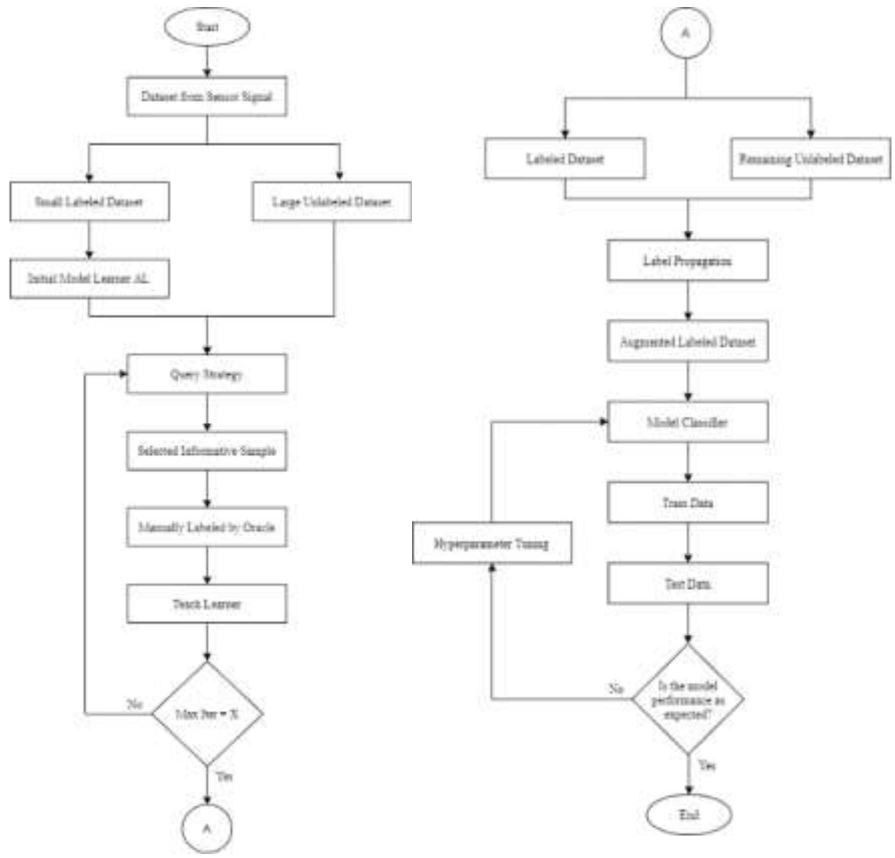

Figure 1. Fault diagnostic methodology

The steps for implementing AL and SSL integration can be seen in Figure 1. First, the multi-sensor machine data that is stored automatically on the main server database is used as raw data, also a small amount of data that has been labeled by expertise (oracle) as supporting data, which both is used as a research samples. Second, the proposed AL scheme selects the most valuable unlabeled samples, which are then labeled and added to the training data set. Third, the SSL scheme to optimize the extraction of information from data, selecting the remaining samples available for labeling. Finally, based on the formed training data, the predictive fault diagnostic model is trained to recognize the failure class mode. If the model performance still does not reach with the desired expectations, then hyperparameter tuning is carried out to determine optimal model parameter changes.

4. Case Study and Analysis

4.1 Case Study

This paper shows how the application of PdM with an active semi-supervised learning approach is carried out to support the fault diagnostic process. To demonstrate its application, it is shown by a case study conducted in a cement processing manufacturing industry. The integration concept was developed to solve labeling problems and support PdM models with a better degree of generalization. Real condition, show that the condition of each machine is monitored using sensor technology that installed on several equipments. The sensor is able to monitor and recording continuously machine activity in every 10 seconds. In this case, the historical data that is owned comes from multi-sensor data which is mostly unlabeled, with a large enough quantity. In addition, it also has limitations on labeled data, which is able to show normal output conditions and engine abnormalities according to the type of equipment failure on the available sensor data input. Information on sensors and machine support equipment are summarized in Table 1.

