IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 9, 2019, accepted December 27, 2019, date of publication January 22, 2020, date of current version February 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2968615

Wind Turbine Fault Diagnosis and Predictive Maintenance Through Statistical Process Control and Machine Learning

JYH-YIH HSU[1,2], YI-FU WANG[1], KUAN-CHENG LIN[1], MU-YEN CHEN[3], AND JENNEILLE HWAI-YUAN HSU[4]

[1]Department of Management Information Systems, National Chung Hsing University, Taichung City 402, Taiwan
[2]Department of Applied Economics, National Chung Hsing University, Taichung City 402, Taiwan
[3]Department of Information Management, National Taichung University of Science and Technology, Taichung City 404, Taiwan
[4]Independent Researcher, San Francisco, CA 94102, USA

Corresponding author: Jenneille Hwai-Yuan Hsu (jenneillehsu@gmail.com)

This work was supported in part by two grants of Taiwan's Ministry of Science and Technology: (1) The Program for Formulation, Maintenance and Operation of Innovative Business Models Integrating Smart Manufacturing and Information System (Grant No. 106-2420-H-005-003), and (2) Establishing and Testing of a 96-Hour Hourly Solar and Wind Energy Production Forecast System in Taiwan (Grant No. 108-3116-F-005-001).

ABSTRACT This study applies statistical process control and machine learning techniques to diagnose wind turbine faults and predict maintenance needs by analyzing 2.8 million sensor data collected from 31 wind turbines from 2015 to 2017 in Taiwan. Unlike previous studies that only relied on historical wind turbine data, this study analyzed the sensor data with practitioners' insight by incorporating maintenance check list items into the data mining processes. We used Pareto analyses, scatter plots, and the cause and effect diagram to cluster and classify the failure types of wind turbines. In addition, control charts were used to establish a monitoring mechanism to track whether operation data are deviated from the controls (i.e., standard deviations) as a mean to detect wind turbine abnormalities. While statistical process control was applied to fault diagnosis, machine learning algorithms were used to predict maintenance needs of wind turbines. First, the density-based spatial clustering of applications with noise algorithm was used to classify abnormal-state wind turbine data from normal-state data. Then, random forest and decision tree algorithms were employed to construct the predictive models for wind turbine anomalies and tested with K-fold cross-validation. The results indicate a high level of accuracy: 92.68% for the decision tree model, and 91.98% for the random forest model. The study demonstrates that, by data mining and modeling, the failures of wind turbines can be detected, and the maintenance needs of parts can be predicted. Model results may provide technicians early warnings, improve equipment efficient, and decrease system downtime of wind turbine operation.

INDEX TERMS Decision trees, fault diagnosis, machine learning, predictive maintenance, random forest, statistical process control, wind energy.

I. INTRODUCTION

Wind energy is a prevailing, potentially low-cost renewable energy technology that holds a key role in clean energy transition. Thanks to its location, Taiwan is well-endowed with abundant wind energy resources and provides a viable home for utility-scale wind farms. According to the 23 Year Average Wind Speed Observation by 4C Offshore, 16 of the world's top 20 places with the most abundant wind resources are located at the Taiwan Strait (4C Offshore 2018). Particularly, southwesterly airstream in summer and northeasterly monsoon in winter along the coast from Taoyuan to Changhua often create strong wind of scale 4 or higher. As a result, Taiwan offers one of the best places in the world to develop wind energy [3].

Most wind turbines in Taiwan are imported and were built to accommodate local environmental and geographical conditions. Hence, wind turbine malfunctions are usually

The associate editor coordinating the review of this manuscript and approving it for publication was Patrick Hung.

time-consuming and costly to fix as technicians need to contact the foreign seller and request maintenance service [3]. If a set of accurate prediction methods is identified and put in place, equipment anomalies may be detected early and then corresponding actions may be taken timely to decrease the downtown of a wind turbine.

For decades, industrial manufacturers have been striving to improve operational efficiency, maximize production, and reduce downtime of equipment. Thanks to the rapid development of the Industrial Internet of Things (IoTs) over the past decade, numerous manufacturers have installed sensors on production equipment to track real-time system performance, monitor equipment degradation, and predict malfunctions. Such fault diagnosis and predictive maintenance mechanisms improve the availability/uptime of a production line and reduce maintenance costs for producers [4], [8].

Wind turbine developers and operators are no exception to the rapid IoT development and have adopted sensors to monitor the state of wind turbines and generation. Currently, most wind turbines rely on an expert control system which is designed based on human experience and stored knowledge. The system does not predict wind turbine faults, and the system solutions for abnormalities are limited and predetermined. By contrast, predictive maintenance models with a large amount of sensor data have the ability to predict and prevent possible faults and reduce maintenance costs [4].

Taipower, the state-owned utility in Taiwan, has installed numerous sensors on wind turbines to monitor real-time rotational speed, temperature, and voltage of wind turbine parts. This study uses wind turbine sensor data obtained from Taipower and employs statistical process control and machine learning techniques to identify attributes that are useful in detecting wind turbine faults and predicting maintenance needs. The sensor data were collected from 31 wind turbines from January 1, 2015 to December 31, 2017.

II. LITERATURE REVIEW

Numerous studies have been published on the topics of wind turbine fault diagnosis and predictive maintenance. The wind turbine prognostics and health management research has mainly focused on vibration analysis, torque analysis, acoustic emission signal analysis, and temperature analysis. While most studies used either control charts or machine learning techniques to predict wind turbine maintenance needs, this study employs both methodologies. In addition, previous studies usually deployed only one of the seven statistical process control methods for fault diagnosis, and this study uses five statistical process control techniques to analyze wind turbine faults. Another contribution of this study is providing clear definitions of wind turbine anomalies from the practitioners' perspectives. Unlike earlier research only analyzed wind turbine anomalies with historical sensor data, this study not only analyzes sensor data but also integrates practitioner's insight into possible causes for wind turbine faults by examining the check list provided by Taipower.

In a previous study, Yang *et al.* [5] proposed a control chart that was based on the residual-based exponentially weighted moving average (EWMA) to remove autocorrelation issues in datasets. By establishing control charts based on EWMA and multivariate EWMA (MEWMA), the authors compared the model robustness across fault diagnosis models and the failure types identified by data analyses. As wind turbine faults might result from multiple causes and variables, the authors found that a MEWMA control chart was able to detect failure faster than, or at least as fast as, an EWMA control chart.

Wang *et al.* [7] modeled the damage of wind turbine blades and found analytical evidence that irregular cracks occurred on blades prior to breaking. Then, the authors employed Deep Autoencoder (DA) models to predict impending ruptures of a blade by using monitoring data. The DA model is a neural network of multiple hidden layers of a symmetric organization. To train the DA model, Boltzmann Machine was utilized and restricted to initialize weights and deviations. Backpropagation was then used to further optimize the training structure. By validating the monitoring data, the authors found that the trends of reconstruction error (RE) were associated with blade breakage. Next, Wang *et al.* [7] established an EWMA control chart to further detect changes in RE. The data of broken blades that were collected from China's wind farms were used to validate the models and showed support to the proposed method. The results demonstrated that the monitoring method proposed by Wang *et al.* [7] effectively identify impending wind turbine blade damage.

Liu *et al.* [13] used K-means and artificial neural network classification to predict failure of wind turbine gear box systems. First, after data pre-processing, K-means clustering was used to cluster data with similar characteristics into multiple clusters. The results of data clustering were then predicted by the artificial neural network classification. Compared with the traditional neural network classification, this method improved the accuracy of prediction by 3.5%. With the proposed model, mechanical failure could be detected and determined more accurately and timely [13].

Yampikulsakul *et al.* [9] found that wind turbine operation was significantly affected by weather conditions. Their study demonstrates that, by monitoring wind turbines' performance, the operation and maintenance costs can be reduced. Yampikulsakul *et al.* [9] employed least squares support vector regression to model climatic factors and turbine operation parameters to predict wind turbine faults. The authors examined the operational variations that were affected by weather conditions to determine the decision boundary of the model. The decision boundary was also set as the central baseline for the subsequent analyses of control charts. The baseline was used to identify the conditions that lead to wind turbine abnormalities.

