RESEARCH ARTICLE OPEN ACCESS

Dataset analysis for energy efficiency to use Machine Learning

Vázquez Rufino Erick
Posgrado CIATEQ A.C., Av. del Retablo 150, Col Constituyentes Fovissste, Querétaro, México. 76150

ABSTRACT
The analysis of the cooling and heating load through software tools such as Weka through classifiers is essential to design efficient, sustainable and economical air conditioning systems in buildings, which could have also an impact to reduce a high initial investment.

***Keywords* –** Analysis, Efficiency, Energy, Multilayer, Perceptron.

Date of Submission: 09-07-2024 Date of acceptance: 23-07-2024

I. INTRODUCTION

The interest in carrying out this research arose from various projects developed for different commercial or industry clients, where historical information has been generated.

Currently, the analysis to reduce the cooling or heating load in buildings is necessary to mainly lower the costs associated with electrical energy consumption and to reduce the carbon footprint produced by the equipment used for these processes. It is also very important to calculate the loads to avoid oversizing that could impact the design and a high initial investment [1].

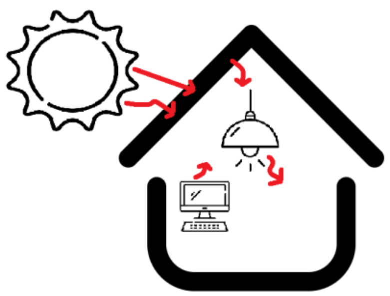

Fig. 1. Example of internal and external loads.

II. RELATED WORKS

The purpose of detection algorithms is to identify where an anomaly has occurred and classify or infer the cause.

[2] Anomaly detection is a Data Mining technique that allows the recognition of patterns that do not behave in the expected way in the data with a wide spectrum of applications, they can be classified according to the nature of the input, the type of the anomaly, the labeling of the data, or the type of output returned by the method. The input to these methods can be univariable or multivariable. There are three types of anomalies: specific, contextual and collective. Depending on the degree of tag availability, anomaly detection methods can also be classified into one of the following three modes: supervised, semi-supervised and unsupervised anomaly detection.

[3] Predictive maintenance has emerged as an ideal approach to save costs and prevent equipment failure in the industry. Traditional reactive maintenance only carries out maintenance activities after fault detection. Generalized preventive maintenance involves periodic maintenance activities based on previous experience about the frequency of failure. The predictive maintenance approach is the alternative of Industry 4.0, failures are predicted based on the information received in real time from sensors in the industry.

[4] Supervised neural network (NN): It is derived from the mechanism of neural network biology and is capable of revealing complex nonlinear relationships between input and output pairs. A neural network has an input layer, an output layer, and one or more hidden layers, with different numbers of neurons in each layer.

[5] Artificial Neural Network (ANN): An ANN consists of different interconnected sections, which can be single or multiple layers. In ANN, input data is sent to input neurons (synapses), where these inputs are weighted by the software. The weighted sum is then operated by an activation function and the data output is fed to other neurons in the network. All neurons are highly connected, therefore activation values can build the result or can

even feed into a following model. These connection weights are modified during training of neurons to obtain an optimal model, and the interpolation of training patterns is presented to the network during training to obtain the desired accuracy.

[6] A set of wavelet features are inserted into the algorithm that are extracted from the vibration signals. The results obtained from the Support Vector Machine (SVM) classifier, uses the continuous wavelet transform of several families (Daubechies, Coiflet, biorthogonal, inverse biorthogonal, Symlets, Meyer wavelet, Morlet and Gaussian), the best versions are compared to use the best global wavelet and family, in this case the Daubechies, with an efficiency of 99.84%. It is tested as normal and faults like cavitation, bearing failure, impeller failure, impeller and bearing failure are injected together. Almost all the characteristics of the wavelet families show excellent results when using the SVM algorithm, so it can be considered a solid candidate for the diagnosis of failures in centrifugal pumps. We have to review this methodology since a similar study shows another percentage of efficiency, so similar experiments should be done to verify the efficiency and the reported results.