Table 1. Information on sensors and machine support equipment

Equipment	**Name of Sensor**	**Numerical Code of Sensor**	**Fault Class Code**	**Unit**
EP FAN	Sensor Speed EP FAN	533FN01U01S01	X1	RPM
ID FAN 1	Sensor Speed ID FAN 1	533FN02U01S01	X2	RPM
ID FAN 2	Sensor Speed ID FAN 2	533FN02U01S01	X3	RPM
Kiln Feed	Sensor Kiln Feed	541KF01A01F01X01	X4	Ton per Hours
Calciner	Sensor Temperature	542CL01N03T01	X5	Derajat Celcius
Maindrive Kiln	Sensor Speed Maindrive	543MD01U01S01	X6	RPM
Coal Feeder	Sensor Coal Feed	545RL03A01F01	X7	Ton per Hours

Based on the sensor data, the screening, processing processes, and analyzes are carried out according to the basic steps in the analytic data. In this study, to support the machine learning process to build a classifier model, used about 10% of the data that was successfully pre-processed, that around 213,847 data from 2,138,470 available data.
Regards to the AL process mechanism, from the 400 oracle data that is owned, later the model will be given learning about 200 queries, as well as the achievement for AL labeling, namely 50 - 60 times the result of AL learning data. Based on the results of the research, only about 40 queries can produce an accuracy rate of about 73.55% in just one single execution. Because the parameters are still not as expected, it is enhanced by a pool-based sampling scenario using uncertainty sampling as a query strategy with Random Forest (RF) as the chosen estimator. Through this strategy, the scenario iterates as many as 200 queries. After a series of iteration processes are carried out, the accuracy at the 200th iteration reaches 86.80%, which this achievement can be used for the labeling process on 50-60x data. Figure 2 shows a graph of the progress in the accuracy of the results of the first to last iterations performed. The learning outcomes of the AL process were then able to predict around 13,847 unlabelled data. Thus, the number of labeled data that is owned is 14,247 labeled data, with details of 13,847 data from the AL process added by 400 oracle data.

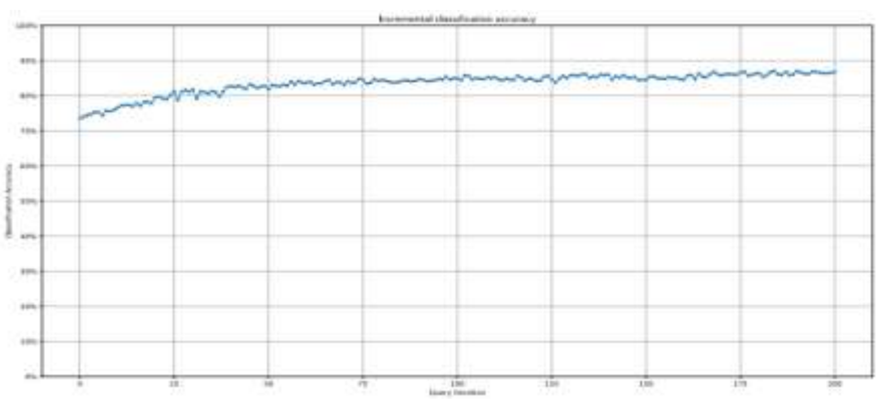

Figure 2. Graph of the development of learning classification accuracy using active learning

Furthermore, the SSL process mechanism uses labeled samples from AL to process unlabeled data that is still available, which in the mechanism uses the label spreading method for the labeling of unlabelled data. In this case, the remaining 200,000 unlabeled data will transformed as a labeled data using based knowledge form the learning process of labeled data in AL process and the owned oracle data that have. The model is processed using {sklearn.semi_supervised.LabelSpreading} with the k-nearest neighbors (k-NN) estimator. After going through the addition of labeled data from the SSL process that was carried out, the number of labeled data that was owned was a total of 214,247 data, with details of the results of the AL learning process obtained by 14,247 data and 200,000 data. After the augmented training data set is formed through the AL and SSL learning process, then the development of a fault diagnostic prediction model is carried out using the distribution of 80% training data (for model learning) and 20% testing data (for testing data) from a number of 214,247 labeled data owned. Initially, the prediction model was built using the Support Vector Classification (SVC) algorithm selection, but as a comparison to find out whether there

are differences in results when using other algorithms, this study compares it with the the Random Forest Classification (RFC) algorithm. In addition, there are several scenarios used in modeling, that is SVC by considering data outliers and or by eliminating data outliers, and RFC by considering data outliers and or eliminating data outliers.
The results of the research conducted show that the performance of the fault diagnostic classifier model that is formed shows a high enough achievement, approximately above 90%, and the achievement also has exceeded the desired expectations. The comparison of model accuracy levels is summarized in Table 2.