Zhang *et al.* [15] proposed seven machine learning algorithms: 1. Neural Network, 2. Neural Network Ensemble, 3. Support Vector Machine, 4. Boosting Tree, 5. Random Forest, 6. K-Nearest Neighbour, and 7. Classification and Regression

Tree for fault diagnosis. The model inputs include the total amount of wind energy generated, motor rotational speed, torque, and temperature of wind turbine parts. Then, mean absolute percent error was used to measure prediction errors. As a result, neural network and neural network ensemble models reported the lowest error (0.01), while the random forest model was the second fitting algorithm with a reported error of 0.06.

Wang *et al.* [11] proposed the Multivariate Time Series Processing Method based on Riemannian Manifold. Through statistical process control charts, the wind turbines' multivariate time-series data of abnormal features were tested. Next, the authors visualized the data of equipment anomalies through covariance matrix distribution for fault diagnosis. This research focuses on the gearbox failure of wind turbines, and its troubleshooting method was validated and found effective.

Kusiak and Verma [2] proposed a three-stage process to predict wind turbine faults: predict any type of faults, predict system-specific faults, and then identify potential faults that have not yet occurred. After data pre-processing, five machine learning algorithms (1. Neural Network, 2. Support Vector Machine, 3. Random Forest, 4. Boosting Tree, and 5. General Chi-square Automatic Interaction Detector) were used to predict the values of four events (1. Turbine OK, 2. Fault, 3. Weather Downtime, 4. Maintenance Downtime). For the final outcomes, random forest reported the best result with a 98% of successful categorization rate.

Marquez *et al.* [4] proposed 11 methods to predict wind turbine faults (1. Vibration, 2. Acoustic Emission, 3. Ultrasonic Techniques, 4. Oil Analysis, 5. Strain, 6. Electrical Effects, 7. Shock Pulse Methods, 8. Process Parameters, 9. Performance Monitoring, 10. Radiographic Inspections, and 11. Thermography). By monitoring wind turbine parts and systems, operators can identify possible wind turbine faults and take actions before abnormal conditions occur. In addition, Marquez *et al.* [4] noted that thermography were often used for monitoring operation condition of wind turbines. As the First Law of Thermodynamics suggests, an increase in temperature is associated with a decrease in energy, implying that malfunctioning parts that create friction and heat are likely to reduce wind energy outputs.

Hameed *et al.* [14] reviewed different methodologies and algorithms developed to monitor the performance of wind turbines and found that the use of condition monitoring systems and fault detection systems are crucial to maintaining wind turbine health.

In summary, there are two bodies of literature related to predictive maintenance of wind turbines: first, monitoring real-time wind turbine operation with post-event analyses based on the control charts of statistical process control; and second, using statistical analyses and machine learning algorithms to predict equipment abnormalities based on historical records. In the second body of the literature, the common algorithms that were used include neural network, support vector machine, random forests, and decision trees. The equipment parts that were commonly analyzed include fan blades, gearbox, turbine, and generator.

TABLE 1. Sensor data that were collected by Taipower.

Data	Unit
Total wind energy generated	KWH
Total hours when wind speed was in range	Hr
Rotor revolutions per minute	Rpm
Total hours of maintenance	Hr
Total hours of yaw control operation	Hr
Hydraulic oil pressure	Bar
Gear bearing temperature	°C
Transmission frequency	Hz
Gear oil temperature	°C
Bus bar temperature	°C
Hydraulic oil temperature	°C
Axle top temperature	°C
Wind turbine seals temperature	°C
Nacelle temperature	°C
Impeller axle temperature	°C
Ambient temperature	°C
Spinner temperature	°C
Wind speed	m/s
Rotor rotational speed	rpm
Direction of the nacelle	°
Blade pitch angle	°

III. METHODOLOGY

A. DATA ANALYSIS

Provided by Taipower, the data were collected from 31 wind turbines in the Changhua Binhai Industrial Zone Phase 1 (23 units) and Phase 2 (8 units) from January 1, 2015 to December 31, 2017. Data were recorded with 21 attributes every 10 minutes. Each wind turbine unit had 52,561 observations, and 31 units totaled 2,815,104 observations for the analyses.

The sensor data that were monitored and collected by Taipower are listed in Table 1.

1) DATA PRE-PROCESSING

Wind turbines often operate intermittently, given the nature of wind not blowing continuously. To tackle this analytical challenge, we first identified the suitable data interval for this study. We found that the temperature change of generation equipment and the generator rotational speed only partially correlates with the amount of wind energy generated. While the amount of wind generation changes immediately when wind speed and generator rotational speed change, the change in the amount of wind generation takes a longer time to respond to temperature change. The slow reaction of wind generation to temperature change leads to the outcome that a single temperature value corresponds to multiple speed values and the amounts of wind generation in the raw dataset. To resolve this analytical issue, the study first attempted to average all attribute variables by various hour intervals (including 1 hours, 2 hours, ..., to 12 hours), but the outcomes were not satisfying. Then, the authors found the daily average of variables was the most robust way to eliminate the imbalanced data issue. As a result, this study set "day" as the data internal and pre-processed the recorded wind turbine data accordingly.

2) DATA CLEANSING

Some of the wind turbine data were lost in the events of sensor malfunction, and no backup data were available at Taipower. As a result, we performed data cleansing to remove blank fields in the original data. To do so, we first selected the attributes that are highly correlated with the amount of wind energy generated. Taking the CK20 wind turbine unit for example, a total of 21 attributes were selected (including the temperature of high-speed bearings, the temperature of wind turbine gearbox lubricant, and generator speed), and blank fields were removed. Specifically, data cleansing was proceeded with 5 steps to maximize the number of observations retained in the dataset. After each step, the correlation coefficients of each field were re-examined, and the correlation scatter plots were compared to ensure the validity of the data cleansing.

(1) We cleared the character strings of all fields to avoid computer recognition error and set them as null.
(2) After removing the data with incomplete time points, we used the remaining data with complete points of time to establish a correlation coefficient table.
(3) We selected variables that have high correlation coefficients (greater than 0.5 or less than -0.5) and classified them into four groups. In addition, only wind turbine attributes that correlate the amount of wind generation were retained for further analyses.
(4) We used scatter plots to visualize the correlation between each attribute and the amount of wind energy generated. We also validated the correlations of attributes with wind generation based on the advice given by Taipower technicians in charge of wind turbine maintenance.
(5) After extracting highly-correlated wind turbine attributes with the amount of wind generation, we deleted all incomplete data points and created a new, complete dataset.

B. PRELIMINARY ANALYSIS OF ABNORMALITIES

By comparing the correlation coefficient table and scatter plots of different wind turbines, we found that the correlation coefficients of some wind turbines were significantly lower than those of other wind turbines. To explore the abnormal correlation coefficients, we first calculated the total amount of wind energy generated by each wind turbine and created a baseline by averaging the correlation coefficients of 30 wind turbines to identify the correlation coefficients that were below the average. Lastly, this study examined the scatter plots of each marked wind turbine attribute and compared it with the scatter plot of each unmarked attribute. The results of abnormalities analyses will be discussed in IV. Empirical Analysis.

C. ANALYSIS OF WIND TURBINE DATA THROUGH STATISTICAL PROCESS CONTROL

This study conducted fault diagnosis based on the wind turbine data provided by Taipower through five analytical tools in statistical process control: 1. check list, 2. Pareto charts, 3. cause and effect diagram, 4. scatter plots, and 5. control charts. First, this study examined Taipower's check list for wind turbine maintenance. The check list includes information such as the type of wind turbine faults, the duration of faults, causes, and repair events. We ranked wind turbine repair events by frequency of anomalies in the dataset. Second, a Pareto chart was created based on the ranked check list items and displayed the repair events in terms of cumulative percentage. Events with higher occurrence percentage are considered the most frequent causes for wind turbine faults and should be checked first by technicians in the occurrence of abnormal states, while events with lower occurrence percentage are secondary contributors to equipment faults and may be maintained in a mid- or long-term time horizon.

