The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2022

Volume 21, Pages 55-63

IConTES 2022: International Conference on Technology, Engineering and Science

# Learning-Based Algorithm for Fault Prediction Combining Different Data Mining Techniques: A Real Case Study

**Laura LUCANTONI**
Universita Politecnica Delle Marche

**Filippo Emanuele CIARAPICA**
Universita Politecnica Delle Marche

**Maurizio BEVILACQUA**
Universita Politecnica Delle Marche

**Abstract:** In recent years, new Data Mining (DM) algorithms and methodologies are increasingly used as an industrial solution for manufacturing improvements. In this context, new techniques are widely required by companies in the field of maintenance due to the need to reduce breakdowns intervention and take advantage of the increasing availability of data. This paper aims to propose a new learning-based algorithm to improve knowledge extraction by combining different DM techniques from a predictive maintenance perspective. First, the J48 algorithm and Random Forest (RF) are used as a predictive model to classify a set of failure modes according to their influence on the Overall Equipment Effectiveness (OEE). Then, the Apriori algorithm is used to identify the relationship among failure events belonging to the lowest OEE range for which, therefore, a predictive maintenance strategy should be defined. In order to describe the learning-based algorithm proposed in this paper, a real case study is presented and detailed. The experimental results showed a valuable tool for knowledge extraction and the definition of a set of predictive maintenance strategies for those failures most affecting the process. In this way, the complexity of decision-making on maintenance strategies can be reduced mainly when dealing with a large amount of information or a challenging dataset.

**Keywords:** Algorithms, Knowledge extraction, Failure prediction, Data mining, Case study.

## Introduction and Background

In the current scenario, data-driven approaches are used by companies to support decision-makers in managing the wide amount of data in multiple contexts: energy consumption (Mugnini et al., 2021), industrial operations productivity (Antomarioni et al., 2021), efficiency (Görür et al., 2021), sustainability (Linke et al., 2019) and so on. Due to the integration of data analytics techniques and the advancement of information technologies, both efficiency and productivity areas have demonstrated the most potential improvement in manufacturing (Mansouri et al., 2020), mainly for failure detection in the field of maintenance (Görür et al., 2021).

Earlier, the term “maintenance” has referred to Breakdown or Time-based maintenance policies, thus attributing a cyclic nature to the fault occurrences and neglecting aspects such as the stochastic unavailability, due to the lack of quantity and quality data.

Since over the years a proper maintenance plan resulted in management, performance, and profit benefits (Lucantoni et al., 2019), the idea of maintenance has moved to a more scientific policy, namely Preventive maintenance, for failure occurrences prevention and proper operating condition restoration in manufacturing

(Ding & Kamaruddin, 2015). Among strategies, Predictive Maintenance (PdM) is considered one of the most interesting in the field of fault prediction, continuously growing with the technological advancement in manufacturing: due to the growing number of machines, systems, equipment, and products, DM models for data management through algorithms analytics (Frontoni et al., 2017) has demonstrated interesting prediction capabilities and decision making from a PdM perspective (Paolanti et al., 2017). In this context, the current paper aims to develop an algorithm for failure prediction. In particular, the algorithm proposed in this work is based on detecting hidden relationships within those failure events most influencing the OEE value using the integration of descriptive and predictive DM techniques.

In the literature, several contributions deal with the development of fault prediction strategies through DM techniques: some authors proposed the use of regression tree models for failure classification (Mohamed et al., 2019; Sezer et al., 2018) while, as the failure event is rarely random, Association Rules techniques are used for recognizing hidden patterns among failures (Antomarioni et al., 2022b) and Support Vector Machine for recognizing special patterns from signals (Gharoun et al., 2019; Widodo & Yang, 2007). Researchers also considered mathematical programming for predicting breakdowns in complex systems (Pisacane et al., 2021); while recently deep learning predictive models for selective maintenance optimization are proposed with successful results (Hesabi et al., 2022). An overview of the literary contributions dealing with the application of J48, RF, and Apriori algorithms for failure and fault prediction is also provided and summarized in Table 1.

