# Automatic Anomaly Detection in Vibration Analysis based on Machine Learning Algorithms

Pedro Torres[1,2], Armando Ramalho[1,3], and Luis Correia[1]

[1] Instituto Politécnico de Castelo Branco, Castelo Branco, Portugal
{pedrotorres,aramalho,lcorreia}@ipcb.pt,
[2] SYSTEC - Research Center for Systems & Technologies, Porto, Portugal
[3] CEMMPRE - Centre for Mechanical Engineering, Materials and Processes, Coimbra, Portugal

**Abstract.** This paper presents an approach for automatic anomaly detection through vibration analysis based on machine learning algorithms. The study focuses on induction motors in a predictive maintenance context, but can be applied to other domains. Vibration analysis is an important diagnostic tool in industrial data analysis to predict anomalies caused by equipment defects or in its use, allowing to increase its lifetime. It is not a new technique and is widely used in the industry, however with the Industry 4.0 paradigm and the need to digitize any process, it gains relevance to automatic fault detection. The *Isolation Forest* algorithm is implemented to detect anomalies in vibration datasets measured in an experimental apparatus composed of an induction motor and a coupling system with shaft alignment/misalignment capabilities. The results show that it is possible to detect anomalies automatically with a high level of precision and accuracy.

**Keywords:** Industry 4.0, Anomaly Detection, Isolation Forest, Vibration Analysis, BigML

## 1 Introduction

According to the Europe Predictive Maintenance Market – Industry Trends and Forecast to 2027 report [1], the predictive maintenance market is expected to growing market with a CAGR (Compound Annual Growth Rate) of 39.6% in the forecast period 2020 to 2027. The increased use of new and emerging technologies to gain valuable insights into decision making has contributed to the growth of the predictive maintenance market. Several vertical end users are increasingly in need of cost reduction and downtime, which has spurred the growth of the predictive maintenance market.

Industry 4.0 marks the revolution of digitization of traditional manufacturing industries supported by modern technologies such as automation, interconnectivity, real-time data processing and intelligence based on machine learning techniques. With the growth of automation technologies, the sensing component is the basis of perception, fundamental to the smart factory concept. Condition monitoring techniques are maintenance methodologies to monitor the operating conditions of an equipment in real time, through measurement and extraction of information that allow understanding its health, wear, degradation or significant changes in operation. The collected data is used to find trends, predict failures and estimate the remaining lifetime of an asset.

Through vibration analysis, by analysing the frequency, amplitude, phase, position and direction of vibrations in machinery, it is possible to identify many common faults. For example it is possible differentiate between wear on a specific gear or bearing, a lack of lubrication, an imbalance, a misalignment, a loose mounting or an electrical fault. Fault detection can be carried out before a machine is stopped, reducing downtime to its bare minimum. Early detection and predictive maintenance can also prevent more serious faults from developing.

Machine learning algorithms have the ability to analyze large amounts of data, and automatically perform diagnoses, without human intervention, based on historic and correlations with failure situations, but also by self-learning. Increasingly, they are the appropriate tool for decision making with high levels of accuracy.

The Figure 1 shows a generalist architecture supported in the context of Industry 4.0, which illustrates the overview of fault detection systems in electrical machines based on vibration analysis and on which we base this work. The Operational Technology (OT) and Information Technology (IT) parts were aligned to design an Industrial Control System (ICS) in the laboratory, for acquiring, controlling and monitoring the operating status of rotating machines, producing reports, automatic alerts and recommending actions to take as a prescriptive maintenance system.

