POLITECNICO DI TORINO
Repository ISTITUZIONALE

Machine Learning Framework for the Sustainable Maintenance of Building Facilities

Original
Machine Learning Framework for the Sustainable Maintenance of Building Facilities / Villa, V., Bruno, G., Aliev, K., Piantanida, P., Corneli, A., Antonelli, D.. - In: SUSTAINABILITY. - ISSN 2071-1050. - 14:2(2022), p. 681. [10.3390/su14020681]

Availability:
This version is available at: 11583/2956607 since: 2022-02-28T11:16:36Z

Publisher:
MDPI

Published
DOI:10.3390/su14020681

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

24 November 2026

sustainability

MDPI

Article

Machine Learning Framework for the Sustainable Maintenance of Building Facilities

Valentina Villa [1,*], Giulia Bruno [2], Khurshid Aliev [1], Paolo Piantanida [1], Alessandra Corneli [3] and Dario Antonelli [2]

[1] Department of Structural, Geotechnical and Building Engineering , Politecnico di Torino, 10129 Turin, Italy; khurshid.aliev@polito.it (K.A.); paolo.piantanida@polito.it (P.P.)
[2] Dipartimento di Ingegneria Gestionale e della Produzione, Politecnico di Torino, 10129 Turin, Italy; giulia.bruno@polito.it (G.B.); dario.antonelli@polito.it (D.A.)
[3] Dipartimento di Ingegneria Civile, Edile e Architettura, Università Politecnica delle Marche, 60121 Ancona, Italy; a.corneli@univpm.it
* Correspondence: valentina.villa@polito.it

Abstract: The importance of sustainable building maintenance is growing as part of the Sustainable Building concept. The integration and implementation of new technologies such as the Internet of Things (IoT), smart sensors, and information and communication technology (ICT) into building facilities generate a large amount of data that will be utilized to better manage the sustainable building maintenance and staff. Anomaly prediction models assist facility managers in informing operators to perform scheduled maintenance and visualizing predicted facility anomalies on building information models (BIM). This study proposes a Machine Learning (ML) anomaly prediction model for sustainable building facility maintenance using an IoT sensor network and a BIM model. The suggested framework shows the data management technique of the anomaly prediction model in the 3D building model. The case study demonstrated the framework's competence to predict anomalies in the heating ventilation air conditioning (HVAC) system. Furthermore, data collected from various simulated conditions of the building facilities was utilized to monitor and forecast anomalies in the 3D model of the fan coil. The faults were then predicted using a classification model, and the results of the models are introduced. Finally, the IoT data from the building facility and the predicted values of the ML models are visualized in the building facility's BIM model and the real-time monitoring dashboard, respectively.

Keywords: Internet of things (IoT); sustainable buildings maintenance; management; smart buildings; artificial intelligence

check for updates

Citation: Villa, V.; Bruno, G.; Aliev, K.; Piantanida, P.; Corneli, A.; Antonelli, D. Machine Learning Framework for the Sustainable Maintenance of Building Facilities. *Sustainability* **2022**, *14*, 681. https://doi.org/10.3390/su14020681

Academic Editor: Jaejun Kim

Received: 30 November 2021
Accepted: 5 January 2022
Published: 8 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The environmental impact of buildings is astonishing. Every year, building construction consumes 25% of world wood harvest, 40% of materials entering the global economy, 3 billion tons of raw materials transformed into foundations, walls, pipes, and panels, and 50% of copper utilized in the United States [1]. Building construction accounts for half of the worldwide output of greenhouse gases and acid rain agents. Buildings, as a critical component of a habitat, have an influence on their local and surrounding areas, which can have unintended consequences for people and the community. Addressing the sustainability in building can significantly reduce these negative effects [2].

In Europe, the building maintenance industry accounts for the same amount of building production as the new construction market. Due to the aging buildings, the maintenance industry grows at a rate of 1.5 percent each year [3]. The size and expansion of the existing market needs a greater focus on sustainable building maintenance [4].

Sustainable buildings may ensure that their constructions and services are suitable for living, working, and other daily activities [5]. Building maintenance is essential to ensuring

the building's long-term sustainability [6]. Maintenance is defined by BS 3811:1984 and ISO 15686-1 as "a collection of operations required to keep an item in, or restore it to, an acceptable condition" [7].

Another facet of enhancing quality is a construction component's "repairability" [8]. In terms of sustainability, there will be a significant preference for repairing rather than replacing components. This repairability has shown to be a barrier to achieving a lengthy life duration for a number of ostensibly "low-maintenance" items. Those items have a minimal maintenance requirement during their lifetime, but if a failure occurs by accident, this defect often cannot, or only at a very high price, be repaired. As a result, a fatal problem might lead to fill replacement or demolition, resulting in the waste of high-quality components. Paying attention to repair possibilities in a timely manner would have extended the component's life. In this regard, the accessibility of components for maintenance operations should also be addressed [9].

Facility managers (FM) rely on real-time accurate data to execute maintenance tasks and deliver correct information to senior managers [10]. However, the operations of inspecting buildings, analyzing maintenance, and collecting data are time-consuming and labor-intensive [11]. Furthermore, the budget and resources provided for building maintenance are restricted, and maintenance workers claim that their budget and resources are insufficient and fall short of their requirements [12]. This trade-off affects the quality and relevance of maintenance operations and inspections, resulting in inadequate facility maintenance and quality management policies [13].

Present building facility maintenance procedures are mostly centered on corrective maintenance: delayed operations are done in response to a user complaint or an unforeseen breakdown [14]. Due to a shortage of funds and human resources, preventive maintenance is limited to the bare minimum of obligatory inspections in key facilities [15]. Furthermore, predictive maintenance is built on applying data analytics tools to analyze operating data supplied via sensors [16].

Sensing technologies, identification and recognition technologies, hardware, software and cloud platforms, communication technologies and networks, software and algorithms, position technologies, data processing solutions, power and energy storage, and security mechanisms etc., are examples of IoT enabling tools and technologies [17]. According to [18], IoT is defined as "interconnection of sensing and actuating devices enables information sharing across platforms via a unified framework, resulting in the development of a common operational picture for allowing novel applications". IoT is mainly based on standard protocols and technologies, and IoT devices that are a significant component. Intelligent gadgets, smart mobile devices, single board computers, and other sorts of sensors and actuators are examples of typical IoT devices [19].

Previous works proposed IoT and BIM integrated platform based on open standards by demonstrating the utility of such system in real-life case studies [20]. Another study presents a platform composed of IoT and BIM that acquires real-time data and performs analytics during the working processes of on-site assembly of pre-fabricated constructions using radio frequency identification (RFID) technology [21]. The recorded data is transferred to the cloud in real-time to be processed and analyzed providing decision support for the site managers and workers involved. IoT based real time warning system is also integrated in the underground construction of China (Wuhan) [22]. The system informs managers about warning in real time and prevents passengers of the Yangtze Riverbed Metro from accidents and improves the safety management in underground construction. Research presented in a cloud-based building management platform for green buildings was also proposed [23]. The research objective is to realize building operation management through the use of cloud computing and IoT technology. Cloud servers are used for data storage, processing, and hosting. The software is in charge of visualization interfaces and modularization services. Hardware development creates devices and objects that link the network. The building facility maintenance management based on IoT and BIM platform

includes several functions such as automatic diagnostics, monitoring, controlling, data processing, management, and service customization in the following articles [24,25].

ML techniques are applied in the construction as data-driven approach to predict life cycle cost of the buildings [26]. Some works used ML technique specifically Deep learning models to predict HVAC failures, to forecast building occupants' complaints or thermal comfortability [27], Other works used artificial neural networks models such as Multi-layer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA) [28]. Similarly, statistical models such as linear and nonlinear regression were applied in HVAC units for problem identification and diagnostics [29].

Despite the potential benefits of the works listed above, none has addressed the integration of IoT and ML models into BIM models presenting predicted building facility anomalies within the building's 3D model.

