

# Anomalies Detection in records of IoT devices using data mining

Izaque Esteves<sup>1</sup>, Regina Braga<sup>1</sup>, José Maria David<sup>1</sup>, Victor Stroele<sup>1</sup>

<sup>1</sup>Instituto de Ciências Exatas – Universidade Federal de Juiz de Fora (UFJF)  
Juiz de Fora – MG – Brazil

{regina.braga, jose.david, victor.stroele}@ufjf.edu.br

izaque.esteves@estudante.ufjf.br

**Abstract.** *The industry underwent several transformations initiated by the first Industrial Revolution at the end of the 18th century. Today we are experiencing the Fourth Industrial Revolution, where equipment is capable of processing data and connect to communication networks. Maintenance planning can use the large volume of data generated by IoT devices to act preventively. This work aims to propose an architecture that uses an outlier detection algorithm, Local Outlier Factor, to detect anomalies in machine failure records, producing information to support equipment maintenance decisions.*

## 1. Introduction

Throughout the industrial evolution, there was a need to increase capacity and flexibility to meet the consumers' needs [Cheng et al. 2015]. The Industry 4.0 stage gave rise to technologies such as sensor networks, artificial intelligence, decision support systems, big data, and analytics [Santos et al. 2016] among others. These technologies are considered enabling technologies in this scenario. Therefore, the maintenance of machinery becomes a strategic business element.

The macro-region of Zona da Mata, Minas Gerais, is an important industrial region, with the city of Juiz de Fora as a pole, considering its industries and workers. Despite economic crises, the textile industry remained operating and today invests in the automation of production and has as new challenges the analysis of large volumes of information and decision-making processes. In the textile industry, in the rewinding sector, machines have low criticality failures during their operation, such as needle breaks or the breaking of weaving threads, which are resolved quickly and then returning to the normal operating condition. However, low criticality faults with longer than normal duration may indicate the need for maintenance.

This article discusses the use of machine state data from sensors and how data analysis can support decision-making in the maintenance processes. This approach is promising, and its strategy is to encourage and leverage the textile industry in general and in Juiz de Fora city in particular. Based on innovative technologies, we seek to reduce costs, improve the quality of work of the local labor force and leverage the potential of software development companies in the use of innovative solutions for this type of industry. In this context, this work proposes an architecture that collects and analyzes data related to failure events in industrial equipment. The proposal has generic architectural

components that can be adapted to different contexts. In this paper, the components were adapted for the use of the textile environment. Outliers can be identified from the analysis of fault records through machine learning algorithms, generating alerts, or even direct interventions on devices. As a proof of concept, the architecture was used in the textile machines context. The initial results point to the help in predicting machinery with anomalous behavior related to operating failures and consequent reduction in operating costs.

The article is organized in the following sections, including the Introduction. Section 2 presents the theoretical background. Section 3 presents related works. In section 4, the proposed architecture is detailed. Section 5 discusses a proof of concept related to the textile industry. Section 6 presents the conclusions and future work.

## 2. Theoretical Background

The use of advanced technologies characterizes the industry's current stage of development. In this vein, we have digital technologies applied to collecting and analyzing data in real time through big data and analytics, IoT, and services mainly offered in the cloud [Frank et al. 2019]. The integration between IoT and IoS (Internet of Services) in the manufacturing process was one of the fourth industrial revolution's driving forces [Kagermann et al. 2013]. This technology allows "things" and "objects" such as sensors, actuators, and cell phones to interact with each other and cooperate with "smart" components to achieve a common goal. Industries can use IoT and IoS in various applications, including industrial / manufacturing, smart cities, transport, digital commerce, and health care.

The current panorama of the industry includes several digital technologies associated with the acquisition and analysis of data in real time, such as internet of things, cloud-based and big data and analytics [Frank et al. 2019]. Through the Internet of Services - IoS we have sensors, actuators and IoT devices integrated by a common industrial communication network [Kagermann et al. 2013] and we have production systems connected to the Internet, figuring as intelligent applications [Drath and Horch 2014]. This concept goes beyond the limits of the industry, allowing the existence of smart cities, electronic commerce and health care, among other applications.

