ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 11, Issue, 10, pp. 51346-51350, October, 2021
https://doi.org/10.37118/ijdr.23097.10.2021

RESEARCH ARTICLE **OPEN ACCESS**

DATA MINING APPLIED TO ABNORMALITY PREDICTION IN ELECTRICAL SUBSTATION TRANSFORMERS

Matheus José da Silva[1,*], Starch Melo de Souza[2], Israel Cavalcante de Lucena[3], Hemir da Cunha Santiago[2], Eldrey Seolin Galindo[3], Antônio Janael Pinheiro[3], Luciana de Albuquerque Romeiro França[3], Patricia Drapal da Silva[2] and Lorrany Fernanda Lopes da Silva[1]

[1]Integração Transmissora de Energia S.A. (INTESA)
[2]In Forma Software
[3]Centro de Estudos e Sistemas Avançados do Recife (CESAR)

ARTICLE INFO

Article History:
Received 17th August, 2021
Received in revised form
20th September, 2021
Accepted 19th October, 2021
Published online 30th October, 2021

Key Words:
Substation,
Predictive Maintenance,
Machine Learning,
Artificial intelligence.

***Corresponding author:**
Matheus José da Silva

ABSTRACT

Ensuring that assets in an electrical substation are managed only when needed is essential to minimizing maintenance costs. In this work, we present a system based on Artificial Intelligence (AI) to contribute to decision making about the performance of predictive maintenance. This system will provide substation operators with additional indicators of the operational condition of the transformers, determining which equipment is most likely to fail in the short term, based on analog and digital data from the substation supervision system. Since AI techniques require that the data obtained have the quality to provide satisfactory results, data mining techniques were applied to the records of equipment in a substation. During the analysis of the data provided, several signs of unwanted events were observed, such as abrupt fluctuations in oil temperature and variations in the frequency of alarms. In the analog data, we have records of fluctuations in voltage, current and oil temperature that are relevant to indicate the operating status of the equipment, in addition to the digital data provide information on alarms related to the equipment, which are records when the equipment exceeds some predefined safety threshold. This information was used to provide a broader view of the behavior of equipment during the construction of the predictive model, improving the detection of abnormalities. The system presented in this work models the typical behavior of the equipment, through the information mentioned in the previous paragraph. When predictions differ significantly from expected values, the system signals the operator of the presence of potential anomalies. The results obtained reveal the potential of the system, helping decision making in predictive maintenance of transformers. The development of an innovative methodology to extract knowledge from the supervisory system records proved to be a favorable field for future investigations.

Citation: ***Matheus José da Silva, Starch Melo de Souza, Israel Cavalcante de Lucena, Hemir da Cunha Santiago et al.*** "Data Mining Applied to Abnormality Prediction in Electrical Substation Transformers", *International Journal of Development Research*, 11, (10), 51346-51350.

INTRODUCTION

Electric power systems (EPS) are a vital component in the infrastructure that sustains modern society, providing numerous benefits, however, their lack produces several negative impacts in multiple areas, such as social, environmental and economic (Weedy, 1987). Also, according to (Zampolli, 2019), the increase in the period of activity of the equipment brings benefits to the company and society. Therefore, it is essential to avoid unscheduled interruptions in the supply of electricity, reducing the impacts on society and on concessionaires, since these are remunerated based on the availability of electrical assets. The substations are part of the EPSfacilities, being responsible for the transformation, protection, control and electrical maneuvering. According to (Weedy, 1987), the main assets within a substation are: transformers, circuit breakers, lightning rods, selector switches and reclosers.