To assist BIM and to develop sustainable building facility maintenance, this article offers an ML-based anomaly prediction system and an IoT wireless sensor node. In the system's application layer, ML and IoT integrated BIM is developed for Facility Managers to monitor HVAC system components that identifies anomalies using ML models. The specific goal of this article is to find the best ML models and data management approaches for IoT facilities to connect with BIM that are suited for FM. The BIM model user interface (UI), which is linked with the IoT and ML model results, provides information about fan coil sensors data and real-time condition to the building 3D model.

2. Machine Learning based Framework

The proposed system classifies and predicts anomalies of building facility using ML methods and visualizes building facility data on the monitoring dashboard in real time. Figure 1 summarizes the anomaly prediction framework of the building facilities. ML based building facility anomaly prediction framework composed of different integrated parts: data acquisition from building facilities; communication layers, ML and deployment and integration on the BIM model.

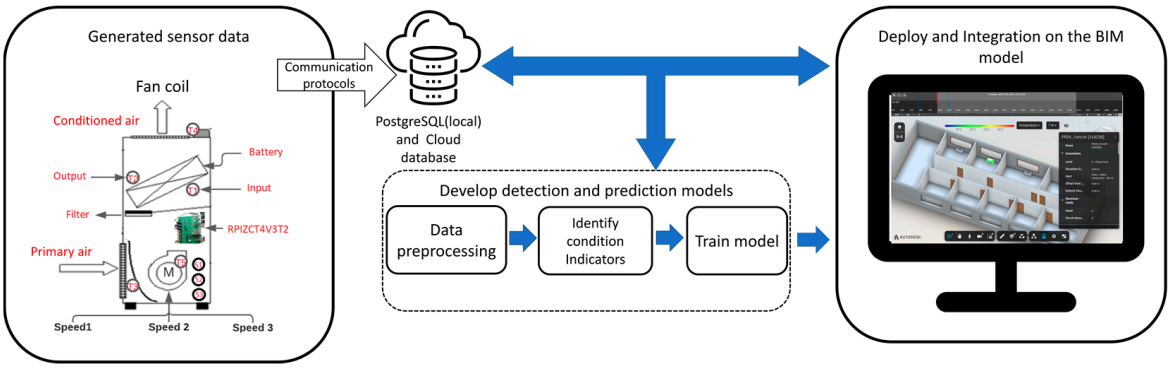


Figure 1. ML-based building facility anomaly prediction framework.

On the framework, generated sensor data of the building facility is driven to the ML models to predict building facility anomalies. The framework's application layer that is deployment and integration, displays IoT data and ML prediction results, as well as integrating the BIM model, to enable facility managers and operators to execute on-time building facility maintenance. Maintenance department can receive anomaly warnings and remotely monitor the location of the building facilities via the dashboard.

2.1. Utilized Components to Identify Anomalies of Indoor Facilities

The proposed framework requires number of specific sensors that is connected to the development board RPIZCT4V3T2 and then to the fan coil (FC), which is wirelessly connected to the gateway. The data coming from sensors are acquired and stored on the Raspberry Pi 3B local memory database (PostgreSQL) and on the cloud with wireless

communication in different sampling frequencies. The database provides raw data to the ML models and the results of the ML and specific data from the database such as temperatures and power consumptions will be visualized on the dashboard of the BIM model. The communication and hardware block diagram of the systems is depicted in Figure 2.

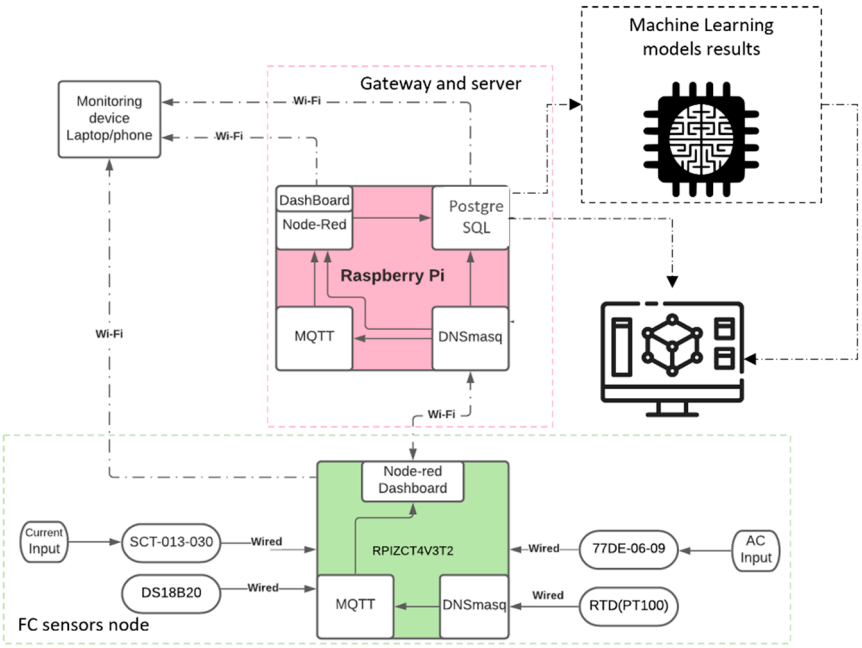


Figure 2. The connectivity block diagram of the proposed system.

Node-Red is installed on the Raspberry Pi 3B system allows to access all sensor variables through serial protocols and display them on its own customized dashboard. A topic is assigned to each sensor, which is responsible for publishing a message to the Raspberry Pi local server, which will function as a receiver (subscriber) via the (Message Queue Telemetry Transport) MQTT protocol. Furthermore, DNSmasq open-access software is installed on the Raspberry Pi 3B to make the board as a router and to create a communication bridge between the sensor board and global network components through Internet Protocol (IP) addresses. The DNSmasq provides the possibility to configure the SSID, password, and IP address on the Raspberry Pi 3B that makes the system visible on the network to the publisher and subscriber.

The PostgreSQL database is used to store sensor data locally on the Raspberry Pi 3B. PostgreSQL obtains an IP address with a specified port number when the Raspberry Pi 3B is powered on and waits for Node-Red to deliver the data to be gathered. The functional flowchart of the system is shown in Figure 3.

In the flowchart, PostgreSQL assigns the incoming data to the associated tables. MQTT, PostgreSQL, and Node-Red later use the same credentials to execute in the background and connect to the network, obtaining the IP address via DNSmasq. Communication is established between the RPIZCT4V3T2 sensor board and the Raspberry Pi 3B through the MQTT protocol on the specified port of the server, which receives all subscribed topic data from publishers via the TCP protocol and enables publishing devices access to the port. Simultaneously, Node-Red begins on port 1880 with the same IP address to control and monitor data flow of the server and database. The BIM dashboard receives machine learning predictive data and sensor node data in CSV format using a data-driven approach or directly from PostgreSQL through TCP/IP communication. As a result, sensor data will be shown on the Raspberry Pi 3B dashboard and via MQTT on the internet, and PostgreSQL can be monitored and seen from any device linked to the same network by opening the IP address followed by the port number.

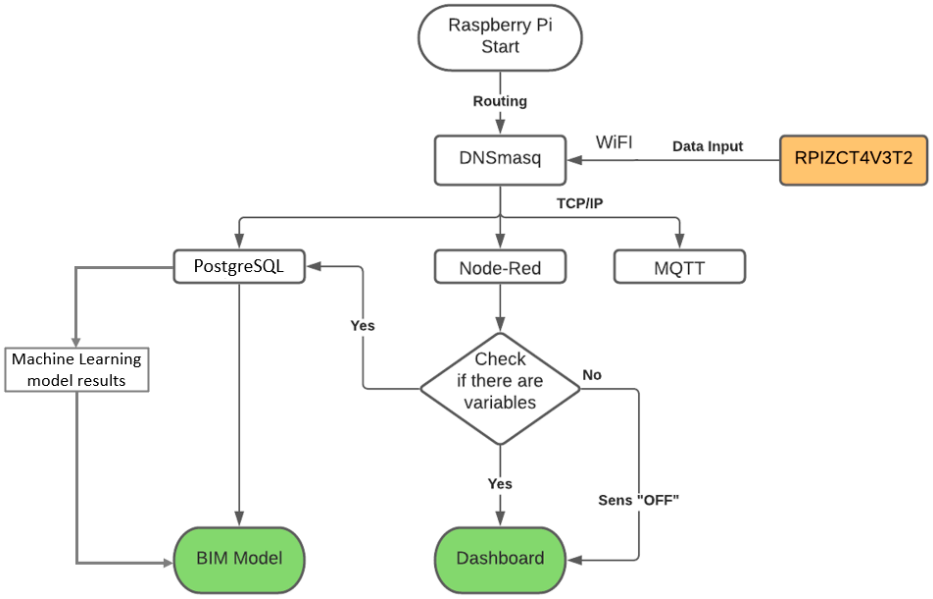


Figure 3. Functional flowchart of the framework.

