Efficient Machine Learning Methods for Hard Disk Drive Yield Prediction Improvement

Anusara Hirunyawanakul, Nittaya Kerdprasop, and Kittisak Kerdprasop

Abstract-Deployment of machine learning techniques are prevailing in world-wide problem solving. Hard disk drive manufacturing is another prominent field seeking for the application of these knowledge intensive techniques. The manufacturing tasks that urgently require support from machine learning are in the portions of failure analysis and yield improvement. We focus our research on the yield improvement sector. Manufacturing yield prediction opens big opportunity for machine learning application because yield is a very important metric in many parts of manufacturing process. But, there rarely is research work about yield prediction in hard disk drive manufacturing found until today. So, we introduce yield prediction improvement by statistical analysis and machine learning methods including the multiple linear regression (MLR), artificial neural networks (ANN), classification and regression tree (CART). Moreover, we introduce technique to group quantity of data for yield prediction by considering consistency number, instead of grouping by calendar period as used in traditional method. The result of our technique shows the better performance. Means absolute error (MAE) of our proposal is 0.010 with a tide error rate produced by MLR and CART algorithms. The best performance from traditional calendar-based grouping is ANN algorithm with the error metric 0.017 MAE.

Index Terms—Yield prediction, hard disk drive (HDD), multiple linear regression (MLR), artificial neural network (ANN), classification and regression tree (CART).

I. INTRODUCTION

Current Hard Disk Drive (HDD) manufacturing process is characterized by capital intensity, customer reliability, and technology migration. As a result, the HDD manufacturers really need "Customer shipment planning", "Material usage planning", "Production planning", "Test capacity planning", and "Product Pricing" accurately [1]-[6]. All of these activities have "Yield Prediction" as the main concern. For example, the precise yield prediction leads to the optimal stocking of material supply quantity.

Suppose yield projection is too high from the optimistic point of view regarding the quality of the production process. Based on this yield projection scenario, we will order material supply lower than the real needs. It turns out that at the end of the production line, the number of qualified products is much less than the amount expected. The serious impact is that we will not able to supply product shipment to our customers as we had committed. On the contrary, if yield projection is too low, that mean we are pessimistic against the production quality. We thus end up over-ordering material supply. The even more serious impact is that it will reduce the cash flow of the company from inventory sink cost [7]-[9].

The method in current practice for yield prediction in HDD manufacturing is based solely on expertise of the process engineers. In recently years, the maturity of machine learning and the introduction of new techniques such as deep learning have captured the interest of engineers worldwide including those in electronics and computer parts industries like HDD manufacturing. From the literature review, we found that those state-of-art techniques are used mostly in the tasks such as failure root-cause analysis and yield improvement, while yield prediction is quite rare.

Yield prediction task in the HDD manufacturing is still using traditional method. As far as we know, there is no research work on yield prediction with machine learning technique. From the literature review, there exist many research papers making yield prediction by means of machine learning or data mining techniques in several fields such as semiconductors manufacturing, PCB (Printed Circuit Board) manufacturing, crop yield prediction, and other agricultural product yield prediction [10]-[15]. We thus propose the initiation of applying machine learning to HDD yield prediction.

In this paper, we focus on generating the HDD yield prediction models that are based on the three prominent algorithms: multiple linear regression (MLR), artificial neural network (ANN), and classification and regression tree (CART). The difficult part on modeling the HDD manufacturing yield prediction is the numerous parameters (or attributes) obtained from many processes along the production line of HDD. It is almost impossible to consider all attributes (about 400 attributes) for yield prediction task. So, we seek for collaboration from the engineering expert in real HDD industry to select from 400 attribute to be only 5 key attributes that relate to yield and failure rate concern.

The traditional yield calculation or forecasting is always based on grouping by calendar periodic such as "by day", "by week", or "by month". This method results in poor accuracy from inconsistency of quantity in each group. The products from some weeks may be of high quantity, whereas those from other weeks may be of low quantity. This inconsistency issue may influence the yield-by-week scale that we use as reference to compare with yield prediction of traditional method.