Table 2. The comparison of model accuracy level

Code	**Evaluation**	**Train Score**	**Test Score**
A	SVC by considering *outlier data*	98%	97,92%
B	SVC with *outlier data* elimination	97,97%	98%
C	RFC by considering *outlier data*	100%	99,85%
D	RFC with *outlier data* elimination	100%	99,85%

In this case, to consider which model to be selected, an evaluation of the overall performance of the model is carried out to find out in how the classifier model was able to predict the target class of failure well. Evaluation of model performance using the classification report that is summarized in Table 3. Meanwhile, the results of the confusion matrix evaluation are presented in Figure 3 for the evaluation of model classifiers using SVC and Figure 4 for RFC models.

Table 3 Summary of evaluation classification report model fault diagnostic

Code	**Evaluation Status**	**Coal Feeder**	**EP FAN**	**ID FAN 1**	**ID FAN 2**	**Kiln Feed**	**Maindrive**	**Normal**
A	Precision	0,51	0,91	0,33	0,69	0,45	0,78	1
	Recall	0,85	0,20	0,28	0,42	0,22	0,65	1
	F1 Score	0,64	0,32	0,31	0,52	0,29	0,71	1
	Accuracy	0,98						
B	Precision	0,53	0,81	0,34	0,71	0,40	0,83	1
	Recall	0,86	0,20	0,31	0,42	0,21	0,67	1
	F1 Score	0,65	0,32	0,32	0,53	0,28	0,74	1
	Acuracy	0,98						
C	Precision	0,98	1	0,97	0,97	0,94	0,94	1
	Recall	0,98	0,98	0,94	0,98	0,92	0,95	1
	F1 Score	0,98	0,99	0,95	0,98	0,93	0,95	1
	Acuracy	1						
D	Precision	0,97	0,94	0,96	0,98	0,93	0,99	1
	Recall	0,98	0,96	0,96	0,97	0,93	0,96	1
	F1-Score	0,97	0,95	0,96	0,98	0,93	0,97	1
	Acuracy	1						

In this study, it is considered that it would be much more risky if the classifying model predicts no failure when in fact there is a failure occured. Therefore, the model should have high recall performance. However, if both of the precision and recall values show the same high results, then the F1-Score assessment can be used as a consideration because the evaluation considers the harmonic mean of recall and precision. If the F1-Score has good evaluation results, it indicates that the classifier has good precision and recall. The comparison of the two evaluation methods shows that the overall RFC is superior to the SVC because the consideration of the F1-Score assessment indicates that the RFC is good enough with the achievement of both precision and recall.

	Coal Feeder	EP Fan	ID Fan 1	ID Fan 2	Kiln Feed	Main Drive	None
Coal Feeder	619	0	0	73	32	1	0
EP Fan	57	30	39	12	0	14	0
ID Fan 1	116	0	53	1	1	17	0
ID Fan 2	383	3	35	192	21	3	0
Kiln Feed	123	0	34	0	45	5	0
Main Drive	76	0	0	0	0	142	0
None	29	0	0	0	0	0	48876

	Coal Feeder	EP Fan	ID Fan 1	ID Fan 2	Kiln Feed	Main Drive	None
Coal Feeder	632	1	0	71	32	0	0
EP Fan	54	30	38	8	0	18	0
ID Fan 1	106	0	56	0	6	10	0
ID Fan 2	194	6	38	191	23	1	0
Kiln Feed	119	0	35	0	42	1	0
Main Drive	71	0	0	0	1	147	0
None	23	0	0	0	0	0	48744

Figure 3. Confusion matrix model SVC (left: code A & right: code B)

	Coal Feeder	EP Fan	ID Fan 1	ID Fan 2	Kiln Feed	Main Drive	None
Coal Feeder	711	0	0	6	4	3	1
EP Fan	0	149	1	1	0	1	0
ID Fan 1	1	0	174	0	5	6	0
ID Fan 2	5	0	1	449	1	1	0
Kiln Feed	7	0	2	4	190	2	2
Main Drive	4	0	1	3	3	207	0
None	0	0	0	1	0	0	40904

	Coal Feeder	EP Fan	ID Fan 1	ID Fan 2	Kiln Feed	Main Drive	None
Coal Feeder	721	0	1	2	8	2	1
EP Fan	2	137	0	1	2	0	0
ID Fan 1	1	3	172	1	2	0	0
ID Fan 2	8	2	0	417	1	0	0
Kiln Feed	3	1	3	2	198	0	0
Main Drive	2	2	3	2	1	218	0
None	1	0	0	0	0	0	40569

Figure 4. Confusion matrix model RFC (left: code C & right: code D)

Based on the results of the confusion matrix, the SVC classifier was able to predict the normal class well, but there are still many prediction errors in some machine abnormalities. Meanwhile, the RFC classifier was able to predict the normal class well with a few prediction errors to predict other machine abnormalities. Thus, from the results of the classification report and confusion matrix analysis, the RFC code D classifier was chosen as a prediction model for fault diagnostic.