The FC is equipped with DS18B20 and RTD PT100 sensors to measure the temperature of various sections of the FC, to measure the current sensor SCT-013-000, and to monitor the voltage 77DE-06-09 sensor. The measurement settings and ranges of the sensors vary depending on the internal elements of the fan coil. Taking these features into account, specific sensors and sensor board including all connected sensors for this study is shown in Table 1.

Table 1. Utilized sensors specifications and allocations.

Building facility Sensor Node	Sensor Name	Variable	Operating Range	Accuracy	Unit	Sensor Allocation
RPIZCT4V3T2	Current SCT-013-000	i1, i2, i3	0−100 A	±3	A	Motor current
	Voltage 77DE-06-09	v1, v2, v3	0–230 (50Hz)	±5	V	motor voltage
	Temperature DS18B20	T1	0–90 °C	±0.5	°C	Delivery pipe
	Temperature DS18B20	T2	0–90 °C	±0.5	°C	Return pipe
	Temperature DS18B20	T3	0–90 °C	±0.5	°C	Air intake
	Temperature DS18B20	T4	0–90 °C	±0.5	°C	Air outlet
	Temperature RTD(PT100)	T5	−200–550 °C	±0.05	°C	Motor case

The sensors presented in Table 1 were connected to the fan coil type FC83M-2014/1 at three different speeds and placed around the room to collect data for the ML tools and display real-time conditions on the dashboard. Temperature sensors T1, T2, and T4 on the FC, which are connected to the RPIZCT4V3T2 integrated system, monitor the condition temperature in the 0–90 °C range. T3 is in charge of monitoring the air temperature in the 0–50 °C range, whereas T5 is linked to the motor case and measures in the 0–200 °C range. Voltage sensors (v1,v2,v3) and current sensors (i1,i2,i3) are responsible to monitor the behavior of the motor.

The RPIZCT4V3T2 sensor node attached to the FC supports sensors described in Table 1 and contains a Raspberry Pi zero W to store data locally and sends measured data over Wi-Fi. A simplified block diagram of the RPIZCT4V3T2 board is presented in Figure 4.

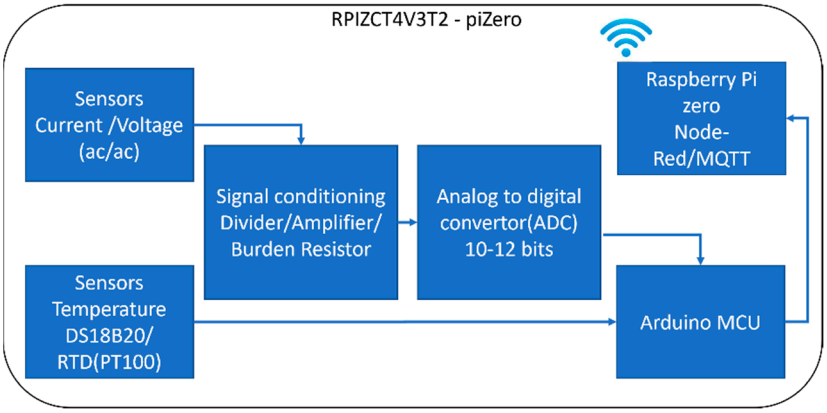


Figure 4. Simplified block diagram of RPIZCT4V3T2-piZero.

The RPIZCT4V3T2 board has an Arduino microcontroller (MCU) that is connected to two types of temperature sensors as well as current/voltage sensors that are linked to the MCU through an amplifier and an analog to digital converter (ADC). The MCU receives all raw data from sensors, computes required values such as Real Power, Power Factor, etc., and sends the final computation to the Raspberry pi Zero W through the universal asynchronous receiver-transmitter (UART) serial port; the Raspberry pi Zero W supports Wi-Fi, and the board connects to the server Rpi3B via IP address.

The board RPIZCT4V3T2 measures current using adjusted burden resistor. It is connected directly to the ADC and Vref/2. The burden resistor can be calculated with the equation as follows:

$$R = \frac{\left(V_{ref} \ * \ N_{turns}\right)}{\left(2\sqrt{2} \ * \ I_{rms}\right)}, \tag{1}$$

where, R is burden resistor value in Ohm; V_{ref} is ADC max voltage range; N_{turns} is number of CT sensor (2000) turns; $Irms$ is maximum current.

The voltage line has a divider as well as an amplifier. When utilizing an ac/ac adapter, the amplifier gain is set to 1 and the signal amplitude is decreased using a divider. Figure 5 depicts current and voltage connection schematics to the ADC of the MCU.

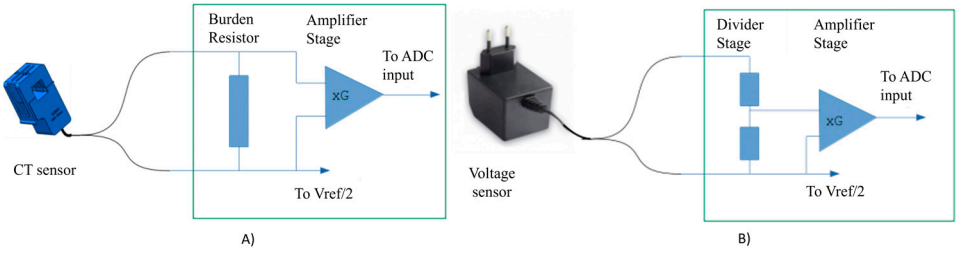


Figure 5. (**A**) Current sensor with burden resistor connected to the ADC through amplifier; (**B**) Voltage sensor with voltage divider connected to the ADC through amplifier.

The RPIZCT4V3T2 board typically reads current $I(t)$ and voltage $V(t)$ values instantaneously at a given time t. Thus, the signals are used in ADC are in the discrete-time domain. Hence t being the iterative time sample. Instantaneous values $I(t)$ and $V(t)$ are used and are adopted to their Root Mean Square (RMS) values. Given that the ADC has direct access to the $I(t)$ and $V(t)$, the RMS value of the current and voltage and formulas

to compute the power is given and described in Table 2. The equations described in Table 2 is performed in hardware level either in MCU or RPIZCT4V3T2 board.

Table 2. Formulation of computed values.

Computed values	Formulation	Description
Current (I_{rms})	$I_{rms} = \sqrt{\left(\frac{1}{N}\sum^{t=N} I^2_{[k]}\right)}$	where N is the number of sample acquired by the ADC
Voltage (V_{rms})	$V_{rms} = \sqrt{\left(\frac{1}{N}\sum^{t=N} V^2_{[k]}\right)}$	
Estimated Power (EP)	$P_{est} = V_{est} * I_{rms}$	where V_{est} is usually 230 V or 120 V
Real Power(RP)	$RP = \frac{1}{N}\sum^{t=N} V_{[k]} * I_{[k]}$	Computation is performed in RPIZCT4V3T2 board
Apparent Power(AP)	$AP = V_{rms} * I_{rms}$	Computation is performed in MCU
Power Factor(PF)	$PF = \cos(\varphi) = \frac{RP}{AP}$	
Reactive Power(RP)	$Q = \sqrt{(AP^2 - RP^2)}$	

In the next chapters, methods applied for ML models and data preparation are described.

2.2. Machine Learning Models

This subsection discusses how Machine learning models were used to monitor the condition of the building's facilities. To commence, a correlation analysis was conducted to determine which variables are significant. The building facility's most closely correlated variables were then loaded into an ML tool to implement prediction analyses. To optimize the efficiency and output of ML, relevant characteristics were extracted from a dataset of building facilities using initial data preparation procedures.