Therefore, we introduce the new quantity grouping method to improve yield prediction by grouping quantity of data into consistency numbers i.e. group of 1,000 rows, group of 5,000 rows and group of 10,000 rows (from total data containing 4,192,000 rows). According to this proposed method, we expect to mitigate the problem of quantity fluctuation.

The rest of this paper is organized as follows. In Section II, we describe the HDD background and details of yield

Manuscript received August 5, 2019; revised January 12, 2020.

The authors are with the School of Computer Engineering, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand (e-mail: anusara.hi@gmail.com, nittaya@sut.ac.th, kerdpras@sut.ac.th).

calculation in HDD production process, as well as brief introduction of the three machine learning algorithms (MLR, ANN, and CART). In Section III, we explain the material and method that we use to develop the yield prediction models. The third section also contains details of dataset, variable selection method, research framework, and research workflow. Section IV is experimentation and results. The conclusion is presented in Section V. In Section VI, we provide discussion and suggestion for future research.

II. BACKGROUND THEORY AND LITERATURE REVIEW

A. Hard Disk Drive

Hard Disk Drive (HDD) is a kind of digital data recording devices. This device stores data on the durable material hard disk (or platter) by magnetic storage technology. HDD is a non-volatile storage in that stored data are still retained even when power is off [16]-[18]. Disks are paired with the heads which are used for reading and writing data on the disks. HDD is always claimed that it is the most reliable storage among various existing technologies in the data storage industry. Conventional HDD hardware consists of many components (as schematically shown in Fig. 1) and the key components are as follows.

- HSA (Head Stack Assembly) [19] is the base plate of reader/writer heads. HSA moves synchronously with the rotation speed of disk to bring the head to the location that we need to read or write the data on disk. This movement really needs the accurate calculation of movement due to the high rotation speed of disk at 5000 RPM (Round per Minute) up to 10000 RPM.
- Media, Platter or Disk [20] is the main part for storing data on the magnetic layer. The substrate layer of disk is made from materials like aluminum or glass that are high tolerant from deformation. The surface of disk is needed to be smoothening as much as possible for data recording at a very small scale.
- VCM (Voice Coil Motor) [21] is a permanent magnet component working in accordance with HSA. It is the key part that moves HSA to the desired location based on the working principle of the magnetic field.
- MBA (Motor Base Assembly) [22] is the component composing of motor and motor base plate hub. The motor is used for rotating disks with the high speed. The motor base plate hub is the strongest part of HDD used for protecting other components in the HDD from external impact force.
- PCBA (Printed Circuit Board Assemble) [23] is the key controller of HDD composed from several wired copper, controllers, and ports to connect with computer or HDD tester.

After process of assembling all components with the five steps as depicted in Fig. 1, the HDD product unit is complete. The next step is to test the mechanical and functional operation performance of HDD. In some product series that require high capacity unit, flow of test time can be as long as 1 month with more than 10 operation steps of performance test. This long testing time is because there are numerous of components in an unit of HDD and the factory must ensure that every component properly functions and operates synchronously with other components [24]-[27].

Reliability is the most important quality criteria in data storage manufacturing. Therefore, HDD makers must test every single block (the smallest unit of data storage) to ensure that it is in good condition until the end of product lifetime. The HDDs that pass the test process are called "pass units" and those that cannot pass the test process are called "fail units".



Fig. 1. Key components of hard disk drive.

B. Yield in Manufacturing

Yield in HDD manufacturing is the ration between "output units" and "input units" in the particular process. The term "units" refer to the amount of HDDs. The calculation of yield can be describe by equation 1.

$$Yield = \frac{Output uantity}{Input Quantity}$$
(1)

Yield in equation 1 is the calculation method of only one test operation. In real life HDD production manufacturing, there are many test operations. Diagram in Fig. 2 shows the calculation method for 3 test operations.

		85.00%		
	Output Qty	1,700	Test Operation#3 Yield	97.14%
TEST OPERATION #3	Input Qty	1,750		A to be and being to be
↓	Output Qty	1,750	Test Operation#2 Yield	97.22%
TEST OPERATION #2	Input Qty	1,800		
•	omput Qij	1,000	rest operations? There	2010070
TEST OPERATION #1	Output Oty	1.800	Test Operation#1 Yield	90.00%
	Input Qty	2,000		

Fig. 2. The example of multi-operations yield calculation.