4.2 Analysis

The development of the concept of active semi-supervised learning to support PdM activities in research provides an understanding that exploration and optimization of extracting data information are important. The proposed method is able to overcome labeling problems and support the development of a fault diagnostic prediction model with a good generalization level. Based on case studies in research, improving the quality of training data through the addition of data labeled to support learning process makes the model obtain the good predictions as well. In this case, to support the learning process of a better fault diagnostic model, through improving the quality of the training data, from the original 400 labeled data, then by using the AL and SSL processes, generate and produce labeled data about 214,247 labeled data that can be added. Thus, the classification model using the RFC algorithm is able to classify the failure class with an accuracy value of 99.85% on the evaluation of the test score and from the evaluation of the F1-score assessment which is based on the harmonic mean of the precision and recall values, indicating that the classifier has good precision and recall results with an average achievement of above 95%. In addition, in this study, it is preferable if the prediction model is able to predict a failure event, even though the actual failure does not exist, this is because it minimizes the risk of failure. Based on the results of the confusion matrix by using the selected model shows that if there is a 1x wrong predicted the normal class when actually experiencing a failure, but overall the model is able to guess the normal class well with only some wrong prediction to predict other engine abnormalities (i.e. Coal Feeder failure, EP FAN failure, ID FAN 1, ID FAN 2, etc).

In a previous study conducted by Chen et al., (2019) to validate the superiority and effectiveness of the proposed method, it was carried out by comparing the proposed AL learning method scenario with the supervised learning method. The study explained that as the number of samples selected by the proposed active learning strategy increases, the diagnostic accuracy is greater, indicating that the proposed active learning has the ability to select the most discriminating sample for failure diagnosis. In this case, the insights expressed are consistent with the results made in this study. In addition, with regard to the ability of AL when compared to the supervised learning method, Chen et al., (2019) comprehensively verified that the proposed active learning approach can obtain the best results compared to the supervised learning method. The supervised learning method has a failure diagnostic accuracy of about 70% -80%, on the contrary, all failure diagnostic results obtained by the proposed AL method have an average accuracy rate of about 80%, and the best results reach 90%, which means that it can correctly identify several fault mode on the object being observed. However, from the results of research conducted at this time, the learning process and the performance achievement of the classifier model that was successfully formed were able to validate the effectiveness of using AL and SSL integration, with higher results when compared to previous studies. Based on the research results, the integration of the two learning methods can improve the quality of training data more optimally, and produce a fault diagnostic accuracy that is above 95%. Thus, the use of AL and SSL in failure diagnostics is able to solve labeling problems more effectively so as to optimize the use of data and support the development of prediction models with a good degree of generalization. In this case, this method plays a role in overcoming labeling problems, which can improving the quality of training data by the addition of data labeled to supprt machine learning process.

5. Conclusion

In this paper, we develop the concept of active semi-supervised learning (AL and SSL) to support PdM activities, especially for the formation of a fault diagnostic model that is able to recognize machine equipment failure modes. The main contribution of this research is to provide the development of research theories and methods through the integration of AL and SSL. AL learning makes expertise only performs labeling on the most informative samples, which can save on the labeling process and can improve the quality of better training data. Furthermore, from the unlabeled data that is still available, to optimize the extraction of information on the data, SSL learning is carried out, in which the labeling mechanism is carried out automatically by machine annotators without the involvement of expertise anymore. In addition, this study contributes to forming a failure diagnostic model more effectively and efficiently, as well as achieving a more accurate classification performance using only a small amount of labeled data. Thus, the integration of the AL and SSL approaches has advantages when compared to other ML strategies. In this case, with the implementation of the existing case studies, the effectiveness and superiority of the proposed method are proven, and the results show that this method can improve the accuracy of the fault diagnostic model by using a small number of labeled samples.