2.2.1. Data Preprocessing

Data preprocessing which is also known as data preparation improves data quality and allows for the extraction of advanced analytics. Data obtained from building facilities sensors is typically inaccurate, biased, and noisy. Raw data feature extraction can enhance the efficiency and accuracy of ML operations. Provided raw data must be cleaned and integrated [30–32], feature scaled, extracted, and selected before it can be used in ML.

Feature scaling often known as data standardization which refers to the standardization of a data set's range of features, which entails altering the values of numerical columns recorded on multiple scales to a formal common scale without affecting the ranges of the values or losing information. Data normalization entails rescaling data dimensions and avoiding over-weighting values that contributes to the overall quality of a data set [33]. As scaling intervals, [0,1] and [1,1] are commonly employed, as indicated in equations:

$$interval[0,1] = \frac{originalvalue - \min(allvalues)}{\max(allvalues) - \min(allvalues)}, \quad (2)$$

Feature extraction takes into account a raw data that is made up of irrelevant and/or duplicate features that might have a significant impact on the trading activity's success. Authors [34] offer several feature selections such as multicollinearity, correlation coefficients, and Variance Inflation Factors (VIF) to increase the performance of ML model outputs. According to the authors [35], correlation coefficients, such as Pearson's for a linear correlation, or rank-based approaches for a nonlinear correlation, are the most widely employed methodologies for numerical input and output models. If the data coming from building facilities and predicting factors is numerical, the approaches used in the current framework are correlation coefficients. The multicollinearity feature selection approach is used to predict categorical variables of building facilities. Conditional data variables received from building facilities were analyzed using a correlation matrix, and the easiest

method to discover collinearity is to examine the predictors' correlation matrix. A significant absolute value element in this matrix denotes a pair of strongly linked variables, and thus a collinearity problem in the data. Correlation analysis of building facility was used to assess the significance of the relationship between predictors X and response Y. Correlation analyses provide insight into the linearity of paired variables. As a parameter of the linear dependency, the correlation coefficients between two random variables are generated for all model variables [36]. The sample estimate of the correlation coefficient r_{xy} is calculated for two variables, X and Y, in the dataset as follows:

$$r_{XY} = \frac{cov(XY)}{\sigma_X \sigma_Y}, \tag{3}$$

Where σ_X and σ_Y are the standart deviations of X and Y.

$$r_{XY} = \begin{cases} 0, \; if \;\; X \; and \; Y \; are \; linearly \; uncorrelated \\ 1, \; if \;\; X \; and \; Y \; have \; perfect \;\; linear \; positive \; relationship \\ -1, \; if \;\; X \; and \; Y \; have \; perfect \; negative \; linear \; relationship \end{cases}, \tag{4}$$

Data transformation is another process of converting a numerical format to a categorical format and removing unnecessary data [37] that used in this research.

To improve the efficiency and output of ML, the aforementioned data preparation approaches were used to extract significant features from a dataset of building facilities. Acquired features and variables were then provided to the ML processes.

2.2.2. Auto Classification Model to Predict Qualitative Variables

To identify the optimal ML model for detection building facility anomalies, the Automatic machine learning (AutoML) method is applied. H2O AutoML is an open source, user-friendly machine learning platform that was created for both advanced and intermediate machine learning users. According to recent research, H2O AutoML [38] outperforms other competing automatic ML tools. The authors of [39] evaluated AutoML's robustness and efficiency in comparison to other automated models such as TPOT [40] and AutoKeras [41]. The authors [42] tested the tool's resilience using fuzzy, clean, and noisy data sets. Similar research [43] have assessed the efficiency of the AutoML method in accordance with other tools, such as auto-sklearn [44] and Auto-WEKA [45] using open source datasets. To produce a large number of ML models in a short period of time, AutoML is the fastest tool to train H2O machine learning algorithms. On tabular datasets, H2O AutoML allows supervised training of regression, binary classification, and multi-class classification models. H2O AutoML platform is available in different programming tools such as Python, R, Java, and Scala, as well as via a web interface. R is considered in this research to perform ML-based classification analyses.

Generalized Linear Models (GLM), Distributed Random Forests (DRF), XGBoost, Gradient Boosting Machines (GBM), and Deep Learning are the H2O AutoML platform essential models. H2O AutoML platform selects one of the three alternative models. It can deploy only one of the base models or their hyperparameter-optimized variants. It may also select a Best Of Family Stacked Ensemble model, which contains one model from each category, and the All Models Stacked Ensemble pipeline.

The evaluation metrics used for the classification models are listed and described in Table 3.

Table 3. Evaluation metrics of the linear regression and classification models.

Type of models	Model metrics	Formulation	Description
Classification model	AUCPR (Area Under the Precision-Recall Curve)		This model metric is used to evaluate how well a binary classification model is able to distinguish between precision recall pairs or points. These values are obtained using different thresholds on a probabilistic or other continuous-output classifier. AUCPR is an average of the precision-recall weighted by the probability of a given threshold.
	AUC (Area Under the ROC Curve)		This model metric is used to evaluate how well a binary classification model is able to distinguish between true positives and false positives. An AUC of 1 indicates a perfect classifier, while an AUC of 0.5 indicates a poor classifier, whose performance is no better than random guessing.
	Accuracy	$Accuracy = \left(\frac{correctly\ predicted}{observations}\right)$	Accuracy is the number of correct predictions made as a ratio of all predictions made.
	Log loss	$Logloss = -\frac{1}{N}\sum_{i=1}^{N} w_i(y_i \ln(p_i) + (1-y_i)\ln(1-p_i)$	The logarithmic loss metric can be used to evaluate the performance of a binomial or multinomial classifier. In the equation, N is the total number of observations; w is the per row user-defined weight; C is the total number of classes; p is the predicted value; y is the actual target value.
	MSE	$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$	The mean squared error (MSE) metric measures the average of the squares of the errors or deviations. MSE takes the distances from the points to the regression line (these distances are the "errors") and squaring them to remove any negative signs. MSE incorporates both the variance and the bias of the predictor.
	RMSE	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$	The root mean square error (RMSE) metric evaluates how well a model can predict a continuous value. The RMSE units are the same as the predicted target, which is useful for understanding if the size of the error is of concern or not. The smaller the RMSE, the better the model's performance. In the equation, N is the total number of observations; y_i is the actual target value $\hat{y}$ is the predicted target value.
	Variable importance	$VAR = \frac{1}{N}\sum_{i=0}^{N}(y_i - \overline{y})^2$	Variable importance's represent the statistical significance of each variable in the data in terms of its affect on the model. Variables are listed in order of most to least importance. The percentage values represent the percentage of importance across all variables, scaled to 100%. The method of computing each variable's importance depends on the algorithm.

3. Deployment and Integration of IoT and Machine Learning Data on the BIM Model

The BIM model of the experimental building and the FC utilized in this study were both created using the Autodesk Revit software. Navisworks from Autodesk is used to integrate IoT sensor data and ML predicted results with a BIM model. Users can benefit from Navisworks to open and merge 3D models, navigate around them in real-time, and evaluate the model using a variety of features such as comments, redlining, viewpoint, and measurements. Interference detection, 4D time simulation, photorealistic rendering, and data management are among the plugins that enhance the program. In this study, Navisworks iConstruct plugin is used to manage IoT and ML data to the BIM model through Excel sheet files and/or PostgreSQL data sources, allowing users to expand the information in the BIM model. Schematic diagram of the data management and transformation into the BIM model is depicted in Figure 6.