Fig. 2 demonstrates the example of yield computation in the multi-operation process. Input of test operation#1 (initial operation) is 2,000. Yield of test operation#1 is 90% (that is, 1,800 divided by 2,000). The input of operation#2 is exactly the number of output quantity from operation#1, which is 1,800, and the output of operation#2 is 1,750. Therefore, yield of operation#2 is 1,750/1,800 = 97.22%. In the same manner of computation, yield of operation#3 is 97.14%

(computed from 1,700 divided by 1,750). Cumulative yield of the whole test process is computed from the output quantity that pass the last test operation (that is, 1,700) divided by the input quantity of the first operation (that is, 2,000).

With this diagram, we can see that the operation test yield in one step has strong influence to operation yield in the next step. Typically, the early test operations (or first operation) almost always relate to key functional performance like mechanical test, electrical test, and basic test of read-write performance. So, the first operation always shows the lower yield than later operations. Accurately yield prediction in earlier operations certainly impacts efficiency and benefit in the planning activities of manufacturing [7]-[15].

Yield prediction is one of high concerns in HDD production and test capacity planning [28]. Precise yield prediction positively affect many activities including financial budgeting for material, production and tester capacity planning, machinery downtime schedule planning, shipment and customer delivery planning, and pricing plan in each lot of HDD products.

C. Algorithms for Yield Prediction

In the literature, linear regression is a statistical-based method popularly applied for yield prediction because of its simplicity and acceptable efficiency. However, with the advancement of machine learning technology, we consider this new technology as an interesting alternative for yield prediction in HDD industry. The two popular machine learning techniques used in our yield prediction are artificial neural network (ANN) and classification and regression tree (CART).

Linear regression analysis [29-31] is the statistical method for studying quantitatively correlation among two or more variables. One variable is defined as a target of analysis; this variable is called dependent variable, Y. Other variables are used for predicting the value of a target variable; these variables are called independent variables, X. The regression modeling is bases on mathematical calculation as shown in equation 2. The computation process is to find the best coefficient of variable X and some constant value that altogether can predict the Y value with minimal error. The equation can be plotted with linear graph. So, we called this algorithm simple linear regression analysis.

$$\widehat{Y} = a + bX \tag{2}$$

when

- \widehat{Y} is dependent variable (or target variable for prediction)
- *X* is independent variable
- *a* is constant of regression (or cutting point on Y axis)
- *b* is slope of line (or regression coefficient of *X*)

In case of multiple input values (or multiple independent variables), the modeling will be called multiple linear regression (MLR) analysis. The computation of MLR can be described with equation 3.

$$\widehat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \tag{3}$$

when

 \widehat{Y} is dependent variable

 $X_1, X_2, X_3, \dots, X_k$ is a set of k independent variables b_0 is a constant of regression (or cutting point on Y axis) $b_1, b_2, b_3, \dots, b_k$ is a set of line's slopes (or regression coefficients of the k independent variables)

ANN or Artificial Neural Network is the machine learning algorithm that is inspired by the biological neural networks that constitute human brains [32]-[36]. There are numerous small size neural nodes in human brain connecting together to construct the big networks with complex relation and very detailing. ANN resembles this scheme; it consists of many nodes connected with lines to compute, learn, and perform tasks. The learning is done through considering examples then adjusting weight in each connecting line to best fit the examples. The learning process can be done without programming task-specific rules. The general architecture of ANN is shown in Fig. 3. There are 3 main levels of ANN, that are, input, hidden, and output layers.

Input layer consists of input nodes and connected lines to hidden layer. The number of input nodes is equal to number of features or attributes of dataset.

Hidden layer consists of hidden nodes and connected lines to the next level. There can be more than one level in this hidden layer. Hidden layer is provided information from previous hidden layer or input layer.