III. DESCRIPTION OF THE EXPERIMENTATION CARRIED OUT

Dataset description

The dataset is multivariable, it has 8 attributes with integer or real values (from X1 to X8), it has 2 classes or dependent variables (y1 and y2), it has 768 instances and has no missing values.

X1	Relative Compactness
X2	Surface area
X3	Wall area
X4	Roof area
X5	Total height
X6	Orientation
X7	Glazed area
X8	Distribution of the glazed area
y1	Heating load
y2	Cooling load

Table 1. Description of the attributes and classes of the dataset.

To create the dataset, energy analyzes were carried out using 12 different building shapes simulated in the Ecotect software. The buildings differ in terms of the glass surface, the distribution of glass surface and its orientation, among other parameters. Various scenarios were simulated based on the characteristics to obtain 768 building shapes.

The data set comprises 768 samples and 8 features, with the goal of predicting two real-valued responses. [7]

Presentation of the evaluation metrics used

The available classifiers were analyzed with cross validation and 10 folds, the best 3 were changed configuration parameters to find the one that gives the best correlation coefficient percentage and lowest error.

The best is Multilayer Perceptron, the 3 experiments were tested using training set, increasing from 500 to 1000 in training time and only using as attribute class the heating load y1.

All the tests are carried out using Weka software. Which is very easy to use and contains many options for analysis and experimentation [8].

General description of the algorithm and/or model proposed for the solution

The Multilayer Perceptron consists of different interconnected sections, which can be single or multiple layers, the input data is sent to the input neurons (synapses), where they are weighted. The weighted sum is then operated by an activation function and the data output is fed to other neurons in the network. All neurons are highly connected, therefore activation values can build the result or feed into a next model. These connection weights are modified during training to obtain an optimal model [9].

The objective of this type of analysis is to visualize and evaluate the performance of heating and cooling operations [10].

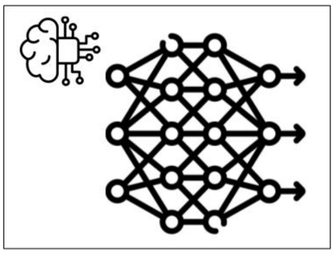

Fig. 2. Example of multilayer perceptron.

Comparison of experimental results with other two algorithms or parameter settings.

The first experiment using a Multilayer Perceptron using a training set and a training time of 500.

Time taken to build model: 0.15 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correlation coefficient	0.9968
Mean absolute error	0.6452
Root mean squared error	0.821
Relative absolute error	7.0555 %
Root relative squared error	8.1422 %
Total Number of Instances	768

The second experiment using a Multilayer Perceptron using a training set and a training time of 1000.

Time taken to build model: 0.34 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correlation coefficient	0.9968
Mean absolute error	0.6409
Root mean squared error	0.8246
Relative absolute error	7.0087 %
Root relative squared error	8.1777 %
Total Number of Instances	768

The third experiment using a Multilayer Perceptron using a training set and a training time of 1000 and with erroneous data.

Time taken to build model: 0.35 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correlation coefficient	0.9716
Mean absolute error	2.47
Root mean squared error	3.1742
Relative absolute error	27.011 %
Root relative squared error	31.479 %
Total Number of Instances	768

IV. RESULTS AND DISCUSSION

The analysis of thermal load is a fundamental stage in the design of energy efficient and environmentally responsible buildings. Allows you to optimize air conditioning systems, reduce operating costs and minimize environmental impact. By visualizing satisfactory results with the Weka software, it allows us to test the different models and those that are most appropriate depending on the type of historical data set we have to carry out any experimentation. Therefore we focus on the most accurate models available.