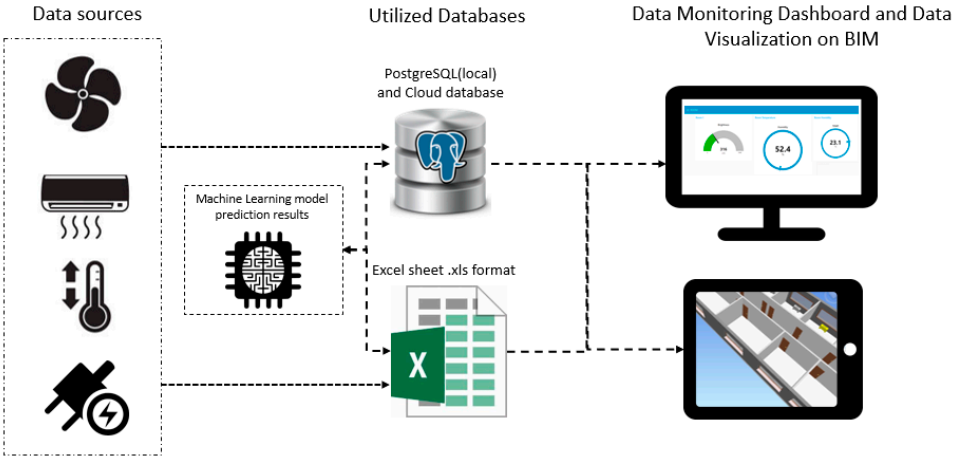


Figure 6. IoT and ML data integration diagram into the BIM model.

The integration of sensors and ML model results started with a static data on the excel sheet, which was then imported into Navisworks through the iConstruct external plugins. To link data to the BIM model, Data Links tab is configured. To link the data users must first identify an attribute that connects both the BIM model's information and an external data source. Once a relationship is established, Data Links can be mapped for additional information between the two data sets and propagate onto the user-defined data tab. The user-defined tab then displays the external information and attributes. iConstruct allows users to customize the color of the fan-coil based on the values of the variables.

Second possibility is to visualize external data in real time on the BIM model. In order to integrate sensor data in real time through a PostgreSQL database, a node-red flow is created to transmit coming sensor data.

After that in Navisworks PostgreSQL database is configured using Data tools settings then PostgreSQL Ansi 64 driver is installed to fill Navisworks columns with a corresponding columns of the sensor variable database.

Thus, an ML-based framework for the maintenance of building facilities that visualize anomalies on the BIM is ready to conduct an experiment in a real case.

To show the reliability and applicability of the framework, the experiment was conducted at the Politecnico di Torino's DISEG laboratory. The experimental laboratory room is located in the building's basement. One of the fan coil units (FCU) of the room that is positioned under the windows is used as a facilities building. The FCU is a part of the HVAC system. It uses a coil and a fan to heat or cool the building's rooms which is considered one of the important indoor facilities. For this study, FC83M-2014/1 FC with four speeds was used. The FC has a motor that produces 1100 RPM in an anti-clockwise direction. Furthermore, FC has a cooling and heating battery as well as filters that must be checked on a regular basis. Technical specifications of the FCU used for this study is provided in Table 4.

Table 4. FC specifications used in case study.

	Metrics name	Operating Range	Unit
FC 83M-2014/1 technical specifications	Working range of voltage/frequency	230-240/50	V/Hz
	Ampere	0.23	A
	Number of velocity	4	
	Power consumption	14/53	W
	Revolutions per minute (RPM)	1100	Min^{-1}

Expanded framework of the experimental setup including data acquisition, machine learning specifications and building 3D model integration is demonstrated in Figure 7. On the framework, building facility specifically FC 83M-2014/1 is equipped with sensors and collects data locally and on the server. On the framework, the building facility fan coil (FC) is equipped with sensors that interact with the local server and the gateway cloud server through TCP/IP and MQTT protocols, performing data gathering and storing essential data for anomaly prediction models. Data collection from sensors is acquired with a memory of ~30GB that means a proposed system able to collect data from installed sensors for 26 days with a sampling frequency of 1 Hz. Furthermore, the collected data is sent into the online condition monitoring dashboard. Data preprocessing allows for the extraction of meaningful features from a dataset and their transfer to ML models.

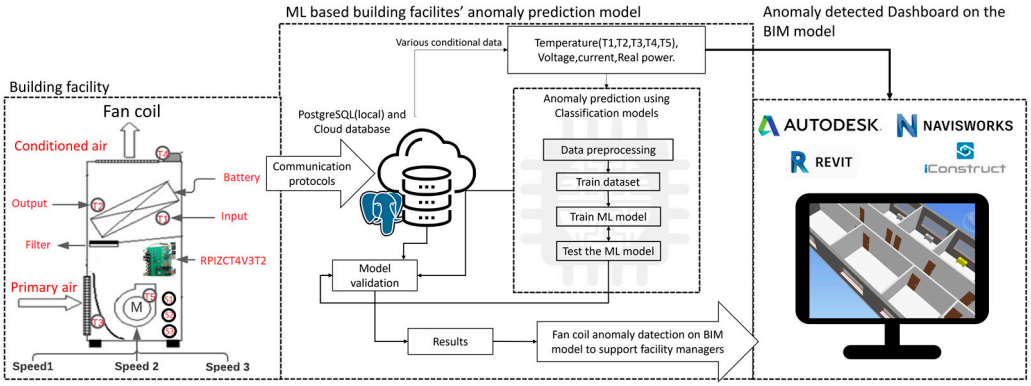


Figure 7. Expanded ML-based anomaly prediction framework for the experiment.

Figure 8 depicts the FC, which has been equipped with RPIZCT4V3T2 board and sensors and is ready to conduct an experiment.

The experiment was divided into two parts: the first was to collect balanced (normal) conditional data from the FC motor, and the second was to simulate an unbalanced condition by attaching 15g of mass to one of the FC motor's blades. The experiment was conducted three times at three different speeds. Throughout the experiment, all essential raw data was collected and saved on the database for later analysis. The collected dataset from the experiment is labeled according to the condition of the FC. The labeled sensor dataset is then trained and tested using Automatic ML classification models and the best performed model is selected for deployment on the application. The classification model results such as balanced, unbalanced or off conditions of the FC is presented on the BIM model of the construction.

In the next chapter the results of the experiment, building facility sensors' data within the ML results integrated into BIM model are introduced.

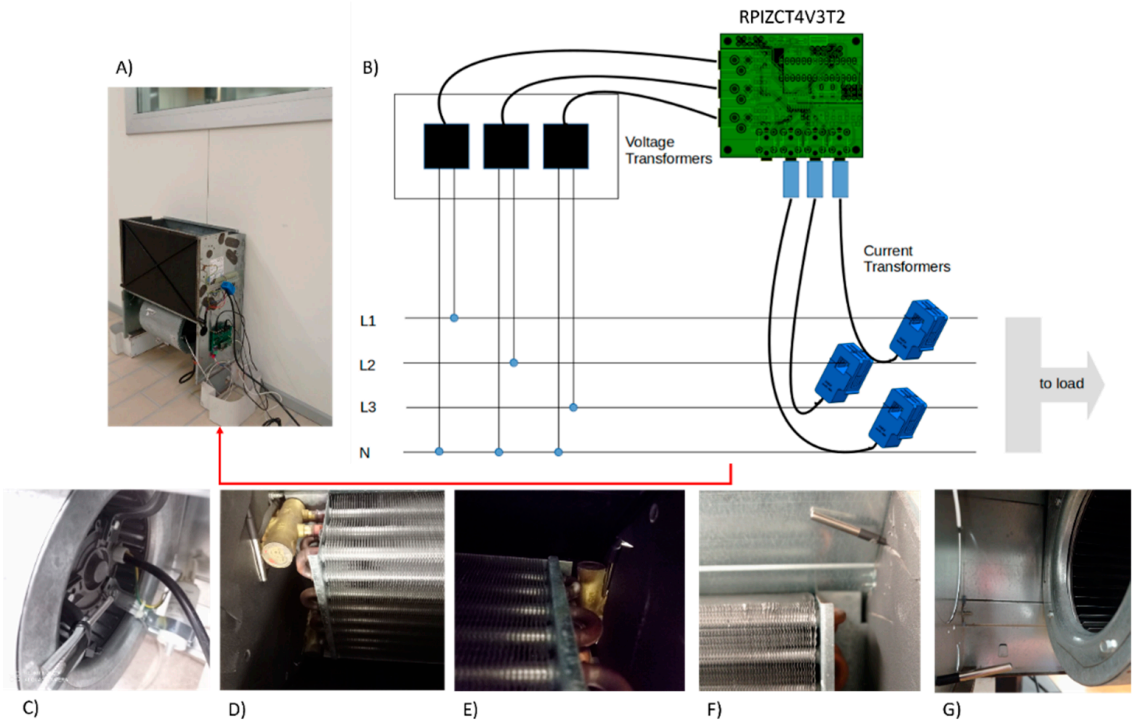


Figure 8. Sensors allocation on the FC components: (**A**) FC equipped with sensors; (**B**) Connecting and measuring diagram of the current and voltage of the FC motor; (**C**) T5 sensor connected to the motor case; (**D**) T2 return pipe sensor; (**E**) T1 delivery pipe sensor; (**F**) T3 air intake sensor; (**G**) T4 air outlet sensor.