For the proper functioning of substations, the management of these assets is essential, thus corrective, preventive and predictive maintenance are part of the assets' routine (Carvalho, 2019). In particular, the potential transformer is one of the most important pieces of equipment in electrical substations, since its failure causes serious damage, consequently, its predictive maintenance becomes extremely relevant for the sector (Liao, 2011). This importance motivated the choice of the transformer as the target asset of the solution presented in this work. In order to reach the full potential of managing these transformers, it is necessary to collect data that reflect the operational condition of these equipment, which includes the typical and abnormal situation, allowing the prediction of failures or even complete shutdowns of the equipment (Ravi, 2019). Among the main electrical quantities of transformers, current, voltage and temperature of the winding and oil stand out, which are collected through Supervisory Control And Data Acquisition systems (SCADA) (Contreras-Valdes, 2020; Peharda, 2017). The data collected by SCADA is very valuable for asset management (Peharda, 2017), as long as this information is properly handled. To find the value contained in the collected data, it is necessary to use techniques to build knowledge from them, such as Data Mining and Artificial Intelligence (AI). These techniques support expressive innovations in different areas, while presenting countless opportunities in the management of electrical equipment maintenance. Despite the transformational potential, AI introduces significant challenges to its adoption in electrical substations. For example, information from different sources, such as SCADA, management and operation systems, communication networks and equipment robustness. To deal with these challenges, this work presents a system for predicting abnormalities occurring in electrical substation transformers, this system incorporates information collected from the equipment, applies the necessary procedures to ensure data quality and, finally, uses this information in construction of predictive models based on AI.

The results presented in this work demonstrate the great potential of the presented system, being able to predict equipment temperature alarms and failures in its protection system. These findings can help operators make decisions about conducting predictive maintenance.
This paper is organized as follows: Section II presents the theoretical guidelines used in this work and describes the most relevant models used for the proposed technical solution. Section III presents the materials and methods from the innovative analytical tool to optimize energy asset renewal decisions. Section IV presents the results obtained using the tool developed in this work to the maintenance management of power transformers. Finally, Section V presents this work's conclusions.

THEORETICAL GUIDELINES

In this section, the theoretical guidelines used in the model proposed by this work will be described, in addition to the general specification of the created solution.

Artificial Intelligence: Artificial Intelligence (AI) is an area of computing that seeks to develop algorithms capable of making decisions or assisting in decision making. AI is divided into three main areas: supervised, unsupervised and reinforcement learning (Patel, 2019). Each of these areas has its advantages and disadvantages, and the choice to adopt them in a project depends on the problem addressed. These areas are further divided into several fields that seek to solve specific objectives, such as: object detection, event prediction, among others. These solutions are mostly based on statistical methods and mathematical optimizations (Nilsson, 1980). For predicting abnormalities, unsupervised learning has shown very promising results in different areas of knowledge. Among the fields of unsupervised learning are: clustering, dimensionality reduction, attribute extraction and time series analysis (Patel, 2019).

Time Series: Time Series is one of the subfields of statistics, widely used in several areas, such as: economics, marketing, social sciences, modeling focuses on statistical and mathematical analysis of the behavior of phenomena or objects. A time series can be considered any sequence of observations of one or more variables (sensors) over time. Generally, these observations have a regularity, being observed at constant intervals. Figure 1 illustrates a time series of the variable oil temperature of a power transformer, where we can observe that in a given period of time there are continuous records of temperature and that these oscillate in a way to present typical and abnormal patterns.

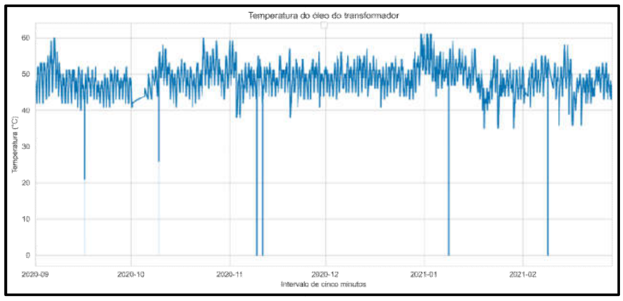

Figure 1. Time series of transformer oil temperature (Authors)

In the context of time series, an abnormality consists of a measurement that departs significantly from other observations, which produces evidence of the occurrence of an atypical event in the process that generates the data (Blázquez-García, 2020). The abnormal oscillations illustrated in Figure 1 suggest that some unwanted event induced abrupt changes in the temperature of a transformer's oil, where in the prediction of these variations being one of the use cases of the time series.