Table 1. Literary contributions overview dealing with fault prediction

| Algorithms | # of papers | # of relevant papers |
|---|---|---|
| Decision Tree J48 | 131 | 12 |
| RF | 118 | 35 |
| Apriori | 10 | 5 |
| RF and Apriori | - | - |
| Decision Tree J48 and RF | 89 | 8 |
| Decision Tree J48 and Apriori | 1 | 1 |
| Decision Tree J48, RF and Apriori | - | - |

Although valuable research is available in the literature, researchers usually consider the integration of one or two of those techniques to the best of the author's knowledge. However, several contributions integrated J48, RF, and Apriori algorithms in the educational and medical fields (Hussain et al., 2018; Kiu, 2018); while 31 research mainly used RF and Apriori algorithms in the technological fields, e.g., for software defects prediction (Abualghanam et al., 2022; Thapa et al., 2020). In particular, it should be highlighted that 70% of papers regarding the integration of J48 and RF algorithms, occurred only substituting “J48” with “Decision Tree”. Likewise, Among papers using the Apriori algorithm for fault prediction, only one also applied the Decision Tree model for the decision-making process in the field of telecommunications (Yang et al., 2017).

To sum up, even though all the methodologies have been successfully investigated in several contexts, there is no evidence of scientific paper dealing with fault prediction through the integration of J48, RF, and Apriori algorithms in the field of fault prediction. For this reason, the aim of the learning-based algorithm proposed in this paper is to reduce this research gap by developing and testing a new decision-making procedure that is based on the integration of three different DM techniques: J48, RF, and Apriori algorithms. In addition, the purpose of the current paper is to support maintainers in their decision-making process by providing reliable knowledge regarding the most appropriate maintenance strategy for the management of each failure mode (FM). In particular, the proposed algorithm enables OEE improvement by exploiting the strengths of different algorithms: Random Forest and J48 algorithms are used to classify failure events according to the OEE labels. Then, the Apriori rule is used to predict the hidden relationship between failure events belonging to the lowest OEE label to enhance failure prevention. The rest of the paper is as follows: this introduction is followed by the description of the learning-based algorithmic model description and implementation (Section 2). The results and discussion about theoretical and practical contributions are detailed in Section 3, while the conclusions and possible feature research topics are summarized in Section 4.

## Algorithm Implementation

The algorithm developed in this paper has been applied and tested in a fully automatic assembly line within the automotive sector in order to reduce failure occurrences and improve the maintainer's decision-making process for fault prediction. The learning-based algorithm is structured and logically detailed as Figure 1 leveraging the availability of a huge amount of data. The methodology is summarized as follows:

1) *Preliminary analysis* for data gathering, pre-processing, and OEE calculation and classification. The main objective is to understand "what" kind of failure events has occurred.
2) *Data Mining* for analyzing data through DM techniques and checking the most impacting FMs relationships. The main objective is to understand "why" a failure event has occurred.
3) *Decision-making* to identify the most suitable maintenance policy and strategy for each major fault relationship. The main aim is to "predict" further occurrences by differentiating maintenance policies and actions.