With automation, we have the challenge of extracting information to assist in planning maintenance processes, avoiding machine downtime, and minimizing operating costs. Data mining, which consists of a set of techniques for analyzing large data sets to find unexpected relationships and summarize the data to be useful and understandable [Gorunescu 2011], allows extracting knowledge from existing data and support decision making.

According to this work proposal, one of the data mining tasks that are very adherent to the problem is detecting deviations since there is a need to map the dynamics of machine failures, detect possible anomalies, and provide mechanisms that allow the triggering of alerts or preventive actions. For this latter, the use of autonomous agents seems to be an appropriate approach. Autonomous agents are defined by Russell and [Norvig and Intelligence 2002] as software that has a degree of Intelligence, and that this Intelligence is highly linked with rational actions, which means acting to reach the objectives defined by someone, according to their beliefs. Autonomous agents are software

that can act on an environment through their perception of that environment without human intervention. In this paper, we consider that the artificial Intelligence associated with data mining allows new information to be discovered considering the data produced by machines. Through autonomous software agents, it is possible to make decisions based on these data analyses.

### **3. Related Works**

Some works related to deviations or anomaly detection are presented in the literature, considering different scenarios and applications. The work presented by [Ma et al. 2013] proposes an algorithm for detecting local outliers in time intervals to monitor industrial processes. The proposal includes a Local Outlier Factor (LOF) detection algorithm that deals with scenario variations, maintaining the monitoring efficiency. In [Lee et al. 2011], a new monitoring method that integrates independent component analysis (ICA) and local outlier factor (LOF) is discussed.

Using a data set that represents normal conditions, it is possible to observe the LOF value in each monitoring interval, which allows detecting if the observed value is a failure or not. Thus, the proposed scheme improves the use of the two analysis schemes in isolation. [Arulanthu and Perumal 2020] proposes an online decision support system to prevent chronic kidney disease. The model considers IoT devices and cloud platforms and a regression model. In [Ma et al. 2013a], the authors proposed a new method called the Neighborhood Standardized Local Outlier Factor method (NSLOF). The method considers changes in production conditions and Gaussian and non-Gaussian distributions of process data. In the new method, the LOF factor of each sample, which means its degree of being an outlier, is used as statistics for monitoring. A new Euclidean distance is normalized based on the restricted location strategy and used during the monitoring index calculation.

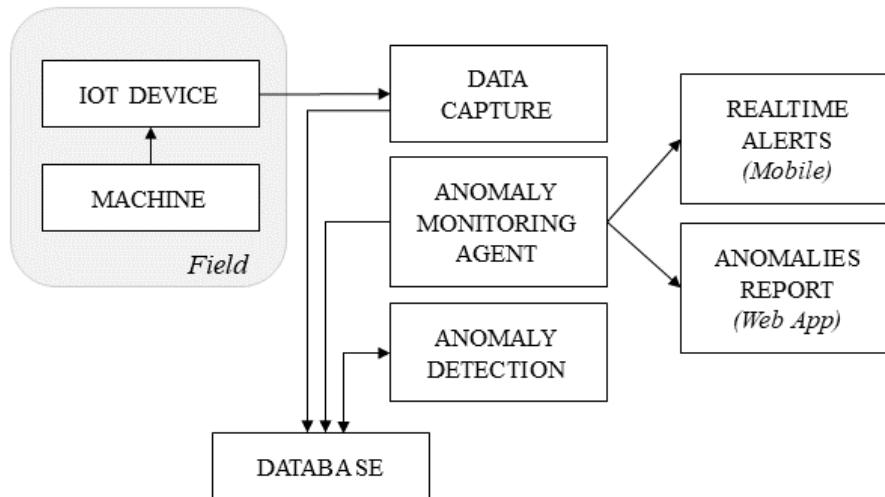
These works use detecting outliers' mechanisms considering scenarios where information distribution does not follow a standard. Among the works presented, [Ma et al. 2013a] is similar to our approach, considering the sample's neighborhood as a determining factor, which will be used as a reference in detecting possible deviations in the analyzed samples. Our approach differs from the others concerning the ability to support decision-making in preventive maintenance processes using autonomous agents capable of detecting anomalous behaviors and automatically alerting about such occurrences.