The third analytical tool is the cause and effect diagram. Wind turbine anomalies are placed on the right side of the diagram, and the four primary classification of the check list items (rotating blade anomalies, gearbox anomalies, generator anomalies, and hydraulic oil system anomalies) are placed on the left side of the diagram. Under each primary classification of anomalies, secondary classification of anomalies are present, including high temperature, insufficient gear oil, and abnormal speed. A cause and effect diagram helps identify the underlying causes for major mechanical problems and make recommendations to technicians for maintenance.

Forth, scatter plots were used to explore the relationship between attributes and identify data that are abnormal. Lastly, control charts were used and offer the strength to show changes and variation in observed data over time.

This study analyzed the continuous, time-series data of a single wind turbine to examine the changes of standard deviation and whether the wind turbine performance data met regulatory standards. This study also identified abnormal variation in the control chart and analyzed the causes of the unusual variations.

D. CLASSIFYING ABNORMAL AND NORMAL DATA

The study used the density-based spatial clustering of applications with noise (DBSCAN) algorithm to illustrate the relationship between the total amount of wind generation and five attributes as well as classify normal and abnormal data which were first clustered through pattern geometric correlation. Two parameters were used in the DBSCAN algorithm: ε (eps) and minPts, namely the radius of neighborhood around a point and the minimum number of data points to be included to form a dense region. The DBSCAN clustering starts with an arbitrary starting point. The starting point is identified as a core point if its radius of ε contains at least minPts number of points, and then a cluster is formed. Points beyond the distance of ε are labeled as noise. The steps repeat until all points have been assigned to a cluster or labeled as visited.

The authors determined the values of ε and minPts based on the results of scatter plots.

E. PREDICTIVE MAINTENANCE MODELS

Two prediction models were built through random forest and decision tree algorithms and both validated by k-fold cross-validation five times. The random forest algorithm was favored by Breiman [6] over the decision tree algorithm as random forest is an ensemble of multiple decision trees. By consisting of multiple weak learning tools, the random forest constructs a strong learning tool which limits overfitting.

On the other hand, while the decision tree algorithm may overfit and be less accurate for prediction, it performs binary splits in the classification process until the final outcomes were yielded. The decision tree offers the advantage of explaining and identifying the features that predict wind turbine maintenance need. The authors created a visualized decision tree to select features that contribute to the prediction outputs and avoid overfitting. By incorporating the decision tree algorithm with feature selection, we eliminated two inferior attributes (rotor speed and wind generation) and retain four influential attributes as well as mitigate the overfitting problem for the predictive models.

In this study, the random forest algorithm was implemented under Python environment. The parameters setting is discussed in the follows. Entropy was used to measure the quality of a split for the information gain. The minimum number of samples required to be at a leaf node was 1, the minimum number of samples required to split an internal node was 2, and the number of the trees in the forest was 10. The number of jobs to run in parallel for both fit and predict was 2.

IV. EMPIRICAL ANALYSIS

A. DATA PRE-PROCESSING

1) DATA CLEANSING

The raw data of wind turbines contain the device number (i.e., Chang Kong, CK) and a string of status number. We first converted the strings into null values. In addition, we selected attributes that are highly correlated with the amount of wind energy generated and removed insignificant attributes from the empirical analyses. The method was described step by step in the following.

First, we calculated the average correlation coefficients of the attributes of the 31 wind turbine units. Then, we only retained 9 attributes that have a correlation coefficient greater than 0.5 with the amount of wind energy generated. The 9 attributes are: 1. generator speed, 2. high-speed bearing temperature, 3. gearbox lubricant temperature, 4. wind speed, 5. wind turbine rotor speed, 6. impeller shaft temperature, 7. ambient temperature, 8. nose tip temperature, and 9. transmission busbar temperature. The splitting ratio for training and testing dataset was 70:30.

Next, scatter plots were created to illustrate the relationship between the amount of wind generation and each selected attribute as well as the correlation coefficients between every two attributes. Among the 9 attributes that are highly correlated with the amount of wind energy generated (>0.5), the scatter plots of some attributes fail to show a clear pattern and are unable to provide evidence that these attributes are highly correlated with the amount of wind energy generated. Therefore, we removed these attributes from the analyses, which were 1. impeller shaft temperature, 2. ambient temperature, 3. nose tip temperature, and 4. transmission busbar temperature.

As a result, the study finally retained five attributes that are highly correlated with the amount of wind energy generated for further analyses: 1. wind turbine generator speed (0.79), 2. high-speed bearing temperature (0.79), 3. gearbox lubricant temperature (0.72), 4. wind speed (0.92), and 5. wind turbine rotor speed (0.80).

Finally, this study extracted the raw data of the five selected attributes and the amount of wind energy generated for further analyses. The authors removed the observations with a null value in the time field and generated the daily average of these data for the reason specified in the "Data Pre-Processing" section.

2) RESULTS OF DATA CLEANSING

After data cleansing, the study compiled the remaining attribute data of 31 wind turbine units. As Table 2 shows, No. CK31 only had 1,905 observations left after the data cleansing, accounting for 1.21% of the total observations (157,438), far less than the average number of observations of other 30 wind turbines. Therefore, the data of No. CK31 were removed from the analyses due to validity concerns.

In addition, CK24-30 (7 units) were built in the Phase 2 of Changhua Binhai Industrial Zone. The data of "Wind Turbine Generator Speed" from the 0:00 a.m. on January 1, 2015 to 4: 40 p.m. on March 16, 2015 were lost, a total of 3 months and 16 days. Despite the incomplete data, the data of CK24-30 were retained to improve the statistical robustness of the correlation estimates between each attribute and the amount of wind generation.

B. PRELIMINARY ANALYSIS OF ABNORMALITIES

First, the study calculated the average of the correlation coefficients of 30 wind turbines and set them as the baselines for the abnormality analysis. Then, the authors compared the correlation coefficients and the scatter plots to identify abnormalities of wind turbine attributes. In particular, the results show that the correlation coefficients of wind turbines CK04, CK05, CK10, CK19, CK20, CK23, CK24, and CK27-30 were abnormal and significantly lower than the average correlation coefficients. The attributes that were found abnormal include: 1. generator speed, 2. high-speed bearing temperature, 3. gearbox lubricant temperature, and 4. rotor speed.

TABLE 2. Results of data cleansing (complied by the authors).

Proportion of Residual to Total Data after Data Cleansing				
No.	CK01	CK02	CK03	CK04
Percentage	40.65%	65.76%	68.20%	66.86%
Quantity	64,102	103,688	107,534	105,420
No.	CK05	CK06	CK07	CK08
Percentage	68.99%	69.18%	70.00%	71.74%
Quantity	108,777	109,080	110,378	113,121
No.	CK09	CK10	CK11	CK12
Percentage	70.65%	69.14%	68.18%	70.72%
Quantity	111,396	109,027	107,499	111,516
No.	CK13	CK14	CK15	CK16
Percentage	68.53%	64.05%	60.17%	68.18%
Quantity	108,064	100,987	94,880	107,506
No.	CK17	CK18	CK19	CK20
Percentage	68.21%	67.53%	60.77%	68.98%
Quantity	107,547	106,480	95,820	108,773
No.	CK21	CK22	CK23	CK24
Percentage	68.03%	70.72%	69.25%	35.96%
Quantity	107,269	111,517	109,192	56,707
No.	CK25	CK26	CK27	CK28
Percentage	35.44%	35.78%	35.68%	35.92%
Quantity	55,886	56,414	56,260	56,645
No.	CK29	CK30	CK31	
Percentage	34.66%	36.19%	1.21%	
Quantity	54,646	57,068	1,905	

C. STATISTICAL PROCESS CONTORL

Based on the monitoring data of wind turbines and the repair records from 2015 to 2017 provided by Taipower, the study investigates the causes of wind turbine equipment abnormalities through five analytical tools in statistical process control: check list, Pareto chart, the cause and effect diagram, scatter plots, and control charts. The results of scatter plots were already discussed in Preliminary Analysis of Abnormalities. The analysis results of the check list, Pareto chart, and the cause and effect diagram are presented in tables and figures in this section. The control charts presented in this section only list the analysis results of No. CK20, as its data had the most exceptional distributions among 31 wind turbine units. The analyses of other 30 units were not presented in the study, as the methodology is identical and the results would be repetitive.