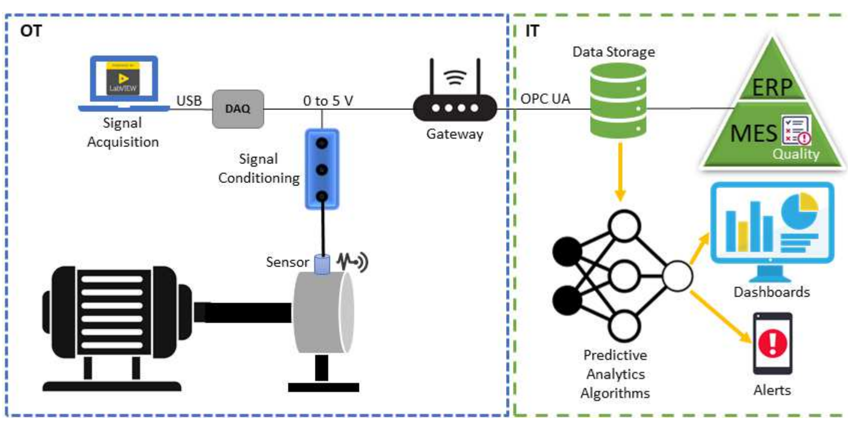


**Fig. 1.** Architecture overview for data acquisition and its interconnection with high-level industry management systems.

The remainder of this paper is organized as follows: Section 2 describes the state of the art and related work. Section 3 describes the materials and methods for the automatic anomaly detection. Section 4 presents the experimental results achieved and, finally, Section 5 presents the conclusions and future work.

## 2 Related Work

This section presents some scientific works related to vibration analysis in rotating machines using different machine learning techniques that help to support the validity of the work of this paper and its importance in the current context of the industry. Any vibration measurement experiment for diagnosing machine operation must be in line with ISO 22096:2007.

Vibration signals carry very important information for predictive maintenance applications, which is why it is widely used. The paper [2] presents and describes some condition maintenance techniques in a predictive maintenance context, in particular it shows an overview of some vibration analysis tools, such as ICA (independent component analysis), TFA (time-frequency analysis), ED (energy distribution) and CD (change detection).

In [3], anomaly detection techniques using machine learning models such as K- Nearest Neighbour (KNN), Support Vector Regression (SVR) and Random Forest (RF) have been applied to vibration data for early fault detection of industrial electric motors. According to the authors the Random Forest presented the best performance compared to SVR and KNN, based on less number of false positives and the detection time.

In [4], is proposed a method for providing a visual explanation of the predictions of a convolutional neural network (CNN)- based anomaly detection system. The CNN takes the monitoring target machine's vibrational data as input and predicts whether the target's state is healthy or anomalous.

For a relation between the motor speed and vibration signals, [5] proposes a CNN based deep learning approach for automatic motor fault diagnosis. In the same research line [6] establishes a comparison of fault motor diagnosis using RNN (Recurrent Neural Networks) and k-means in vibration analysis.

For accurate detection, it is important that the acquisitions and the entire experimental acquisition chain are also calibrated and that the best sensors are used for the best application. Accelerometers are widely used in this type of applications, however, other types of sensors such as magnetoresistive sensors have shown their potential due to their high sensitivity. In [7] was studied the comparison between magnetoresistive sensors and accelerometers in the acquisition of vibration signals to validate the accuracy in vibration analysis condition monitoring systems.

## 3 Materials and Methods

According to [8], vibration analysis methodology in rotating machines help to determine, unbalanced, misalignment, looseness, bearing faults, gear defects, belt wear an tear, pum cavitation and others. Analysis is usually performed by measuring mechanical movements with accelerometers or magnetic sensors. Typically the signal amplitude is analyzed on the time and frequency through a FFT computation. Based on the ISO-10816 Vibration Severity Chart standard, it is possible to categorize the severity of the problem into 4 classes depending on the power and size of the rotary machine. According to the same standard, typical faults produce unusual low-frequency vibrations, so the analysis range should be from 10 to 1000 Hz. To be more precise in analysing the data, it is known that imbalances, misalignments and looseness are recorded at frequencies up to 300 Hz. The relationship between the failures that occur in rotating machines and the frequency spectrum is illustrated on Figure 2.