### **4. An architecture for detecting anomalies in data about failures**

This section presents an architecture that collects and analyzes information related to failure events in equipment connected to sensors. Through AI techniques, sensors (IoT), the data are analyzed, and the operational conditions that require maintenance are detected, triggering preventive actions. Autonomous agents using intelligent algorithms are responsible for supporting decision-making related to the predictive maintenance of equipment.

This architecture takes advantage of the increasingly frequent use of machines connected to a data communication network, which allows obtaining information about the functioning of each device. Even when a device cannot provide the network information about its status, it is possible to capture this information by sensors connected to a programmable logic controller, using protocols already established and known by the

networks. The proposed architecture considers the machine's state data as the initial input for the anomaly detection process, as shown in (Figure 1). The equipment can manually provide this information through human-machine interfaces or IoT devices that receive signals from sensors and communicate through the data network.



**Figure 1. Architecture for detecting anomalies in failure data**

The acquisition of information from devices is processed through data capture component, collecting the data provided by sensors or programmable logic controllers, since IoT devices implement communication protocols that are easily connected to software solutions, such as the MQTT protocol that offers a system to subscribe to topics by which data is published [Jaloudi 2019]. The collected data represent the failures that occurred and identify an event with its start and end timestamp, machine identification, and the type of event being stored through the database component, which provides access to the collected data.

The anomaly detection component performs intelligent analysis on fault data in a defined timestamp, using intelligent techniques and configurable AI algorithms. Through data analysis, it is possible to highlight different variations in the data set. The variations are detected as a function of the time the machine remains in an anomalous state. The anomaly monitoring agent is responsible for reporting in real-time the anomalies detected considering each execution of the intelligent algorithms. Finally, we have the component responsible for communicating the result of anomaly detection: real-time alerts, which trigger real-time alarms reporting the detection of anomalies to mobile devices. On the other hand, the anomalies report component provides a detailed report of anomalies, allowing the analysis of the failure machine.

## 5. Proof of Concept

### 5.1. Application Scenario

As stated in the introduction, machines in the textile industry, especially in the rewinding sector, have errors during their operation that are usually solved quickly, returning the machine to operational status, but which in the medium term may point to operational

inefficiency. Thus, identifying events that stand out from the majority is important. That is, realizing the deviations in the data sets about failures helps in decision making.

According to [Barnett and Lewis 1984] deviations or outliers are data that are far from the expected standards. Detection of deviations in operational failures highlights machines with different behavior from the majority, indicating the need for evaluation by the maintenance area. Thus, the proposed architecture can assist in the detection of anomalies in failure data. From the intelligent analysis of this data, the software agent highlights the deviations, assisting in the decision-making process.

## 5.2. Execution

To test the architecture deviation detection capacity, a dataset with 10,000 records of operational failures was used, containing: machine identification, sector, the reason for failure, and duration.

The algorithm called LOF (Local Outlier Factor), available in the machine learning library scikit [Trappenberg 2019] was used to detect the outliers in failure events. The LOF algorithm is suitable for the purpose, as it highlights samples that have a distance to other samples. Thus, extended machine downtime shows functional anomalies and deserves attention. A Python script was written to process the dataset, recording the anomalies in fault records in a database.

```

root@izaque-VirtualBox:/var/www/html/monitor# python3 get_outliers.py
[ 0  1 230  4]
[ 0  6 250  5]
[ 0 10 300  4]
[ 0 23 190  6]
[ 0 27 230  1]
[ 3 3997 100  3]
[ 2 4034 130  3]
[ 4 5987 100  2]
[ 4 5988 370  2]
[ 4 5998 240  4]
[ 4 6004 360  6]
[ 2 6008 460  2]
root@izaque-VirtualBox:/var/www/html/monitor#

```

**Figure 2. The output of the outlier's detection script**

The algorithm highlights outliers in the dataset, analyze the data matrix and outputs a new matrix containing the data that has been identified as anomalies. For example, the (Figure 2) shows the output of the processing of an event timestamp, whose size was defined by the size of the dataset, that is, 10,000 records. The columns of the output matrix represent, respectively: the machine identification, the fault event identification, the fault duration time in minutes, and the type of fault.