V. CONCLUSION

It is concluded that the Multilayer Perceptron through the experiment that gives the highest correlation coefficient, we are certain that it is a good model and we subsequently could verify it by inserting error data into the dataset.

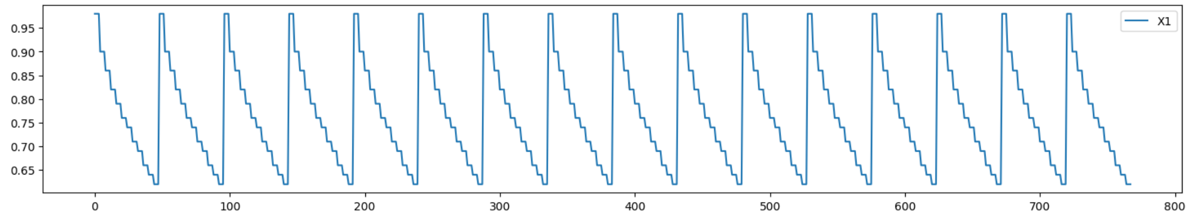

Fig. 3. Dataset X1 component visualization.

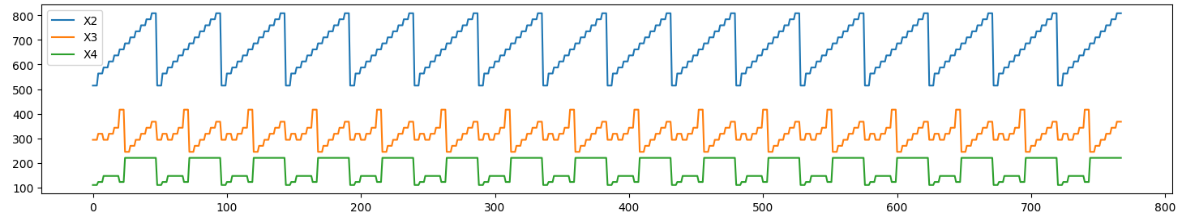

Fig. 4. Dataset X2, X3, X4 component visualization.

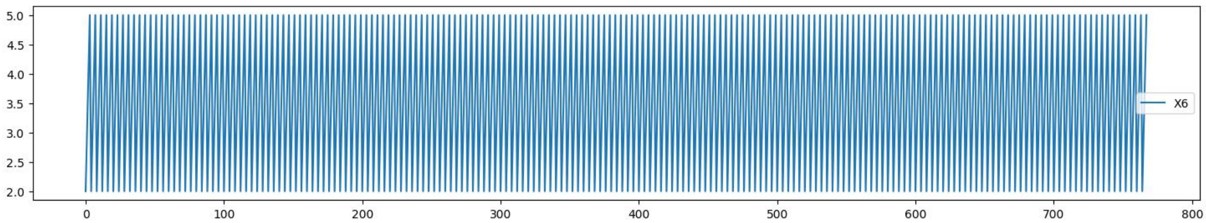

Fig. 5. Dataset X6 component visualization.

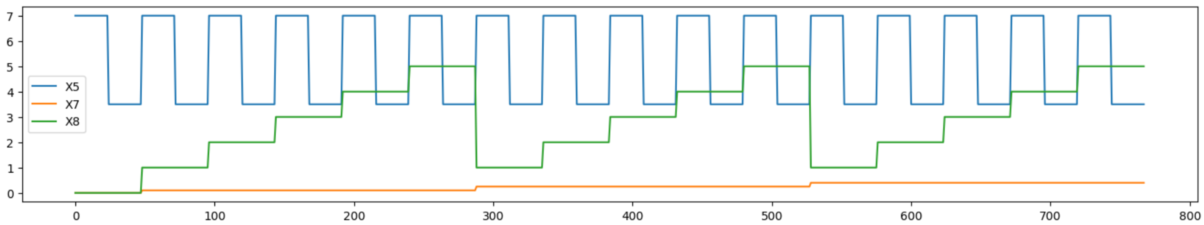

Fig. 6. Dataset X5, X7, X8 component visualization.