1) CHECK LIST CONSOLIDATION

Taipower has a standardized check list for wind turbine repairs. The check list includes 1. repair form number, 2. unit number, 3. rated capacity (KW), 4. the start time of a downtime, 5. the end time of a downtime, 6. the length of a downtime, 7. alert number, 8. alert message, 9. maintenance process, and 10. the status of maintenance.

From 2015 to 2017, the 30 wind turbines had a total of 88 abnormal conditions and 976 repairs. We classified the abnormal alerts and counted the frequency of each classification

TABLE 3. Check list items by frequency (estimated by the authors).

Number	Abnormal Condition	Frequency (times)
1	154 Max rotor RPM	110
2	V-216 Oil leakage in Hub	95
3	917 High temp. Gen bear. DE:96℃	76
4	111 No offline filtrat	74
5	159 External RPM guard	51
6	353 Q8 breaker open	33
7	187 External 24 V power supply	32
8	147 High gear temperature	31
9	EMCV. Pitch min:-5.0°max:5.3°	29
10	165 Low oil-level, hydraulic	26
11	Pitch dev. min:-0.5° max:4.8°	26
12	80 Low gear oil pressure	25
13	163 Low working pressure	22
14	112 Max DP filt	21
15	Thermo error nacelle fan F307	16
16	Feedback = 0, Brake	15
17	Thermo error Int.Gen.Fan F515A	14
18	Low oil pressure:1472RPM/42°	13
19	Q7 breaker open	13
20	ExtHighIRotorInv phase: 0	11
21	158 Rotor:__._RPM, Gen.:____._ RPM	10
22	Ext. High cur. Grid inv. L0	10
23	149 High temperature T53	9
24	311 Trip Q8 Feedback error K 0,0	9
25	189 Feedback=0,Brake	8
26	312 Thermo error,ventilators T53	8
27	315 ExEx low voltage L1: 0V	8
28	327 Grid inv.HW error Lx	8
29	High temp.Gear bearing 1:91°C	7
30	Others	166

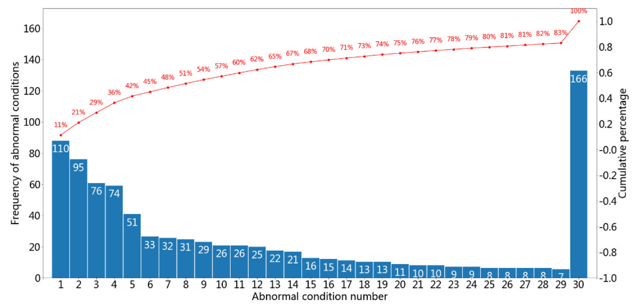

FIGURE 1. The Pareto chart of abnormal frequency statistics (estimated by the authors).

of abnormal alerts. Table 3 lists the 29 most common types of abnormal conditions and consolidated less frequent conditions (less than 7 occurrences) in the 30th item "Others".

2) PARETO ANALYSIS

Based on the frequency of abnormal conditions in Table 3, the Pareto chart was created with the x-axis indicating the abnormal condition number (type 1 to 30 in Table 3). As Fig. 1 displays, the top five abnormal conditions account for 42% of the total occurrence of abnormal conditions, much higher than the rest of 83 conditions. The top eight abnormal conditions account for 51% of all abnormal conditions, and the top 16 abnormal conditions account for 70%. The first-ranked abnormal condition by frequency is "154 Max rotor RPM", accounting for 11%.

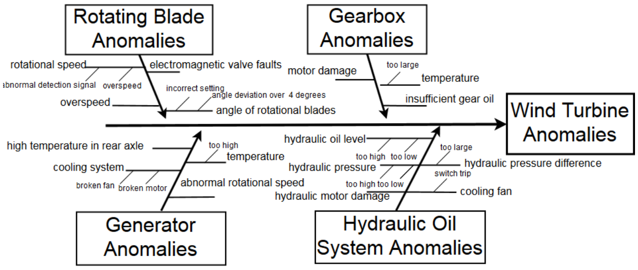

FIGURE 2. Analysis of the causes of wind turbine faults.

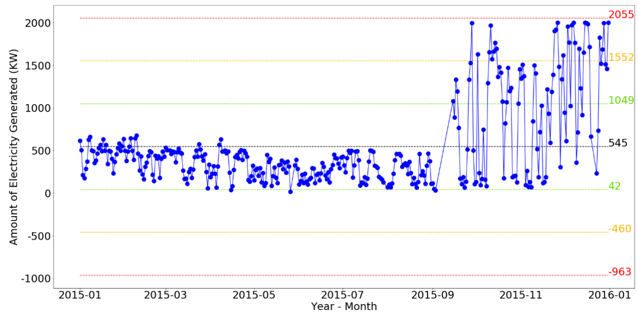

FIGURE 3. The control chart of amount of wind energy generated by No. CK20.

Based on the results of the Pareto analysis, improvement plans with short-, mid-, and long-term goals could be created by Taipower. First, the improvement priority of fault prevention should be given to the first eight abnormal conditions – more than 50% of wind turbine anomalies may be avoided. For the mid-range goal, by improving the number 9 to 16 items, wind turbine faults may be reduced by 70%. Finally, the long-range improvement goal is to eliminate the subsequent abnormal conditions and further advance the operation efficiency of wind turbine equipment.

3) CAUSE AND EFFECT DIAGRAM ANALYSIS

Based on the results of Taipower check list consolidation in Table 3, the author analyzed the locations of faults, the types of equipment faults, and the values of abnormal monitoring conditions and then classified the causes for wind turbine anomalies into four types: rotating blade anomalies, gearbox anomalies, generator anomalies, and hydraulic oil system anomalies. We also referred to the wind turbine prognostics and health management literature and consulted with Taipower experts for appropriate classifications. Under these four major types of anomalies, a total of 30 secondary causes were also classified in Fig. 2.

By analyzing the four major classifications of anomalies, we found that, first, high gearbox temperature was a common cause for maintenance needs. Another frequent cause for equipment faults was a worn-out gear set or gear oil running out. In both scenarios, the wind turbine had to stop operating and wait for cooling down or adding gear oil. The second type of common equipment anomaly that caused wind turbine faults was rotating blades, the part that receives the strongest wind in operation. High wind speed wears and tears electromagnetic valves and may lead to an angle deviation of a rotating blade. The speed of rotor blades can be limited by installing a deceleration valve. Third, the malfunction of a hydraulic oil pressure system, including gearbox and motor steering components, was also found to be one of the major causes for wind turbine faults. A healthy hydraulic oil pressure system can reduce the maintenance needs for wind turbines by keeping hydraulic pressure and oil pressure in the normal range. Generator anomaly is the fourth major cause for wind turbine faults. We found that high temperature of a generator is a frequent cause for wind turbine faults, as high temperature can easily damage the critical mechanic components in a generator, such as gearbox, and low-speed and high-speed bearings. According to the check list records, engineers would stop a faulted generator from operating, wait for equipment temperature to drop, and then restart the generator. However, the standard troubleshooting process significantly reduces the amount of wind generation and would benefit from an improved maintenance process.

Based on the result of the cause and effect diagram analysis, Taipower may install more sensors on the mechanic components of wind turbines to monitor wind speed, module temperature, motor revolution speed, and hydraulic oil pressure system. New sensor data may provide insight and help Taipower design improved maintenance practices that effectively reduce the occurrence of wind turbine anomalies.

4) CONTROL CHART ANALYSIS

In this study, control chart analyses were used to establish baselines that can be used to monitor wind turbine performance and identify potential faults. After analyzing all the historical data of 30 wind turbines, we found that CK20 was a single abnormal case worthy of further analyses in this section. No similar abnormal conditions were found in other 29 wind turbines.