## 5.3. Results

According to the proposed architecture, the anomalies report component was implemented as a panel to display the anomalies detected at each verification cycle, as shown in (Figure 3). The panel obtains data on the anomaly records and allows the manager to verify all anomalies of a machine in the verification cycle. For example, in (Figure 3), it is possible to notice that machine 0 presented a high number of anomalies in relation to its failures.

Machine 3 presented only one anomaly event. Also, in the panel, it is possible to access a ranking that classifies the machines according to the number of anomalies presented in the verification cycle. We can notice that machine 0 deserves attention from the maintenance sector because it presented the largest number of anomalies. From a predictive maintenance point of view, it is necessary to foresee the failures events and thus plan interventions to improve the processes' cost-benefit. The time factor is decisive, from the detection of anomalies to manager action.

## Painel de Monitoramento

É apresentado, para cada máquina, o extrato de tempo em falha fora do comportamento de falha padrão, considerando uma janela de 10.000 eventos.

### Visão geral das anomalias

Após a análise de toda a amostra, abaixo estão listadas todas as anomalias detectadas.

Id	Setor	Tipo de Falha	Duração (min)
0	Remalhção	Falha mecânica	230
0	Remalhção	Falha pneumática	250
0	Remalhção	Falha mecânica	300
0	Remalhção	Falha elétrica	190
0	Remalhção	Quebra de agulha	230
2	Remalhção	Ausência de matéria-prima	130
2	Remalhção	Rompimento de fio	460
3	Remalhção	Ausência de matéria-prima	100

### Ranking: máquinas x anomalias

Ranking que mostra as máquinas e o total de anomalias.

ID Máquina	Qtde de anomalias
0	5
4	4
2	2
3	1

**Figure 3. Anomaly monitoring panel (in Portuguese)**

The anomaly reports component also has a proactive autonomous software agent that monitors anomalies in the records of machine failures and, in real-time, signals, in a classified way, the equipment that in a given time interval presented anomalies. This function is complemented by alarms sent to mobile devices that receive information about anomalies in real-time and can check the machine and act to restore its functioning. Thus, the software agent is responsible for automatically making decisions without prior analysis by the maintenance team.

To monitor the anomalies, two application prototypes were developed, for Android and IOS, that query the autonomous agent regarding detecting anomalies and issue an alert to the user. The solution also includes a web application to monitor the anomalies. (Figure 4) shows the prototypes running on Android and IOS systems. For each alert issued, the user can access an interface for viewing anomalies. Therefore, considering the proof-of-concept preliminary results, we had some evidence of the feasibility of the architecture in the support for decision-making to detect failures in textile machines. Thus, the approach is promising, and advances will be made in the development of architecture and also in the specification of a broader evaluation scenario.