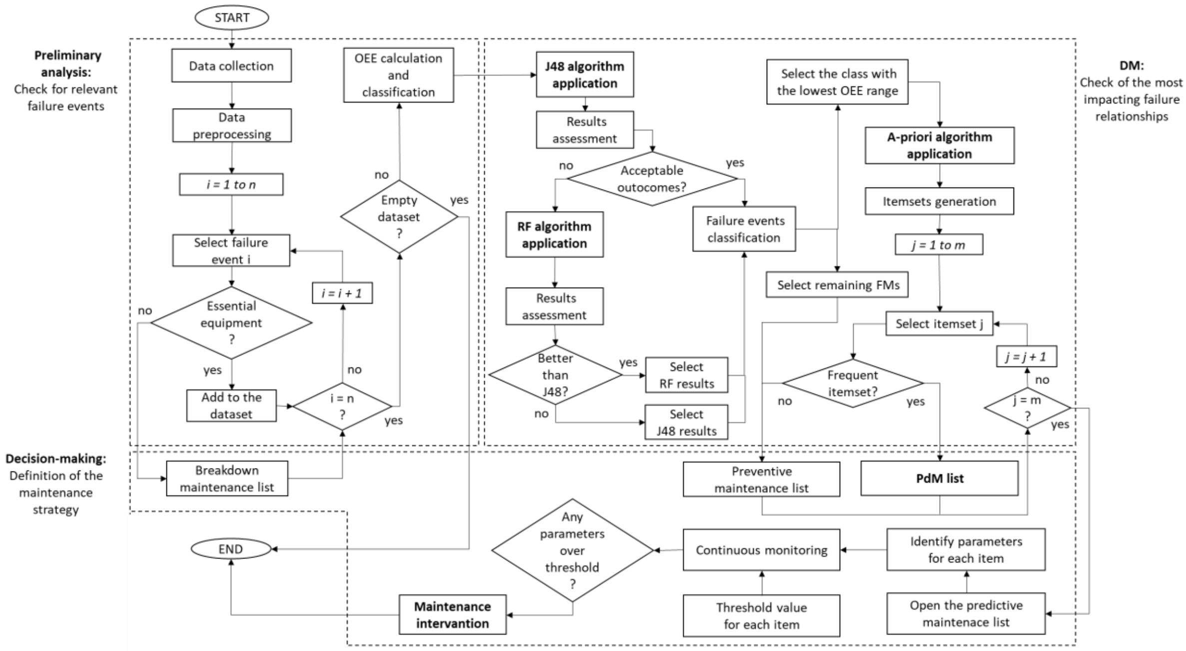


Figure 1. Learning-based algorithm for failure prevention

**Preliminary Analysis**

The Preliminary analysis layer is the foundation of the learning-based algorithm proposed in this paper. Its early activity involves data collection and pre-processing for failure events selection and cause identification. The real case study regards a fully automated line consisting of twelve assembly stations. In this experiment, the Computerized Maintenance Management System (CMMS) and Manufacturing Execution System (MES) are used to collect maintenance and manufacturing process information, respectively.

As a first step, data pre-processing is used for creating a single dataset with relevant information reporting: failure events, the date of the event, the shift in which the event occurred, and the production losses parameters, (availability, performance, quality). All inconsistencies were then removed thus providing a consistent dataset. Next, one-to-one failure modes selection is required for cause identification: if the equipment's or product's failures are not essential for operations or product, the failure mode is added to the *Breakdown maintenance list* (e.g., low-cost equipment repaired when needed); otherwise, it is left in the dataset as input to the second layer. The real dataset consisted of 1020 instances with 139 attributes describing the different failure modes, i.e:

- 143 instances with 32 attributes from the breakdown maintenance list;
- 877 instances with 107 attributes that need in-depth analysis and OEE calculation and classification.

OEE is a performance-measurement indicator with a range of 0 – 1 used to measure different types of production losses such as availability, performance, and quality (Muchiri & Pintelon, 2008). Data conversion and normalization have been applied to the following attributes before the evaluation of the prediction model:

- The failure modes attribute is converted to binary (0 | 1)
- 1st OEE range (f1) and 2nd OEE range (f2) are based on the assessment of attributes of the equipment effectiveness losses (Table ).

Table 2. OEE classification

| OEE range | Class Code | # of instances |
|---|---|---|
| 0< 0EE <0,695 | f1 | 436 |
| 0,695<= OEE <1 | f2 | 441 |

**Data Mining**

The main aim of the DM layer is the processing of observations, data classification, and analytics to discover hidden relationships among the occurrences of different FMs. In particular, the objective of the learning-based algorithm is to identify correlations among the data, parameterize and recognize failure patterns for relevant knowledge extraction through the integration of multiple DM algorithms. It should be noted that the algorithm proposed in this work is designed for *off-line* operation with regular updating. The DM layer mainly consists of two main areas aiming to make the analyses easily accessible and user-friendly:

- *Decision Tree J48* and *RF* algorithm application for predicting OEE labels and values.
- *Apriori* algorithm application for the interpretable pattern recognition among the FMs most influencing OEE and the identification of the best maintenance strategy.