References

Adhikari, P., Rao, H. G., & Buderath, D.-I. M. Machine Learning based Data Driven Diagnostics & Prognostics Framework for Aircraft Predictive Maintenance. 10th International Symposium on NDT in Aerospace, October 24-26, 2018, Dresden, Germany, Ml, 1–15. https://www.ndt.net/article/aero2018/papers/We.5.B.3.pdf, 2018.

Azar, K., & Naderkhani, F. (2020). Semi-Supervised learning approach for optimizing condition-based-maintenance (CBM) decisions. *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*, *2020-June*. https://doi.org/10.1109/ICPHM49022.2020.9187022

Barraza-Barraza, D., Limón-Robles, J., & Beruvides, M. G. Opportunities and challenges in Condition-Based Maintenance research. IIE Annual Conference and Expo 2014, May, 3035–3043.Camargo, G., Bugatti, P. H., & Saito, P. T. M. Active Semi-Supervised Learning for biological data classification. PLoS ONE, 15(8 August), 1–20. https://doi.org/10.1371/journal.pone.0237428, 2020, 2014.

Bekar, E. T., Nyqvist, P., & Skoogh, A. An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. Advances in Mechanical Engineering, 12(5), 1–14. https://doi.org/10.1177/1687814020919207, 2020.

Chen, J., Zhou, D., Guo, Z., Lin, J., Lyu, C., & Lu, C. (2019). An active learning method based on uncertainty and complexity for gearbox fault diagnosis. *IEEE Access*, *7*(2), 9022–9031. https://doi.org/10.1109/ACCESS.2019.2890979

Chuang, S. Y., Sahoo, N., Lin, H. W., & Chang, Y. H. (2019). Predictive maintenance with sensor data analytics on a Raspberry Pi-based experimental platform. *Sensors (Switzerland)*, *19*(18). https://doi.org/10.3390/s19183884

Coleman, C., Damodaran, S., Chandramouli, M., & Deuel, E. (2017). Making maintenance smarter Predictive maintenance and the digital supply network. *Deloitte University Press*, 1–24. https://www2.deloitte.com/us/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html

Fazakis, N., Kanas, V. G., Aridas, C. K., Karlos, S., & Kotsiantis, S. Combination of Active Learning and Semi-Supervised Learning under a self-training scheme. Entropy, 21(10). https://doi.org/10.3390/e21100988, 2019.

Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., & Zhu, X. Semi-supervised Active Learning for sound classification in hybrid learning environments. PLoS ONE, 11(9), 1–23. https://doi.org/10.1371/journal.pone.0162075, 2016.

Huang, S. J., Jin, R., & Zhou, Z. H. Active Learning by Querying Informative and Representative Examples. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(10), 1936–1949. https://doi.org/10.1109/TPAMI.2014.2307881, 2014.

Jardine, A. K. S., Lin, D., & Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012, 2006.

Jimenez-Cortadi, A., Irigoien, I., Boto, F., Sierra, B., & Rodriguez, G. (2020). Predictive maintenance on the machining process and machine tool. *Applied Sciences (Switzerland)*, *10*(1). https://doi.org/10.3390/app10010224

Jimenez, V. J., Bouhmala, N., & Gausdal, A. H. (2020). Developing a predictive maintenance model for vessel machinery. *Journal of Ocean Engineering and Science*, *xxxx*. https://doi.org/10.1016/j.joes.2020.03.003

Kontonatsios, G., Brockmeier, A. J., Przybyła, P., McNaught, J., Mu, T., Goulermas, J. Y., & Ananiadou, S. A semi-supervised approach using label propagation to support citation screening. Journal of Biomedical Informatics, 72, 67–76. https://doi.org/10.1016/j.jbi.2017.06.018, 2017.

Lande, M. S., Sirsat, P. M., & Tupkar, R. S. (2019). *A Case Study On Predictive Maintenance Of Oj / 5522 Dt- 40 Cnc Milling Machine*. *January*, 27–30.

Le, T., Duong, P., & Dinh, M. (2015). *Budgeted Semi-supervised Support Vector Machine*. UAI, 2016.

Lee, C. K. M., Cao, Y., & Ng, K. K. H. Big data analytics for predictive maintenance strategies. *Supply Chain Management in the Big Data Era*, *January*, 50–74. https://doi.org/10.4018/978-1-5225-0956-1.ch004, 2017.

Leng, Y., Xu, X., & Qi, G. Combining Active Learning and Semi-Supervised Learning to construct SVM classifier. Knowledge-Based Systems, 44, 121–131. https://doi.org/10.1016/j.knosys.2013.01.032, 2013.