Output layer consists of output nodes. The number of nodes is equal to number of values of target variable. The output nodes are always provided the information by last hidden layer.





Classification and regression tree (CART) is one of the tree-based algorithms that can perform classification and prediction tasks. This algorithm is introduced by Breiman in 1984 [37]-[39]. CART is a binary decision tree in that the node can split only 2 branches. The tree is consisted of a root node (the node at level 0 in Fig. 4) and two groups of binary subtree called left subtree and right subtree. The nodes are features of data. CART uses the Gini index for feature selection in the classification task and uses sum of squared error for predicting values. The classified target or predicted value is in the leaf node.



Fig. 4. Structure of CART.

D. Performance Measurement

For HDD yield prediction, we adopt mean absolute error (MAE) as the prediction performance measurement. The adopting of MAE is to comply with other yield prediction works appeared in the literature [7]-[15]. The calculation [40] of MAE can be done by averaging gap between real values of target variable and the predicted values as demonstrated in equation 4.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i| \tag{4}$$

when

MAE is Mean Absolute Error

n is numbers of data

- y_i is real value of target variable
- \widehat{y}_i is predicted value from model

III. MATERIAL AND METHOD

A. Dataset and Variable Selection

Our dataset is collected from the real hard disk drive production manufacturing in Thailand. The collected dataset is very big because it covers 12 months of production timeframe. Number of rows is 4,192,000 rows and number of features is more than 100.

From the assistance of engineering expert, number of features are reduced to five as they are expected to be key attributes (or features). These attributes are from type (Prime or Recycle) of key components. (PCBA Type, HSA Type, Media Type, MBA and VCM Type). Type of these key components is the main factor contributing to yield. Type "prime material" always provides the better or higher yield than "recycle material". Definition of "prime material" is the new and fresh material from suppliers and it has never been used to assemble in any HDD. Definition of "recycle material" is part of components that had been installed in some HDD. When that HDD failed in the previous test process, the reusable components will be torn down and input to rework in the recycle process.

In traditional method all rows of data are summarized in group of "date" and "week" for predicting yield. In this research, we try to mitigate the problem of quantity unit inconsistency by grouping rows of data into constant number i.e. "grouping by 1K rows", "grouping by 5K rows" and "grouping by 10K rows". Therefore, in our experiment, there are totally 5 new datasets.

B. Research Framework

Our research steps are explained in Fig. 5. Based on the objective of yield prediction performance comparison between 3 algorithms and 5 types of data grouping method, we design the research experiment according to assumption that the consistency quantity of rows (or data) in each group affects yield prediction performance.

C. Research Workflow

The flow chart in Fig. 6 depicts our research workflow. It starts by aggregating the 4.192 million rows of HDD manufacturing dataset to form the new 5 datasets: "Group of date", "Group of week", "Group of size 1K", "Group of size 5K", and "Group of size 10K". After that, we separate each 5

new datasets into 80% and 20%. The first 80% is for training the model, and the rest 20% is for testing model performance. The training datasets are input into 3 algorithms. In the performance comparison stage, we compare along the two main aspects: comparison between 5 grouping types and comparison between 3 algorithms.



Fig. 6. Research workflow diagram

IV. EXPERIMENTATION AND RESULTS

The original HDD manufacturing dataset contains 4.192 million rows. We perform five different granularity levels of data grouping: group by date, group by week, group by size of 1K, group by size of 5K, and group by size of 10K. All groups have the same set of features and number of selected features is 5. After data preparation, we modeling each dataset using 3 algorithms: MLR, ANN and CART. The

model evaluation results in terms of MAE are shown in Table 1.

On comparing model performance based on data grouping by calendar period like "By date" and "By week", the best performance is ANN with MAE on test data at 0.039 and 0.017, respectively. When considering model performance based on our proposed method for grouping data by quantity consistency with the group size of 1K, ANN shows the best performance with MAE at 0.041. We can notice that when the group size is larger (5K and 10K), the MAE drops significantly (0.019 in 5K group and 0.01 in 10K group). The best models are those from the MLR and CART algorithms, which perform slightly better than ANN.