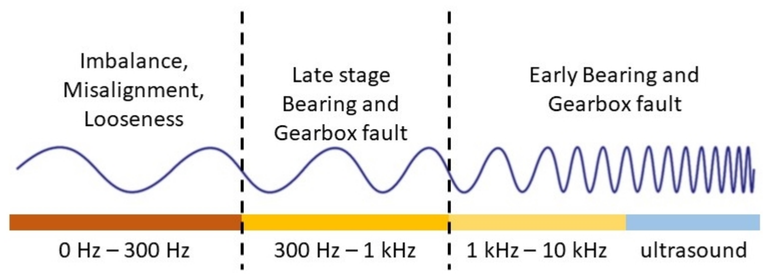


**Fig. 2.** Machines typical faults distributed in the frequency spectrum.

In the state of the art there are some algorithms that are currently widely used to implement anomaly detection solutions. *Robust Covariance* [9] is an algorithm that detecting anomalies and outliers by means of the *Mahalanobis* distance. *One-class SVM* [10] is an unsupervised algorithm that learns a decision function for novelty detection: classifying new data as similar or different to the training set. The *Local Outlier Factor (LOF)* [11] algorithm is an unsupervised anomaly detection method which computes the local density deviation of a given data point with respect to its neighbours. It considers as outliers the samples that have a substantially lower density than their neighbours. For this work, the implementation of the *Isolation Forest* through framework *BigML* was considered.

*Isolation Forest (IF)* [12] [13] is an unsupervised model, without need a predefined labels, based on decision trees, extensively used for outlier detection. In an *Isolation Forest*, randomly sub-sampled data is processed in a tree structure based on randomly selected features. The samples that travel deeper into the tree are less likely to be anomalies as they required more cuts to isolate them. Similarly, the samples which end up in shorter branches indicate anomalies as it was easier for the tree to separate them from other observations.

As illustrated on Figure 3, the algorithm try "isolate" outliers from the normal points. In order to isolate a data point, the algorithm recursively generates partitions on the sample by randomly selecting an attribute and then randomly selecting a split value for the attribute, between the minimum and maximum values allowed for that attribute.

According to the original propose of isolation forest [12], the anomaly score $(s)$ in a instance $(x)$ is calculating by:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{1}$$

where $h(x)$ is the path length of a point $x$ and $E(h(x))$ is the average of $h(x)$ from a collection of isolation trees. $c(n)$ is the constant value to normalize the average path length for $n$ trees.

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \tag{2}$$

where H(i) can be estimated by $ln(i) + 0.5772156649$ (Euler's constant) as the harmonic number.