*J48 and RF application*

For data prediction among the selected failure events, supervised learning, including Classification techniques, is usually used to classify each item in a set of data into one of a predefined set of classes or groups (Kesavaraj & Sukumaran, 2013). To do that, the Weka DM tool (Waikato Environment for Knowledge Analysis) is used to perform the analysis as open-source software. Two supervised DM techniques, namely, Decision Tree J48 (the open-source Java implementation in Weka of the C4.5 algorithm), and the infrequent Random Forest techniques (Mikut & Reischl, 2011) have been selected for the major failure events analysis and classification. Hence, in the real case study, the 107 attributes that need in-depth analysis are used to predict the final OEE label. The main objective is to perform a "pruning" of the tree, removing branches incomprehensible or creating "noise" among the data. Once the J48 algorithm application, the output of Weka is as in Figure 2:

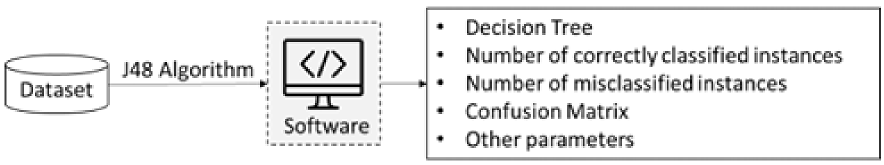


Figure 2. J-48 algorithm output

Decision Trees have a hierarchical structure of arcs and nodes:

- The *internal nodes* represent an FM.
- The *arcs* correspond to the FM's value (i.e., binary value).
- The *root nodes* make a classification by grouping OEE into two labels.

In the current experiment, a binary Decision Tree J48 is provided, i.e., only two conditions can be evaluated in the splitting:

- Different from zero (! =0) if the fault occurs.
- Equal to zero (= 0) if the fault does not occur.

The methodology should be followed until reaching all the leaves and the final classification is thus achieved when the OEE label assignment is completed. Starting with the root node (= FM 5), the splitting is performed. If such a failure event occurs (i.e., arc value " ! = 0 ") the classification proceeds with the branch on the right immediately leading back to a leaf node performing the classification. Otherwise, if it does not occur (i.e., arc value " = 0 ") the classification proceeds with the left branch leading to a different attribute in order to perform the next splitting. This procedure is followed until the final classification. Hence, the assignment of the respective OEE label namely f1 or f2 is achieved.

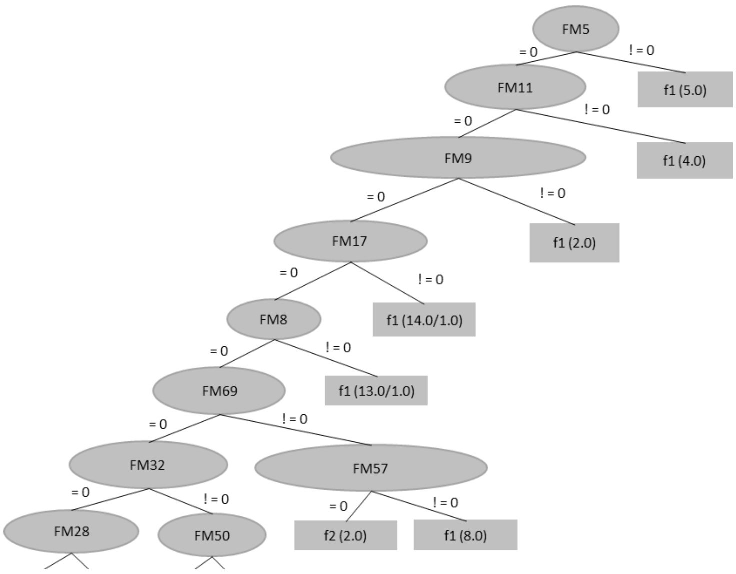

Figure 3. Extraction of the DT J48

Likewise, starting from the “FM 69” node, instead of reaching the leaf with the relevant classification immediately, two other nodes have been provided. This implicates that if a failure event related to “FM 69” occurs then a failure event related to “FM 32” also occurs; otherwise, the failure is related to “FM 57”.