GROUPING METHOD	No.	DATASET	#Rows	LEARNING	MEAN ABSOLUTE ERROR	
				Algorithm	TRAINING	TESTING
Traditional Method:Grouping by Calendar Periodic	1	GROUP BY DATE	315	MLR	0.034	0.041
				ANN	0.032	0.039
				CART	0.033	0.043
	2	GROUP BY WEEK	52	MLR	0.004	0.067
				ANN	0.005	0.017
				CART	0.009	0.027
PROPOSED METHOD: GROUPING BY QUANTITY CONSISTENCY	3	GROUP SIZE = $1K$	4,192	MLR	0.042	0.043
				ANN	0.040	0.041
				CART	0.049	0.049
	4	GROUP SIZE = $5K$	836	MLR	0.017	0.019
				ANN	0.018	0.019
				CART	0.018	0.019
	5	GROUP SIZE = $10K$	416	MLR	0.010	0.010
				ANN	0.010	0.011
				CART	0.011	0.010

TABLE I: YIELD PREDICTION PERFORMANCE COMPARISON BASED ON MEAN ABSOLUTE ERROR (MAE)





A graph on the right hand side of Fig. 7 illustrates error trends of yield prediction models trained with data grouped by quantity consistency method. It can be noticed from the trend line that grouping by 10K rows provides the lowest error rate with almost equal predictive performance among the three learning algorithms. This graph also depicts the observation that with larger scale of grouping by consistency number (from 1K to 5K and then 10K), the trend of MAE is getting lower in accordance with the number of rows in the group. This is in contrast to the grouping by calendar that the prediction results of the models are quite stray and fluctuate.

V. CONCLUSION AND DISCUSSION

This paper introduces the case study of applying machine

learning and statistical analysis techniques to predict yield in the hard disk drive (HDD) industry. We also presents the idea for yield prediction performance improvement by grouping data with consistency of quantity. Our assumption is that the fluctuation of quantity between groups has more or less influence to the low performance on yield prediction. Therefore, we design the empirical research framework by grouping data in 3 granular scales, that is, grouping of 1K rows, grouping of 5K rows, and grouping of 10K rows. The proposed method for data grouping has been experimentally compared against the traditional method that groups data either by date or by week.

The experimentation has been done with data that were collected from real life HDD manufacturing. The dataset contains over 4 million rows covering 1 year of production records. From the experimental results, we can conclude that our proposed method of grouping by quantity consistency of rows shows the better performance of yield prediction when compared against the traditional method that groups data by calendar period. The performance comparison is based on the mean absoluter error measured from yield prediction. We also notice the better trend when number of rows is getting higher. The three learning algorithms depict the same trend of this significant observation. So, these results allow us to conclude that different schemes on data grouping can result in diverse performance of HDD yield prediction.

VI. RECOMMENDATION

This research uses five key attributes as independent variables to train the learning models. The feature selection depends solely on experience of the expert engineer. In our future research work, we plan to make this step more systematic by applying the available feature selection techniques to evaluate each feature and select the most promising ones. Moreover, regarding our proposed scheme of data grouping, we plan to investigate several sizes of data group. However, that means we are challenging by the very big data size that may contain over 10 million rows of data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

The first author is responsible for designing the research framework, organizing the experimentation steps and preparing the manuscript. The second author helps editing the manuscript and validating the research steps. The last author helps confirming the experimental results and discussing the future research trends.

ACKNOWLEDGMENT

This research work has been supported by grants from the National Research Council of Thailand (NRCT). Data and Knowledge Engineering Research Unit has been fully supported by a research grant from Suranaree University of Technology.