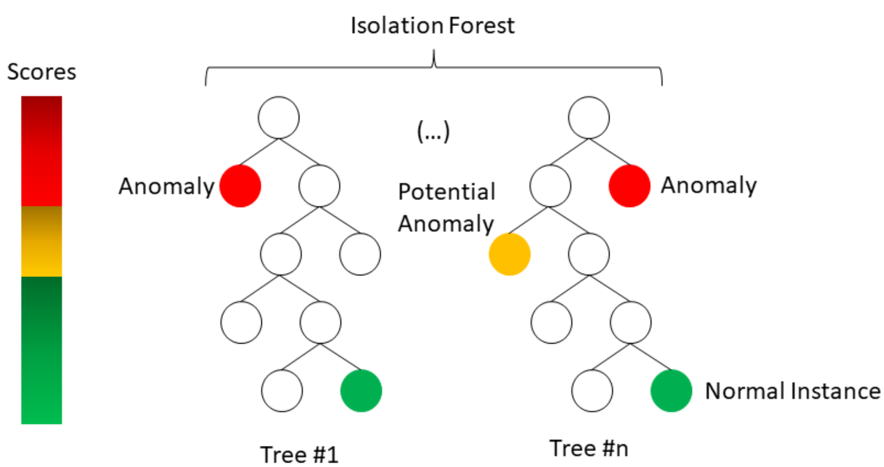


**Fig. 3.** Overview of the isolation forest method.

## 4 Experimental Results

This section presents the experimental results achieved in the laboratory through different tests in an experimental apparatus consisting of a single-phase 0.2 kW / 3000 rpm motor and a shaft alignment/misalignment system, as illustrated on Figure 4. Vibration acquisitions were performed with a DYTRAN model 3134D piezoelectric accelerometer with a sensitivity of 500 mV/g through a PCB PIEZOTRONICS 482A21 ICP signal conditioner, a National Instruments data acquisition board (NI DAQ 6008) and a data acquisition virtual instrument software developed in LabVIEW. The accelerometer and the data acquisition chain was calibrated with a PCB PIEZOTRONICS 394C06 handheld portable shaker that oscillates a 159.2 Hz. The data acquisition setup is aligned with the Operational Technology (OT) part of Figure 1.

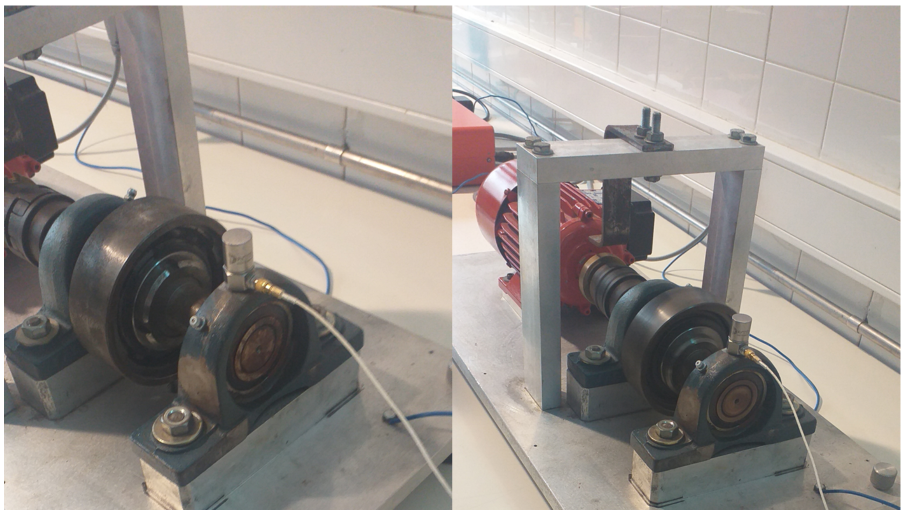

**Fig. 4.** Experimental apparatus.

According to the ISO-10816 standard, all acquisitions were performed up to 1000 $Hz$, with a resolution of 0.5 $Hz$. Different experiments were carried out, namely misalignments and loosening at different rotation speeds, which allowed the creation of a diversity of datasets for analysis. On Figure 5 is represented the frequency spectrum of one acquisition. As expected, signals with information about the motor status operation are identified at low frequencies, until 300 $Hz$.

For better interpretation of the signals, a truncated to 300 $Hz$ representation is performed, as shown in Figure 6. Three groups of signals with greater amplitude are perfectly visible, which means that the detection of any anomaly will have to go through the analysis of these signals. This acquisition was performed with the motor at 570 $rpm$, measured with a strobe lamp. It means that the first peak 9.5 $Hz$ (570 $rpm$/60) corresponds to the motor speed. From the spectral analysis it is also possible to pre-determine a threshold for which we can consider that the value is an outlier, a potential anomaly. This threshold corresponds to the value of mean plus *3 × standard deviation* as shown in Figure 6.