Once the classification is completed, it is important to take into account the misclassification error and possible inconsistencies that could be explained as the presence of a fault with limited or no impact on the performance of the production line under consideration (e.g., if band *f2* is assigned when a fault occurs). Hence, the evaluation of this classification is provided after the explanation of two more different experiments for the best results selection.

- 2nd classification: The day of the week is included among the attributes, from “1 = Monday” up to “7 = Sunday”. The DT J48 has a similar classification to the previous one with the addition, in a few parts, of nodes with the day of the week attribute, which, however, does not significantly affect the whole DT structure. In fact, most of the same classification criteria present in the first tree have been observed.
- 3rd classification: RF, i.e., a multitude of trees into a single model, is provided. The tree-related graph has not been provided since generating a multitude of trees: 150 in the case under examination.

Eventually, results are assessed by the “Confusion Matrix” for the classification error quantification and accuracy evaluation (Figure 4):

| CONFUSION MATRIX | | | CLASSIFICATION ACCURACY |
|---|---|---|---|
| a b <-- classified as | | | |
| 246 190 \| a = f1 | | First classification | = 60 % |
| 163 278 \| b = f2 | | | |
| a b <-- classified as | | | |
| 218 218 \| a = f1 | | Second classification | = 58 % |
| 150 291 \| b = f2 | | | |
| a b <-- classified as | | | |
| 222 214 \| a = f1 | | Third classification | = 59 % |
| 147 294 \| b = f2 | | | |

Figure 4. Confusion matrix and accuracy assessment

The first classification, namely, the DT J48 only considering the FMs attributes, provided 246 instances correctly classified as *f1* and 278 as *f2*. All remaining ones are considered misclassification probably due to the challenging dataset. In the second classification -2% in the accuracy value is highlighted mainly due to a decrease in *f1* correct classification. Then, the classification achieved from the RF is very similar and almost completely comparable with the J48 algorithm in terms of returned parameters.

Then, faults belonging to the lower OEE range should be selected: since the first classification appears to have the best accuracy, all the instances belonging to the f1 class have been selected as at-risk failure events to be better analyzed through the Apriori algorithm application. All the remaining FM not appearing in the lowest OEE class should be added to the preventive maintenance list.

*Apriori Algorithm Application*

The learning-based algorithm proposed in this paper proceeds with the identification of the most frequent itemset among failure events belonging to the lower OEE label. Association rules, i.e., "*if→then*" prepositions formed by an antecedent and consequent for pattern extraction from huge amounts of data. In particular, the current work procedure can be summarized as follows:

a) Scanning the database for frequent itemsets generation using the "*Confidence*" and the "S*upport"* values, which respectively identifies the strength and the statistical significance of the rule (Antomarioni et al., 2022a): itemsets didn't reach the minimum support value as a threshold should be added to the *preventive maintenance* list to follow the existing regular maintenance plan, i.e., every 1000 pieces assembled in the real case study.
b) From the frequent itemsets, association rules are extracted as reported in Table 3 with an average accuracy of 0,59: the hidden patterns discovered among FMs have been added to the *PdM list* to schedule maintenance based on the asset conditions.

Once all itemsets have been analyzed, the FMs typically result in an OEE reduction and relevant hidden patterns among them are detected to enhance the decision-making strategy described in the paragraph below.

**Decision-Making**

The Decision-making layer only concerned the frequent itemsets selected in the previous DM layer, i.e., those FMs for which PdM strategies should be generated. After identifying the critical FMs and their relationships, the most appropriate maintenance policy should be selected for each of the most frequent items. In particular, each FM which is part of the precursors of the rule should be predicted in order to avoid the occurrence of the consequent in the rule. The choice of maintenance policies derives both from the *Preliminary analysis* and *Data mining* layers of the learning-based algorithm in this work.