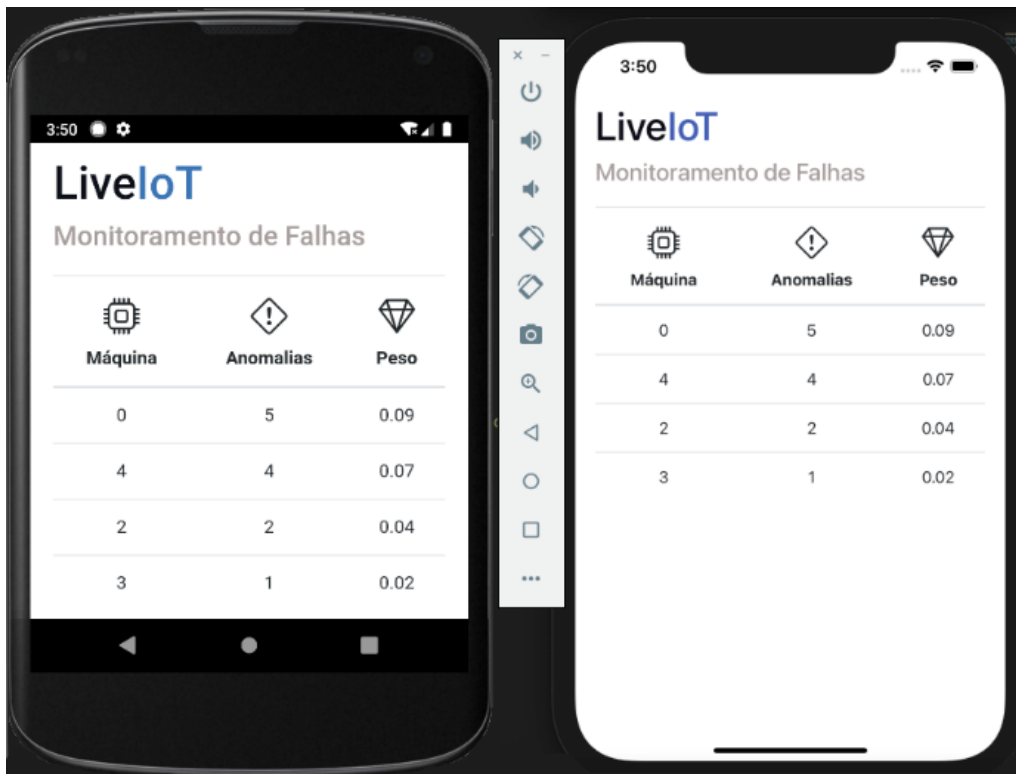


Figure 4. Prototypes created for mobile devices (in Portuguese)

## 6. Conclusion

Industry 4.0 has as one of its enabling technologies, the IoT. The use of this technology is essential for data communication in real-time, enabling fast decision-making. Integrating machines with more than a decade of use, in this new context, is challenging. However, with IoT technologies, together with internet connectivity, this integration is possible. This paper presented an architecture that uses IoT, AI, and data mining techniques to provide a predictive maintenance solution.

With our solution, it was possible to collect and store data and produce knowledge from machine failure data. The outlier detection algorithm was able to identify anomalies with acceptable performance, considering a dataset with 10,000 records and for sample analysis cycles, with an execution time of 2 seconds. The proactive autonomous agent indicated anomalies in real time to mobile devices and presented these anomalies in the panel available in a web application. As a future contribution, this work aims to correlate failures and their causes to build knowledge about anomalies that can generate other anomalies, expanding the predictive capacity of the software agent, creating alerts about probable machine stops.

## References

- Arulanthu, P. and Perumal, E. (2020). An intelligent iot with cloud centric medical decision support system for chronic kidney disease prediction. *International Journal of Imaging Systems and Technology*, 30(3):815–827.
- Barnett, V. and Lewis, T. (1984). Outliers in statistical data. *Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics*.

- Cheng, C.-H., Guelfirat, T., Messinger, C., Schmitt, J. O., Schnelte, M., and Weber, P. (2015). Semantic degrees for industrie 4.0 engineering: deciding on the degree of semantic formalization to select appropriate technologies. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, pages 1010–1013.
- Drath, R. and Horch, A. (2014). Industrie 4.0: Hit or hype?[industry forum]. *IEEE industrial electronics magazine*, 8(2):56–58.
- Frank, A. G., Dalenogare, L. S., and Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, 210:15–26.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques*, volume 12. Springer Science & Business Media.
- Jaloudi, S. (2019). Mqtt for iot-based applications in smart cities. *Palestinian Journal of Technology and Applied Sciences (PJTAS)*, (2).
- Kagermann, H., Wahlster, W., Helbig, J., et al. (2013). Recommendations for implementing the strategic initiative industrie 4.0: Final report of the industrie 4.0 working group. *Forschungsunion: Berlin, Germany*.
- Lee, J., Kang, B., and Kang, S.-H. (2011). Integrating independent component analysis and local outlier factor for plant-wide process monitoring. *Journal of Process Control*, 21(7):1011–1021.
- Ma, Y., Shi, H., Ma, H., and Wang, M. (2013). Dynamic process monitoring using adaptive local outlier factor. *Chemometrics and Intelligent Laboratory Systems*, 127:89–101.
- Norvig, P. R. and Intelligence, S. A. (2002). *A modern approach*. Prentice Hall Upper Saddle River, NJ, USA:.
- Santos, B. P., Silva, L. A., Celes, C., Borges, J. B., Neto, B. S. P., Vieira, M. A. M., Vieira, L. F. M., Goussevskaia, O. N., and Loureiro, A. (2016). Internet das coisas: da teoria à prática. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, 31.
- Trappenberg, T. P. (2019). Machine learning with sklearn. In *Fundamentals of Machine Learning*, pages 38–65. Oxford University Press.