4. Results

H2O AutoML is used to detect anomaly from the data coming from sensors. Data acquired from sensors are trained in an H2O.ai cluster using R studio with a version 3.34.0.3. The AutoML function in H2O automates the process of building a large number of models and finds the most suitable model for a given dataset. Some H2O models yield varying significance for one-hot (binary indicator) encoded categorical columns (e.g., Deep Learning, XGBoost). For the variable significance of categorical columns, a summary of the variable importance across all one-hot encoded features must be compared across all model types and produce a single variable importance for the original categorical feature. The models and variables are ordered by their proximity. The multinomial distribution approach was used to train the dataset as it was categorical.

As the main metrics of the predictive models were used error metrics to evaluate the best performed models. Table 5 demonstrates the results of the best ML models that have been used to classify different conditions of the fan coil. According to the table "Deep learning grid 3 AutoML model 3" is the best performed model for our dataset. The deep learning model parameters are three class classification model, multinomial distribution, 6.053 weights/biases and mini-batch size is 1. The results of the model: RMSE is 0.096372 and MSE is 0.009287 which is very good multinomial classifier. However, training time is slightly higher respect to the other models.

The H2OAutoML leaderboard displays the model results together with corresponding metrics. The H2OAutoML leaderboard graph displays 5-fold cross-validated metrics. Figure 9 depicts the variable significance of the Deep learning model produced by H2OAutoML. According to the plot, Real power (RP1) is the most influenced variable to the model.

Table 5. Metrics results of the different automatic ML models used to predict anomalies of the FC motor.

Model_id	Mean per Class Error	Logloss	RMSE	MSE	Training Time (ms)
DeepLearning_grid__3_AutoML_20211105_090506_model_1	0.008263	0.115798	0.096372	0.009287	11,425
StackedEnsemble_BestOfFamily_AutoML_20211105	0.010678	0.08707	0.106898	0.011427	228
GBM_grid__1_AutoML_20211105_090506_model_2	0.012162	0.052030	0.108987	0.011878	125
StackedEnsemble_AllModels_AutoML_20211105	0.012628	0.151789	0.163482	0.026726	348
DRF_1_AutoML_20211105_090506	0.014111	0.062486	0.119899	0.014376	76
XRT_1_AutoML_20211105_090506	0.014111	0.072696	0.133769	0.017894	67

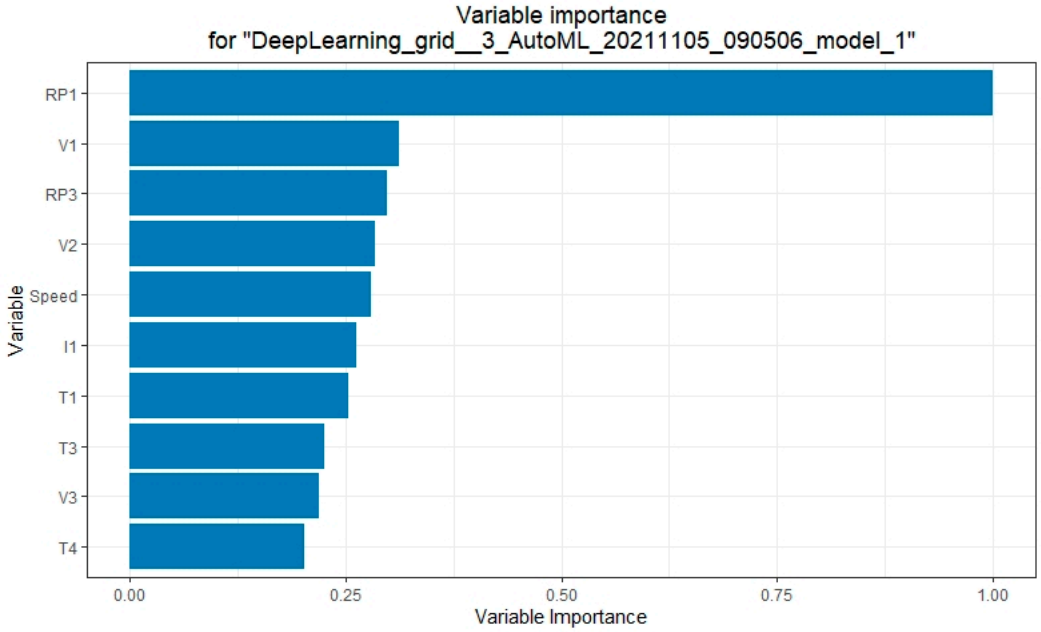


Figure 9. Variable importance of the best performed model of the H2O AutoML.

The confusion matrix in Table 6 represents the measurements of the highest performing model that is predicted class with actual class, indicating that the model identified anomaly with 0.0065 error and only three points out of 463 are misclassified for the provided dataset.

Table 6. Confusion Matrix of the DeepLearning_grid__3_AutoML_20211105_090506_model_1: Row labels: Actual class; Column labels: Predicted class.

	Balanced	Off	Unbalanced	Error	Rate
Balanced	168	0	3	0.0175	=3/171
Off	0	154	0	0.0000	=0/154
Unbalanced	0	0	138	0.0000	=0/138
Totals	168	154	141	0.0065	=3/463

The results of AutoML that classifies different conditions and anomalies of the FC motor in real time are deployed on the BIM model. To understand the behavior of the balanced, off, and unbalanced condition of the FC motor a real time data of the Real Power in three speeds is depicted in Figure 10.

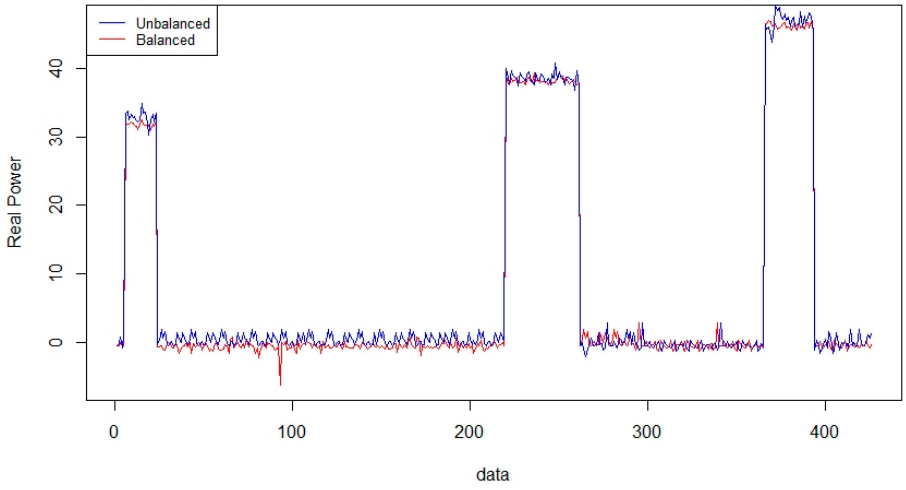


Figure 10. Real Power of the FC motor in three speeds: blue is unbalanced and red is balanced condition.

Finally, Figure 11 demonstrates an application developed on the Autodesk Navisworks that integrates ML results within the BIM model. In the application, the fan coil color changes according to the results of ML data. The color is "Blue" if the fan coil motor is under the balanced condition, "red" if the FC motor is under the unbalanced condition and "Grey" if the FC is turned off.