Here, we only present the analysis results of the wind turbine No.CK20 and display the control charts of its high-speed bearing temperature and the amount of wind energy generated. Although six attributes were analyzed through control charts as they all correlate with the amount of wind generation, the results of control chart analyses for other attributes and other wind turbines are not presented for brevity.

Fig. 3 shows that, from January to September in 2015, the high range of wind generation was approximately 500 KW. During the period, the amount of electricity generated fluctuated within one standard deviation from the mean. However, since October, the high range of the generation outputs increased to about 2,000 KW, and the generation outputs fluctuated greatly within three standard deviations from the mean. During the same period (September 2015 and after), the high-speed bearing temperature of CK20 showed an inverse relationship with the wind turbine's generation outputs in Fig. 4. The amount of wind generation was relatively stable in the first three quarters of the year, while the high-speeding bearing temperature also remained relatively

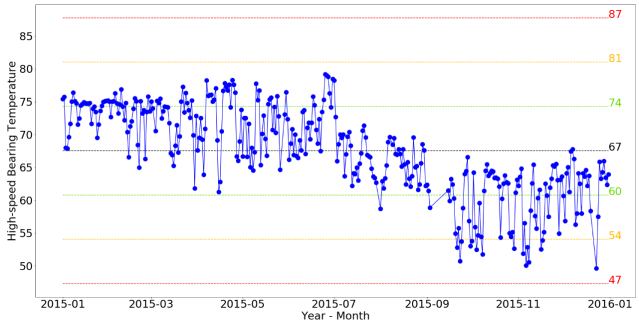

FIGURE 4. The control chart of high-speed end bearing temperature in No. CK20 in 2015.

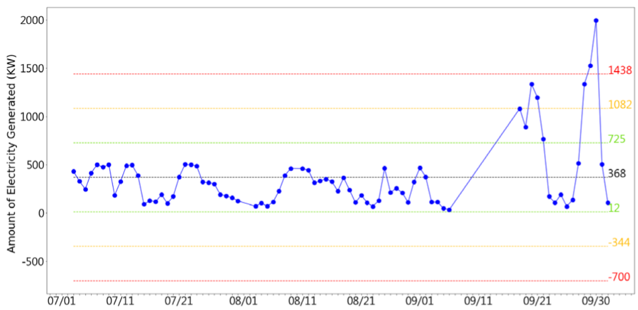

FIGURE 5. The control chart of amount of wind energy generated by No. CK20 (July to September in 2015).

stable, between 65 and 70°C from January to July, within a width of two standard deviations from the mean. However, the high-speed bearing temperature dropped sharply from July to September, and since October, the temperature further dropped below 65°C and fluctuated greatly in the width of three standard deviations from the mean.

In addition, we found that wind turbine data were missing in the period of September 4 to September 17 for CK20. As the repair record indicates, it was because that maintenance was taking place due to the abnormally high temperature of the nacelle. After the maintenance completed on September 17, the amount of wind generation instantly increased and the high-speed bearing temperature reduced significantly, compared to the status prior to maintenance. We presumed that the engineers had adjusted the generation equipment which led to great improvement in generation efficiency.

Fig. 5 shows that two data points of the amount of electricity generated exceeded three standard deviations from the mean and four data points fell between two and three standard deviations from the mean in September 2015. These extreme-value data points might suggest maintenance needs and that a professional technician could visit and check the wind turbine unit to prevent faults. Another abnormality shown in the figure was the significant difference between the trend before September 4 and the trend after September 17. During the same periods, the control chart of high-speed bearing temperature and the amount of electricity generated also coincides an inverse (i.e. downward) trend in Fig. 4.

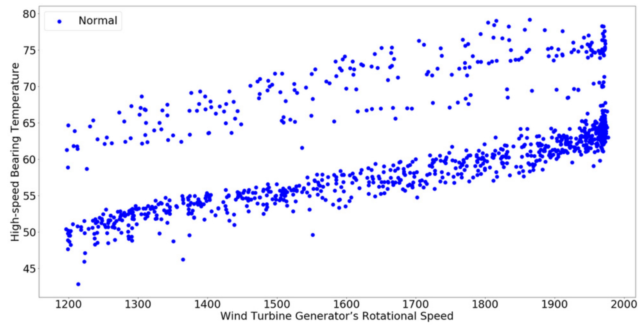

FIGURE 6. DBSCAN high-speed bearing temperature and generator rotational speed of No. CK20.

D. DATA CLUSTERING OF ABNORMAL AND NORMAL STATES

1) DBSCAN CLUSTERING

To allow result comparison, the same wind turbine unit, No. CK 20, was used in DBSCAN clustering as it was in ''Control Chart Analysis''. The scatter plot of high-speed bearing temperature and wind turbine generator speed is shown in Fig. 6, illustrating two clusters of the historical data. We then compared the scatter plot outcomes with Taipower's check list to distinguish the abnormal and normal states of wind turbines. In addition, according to the Taipower staff in charge of wind turbines, ''the normal temperature of an operating high-speed bearing must be lower than 60 to 65°C; therefore, the data points that had high-speed bearing temperature higher than 60 to 65°C and/or the total amount of wind energy generated between 0 and 750 kWh were considered abnormal.''

These criteria were applied to the historical data of No. CK20, and the results are presented in Fig. 6 – the upper cluster of data was supposed to indicate abnormal conditions, and the lower cluster of data was supposed to indicate normal conditions. To verify the result, we sorted the abnormal-state data by time and found that all the abnormal conditions occurred in the same time period. Additionally, the occurrence of abnormal states of wind turbines were also found to correlate with the presence of high-speed bearing temperature. This provides evidence that high temperature of a high-speed bearing was the main factor for reduced generation. Therefore, in order to separate the abnormal-state data from the normal-state data, we optimized the DBSCAN clustering results with two parameters automatically tested: the number of core points and the value of epsilon (eps). We then had experts from Taipower verified the classification results for accuracy.

It was shown in Fig. 6 that the data points of high generator rotational speed (1,950 to 2,000 rpm) were highly concentrated and were not successfully clustered by DBSCAN. To solve this problem, we further clustered the data points in two stages.

First, the observations with a high generator rotational speed between 1,950 and 2,000 rpm were removed from the clustering outcome. With the number of core points set

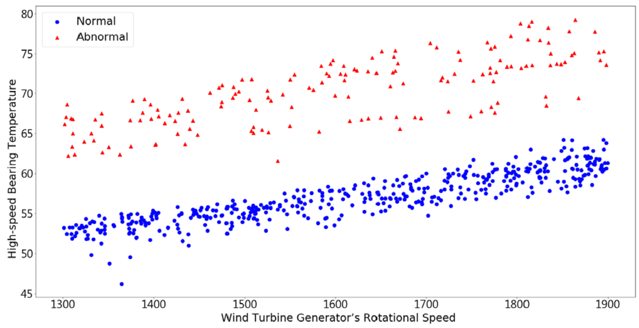

FIGURE 7. **DBSCAN clustering result of high-speed bearing temperature and generator rotational speed of No. CK20 – Stage 1.**

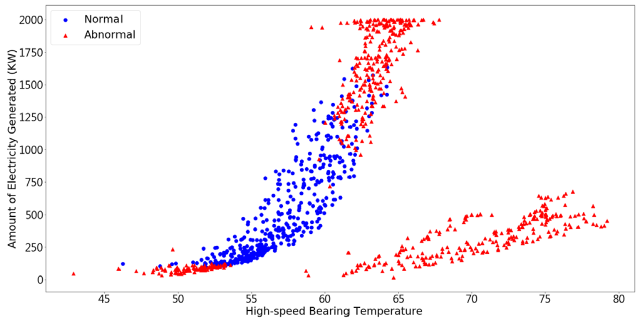

FIGURE 8. **DBSCAN clustering result of total amount of electricity generated high-speed bearing temperature of No. CK20 – Stage 1.**

as 5 and the eps set as 0.5, the abnormal- and normal-state data were successfully clustered as displayed in Fig. 7.