Table 3. PdM strategy

| Failure code to predict | Faulure patterns | Parameters to be monitored |
|---|---|---|
| 5 | 103, 6, 11 | Workstation energy consumption |
| 44 | 7, 4, 12, 55 | Part positioning coordinates |
| 47 | 10, 22, 30, 63 | Tool vibrations |
| 61 | 34, 51, 98 | Workstation energy consumption |
| 69 | 89, 32, 57 | Tool deterioration |
| 90 | 1, 15 | Tool deterioration |
| 102 | 48, 3, 27, 81 | Workstation temperature |

As previously reported, the most suitable maintenance strategy is identified among the main maintenance policies:

- Breakdown maintenance,
- Preventive maintenance,
- Predictive maintenance.

Regarding PdM, the basic idea in the current real experimentation is based on the continuous monitoring of the asset conditions and signals degradation. Hence, in order to provide some predictability to such failures belonging to the PdM list in the real case study, seven predictive maintenance strategies have been identified as shown in Table 3 in order to avoid both the antecedent failure and its consenquents.

## Results and Discussion

After implementing the learning-based algorithm, 139 FMs belonging to the source database have been classified as follows:

- The 23% is related to the corrective maintenance list due to their lack of predictability, as well as, human error without explanation.
- The 55% is assigned to the preventive maintenance list as no frequent itemset was found. This means that maintainers will continue to handle such failure events according to their experience and strategy.
- The 22% of failure modes have been associated with the predictive maintenance list due to relevant patterns among FMs most reducing OEE.

Some consistency in the number of corrective maintenance interventions has been found. Otherwise, an average reduction of 15% in preventive maintenance interventions has been observed, then replaced by PdM interventions as a result of the detection of 11 over-threshold parameter values. In addition, to deal with the last point, 7 PdM actions have been applied by monitoring the assembly line for two weeks. An average improvement in OEE of 0.5 percent is assessed due to the PdM actions taken.

Based on these results, since much research in the literature usually compared different DM techniques identifying the best one for the specific case study without founding one technique universally superior to the others, the implementation of the learning-based algorithm proposed in this work results particularly useful to exploit as many strengths as possible from multiple DM techniques.

It should be stated that the proposed algorithm is beneficial on huge datasets, as homogeneous and consistent as possible with the business reality, thus proving poor applicability to manually manufacturing contexts or challenging datasets. In conclusion, despite the moderate classification accuracy values of 60%, the challenging nature of the source dataset was highlighted thus requiring improvement.

## Conclusion

DM is the process of discovering previously unknown and potentially valuable relationships, patterns, and information within huge databases. In particular, this paper demonstrated how integrating J48, RF, and Apriori algorithms can be significant for early failure prediction in a real case study, supporting the maintenance decision-making process. Hence, early fault prediction with this learning-based algorithm also supports the identification of failure occurrences relationship most influencing the Overall Equipment Effectiveness. In this way, maintainers may be able to perform better in their daily maintenance interventions.

In future development, a more accurate learning-based algorithm for predicting early failure occurrences can be provided by integrating other DM techniques, such as Network Analysis to represent the hidden relationship among failures to facilitate their comprehension. In conclusion, a regular algorithm application should be tested in the long term from a continuous improvement perspective in the field of fault prediction.

## Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

## Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Technology, Engineering and Science ( www.icontes.net ) held in Antalya/Turkey on November 16-19, 2022.