Clustering Algorithms: As mentioned in section II.A, the clustering area is a sub-area of the AI. In this work, clustering algorithms were used to predict transformer abnormalities. This type of algorithm is intended to identify the underlying structure of the data to which it is applied. Thus, these algorithms are capable of grouping data samples without the need for prior knowledge of them. This approach is very useful in detecting abnormalities, because due to the complexity and volume of information analyzed, labels that identify the data as typical or abnormal are rarely available (Blázquez-García, 2020). In the process of building the predictive models, the K-Means algorithm (Likas, 2003) was applied to the training set data (Eke, 2017). This algorithm requires specifying the number of clusters into which measurements should be separated, these clusters are defined by common standards between the data. As the number of groups is the most relevant parameter of the K-Means algorithm, and influences the results of the models, we applied the Elbow method (Kodinariya, 2013) to select the number of groups in which the data were separated.

CRISP-DM Methodology: The CRISP-DM methodology (Cross Industry Standard Process for Data Mining) provides an overview of the lifecycle of a data mining project. This methodology is adopted in projects in various industrial areas, such as the electricity, logistics and food sectors (Wirth, 2000). This methodology proposes the division of a project into six phases. The activities carried out in phases are interactive and a phase can be revisited multiple times, as identified in Figure 2. The adoption of this methodology guarantees the systematization of activities and contributes to the management of a data mining project.

Dataset: For the development of this work, a dataset from the SCADA of a substation of the Integração Transmissora de Energia S.A. (Intesa) company was obtained. This database contains digital (general events, alarms and TRIP) and analog data from a power transformer, namely: winding temperature, oil temperature, A, B and C phase current and tap position. Data is collected by the SCADA system every 5 minutes and is comprised in the period from 09/01/2020 to 02/28/2021. Previous studies indicate that the winding temperature is the most relevant magnitude of a transformer, which may indicate possible equipment downtime (Rodrigues, 2020).

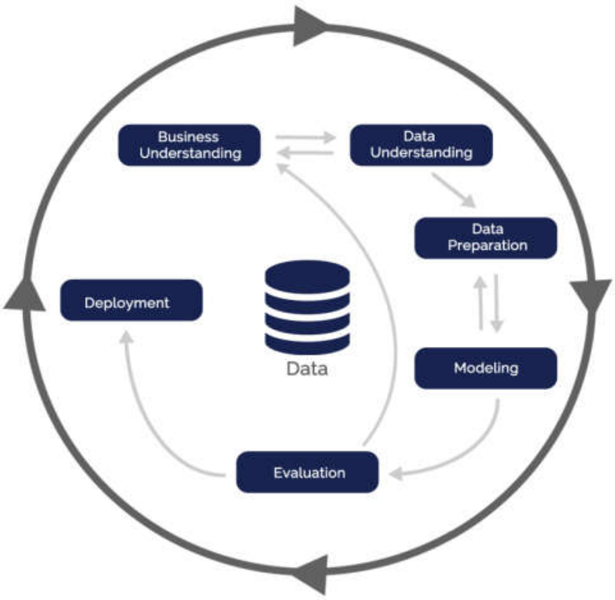

Figure 2. CRISP-DM Diagram (Wirth, 2000)

Therefore, the experiments presented in this article take into account only the winding temperature and winding alarm data, in order to identify possible failures in the equipment or in the temperature monitoring system. When analyzing the winding temperature data, we can see that the measurements behave similarly to the Normal curve, with a maximum value of 65°C, a minimum of 35°C, the mean and median around 51°C, as shown in Figure 3.