Fig. 8 shows the clustering result of Stage 1 and the relationship of total wind energy generated and high-speed bearing temperature. The data were successfully clustered by DBSCAN and labeled as "normal" and "abnormal". In most normal states (the left-handed cluster in Fig. 8), the high-speed bearing temperature fell between 20 to 65°, and total amount of wind energy generated ranged from 0 to 2,000 KW. After testing sets of parameters, we found that with the core points set as 7 and the eps set as 0.3, and then the data were clustered with statistical significance in the second stage. In Fig. 8, some of the classified normal data points in the left cluster are currently marked as abnormal data. Therefore, the classification result of the second stage were used to correct the unfitting classification of the data points.

In the second stage, the unclassified observations in the first stage were analyzed by using the scatter plot of the amount of wind generation and high-speed bearing temperature as references. We then used the AND logic gate to correct Stage 1 classification results with Stage 2 classification results and finally integrate the results of two stages. The corrected classification results are presented in Fig. 9, total amount of electricity generated and high-speed bearing temperature, and Fig. 10, high-speed bearing temperature and wind turbine generator speed. With the two-stage classification method, we were able to successfully classify the data of abnormal and normal states.

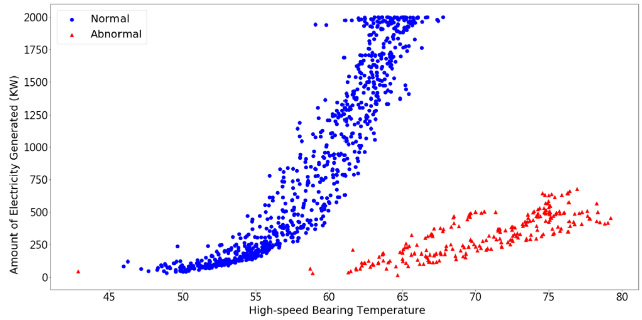

FIGURE 9. **DBSCAN classification result of total amount of electricity generated high-speed bearing temperature of No. CK20 – Stage 2.**

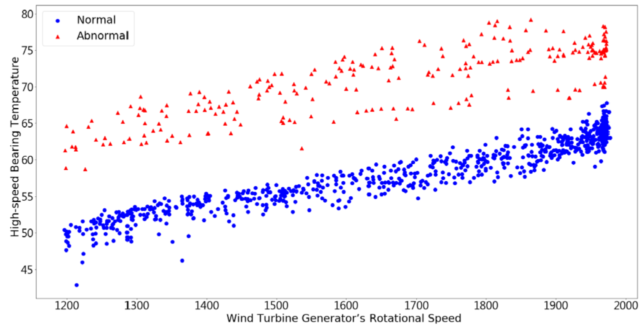

FIGURE 10. **DBSCAN classification result of high-speed bearing temperature and wind turbine generator speed of No. CK20 – Stage 2.**

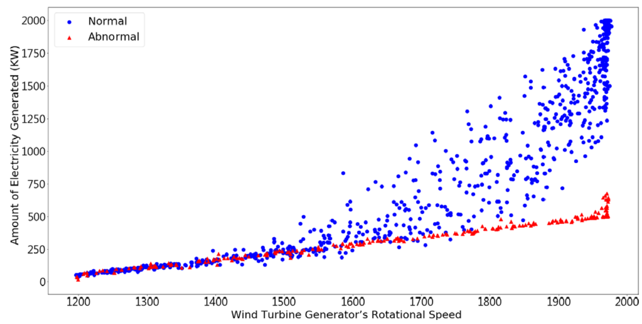

FIGURE 11. **DBSCAN classification result of total amount of electricity generated and wind turbine generator speed of No. CK20.**

2) DBSCAN CLASSIFICATION RESULTS

After the two-stage classification, we examined the scatter plots of total amount of wind energy generated with four attributes: (1) generator speed, (2) rotor speed, (3) gearbox lubricant temperature, and (4) wind speed in Fig. 11-14. Due to the great amount of overlap between the abnormal data and the normal data, the abnormal and normal conditions could not be differentiated effectively prior to the classification. After successful classification, we were then able to identify clear distinct trends of the abnormal- and normal-state data of wind turbines – the normal-state data trend showed an exponential distribution, while the abnormal-state data had a linear relationship with the amount of wind generation and a significantly lower amount of electricity generation compared to normal states.

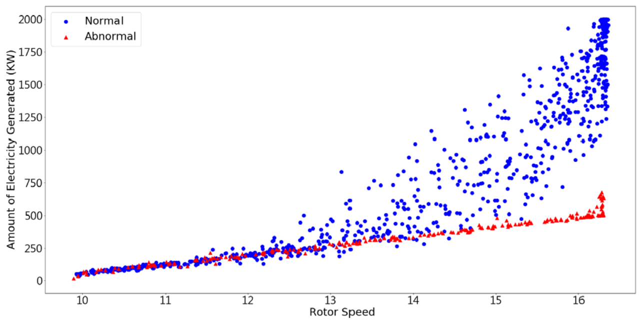

FIGURE 12. DBSCAN classification result of total amount of electricity generated and wind turbine rotor speed of No. CK20.

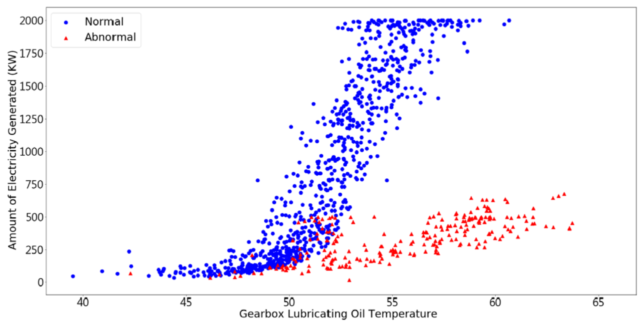

FIGURE 13. DBSCAN classification result of total amount of electricity generated and wind turbine gearbox lubricant temperature of No. CK20.

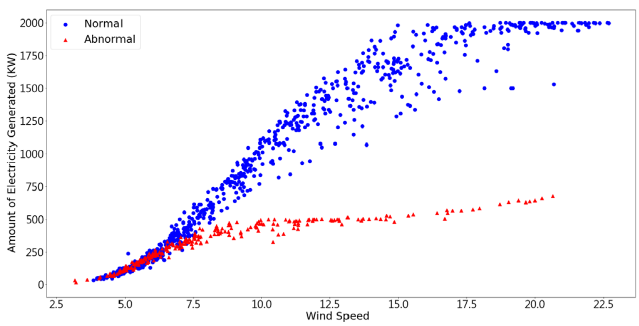

FIGURE 14. DBSCAN classification result of total amount of electricity generated and wind speed of No. CK20.

Based on the classification results, we found the factors that lead to wind turbine abnormalities are closely associated with components' temperature and rotational speed. Holding the amount of electricity generated and generator speed constant, the maximum temperature of the high-speed bearing in abnormal states was found approximately 7°C higher than it was in normal states. Additionally, the maximum temperature of gearbox lubricant was found 6°C higher in abnormal states than it was in normal states.

Based on the above classification results, we found that the high-speed bearing temperature of a wind turbine is a particularly useful attribute to classify abnormal- and normal-state data. In the DBSCAN classification scatter plots of the amount of electricity generated and generator rotational speed (Fig. 9 and 10), the cluster of abnormal-state data are clearly distinguished from the cluster of normal-state data. On the other hand, gearbox lubricant temperature was found to be a less useful attribute in identifying abnormal-state data from normal-state data in DBSCAN classification. Although the abnormal data cluster and the normal data cluster are somewhat distinguished in Fig. 13, nearly half of the abnormal and normal data clusters overlap. Therefore, gearbox lubricant temperature is a less ideal attribute for data classification.

TABLE 4. The training result of decision trees.

First training accuracy rate	92.86%		First test accuracy rate	92.62%	
Five cross-validation accuracy rates					
1st	2nd	3rd	4th	5th	average
92.06%	92.51%	92.79%	93.06%	92.99%	92.68%

The other attributes, including rotational speed (generator speed and rotor speed), environmental factors (wind speed), and output (total amount of wind generation) were not found useful for the classification of wind turbine abnormalities.