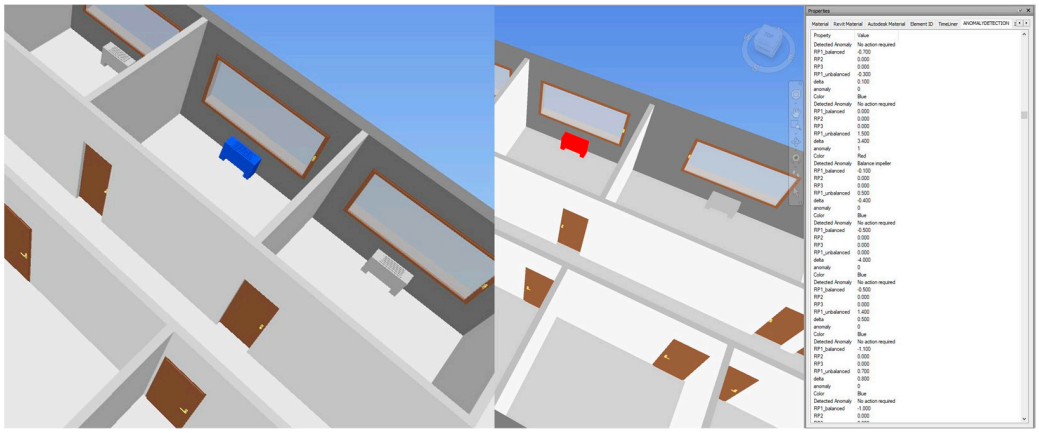


Figure 11. Autodesk Navisworks custom 3D BIM model integrated with IoT and ML results data: on the left, balanced condition is colored with a "Blue" and on the right, the motor is under unbalanced condition and colored with a "Red". Turned off FCs are colored with a "Grey" color.

Moreover, Node-red dashboard displays all coming data from sensors in real time as shown in Figure 12.

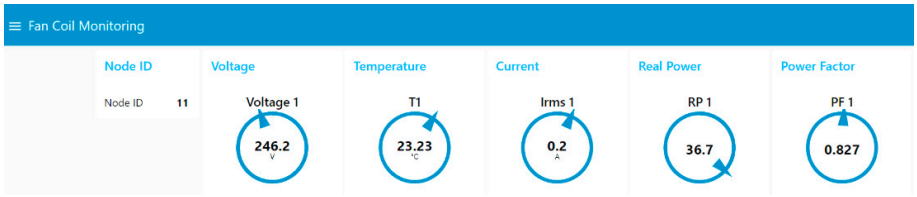


Figure 12. Node-red monitoring dashboard in real time.

5. Discussion

Cloud based IoT and analytical platforms such as Google clouds, Microsoft Azure, IBM Watson, AWS from Amazon, Thingworx, Thingspeak give a possibility easily acquiring IoT data to perform analytics, however, most of the important services such as data storage, extra AI tool usability for prediction analyses are not free and there is no integration tools with a custom based BIM models.

A proposed ML-based framework for building facility monitoring system would enhance the building's maintenance plan by assisting facility managers in evaluating the monitored building conditions within the 3D model. As a result, facility managers may benefit from the proposed framework to solve sustainable maintenance issues in the following: Facility managers may plan and schedule maintenance work in advance by using indoor facility anomaly or failure signals; Data visualization and anomaly data on the dashboard provide the potential of minimizing the danger of devastating failures and reducing unexpected forced outages of building components; ML-based anticipated and real-time data from sensors allow for more accurate maintenance; The use of ML predicted results for room components inside the BIM model simplifies maintenance operations; Finally, identifying the location of the failed component in real time is simple.

Thus, the research presented in this paper attempts to solve the gaps in the following studies: lack of ML framework to develop sustainable building facility maintenance [26], a lack of automation [46], and a need of a data collecting system [47].

However, there are significant limits towards this work, which are as follows:

The framework is quite limited in terms of ML-based building facility maintenance. Investigations are required to conduct sizable research by connecting more building facilities such as sprinkler systems, water pipes, drains, wires, ducts, cables, fire services equipment, and so on, as well as evaluating the system's integration with facility managers and clients to assess its reliability, repeatability, robustness, and ease of use.

More building facility sensors, such as an oxide gas sensor, a particle dust sensor, etc., as well as facility management sensors, including a motion sensor, occupancy sensor, and so on, may be added to the system and evaluated using the proposed framework.

6. Conclusions

In this research an ML framework for the sustainable maintenance of the building facilities is introduced. The proposed framework addresses the gap in maintenance procedures by assisting FM teams in taking early action and avoiding unforeseen failures without the requirement for costly extensive site inspections of the installations. The framework is composed by three main parts: (i) an IoT architecture to collect real-time data from the building, (ii) an ML algorithm to analyze the data and support the maintenance activity, and (iii) a BIM to provide a virtual representation of the building and visualize the maintenance activity.

To prove the applicability and reliability of the framework the experiment has been conducted on the indoor facility FC that is equipped with sensors and sensor node that continuously sends data to server database. On the edge of the framework, the ML model predicts anomalies of the unbalanced condition of the FC.

To classify different conditions including balanced, unbalanced, and turned off conditions, automatic ML platform tool AutoML H2O is used and according to the multinomial classification models, the "Deep learning grid 3 AutoML model 3" is the best performed model for our case study. The results of the best performed Deep Learning model: RMSE 0.096372 and MSE 0.009287 which means very good classifier. However, training time is slightly higher respect to other models. Finally, in the application layer of the framework, IoT data and ML results are integrated with a BIM model. In the 3D model of the BIM, the fan coil color changes according to the anomaly prediction results.

Future research will concentrate on transferring IoT sensor data and ML anomaly information to the three-dimensional BIM models that will be merged with virtual real-

ity (VR), augmented reality (AR), and mixed reality (MR) technologies and visualizing anomalies on the Digital Twin of BIM model.

A proposed ML framework might serve as a starting point for developing collaborative BIM systems based on AR/MR/VR. For example, to enable information flow in facility management, create a framework based on ML, BIM, mixed reality, and a cloud platform.

Such systems are valuable to facility managers because they provide quick remote access to 3D models of buildings and allow them to analyze a scenario before making decisions about building facility operations.

Author Contributions: Conceptualization, V.V.; data curation, K.A.; formal analysis, K.A. and P.P.; funding acquisition, V.V.; investigation, V.V. and G.B.; methodology, G.B., P.P. and D.A.; resources, K.A., A.C. and D.A.; software, K.A.; supervision, V.V.; validation, G.B.; visualization, P.P. and A.C.; writing-review & editing, V.V., G.B., K.A. and D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Italian government, through the PRIN2017 project of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR). The project entitled "Distributed Digital Collaboration Framework for Small and Medium-Sized Engineering and Construction Enterprises" is coordinated at the national level by Berardo Naticchia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CIWMB. *Designing with Vision: A Technical Manual for Material Choices in Sustainable Construction*; California Environmental Protection Agency: Sacramento, CA, USA, 2000.
2. Ding, G.K.C. Sustainable construction—The role of environmental assessment tools. *J. Environ. Manag.* **2008**, *86*, 451–464. [CrossRef] [PubMed]
3. del Río Merino, M.; Izquierdo Gracia, P.; Weis Azevedo, I.S. Sustainable construction: Construction and demolition waste reconsidered. *Waste Manag. Res.* **2010**, *28*, 118–129. [CrossRef] [PubMed]
4. Wilkinson, S.J.; Remøy, H.; Langston, C. *Sustainable Building Adaptation: Innovations in Decision-Making*; John Wiley & Sons: Hoboken, NJ, USA, 2014.
5. Sassi, P. *Strategies for Sustainable Architecture*; Taylor & Francis: Abingdon, UK, 2006.
6. Clements-Croome, D.; Croome, D.J. *Intelligent Buildings: Design, Management and Operation*; Thomas Telford: London, UK, 2004.
7. *ISO 15686-1*; Buildings and Constructed Assets-Service Life Planning: Part 1: General Principle. International Organization for Standardization: Geneva, Switzerland, 2000.
8. Wang, B.; Xia, X.; Zhang, J. A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. *Energy Build.* **2014**, *77*, 227–235. [CrossRef]
9. Hermans, M.H. Sustainable Building Maintenance. In Proceedings of the CIB World Building Congress 1998, Managing for Sustainability—Endurance Through Change, Gävle, Sweden, 7–12 June 1998; pp. 1863–1870.
10. Teicholz, P. (Ed.) *BIM for Facility Managers*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
11. Matarneh, S.T.; Danso-Amoako, M.; Al-Bizri, S.; Gaterell, M.; Matarneh, R.T. BIM for FM: Developing information requirements to support facilities management systems. *Facilities* **2019**, *38*, 378–394. [CrossRef]
12. Zhan, J.; Ge, X.J.; Huang, S.; Zhao, L.; Wong, J.K.W.; He, S.X. Improvement of the inspection-repair process with building information modelling and image classification. *Facilities* **2019**, *37*, 395–414. [CrossRef]
13. Pitt, T.J. Data requirements for the prioritization of predictive building maintenance. *Facilities* **1997**, *15*, 97–104. [CrossRef]
14. Yang, X.; Ergan, S. BIM for FM: Information requirements to support HVAC-related corrective maintenance. *J. Arch. Eng.* **2017**, *23*, 04017023. [CrossRef]
15. Ottoman, G.R.; Nixon, W.B.; Lofgren, S.T. Budgeting for facility maintenance and repair. I: Methods and models. *J. Manag. Eng.* **1999**, *15*, 71–83. [CrossRef]
16. Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; Francisco, R.D.P.; Basto, J.; Alcalá, S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [CrossRef]
17. Čolaković, A.; Hadžialić, M. Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Comput. Netw.* **2018**, *144*, 17–39. [CrossRef]
18. Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things: A vision, architectural elements, and future directions. *Future Gen. Comp. Syst. J.* **2012**, *29*, 1645. [CrossRef]