REFERENCES

- [1] K. N. Malitski, *Method for Shipment Planning/Scheduling*, U.S. Patent 8,244,645, 2012.
- [2] A. K. R. Katta and R. Allgor, "Heuristic methods for customer order fulfillment planning," U.S. Patent 8,352,382, Amazon Technologies Inc, 2013.
- [3] M. Braglia, D. Castellano, M. Frosolini, and M.Gallo, "Overall material usage effectiveness (OME): A structured indicator to measure the effective material usage within manufacturing processes," *Production Planning & Control*, vol. 29, no. 2, pp. 143-157, 2018.
- [4] J. Shi, G. Zhang, and J.Sha, 2011, "Optimal production planning for a multi-product closed loop system with uncertain demand and return," *Computers & Operations Research*, vol. 38, no. 3, pp. 641-650.
- [5] A. P. Rastogi, J. W. Fowler, W. M. Carlyle, O. M. Araz, A. Maltz, and B. Büke, "Supply network capacity planning for semiconductor manufacturing with uncertain demand and correlation in demand considerations," *International Journal of Production Economics*, vol. 134, no. 2, pp. 322-332, 2011.
- [6] S. M. Liozu and A. Hinterhuber, "Industrial product pricing: A value-based approach," *Journal of Business Strategy*, vol. 33, no. 4, pp. 28-39, 2012.
- [7] H. Lee, C. O.Kim, H. H. Ko, and M. K. Kim, "Yield prediction through the event sequence analysis of the die attach process," *IEEE Transactions on Semiconductor Manufacturing*, vol. 28, no. 4, pp. 563-570, 2015.
- [8] J. Li, X. Ji, Y. Jia et al., "Hard drive failure prediction using classification and regression trees," in Proc. 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, 2014, pp. 383-394.
- [9] T. Yuan, S. Z. Ramadan, and S. J. Bae, "Yield prediction for integrated circuits manufacturing through hierarchical Bayesian modeling of spatial defects," *IEEE Transactions on Reliability*, vol. 60, no. 4, pp. 729-741, 2011.
- [10] T.Chen, "A heterogeneous fuzzy collaborative intelligence approach for forecasting the product yield," *Applied Soft Computing*, vol. 57, pp. 210-224, 2017.
- [11] D. An, H. H. Ko, T.Gulambar, J. Kim, J. G. Baek, and S. S. Kim, 2009, "A semiconductor yields prediction using stepwise support vector machine," in *Proc. 2009 IEEE International Symposium on Assembly* and Manufacturing, pp. 130-136.
- [12] K. Roy, A. Mukherjee, and D. K. Jana, "Prediction of maximum oil-yield from almond seed in a chemical industry: A novel type-2 fuzzy logic approach," *South African Journal of Chemical Engineering*, vol. 29, pp. 1-9, 2019.
- [13] L. Kouadio, R. C. Deo, V. Byrareddy, et al., "Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties," *Computers and Electronics in Agriculture*, vol. 155, pp. 324-338, 2018.
- [14] T. Chen, "An ANN approach for modeling the multisource yield learning process with semiconductor manufacturing as an example," *Computers & Industrial Engineering*, vol. 103, pp. 98-104, 2017.