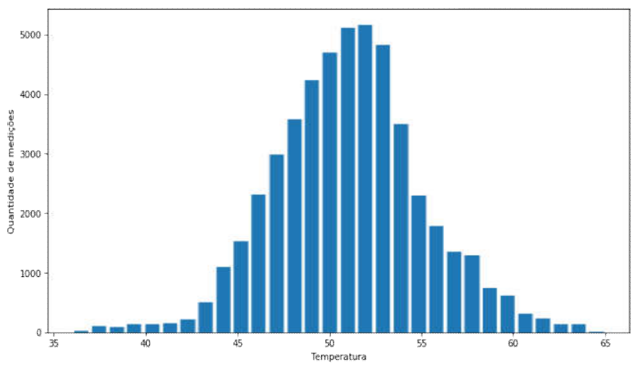

Figure 3. Example of a winding temperature histogram (Authors)

These values are in accordance with the standards ABNT NBR 5416/1997 and ABNT NBR 5440/1999 (ABNT, 1997; ABNT, 1999). According to ABNT's NBR 5416/1997 and NBR 5440/1999 standards, 55 °C class transformers have an average winding temperature of 55 °C and a maximum temperature of 65 °C. Analyzing the alarm data, we identified that there are alarms at low temperature moments (e.g. below 51°C) and high temperature moments without the occurrence of alarms, characterizing a non-compliance of the protection system with the standards and a risk to the transformer. Thus, the temperature limits defined in the aforementioned ABNT standards were considered in this paper.

MATERIALS AND METHODS

For the system presented in this work, a process was defined that consists of a continuous flow of data analysis divided into 4 large steps, as illustrated in Figure 4. The first step consists of collecting SCADA data, which is processed in the second step , which consists of removing noisy data, such as non-numeric values and negative temperature. Another processing done is the standardization of values ensuring that all values of the same variable are in the same format (Ex: dates in Brazilian and American standards).

In the third stage, the analysis and prediction of anomalies is carried out, being considered the main contribution of this work (it will be further detailed in section III.A. Finally, in the fourth stage, the system will produce a predictive alarm, in order to warn in advance of the need for maintenance on the equipment (maintenance parameters defined by the operator), thus allowing the scheduling of maintenance and, if necessary, a scheduled shutdown.

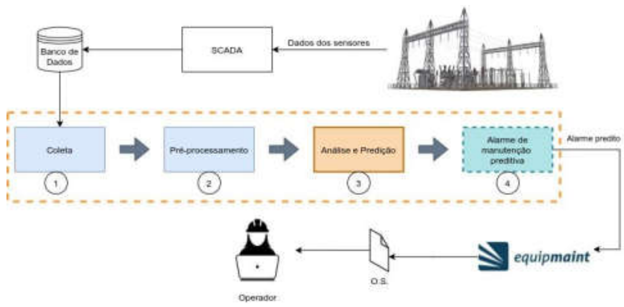

Figure 4. General application flow diagram (Authors)

Analysis and Prediction: The data analysis process starts with the selection of variables used in modeling the typical behavior of transformers. First, we consider the current, the temperature of the oil and the winding. In selecting the relevant variables, we calculated the Pearson correlation coefficient (Sedgwick, 2012), which suggested a strong correlation between the analyzed variables. Due to this result and based on studies carried out, we chose to use the winding temperature to model the typical and abnormal behavior of the transformer. Reducing the amount of variables has numerous benefits, such as faster decision making. After selecting the winding temperature as the target of investigations, we identified missing and outliers, such as winding temperature below ambient temperature and missing measurements at specific times. This type of incongruous measurement can negatively interfere with the performance of AI models. Therefore, we apply techniques to replace missing values and outliers with the average of measurements. It is noteworthy that this procedure is widely used in data mining projects.