In this experiment a misalignment of the motor shaft has been imposed, which implies a significant increase in the amplitude of the frequencies associated to the motor, as better illustrated in Figure 7. This misalignment caused an imbalance in the motor, which is why all the peaks gained amplitude. At the motor speed frequency, the peak reaches a vibration velocity $V_{RMS} = 0.858$ $mm/s$, which

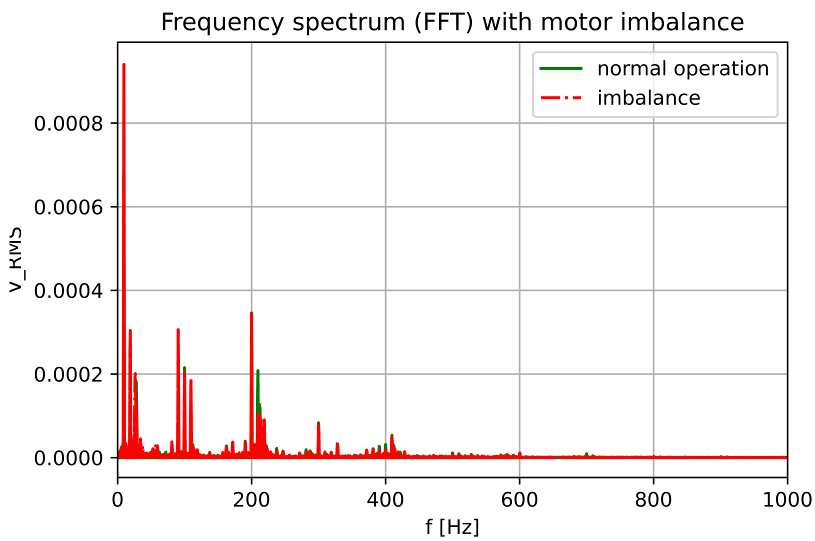


**Fig. 5.** Frequency spectrum of an acquisition.

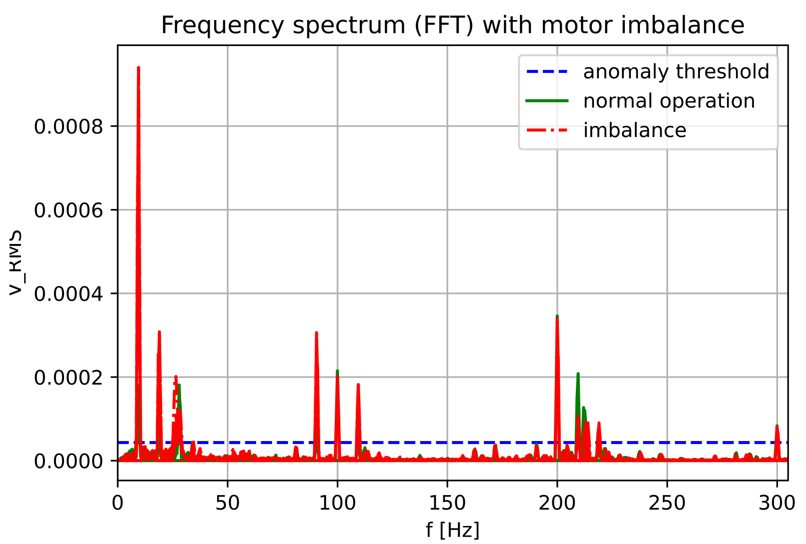


**Fig. 6.** Frequency spectrum [0 - 300 Hz].

according to ISO 10816 vibration severity standards, for small machines which is the case (Class I), corresponds to a **satisfactory** severity.