Montero Jimenez, J. J., Schwartz, S., Vingerhoeds, R., Grabot, B., & Salaün, M. (2020). Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics. *Journal of Manufacturing Systems*, *56*(March), 539–557. https://doi.org/10.1016/j.jmsy.2020.07.008

Namuduri, S., Narayanan, B. N., Davuluru, V. S. P., Burton, L., & Bhansali, S. (2020). Review—Deep Learning Methods for Sensor Based Predictive Maintenance and Future Perspectives for Electrochemical Sensors. *Journal of The Electrochemical Society*, *167*(3), 037552. https://doi.org/10.1149/1945-7111/ab67a8

Orrù, P. F., Zoccheddu, A., Sassu, L., Mattia, C., Cozza, R., & Arena, S. (2020). Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. *Sustainability (Switzerland)*, *12*(11). https://doi.org/10.3390/su12114776

Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A survey of predictive maintenance: Systems, purposes and approaches. *ArXiv*, *XX*(Xx), 1–36.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., & Wang, X. A Survey of Deep Active Learning. 37(4). http://arxiv.org/abs/2009.00236, 2020.

Sakib, N., & Wuest, T. (2018). Challenges and Opportunities of Condition-based Predictive Maintenance: A Review. *Procedia CIRP*, *78*, 267–272. https://doi.org/10.1016/j.procir.2018.08.318

Santosa, B., & Umam, A. Data Mining dan Big Data Analytics (2nd ed.). Penebar Media Pustaka, 2018.

Saravanan, R., & Sujatha, P. (2018). Algorithms : A Perspective of Supervised Learning Approaches in Data Classification. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, *Iciccs*, 945–949. http://link.springer.com/10.1007/978-981-13-7403-6%0Ahttp://link.springer.com/10.1007/978-981-13-7403-6_11

Solhmirzaei, R., Salehi, H., Kodur, V., & Naser, M. Z. (2020). Machine learning framework for predicting failure mode and shear capacity of ultra high performance concrete beams. *Engineering Structures*, *224*(June), 111221. https://doi.org/10.1016/j.engstruct.2020.111221

Susanto, B. S. P., & Kurniati, N. (2020). Multi Sensor-Based Failure Diagnosis using the Mahalanobis Taguchi System. *IOP Conference Series: Materials Science and Engineering*, *847*(1). https://doi.org/10.1088/1757-899X/847/1/012036

Susto, G. A., Schirru, A., Pampuri, S., Pagano, D., McLoone, S., & Beghi, A. A predictive maintenance system for integral type faults based on support vector machines: An application to ion implantation. IEEE International Conference on Automation Science and Engineering, i, 195–200. https://doi.org/10.1109/CoASE.2013.6653952, 2013.

Youn, B. D., & Kim, T. Semi-Supervised Learning with Co-Training for Data-Driven Prognostics, 2012.

Biography

Adelina Zian Andriani is received the B.S. degree at Department of Industrial Engineering from the Surabaya University (UBAYA), Surabaya, Indonesia in 2019. Now is currently pursuing the Master degree in Industrial and Systems Engineering with concern in manufacturing systems engineering and management with Sepuluh Nopember

Institute of Technology, Surabaya, Indonesia. She have research interests include manufacturing, reliability, maintenance management, quality assessment and improvement, also lean.

Nani Kurniati is an active faculty member at Department of Industrial and Systems Engineering Sepuluh Nopember Institute of Technology, Surabaya, Indonesia. She have received the PhD degree in Department of Industrial Management from at National Taiwan University of Science and Technology (NTUST), Taiwan. She is senior researcher in reliability analysis, maintenance, warranty, and quality management.

Budi Santosa an active faculty member at Department of Industrial and Systems Engineering Sepuluh Nopember Institute of Technology, Surabaya, Indonesia. He received the PhD degree in optimization and data mining at University of Oklahoma, AS. He is senior researcher in metaheuristics, optimization, data mining, and artificial intelligent. He also an author of some books, such as Data Mining: Teknik Pemanfaatan Data, Graha Ilmu, 2006; Data Mining Terapan dengan Matlab, Graha Ilmu, 2008; Matlab untuk Statistika dan Optimasi, Graha Ilmu, 2008; and with anothers book related to the his interest topic. He is also wrote many popular articles in Kompas and Jawa Pos Newspapers.