- [15] C. Y. Lee and T. L. Tsai, "Data science framework for variable selection, metrology prediction, and process control in TFT-LCD manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 55, pp.76-87, 2019.
- [16] D. A. Patterson and J. L. Hennessy, "Computer organization and design MIPS edition: The hardware/software interface," *Newnes*, 2013.
 [17] J. S. Domingo, "SSD vs HDD: What's the difference?," *PC Magazine*,
- [17] J. S. Domingo, "SSD vs HDD: What's the difference?," PC Magazine, 2015.
- [18] N. U. Mustafa, A. Armejach, O. Ozturk, A. Cristal, and O. S. Unsal, "Implications of non-volatile memory as primary storage for database management systems," in *Proc. 2016 International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation*, 2016, pp. 164-171.
- [19] G. G. Foisy, N. E. Larson, S. Narayan, A. F. Sancheti, and J. Edwards, "Western Digital Corp," *Head Stack Assembly for a Disk Drive Having a Unitary Molded Plastic e-Block*, 2000.
- [20] K. Takaishi, Y. Uematsu, T. Yamada, M. Kamimura, M. Fukushi, and Y. Kuroba, "Hard disk drive servo technology for media-level servo track writing," *IEEE Transactions on Magnetics*, vol. 39, no. 2, pp. 851-856, 2003.
- [21] T. R. Simon, L. Cong, Y. Zhai, Y. Zhu, and F. Zhao, "A semi-automatic system for efficient recovery of rare earth permanent magnets from hard disk drives," *Procedia CIRP*, vol. 69, pp. 916-920, 2018.
- [22] A. Al Mamun, G. Guo, and C. Bi, Hard Disk Drive: Mechatronics and Control, CRC press, 2017.
- [23] V. W. Santini and A. D. Little, Western Digital Technologies Inc, "Disk drive including a printed circuit board assembly and a PCBA shield with tabs engaged in slots of a disk drive base," U.S. Patent 7,271,978, 2007.
- [24] N. Samattapapong and N. Afzulpurkar, "A production throughput forecasting system in an automated hard disk drive test operation using GRNN," *Journal of Industrial Engineering and Management*, vol. 9, no. 2, pp. 330-358, 2016.
- [25] S. Sankar, M. Shaw, K. Vaid, and S. Gurumurthi, "Datacenter scale evaluation of the impact of temperature on hard disk drive failures," *ACM Transactions on Storage (TOS)*, vol. 9, no. 2, p. 6, 2013.
- [26] W. Song,A. Ovcharenko, B. Knigge, M. Yang, and F. E. Talke, "Effect of contact conditions during thermo-mechanical contact between a thermal flying height control slider and a disk asperity," *Tribology International*, vol. 55, pp. 100-107, 2012.
- [27] Z. S. Ye, M. Xie, and L. C. Tang, "Reliability evaluation of hard disk drive failures based on counting processes," *Reliability Engineering & System Safety*, vol. 109, pp. 110-118, 2013.
- [28] H. U. Nwosu, C. C. Obieke, and A. J. Ameh, "Failure analysis and shock protection of external hard disk drive," *Nigerian Journal of Technology*, vol. 35, no. 4, pp. 855-865, 2016.
- [29] G. A. Seber and A. J. Lee, *Linear Regression Analysis*, vol. 329, John Wiley & Sons, 2012.
- [30] D. C. Montgomery, E. A. Peck, and G. G. Vining, 2012, Introduction to Linear Regression Analysis, vol. 821, John Wiley & Sons.
- [31] J.Neter, M. H. Kutner, C. J. Nachtsheim, and W.Wasserman, Applied Linear Statistical Models, vol. 4, p. 318, 1996, Chicago: Irwin.
- [32] X.Yao, "Evolving artificial neural networks," in *Proc. the IEEE*, vol. 87, no. 9, pp. 1423-1447, 1999.
- [33] A. K.Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 3, pp. 31-44, 1996.
- [34] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35-62, 1998.
- [35] Y. T. Chae, R.Horesh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy and Buildings*, vol. 111, pp. 184-194, 2016.
- [36] L. Bertinetto, J. F. Henriques, J. Valmadre, P. Torr, and A. Vedaldi, "Learning feed-forward one-shot learners," *Advances in Neural Information Processing Systems*, pp. 523-531, 2016.
- [37] L. Breiman, Classification and Regression Trees, Routledge, 2017.
- [38] J. Lepping, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018.
- [39] Y. Y. Song and L. U. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, p. 130, 2015.
- [40] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79-82, 2005.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted

use, distribution, and reproduction in any medium, provided the original work is properly cited ($\underline{CC BY 4.0}$).



A. Hirunyawanakul is a Ph.D. student in the School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. She received her B.E. and M.E. in computer engineering from Suranaree University of Technology, Thailand, in 2006 and 2014, respectively. Her research of interest includes data mining, machine learning, and artificial intelligence.



K. Kerdprasop is an associate professor at the School of Computer Engineering, chair of the School, and the head of Knowledge Engineering Research Unit, SUT. He received his bachelor degree in mathematics from Srinakarinwirot University, Thailand, in 1986, MS in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes machine elligences

learning and artificial intelligences.



N. Kerdprasop is an associate professor and the head of Data Engineering Research Unit, School of Computer Engineering, Suranaree University of Technology (SUT), Thailand. She received her B.S. in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. Her research of interest includes data mining, artificial

intelligence, logic and constraint programming.