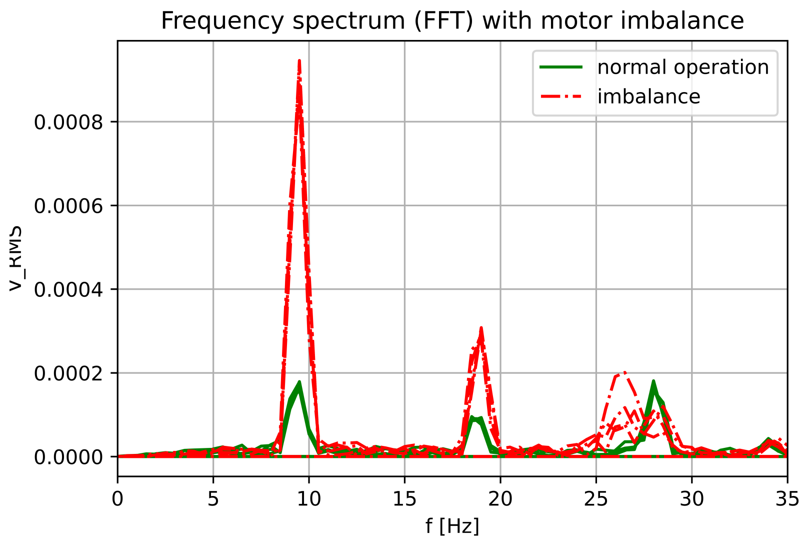


**Fig. 7.** Frequency spectrum [0 - 35 Hz].

The automatic anomaly detection was implemented in the BigML framework, exploring the potentialities of unsupervised learning isolation forest algorithm. Figure 8 shows a snapshot of the BigML output with the identification of the top 5 anomalies detected. The algorithm is configured to search 20 anomalies along the spectrum and is automatically identified 16 anomalies with an accuracy score of 90.61% as presented on Table 1. This results are very promissory and shows that the isolation forest is an adequate algorithm for this kind of analysis.

## 5 Conclusions and Future Work

The work focuses on the automatic detection of anomalies in rotating machines through machine learning algorithms. The baseline of the work is conditioned monitoring through vibration analysis, extensively used in the industry, but with the added value of seeking to detect failures as early as possible, in an intelligent way, indispensable for predictive maintenance scenarios. Machine learning algorithms have the capability to process large amount of data and identify patterns with high levels of accuracy, reason for the important on explore this contribution for the automatic anomaly detection. The paper explores the well-known isolation forest algorithm, due to its outlier detection capabilities, reason for that is considered a good algorithm for detecting anomalies in vibration analysis scenarios. Anomalies cause frequency peaks elevations and abnormal variations can be easily identified as a potential fault. The results show that the algorithm achieves results with scores higher 90% on the detection of potential anomalies.