E. CONSTRUCTING ANOMALY PREDICTIVE MODELS

In this section, the Scikit-Learn package's CART decision tree and a random forest algorithm were used to build the anomaly prediction models[1]. As described earlier, we constructed two prediction models through random forest and decision tree algorithms. We used the classification results in the section of "Abnormal and Normal State Data Classification" to train the models. There were 27,596 wind turbine observations collected from 30 wind turbines, consisting of 22,077 normal-state observations and 5,519 abnormal-state observations. Seventy percent of the sample was used for model training (19,318 observations), while 30% was used for model testing (8,278 observations). The accuracy of single model training results fell between 90% and 95%, indicating that sampling errors might exist. Hence, we further validated the predictive results with K-fold cross-validation.

In the decision tree prediction model, the training model had an accuracy rate of 92.86%, and the testing model had an accuracy rate of 92.62%. There was little difference between the accuracy rates of the two models. Then, after validating the models with K cross-validation five times, the results in Table 4 show that the lowest accuracy rate was 92.06%, the highest accuracy rate was 93.06%, and the average accuracy rate was 92.68%.

Fig. 15 illustrates the decision tree classifications, starting from the root as high-speed bearing temperature to the third-level nodes. In the decision tree analyses, high-speed bearing temperature was used as a classifier three times: 62.986°C, 56.894°C, and 54.755°C. Gearbox lubricant temperature was

[1]The parameters used in the decision tree were criterion = 'entropy', max_depth = 3, random_state = 0. The parameter used by the random forest is criterion = 'entropy', Min_samples_leaf = 1, min_samples_split = 2, n_estimators = 10, n_job = 2, oob_score = False, random_state = 1, warm_start = False.

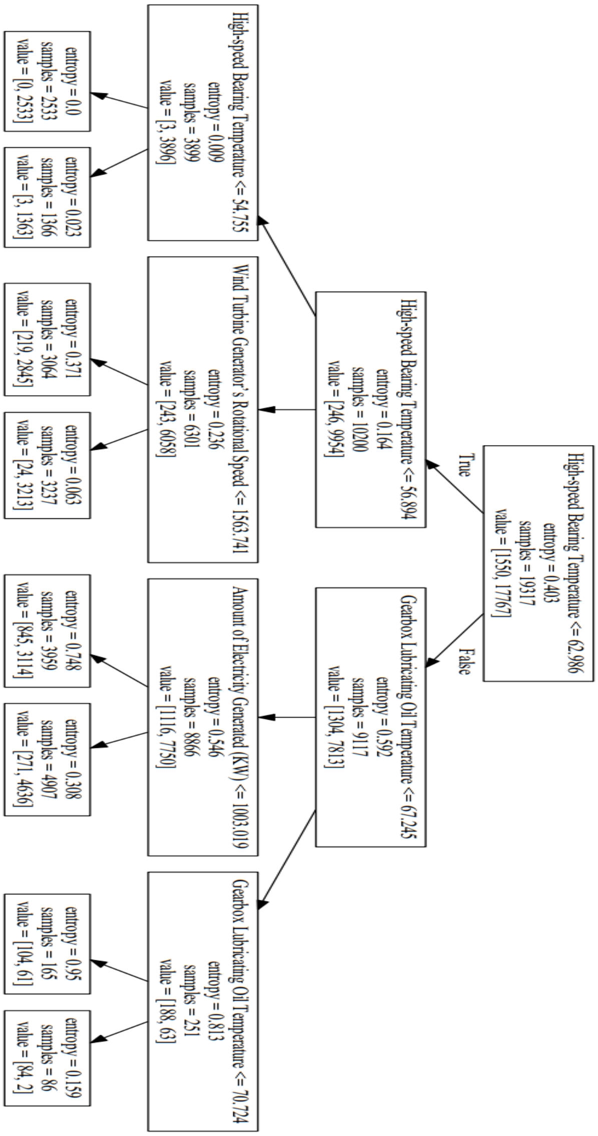

FIGURE 15. Decision tree classification.

used twice as a classifier: 67.245°C and 70.724°C. The classifiers that were only used once were wind turbine generator speed and the total amount of electricity generated, which were 1,563.741 RPM and 1,003.019 KW. The other two attributes – rotor speed and wind speed – were not selected as classifiers. As a result, we removed rotor speed and wind speed as classifiers from predicting anomalies of wind turbines and retained the other four attributes in the predictive models.

Random forest algorithm is an improved model of decision trees and offers less risk of overfitting. Hence, after the decision tree analyses, we further constructed the predictive model with the random forest algorithm. The results show that the first training accuracy rate was 99.71% and the testing accuracy rate was 95.34%. Next, K-fold cross-validation methods were used five times. The results in Table 5 display that the lowest accuracy rate was 88.75%, the highest

TABLE 5. The training result of random forests.

First training accuracy rate	99.71%	First test accuracy rate	95.34%		
Five cross-validation accuracy rate					
1st	2nd	3rd	4th	5th	average
88.75%	89.91%	93.33%	94.78%	94.02%	92.16%

accuracy was 94.78%, and the average accuracy rate was 92.16%. The cross-validation results show that the accuracy rate of the first test was slightly higher by 3.2% to the average cross-validation accuracy rate of 92.16%.

Both the decision tree and the random forest analyses have accuracy rates higher than 90%. In particular, the decision tree model has high explanatory power as high-speed bearing temperature was identified as the most influential predictor, followed by the gear lubricant temperature. The results are also consistent with the findings in the sections of ''Statistical Process Control'' and ''DBSCAN classification results''.

V. CONCLUSION

This study analyzes and predicts maintenance needs of wind turbines by using the wind turbine historical data collected in the ChangHua Coastal Industrial Park, Taiwan. The total observations were 2,815,104 from 31 wind turbines since 2015 to 2017. While previous studies usually adopt statistical process control charts or machine learning techniques for prediction, this study employs both methodologies for robust results.

Statistical program control was used to identify four categories of wind turbine faults (rotary blades, gearboxes, generators, and hydraulic oil systems). The authors then used two machine learning algorithms, decision tree and random forest classifications, to predict wind turbine abnormalities with accuracy rates higher than 92%. Particularly, we analyzed the wind turbine sensor data with the practitioners' insight gained from the maintenance check list. Combined data analytics and firsthand knowledge, we investigated the causes of wind turbine faults, classified abnormal- and normal-state wind turbine data, and constructed predictive models. The results provide Taipower and other wind turbine operators useful indicators to diagnose wind turbine faults and predict future maintenance needs.

After data cleansing, we examined the correlation coefficients of various wind turbine attributes by reviewing scatter plots to identify possible causes for wind turbine faults and abnormal wind turbine units. Then, we analyzed the causes for wind turbine anomalies through statistical process controls and classified abnormal- and normal-state data through the DBSCAN algorithm.

The results of the statistical process controls suggested that No. CK20 wind turbine had irregular performance and was the most suitable sample for faults analyses. We then established frequency tables based on the utility's check list, abnormal-state alerts, the occurrence frequency of faults, and the total duration of anomalies. Based on the frequency of abnormal conditions, the Pareto chart was created and found that the top 15 abnormal conditions for wind turbine faults accounted for 70% of the total occurrence of abnormal conditions. According to the Pareto analysis results, the causes for wind turbine faults were characterized into four primary types and a total of 30 secondary types. In addition, the results of the cause and effect diagram inform wind turbine operators the maintenance priorities of wind turbines and the extents that an abnormal condition may lead to wind turbine faults. Then, control charts were used to analyze wind turbine abnormal conditions by identifying deviated data points that exceeded two or three standard deviations from the mean. The results of control charts can be used as a monitoring mechanism to identify wind turbine's abnormal operation status.