A good practice in time series modeling for predicting abnormalities is to divide the measurements into two groups, training and testing. Data was divided into these groups based on the date when the measurements were recorded. The training data, from the period between 09/01/2020 to 01/31/2021, were used to build the predictive models. In turn, the test measurements, referring to the period between 02/01/2021 and 02/28/2021, were used to assess the capacity of the built model to detect potential abnormalities in the winding temperature. In the clustering step, the K-Means algorithm was configured to divide the training data into five groups, after applying the Elbow method (see subsection II.C. This combination of the K-Means algorithm and training data results in a trained model. Subsequently, this model was applied to test measurements in order to identify the group(s) in which potentially abnormal values reside. The analysis of these groups contributes to the identification of points of attention, such as high temperature and warm-up period. By determining the moments in which these abnormalities occur, the model contributes to the prediction of unwanted events, such as transformer wear due to high temperatures.

RESULTS

The proposed system presents promising results, being able to identify temperature thresholds autonomously (without the need for operator intervention). Therefore, its methodology is considered robust to human error and allows the analysis of a large amount of data effectively. The result of this clustering for data obtained from February 2021 can be seen in Figure 5, where the blue clusters represent the equipment at its normal temperature, while the orange clusters represent heating operation. In turn, the horizontal and vertical lines show the average temperature (58°C) of the high temperature cluster and the occurrences of alarms in the protection system, respectively.

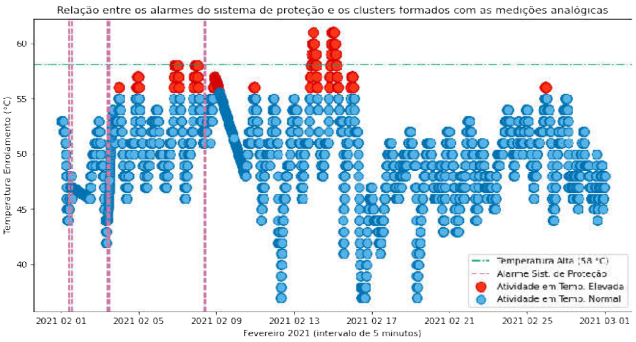

Figure 5. Relation between protection system alarms and two clusters formed with analog measurements (Authors)

From clustering, the method is able to group the temperature based on your measurements. This clustering gathers temperatures close to each other within the same cluster, while highlighting measurements belonging to the group that contain potentially critical values such as high temperatures. Thus, an alarm can be triggered when the temperature is assigned by the model to a group that gathers significantly high values. This alarm allows the operator to make the decision whether or not to initiate predictive maintenance. Literature presents configuration errors generated by human intervention or failures that occurred in the protection system (Bacega, 2009; Lima, 2015). An expert in the electricity sector confirms the occurrence of failures in the substation protection system. He suggests that desynchronization between digital and analog data can occur, but with delays that shouldn't exceed a few seconds. The identified problem suggests a failure in the protective equipment, which is an unwanted situation. The proposed solution will help to identify situations of desynchronization between analog and digital data. The results obtained also help to determine the duration of the moments when the temperature remains in a critical range. An upward trend in the duration of moments with high temperature can be observed, as illustrated in Figure 6. These moments were recorded in February 2021. This figure shows that the equipment remains for longer periods at high temperatures over time . This type of information can help maintenance personnel to determine potential transformer overheating, which accelerates equipment wear. This knowledge contributes to the maintenance management of power transformers.

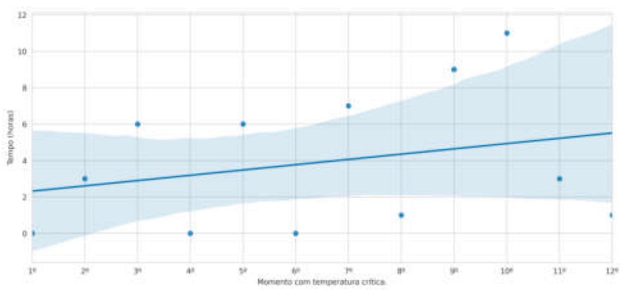

Figure 6. Elevation trend in the duration of the moments when the transformer temperature is high (Authors)