19. Dehury, C.K.; Sahoo, P.K. Design and implementation of a novel service management framework for IoT devices in cloud. *J. Syst. Softw.* **2016**, *119*, 149–161. [CrossRef]
20. Dave, B.; Buda, A.; Nurminen, A.; Främling, K. A framework for integrating BIM and IoT through open standards. *Autom. Constr.* **2018**, *95*, 35–45. [CrossRef]
21. Li, C.Z.; Xue, F.; Li, X.; Hong, J.; Shen, G.Q. An Internet of Things-enabled BIM platform for on-site assembly services in prefabricated construction. *Autom. Constr.* **2018**, *89*, 146–161. [CrossRef]
22. Ding, L.Y.; Zhou, C.; Deng, Q.X.; Luo, H.B.; Ye, X.W.; Ni, Y.Q.; Guo, P. Real-time safety early warning system for cross passage construction in Yangtze Riverbed Metro Tunnel based on the internet of things. *Autom. Constr.* **2013**, *36*, 25–37. [CrossRef]
23. Wang, M.; Qiu, S.; Dong, H.; Wang, Y. Design an IoT-based building management cloud platform for green buildings. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 5663–5667.
24. Aliev, K.; Antonelli, D.; Bruno, G.; Piantanida, P.; Villa, V. Building system diagnostics through a network of smart local sensors. In Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Preveza, Greece, 24–26 September 2021; pp. 1–6.
25. Villa, V.; Naticchia, B.; Bruno, G.; Aliev, K.; Piantanida, P.; Antonelli, D. IoT Open-Source Architecture for the Maintenance of Building Facilities. *Appl. Sci.* **2021**, *11*, 5374. [CrossRef]
26. Gao, X.; Pishdad-Bozorgi, P. A framework of developing machine learning models for facility life-cycle cost analysis. *Build. Res. Inf.* **2020**, *48*, 501–525. [CrossRef]
27. Bouabdallaoui, Y.; Lafhaj, Z.; Yim, P.; Ducoulombier, L.; Bennadji, B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. *Sensors* **2021**, *21*, 1044. [CrossRef]
28. Assaf, S.; Srour, I. Using a data driven neural network approach to forecast building occupant complaints. *Build. Environ.* **2021**, *200*, 107972. [CrossRef]
29. Mattera, C.G.; Quevedo, J.; Escobet, T.; Shaker, H.R.; Jradi, M. A method for fault detection and diagnostics in ventilation units using virtual sensors. *Sensors* **2018**, *18*, 3931. [CrossRef]
30. García, S.; Luengo, J.; Herrera, F. *Data Preprocessing in Data Mining*; Springer International Publishing: Cham, Switzerland, 2015; Volume 72.
31. Twala, B.; Cartwright, M. Ensemble missing data techniques for software effort prediction. *Intell. Data Anal.* **2010**, *14*, 299–331. [CrossRef]
32. Myrtveit, I.; Stensrud, E.; Olsson, U.H. Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods. *IEEE Trans. Softw. Eng.* **2001**, *27*, 999–1013. [CrossRef]
33. Al Shalabi, L.; Shaaban, Z.; Kasasbeh, B. Data mining: A preprocessing engine. *J. Comput. Sci.* **2006**, *2*, 735–739. [CrossRef]
34. Daoud, J.I. Multicollinearity and regression analysis. *J. Phys. Conf. Ser.* **2017**, *949*, 012009. [CrossRef]
35. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013; Volume 26, p. 13.
36. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; Chapman and Hall/CRC: New York, NY, USA, 2015. [CrossRef]
37. Wei, M.; Chow, T.W.; Chan, R.H. Heterogeneous feature subset selection using mutual information-based feature transformation. *Neurocomputing* **2015**, *168*, 706–718. [CrossRef]
38. H2O Platform. Available online: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/performance-and-prediction.html#mse-mean-squared-error (accessed on 1 October 2021).
39. Halvari, T.; Nurminen, J.K.; Mikkonen, T. Testing the Robustness of AutoML Systems. *arXiv* **2020**, arXiv:2005.02649. Available online: https://arxiv.org/pdf/2005.02649.pdf (accessed on 4 January 2022). [CrossRef]
40. Olson, R.S.; Moore, J.H. TPOT: A tree-based pipeline optimization tool for automating machine learning. In *Workshop on Automatic Machine Learning*; Springer: New York City, NY, USA, 24 June 2016; pp. 66–74.
41. Jin, H.; Song, Q.; Hu, X. Auto-keras: An Efficient Neural Architecture Search System. Available online: https://arxiv.org/pdf/1806.10282.pdf (accessed on 4 January 2022).
42. LeDell, E.; Poirier, S. H2o automl: Scalable automatic machine learning. In Proceedings of the AutoML Workshop at ICML, Virtual, 18 July 2020; Volume 2020.
43. Milutinovic, M.; Schoenfeld, B.; Martinez-Garcia, D.; Ray, S.; Shah, S.; Yan, D. On evaluation of automl systems. In Proceedings of the ICML Workshop on Automatic Machine Learning (Virtual Conference), Virtual, 12–18 July 2020. Available online: https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_59.pdf (accessed on 4 January 2022).
44. Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J.T.; Blum, M.; Hutter, F. Auto-sklearn: Efficient and robust automated machine learning. In *Automated Machine Learning*; Springer: Cham, Switzerland, 2019; pp. 113–134.
45. Kotthoff, L.; Thornton, C.; Hoos, H.H.; Hutter, F.; Leyton-Brown, K. Auto-WEKA: Automatic model selection and hyperparameter optimization in WEKA. In *Automated Machine Learning*; Springer: Cham, Switzerland, 2019; pp. 81–95.
46. Natephra, W.; Motamedi, A. BIM-based live sensor data visualization using virtual reality for monitoring indoor conditions. In *Intelligent & Informed—Proceedings of the 24th CAADRIA Conference*; Haeusler, M., Schnabel, M.A., Fukuda, T., Eds.; Victoria University of Wellington: Wellington, New Zealand, 2019; Volume 2, pp. 191–200.
47. Wehbe, R.; Shahrour, I. Use of BIM and smart monitoring for buildings' indoor comfort control. In Proceedings of the MATEC Web of Conferences, Lille, France, 8–10 October 2019; Volume 295.