By testing the number of core points and eps, we used the DBSCAN classification to successfully distinguish abnormal- from normal state- wind turbine data. We then validated the classification results and found the results are consistent with the findings of the statistical process control analyses. Finally, we segmented the classified data into a training dataset (70% of the classified data) and a testing dataset (30% of the classified data) to construct the predictive maintenance models through decision tree and random forest algorithms. The K-fold cross-validation model was used five times to verify whether the prediction models were robust and accurate. The empirical results showed that the accuracy rates of the two predictive models were 92.86% and 95.34%, and the average accuracy rate of the K-fold cross-validation tests were 92.68% and 92.16%. Both prediction models were quite accurate. In addition, the number of predictive attributes were reduced from six to four (the amount of wind energy generated, high-speed bearing temperature, gearbox lubricant temperature, and generator speed) and created a focused list of predictive factors for wind turbine faults.

This study not only informs wind turbine operators about wind turbine fault diagnosis and maintenance needs, but also reduces the sample size that is required for accurate predictive modeling. The study demonstrates that, by modeling, the failures of wind turbines can be detected, and the maintenance needs of parts can also be predicted. Model results may provide technicians early warnings, improve equipment efficient, decrease system downtime, and increase availability percentage or uptime of wind turbines. Improved wind turbine systems not only benefit wind turbine operators but also the society in transition toward clean energy development.

The results also inform Taipower and wind turbine operators possible ways to improve or streamline their maintenance work. The predictive models also predict wind turbine performance and detect if a performance data point is likely to fall into an abnormal range that signals maintenance need. In the future work, we will apply the prediction model to test other wind turbines and/or other times of the selected wind turbines, depending on data availability.

In addition, the prediction model is not designed for the real-time data streaming. The main reason is the wind occurs or wind direction changes always happen so fast, and it is not possible and no need to use real-time data streaming to implement an on-line learning system in this phase. However, online learning approaches can consider to build the prediction model in the future.

REFERENCES

[1] 4C Offshore. (2018). *Global Offshore Win Speeds Rankings*. [Online]. Available: https://www.4coffshore.c om/windfarms/windspeeds.aspx?m=no

[2] A. Kusiak and A. Verma, "A data-mining approach to monitoring wind turbines," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 150–157, Jan. 2012.

[3] C. H. Huang. (2018). *Offshore Wind Power Generation in Taiwan (Chinese)*. Accessed: Jun. 14, 2019. [Online]. Available: http://www.ntuce-newsletter.tw/vol.66/T4_1M.html

[4] F. P. G. Marquez, A. M. Tobias, J. M. Pérez, and M. Papaelias, "Condition monitoring of wind turbines: Techniques and methods," *Renew. Energy*, vol. 46, pp. 169–178, Oct. 2012.

[5] H.-H. Yang, M.-L. Huang, C.-M. Lai, and J.-R. Jin, "An approach combining data mining and control charts-based model for fault detection in wind turbines," *Renew. Energy*, vol. 115, pp. 808–816, Jan. 2018.

[6] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[7] L. Wang, Z. Zhang, J. Xu, and R. Liu, "Wind turbine blade breakage monitoring with deep autoencoders," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2824–2833, Oct. 2018.

[8] N. Amruthnath and T. Gupta. (2018). *Fault Class Prediction in Unsupervised Learning using Model-Based Clustering Approach*. [Online]. Available: https://www.researchgate.net/publication/322900854

[9] N. Yampikulsakul, E. Byon, S. Huang, S. Sheng, and M. You, "Condition monitoring of wind power system with nonparametric regression analysis," *IEEE Trans. Energy Convers.*, vol. 29, no. 2, pp. 288–299, Jan. 2018.

[10] P. L. Wen, S. M. Lin, C. S. Lin, C. C. Feng, and F. K. Ko, "Cost-benefit analysis of Taiwan's land-based wind power generation (Chinese)," *Taiwan's Economic Forecast Policy*, vol. 45, no. 1, pp. 41–76, 2014.

[11] S. Wang, X. Sun, and C. Li, "Wind turbine gearbox fault diagnosis method based on Riemannian manifold," *Math. Problems Eng.*, vol. 2014, pp. 1–10, 2014.

[12] Paris Agreement. (2015). *United Nations Treaty Collection*. [Online]. Available: https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=_en

[13] X. Liu, M. Li, S. Qin, X. Ma, and W. Wang, "A predictive fault diagnose method of wind turbine based on K-means clustering and neural networks," *J. Internet Technol.*, vol. 17, no. 7, pp. 1521–1528, 2016.

[14] Z. Hameed, Y. S. Hong, Y. M. Cho, S. H. Ahn, and C. K. Song, "Condition monitoring and fault detection of wind turbines and related algorithms: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 1, pp. 1–39, 2009.

[15] Z. Zhang, Q. Zhou, and A. Kusiak, "Optimization of wind power and its variability with a computational intelligence approach," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 228–236, Jan. 2014.

JYH-YIH HSU received the B.S. degree in applied economics from National Chung Hsing University, Taichung City, Taiwan, and the master's and Ph.D. degrees from the Natural Resource and Environmental Management Department, University of Hawaii.

He has been a Professor with the Department of Management Information Systems and the Department of Applied Economics, National Chung Hsing University, since 2004. He has also been an Independent Member of the Board Directors, Taiwan Power Company, since 2017. He used to be a Commissioner of the Fair Trade Commission. And before that, he was the Director at the Center of Energy and Environmental Studies, Chung-Hwa Institution of Economic Research, a Government think tank. He once was a consultant to Electric Power Research Institute (EPRI).

YI-FU WANG graduated from the Department of Mechanical and Electro-Mechanical Engineering, National l-Lan University, in 2012, and the master's degree focusing on matching learning and data analytics from the Department of Management Information Systems, National Chung-Hsing University, in 2019. He has worked as a system engineer for many MIS projects, supporting and maintaining corporate mainframe equipment, from 2013 to 2017.

KUAN-CHENG LIN was born in Taiwan, in September 1964. He received the B.S. degree in chemistry from National Taiwan University, in 1988, and the Ph.D. degree in applied mathematics from National Chung Hsing University, Taichung City, Taiwan, in 2000. From 2000 to 2006, he was an Assistant Professor with the Department of Information Management, Northern Taiwan Institute of Science and Technology, Taipei, Taiwan. From 2006 to 2008, he was an Assistant Professor with the Department of Management Information Systems, National Chung Hsing University. From 2008 to 2015, he was an Associate Professor with the Department of Management Information Systems, National Chung Hsing University, where he has been a Professor, since 2015. His current research interests include affective computing, intelligent tutoring system, and data mining.

MU-YEN CHEN received the Ph.D. degree in information management from National Chiao Tung University, Taiwan. He is currently a Professor of information management with the National Taichung University of Science and Technology, Taiwan. His current research interests include artificial intelligent, soft computing, bio-inspired computing, data mining, deep learning, context-awareness, machine learning, and financial engineering, with more than 100 publications in these areas. He has co-edited 15 special issues in International Journals, such as *Computers in Human Behavior*, *Applied Soft Computing*, *Soft Computing*, IEEE Access, *Information Fusion*, *Neurocomputing*, the *Journal of Medical and Biological Engineering*, *Electronic Library*, and *Library High Tech*. He has served as the Editor-in-Chief and an Associate Editor of international journals, such as the *International Journal of Big Data and Analytics in Healthcare*, IEEE Access, *Applied Soft Computing*, *Human-Centric Computing and Information Sciences*, the *Journal of Information Processing Systems*, and the *International Journal of Social and Humanistic Computing* while he is an editorial board member on several SCI journals.

JENNEILLE HWAI-YUAN HSU received the bachelor's degree in economics from National Taiwan University, the master's degree in public policy from the University of California at Los Angeles, CA, USA, and the Ph.D. degree in public policy and management from the Sol Price School of Public Policy, University of Southern California.

She had worked as a Data Science Fellow of the Southern California Association of Governments, Los Angeles, CA, USA, and a Summer Fellow of the Woodrow Wilson Center, Washington, D.C., USA. She is currently an Independent Researcher with the University of Southern California. Her work focus on energy policy, sustainable cities, and solar technology adoption.