CONCLUSION

The predictive system presented in this work models the typical behavior of power transformers, autonomously, based on the winding temperature collected by the substation supervisory system. The results obtained demonstrate that the system is promising to assist operators in decision making and in the maintenance management of substation power transformers, reducing unscheduled equipment shutdowns and increasing benefits for companies and society. The results obtained also show the potential and low cost of extracting knowledge and value contained in analog and digital data collected by SCADA. Since SCADA systems are common in substations, the proposed method establishes a new indicator for decision-making on predictive maintenance, while avoiding introducing costs with the acquisition of sensors, using existing data. Given the necessary adjustments, the predictive system is viable to be replicated in other electrical energy transmission equipment. As future work, the system will be extended to other equipment and substations, enhancing the useful life of the substation equipment as a whole.

Acknowledgment

The authors acknowledge the financial support of Integração Transmissora de Energia S.A. (INTESA) through the ANEEL R&D Program for the development of the research project entitled: "PD-05456-0002/2019 Intelligent predictive maintenance system based on the automation of the electrical testing process in substation equipment without sensing".

REFERENCES

ABNT – Associação Brasileira de Normas Técnicas; NBR 5416 – Aplicação de cargas em transformadores de potência – Procedimento; Julho, 1997.

ABNT – Associação Brasileira de Normas Técnicas; NBR 5440 Transformadores para Redes Aéreas de Distribuição. Características Elétricas e Mecânicas - Padronização, ABNT, 1999.

Bacega, W. R. (2009). Modernização Da Proteção Térmica De Transformadores De Potência.

Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2020). A review on outlier/anomaly detection in time series data. arXiv preprint arXiv:2002.04236.

Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024.

Contreras-Valdes, A., Amezquita-Sanchez, J. P., Granados-Lieberman, D., & Valtierra-Rodriguez, M. (2020). Predictive data mining techniques for fault diagnosis of electric equipment: a review. Applied Sciences, 10(3), 950.

Eke, S., Aka-Ngnui, T., Clerc, G., & Fofana, I. (2017, August). Characterization of the operating periods of a power transformer by clustering the dissolved gas data. In 2017 IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED) (pp. 298-303). IEEE.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.

Liao, R., Zheng, H., Grzybowski, S., Yang, L., Zhang, Y., & Liao, Y. (2011). An integrated decision-making model for condition assessment of power transformers using fuzzy approach and evidential reasoning. IEEE Transactions on power delivery, 26(2), 1111-1118.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. Pattern recognition, 36(2), 451-461.

Lima, G. S. D. (2015). Análise da importância de controladores digitais na medição da temperatura do enrolamento em transformadores a óleo (Bachelor's thesis, Universidade Tecnológica Federal do Paraná).

Nilsson, N. J. (1982). Principles of artificial intelligence. Springer Science & Business Media.

Patel, A. A. (2019). Hands-on unsupervised learning using Python: how to build applied machine learning solutions from unlabeled data. O'Reilly Media.

Ravi, N. N., Drus, S. M., & Krishnan, P. S. (2019, April). Data mining techniques for transformer failure prediction model: A systematic literature review. In 2019 IEEE 9th Symposium on

Computer Applications & Industrial Electronics (ISCAIE) (pp. 305-309). IEEE.

Rodrigues, G. A., Araujo, B. V. S., & Ferreira, T. V. (2020, December). Gêmeos Digitais e Método dos Elementos Finitos, um Estudo de Caso: Mapeamento Térmico de Transformador. In Congresso Brasileiro de Automática-CBA (Vol. 2, No. 1).

Sedgwick, P. (2012). Pearson's correlation coefficient. Bmj, 345.

Weedy, B. M., Cory, B. J., Jenkins, N., Ekanayake, J. B., & Strbac, G. (2012). Electric power systems. John Wiley & Sons.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1). London, UK: Springer-Verlag.

Zampolli, M. (2019). Gestão de Ativos: guia para a aplicação da norma ABNT NBR ISO 55001 considerando as diretrizes da ISO 55002: 2018. International Copper Association Brazil, 2.

****