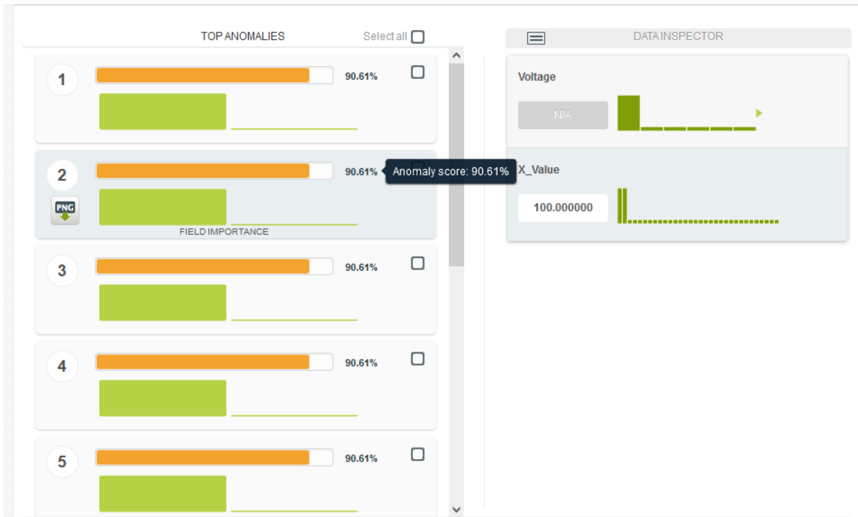


**Fig. 8.** BigML - Automatic anomaly detection results.

**Table 1.** Anomalies automatic detected.

| Frequency [Hz] | Anomaly Score [%] |
|---|---|
| 9.0 | 90.61 |
| 9.5 | 90.61 |
| 10.0 | 90.61 |
| 18.5 | 90.61 |
| 19.0 | 90.61 |
| 28.0 | 90.61 |
| 90.5 | 90.61 |
| 91.0 | 90.61 |
| 100.0 | 90.61 |
| 100.5 | 90.61 |
| 109.0 | 90.61 |
| 109.5 | 90.61 |
| 199.5 | 90.61 |
| 200.0 | 90.61 |
| 200.5 | 90.61 |
| 209.5 | 90.61 |

Results compared by a human analysis to the frequency spectrum that confirm the frequencies correspondent to misalignments and imbalances done on the motor.

As future work is expected integrate this approach in a real time monitoring framework, aligned with the Reference Architectural Model for Industrie 4.0 (RAMI 4.0), with online reports through dashboards and alerts on smart devices. Other machine / deep learning algorithms will be explored to establish comparisons of precision and robustness.

## References


1. Europe Predictive Maintenance Market – Industry Trends and Forecast to 2027, accessed on: 23/01/2022. https://www.databridgemarketresearch.com/reports/europe-predictive-maintenance-market
2. Popescu, T.D., Aiordachioaie, D. & Culea-Florescu, A. Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: an overview. Int J Adv Manuf Technol (2021). https://doi.org/10.1007/s00170-021-07703-1.
3. O. A. Egaji, T. Ekwevugbe and M. Griffiths, "A Data Mining based Approach for Electric Motor Anomaly Detection Applied on Vibration Data," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020, pp. 330-334, doi:10.1109/WorldS450073.2020.9210318.
4. M. Saeki, J. Ogata, M. Murakawa and T. Ogawa, "Visual explanation of neural network based rotation machinery anomaly detection system," 2019 IEEE International Conference on Prognostics and Health Management (ICPHM), 2019, pp. 1-4, doi:10.1109/ICPHM.2019.8819396.
5. J. -H. Han, D. -J. Choi, S. -K. Hong and H. -S. Kim, "Motor Fault Diagnosis Using CNN Based Deep Learning Algorithm Considering Motor Rotating Speed," 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), 2019, pp. 440-445, doi:10.1109/IEA.2019.8714900.
6. D. -J. Choi, J. -H. Han, S. -U. Park and S. -K. Hong, "Comparison of motor fault diagnosis performance using RNN and K-means for data with disturbance," 2020 20th International Conference on Control, Automation and Systems (ICCAS), 2020, pp. 443-446, doi:10.23919/ICCAS50221.2020.9268271.
7. Dionisio, R.; Torres, P.; Ramalho, A.; Ferreira, R. Magnetoresistive Sensors and Piezoresistive Accelerometers for Vibration Measurements: A Comparative Study. J. Sens. Actuator Netw. 2021, 10, 22. https://doi.org/10.3390/jsan10010022.
8. SenseGrow, accessed on: 19/01/2022. https://www.sensegrow.com/.
9. Robust covariance estimation and Mahalanobis distances relevance, accessed on: 23/01/2022. https://scikit-learn.org/stable/auto_examples/covariance/plot_mahalanobis_distances.html
10. One-Class SVM, accessed on: 23/01/2022. https://scikit-learn.org/stable/auto_examples/svm/plot_oneclass.html.
11. Local Outlier Factor, accessed on: 23/01/2022. https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html
12. Liu, Fei Tony & Ting, Kai & Zhou, Zhi-Hua. (2009). Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining: 413 - 422. doi:10.1109/ICDM.2008.17.
13. S. Buschjäger, P. -J. Honysz and K. Morik, "Generalized Isolation Forest: Some Theory and More Applications Extended Abstract," 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 2020, pp. 793-794, doi:10.1109/DSAA49011.2020.00120.