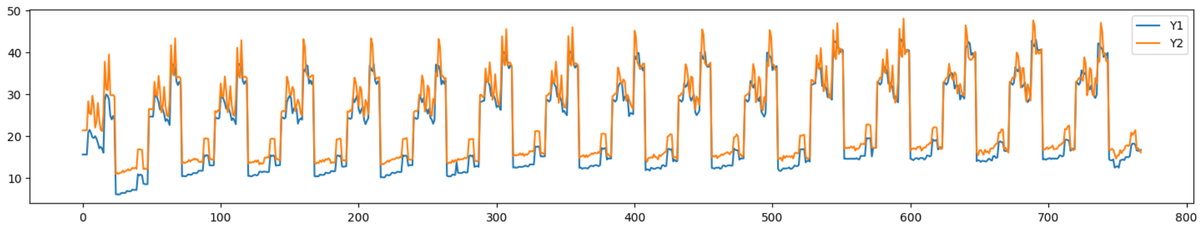

Fig. 7. Dataset Y1, Y2 component visualization.

REFERENCES

[1]. “Balance térmico para sistemas HVAC. Importancia y requerimientos | KINENERGY”. https://www.kin.energy/blogs/post/balance-t%C3%A9rmico-para-sistemas-hvac.-importancia-y-requerimientos (consulted on February 7, 2023).
[2]. L. L. A. y Niusvel Acosta Mendoza y Andrés Gago Alonso, “Detección de anomalías basada en aprendizaje profundo: Revisión”, Revista Cubana de Ciencias Informáticas, vol. 13, núm. 3, 2020, https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=1874
[3]. L. Magadán, F. J. Suárez, J. C. Granda, y D. F. García, “Real-Time Monitoring of Electric Motors for Detection of Operating Anomalies and Predictive Maintenance”, en Science and Technologies for Smart Cities, H. Santos, G. V. Pereira, M. Budde, S. F. Lopes, y P. Nikolic, Eds., Cham: Springer International Publishing, 2020, pp. 301–311.
[4]. Z. Ding, W. Zhang, y D. Zhu, “Neural-network based wind pressure prediction for low-rise buildings with genetic algorithm and Bayesian optimization”, Eng Struct, vol. 260, p. 114203, jun. 2022, doi: 10.1016/J.ENGSTRUCT.2022.114203.
[5]. J. F. Saldarriaga, “Application of an artificial neural networks for predicting the heat transfer in conical spouted bed using the Nusselt module”, Heliyon, vol. 8, núm. 11, p. e11611, nov. 2022, doi: 10.1016/J.HELIYON.2022.E11611.
[6]. V. Muralidharan, V. Sugumaran, y V. Indira, “Fault diagnosis of monoblock centrifugal pump using SVM”, Engineering Science and Technology, an International Journal, vol. 17, núm. 3, pp. 152–157, sep. 2014, doi: 10.1016/J.JESTCH.2014.04.005.
[7]. “UCI Machine Learning Repository: Energy efficiency Data Set”. https://archive.ics.uci.edu/ml/datasets/Energy+efficiency (consulted on February 7, 2023).
[8]. E. Frank, M. A. Hall, I. H. Witten, y M. Kaufmann, “WEKA Workbench Online Appendix for ‘Data Mining: Practical Machine Learning Tools and Techniques’”, 2016.
[9]. J. F. Saldarriaga, “Application of an artificial neural networks for predicting the heat transfer in conical spouted bed using the Nusselt module”, Heliyon, vol. 8, núm. 11, p. e11611, nov. 2022, doi: 10.1016/J.HELIYON.2022.E11611.
[10]. M. Taha, Y. Id, S. Basurra, y M. M. Gaber, “Edge Machine Learning: Enabling Smart Internet of Things Applications”, doi: 10.3390/bdcc2030026.