## References


Abualghanam, O., Al-Khatib, S. N., & Hiari, M. O. (2022). Data mining model for predicting customer purchase behavior in e-commerce context. *International Journal of Advanced Computer Science and Applications*, *13*(2), 421. https://doi.org/10.14569/IJACSA.2022.0130249

Antomarioni, S., Ciarapica, F. E., & Bevilacqua, M. (2022a). Association rules and social network analysis for supporting failure mode effects and criticality analysis: Framework development and insights from an onshore platform. *Safety Science*, *150*. https://doi.org/https://doi.org/10.1016/j.ssci.2022.105711

Antomarioni, S., Ciarapica, F. E., & Bevilacqua, M. (2022b). Data-driven approach to predict the sequence of component failures: a framework and a case study on a process industry. *International Journal of Quality & Reliability Management*. https://doi.org/10.1108/IJQRM-12-2020-0413

Antomarioni, S., Lucantoni, L., Ciarapica, F. E., & Bevilacqua, M. (2021). Data-driven decision support system for managing item allocation in an ASRS: A framework development and a case study. *Expert Systems with Applications*, *185*, 115622. https://doi.org/10.1016/J.ESWA.2021.115622

Ding, S. H., & Kamaruddin, S. (2015). Maintenance policy optimization—literature review and directions. *Int J Adv Manuf Technol*, *76*, 1263–1283. https://doi.org/10.1007/s00170-014-6341-2

Frontoni, E., Pollini, R., Russo, P., Zingaretti, P., & Cerri, G. (2017). Hdomo: Smart sensor integration for an active and independent longevity of the elderly. *Sensors*, *17*(11).

Gharoun, H., Keramati, A., Nasiri, M. M., & Azadeh, A. (2019). An integrated approach for aircraft turbofan engine fault detection based on data mining techniques. *Expert Systems*, *36*(4). https://doi.org/10.1111/exsy.12370

Görür, O. C., Yu, X., & Sivrikaya, F. (2021). Integrating predictive maintenance in adaptive process scheduling for a safe and efficient industrial process. *Applied Sciences*, *11*(11), 5042.

Hesabi, H., Nourelfath, M., & Hajji, A. (2022). A deep learning predictive model for selective maintenance optimization. *Reliability Engineering & System Safety*, *219*. https://doi.org/https://doi.org/10.1016/j.ress.2021.108191

Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447–459. https://doi.org/10.11591/ijeecs.v9.i2.pp447-459

Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. *Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–7. https://doi.org/10.1109/ICCCNT.2013.6726842

Kiu, C.-C. (2018). Data mining analysis on student's academic performance through exploration of student's background and social activities. *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*, 1–5. https://doi.org/10.1109/ICACCAF.2018.8776809

Linke, B. S., Garcia, D. R., Kamath, A., & Garretson, I. C. (2019). Data-driven sustainability in manufacturing: selected examples. *Procedia Manufacturing*, *33*, 602–609. https://doi.org/https://doi.org/10.1016/j.promfg.2019.04.075

Lucantoni, L., Bevilacqua, M., & Ciarapica, F. E. (2019). An enhanced approach for implementing risk-based maintenance in a total productive maintenance perspective. *XXIV Summer School "Francesco Turco" – Industrial Systems Engineering*, 53–59. http://www.summerschool-aidi.it/edition-2019/cms/extra/papers/211.pdf

Mansouri, S., Castronovo, F., & Akhavian, R. (2020). Analysis of the synergistic effect of data analytics and technology trends in the AEC/FM ındustry. *Journal of Construction Engineering and Management*, *146*(3).

Mikut, R., & Reischl, M. (2011). Data mining tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*, 431–443. https://doi.org/10.1002/widm.24

Mohamed, A., Hamdi, M. S., & Tahar, S. (2019). Decision tree-based approach for defect detection and classification in oil and gas pipelines. In *Advances in Intelligent Systems and Computing* (Vol. 880, pp. 490–504). https://doi.org/https://doi.org/10.1007/978-3-030-02686-8_37

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. *International Journal of Production Research*, *46*(13). https://doi.org/https://doi.org/10.1080/00207540601142645

Mugnini, A., Polonara, F., & Arteconi, A. (2021). Energy flexibility curves to characterize the residential space cooling sector: The role of cooling technology and emission system. *Energy and Buildings*, *253*. https://doi.org/https://doi.org/10.1016/j.enbuild.2021.111335

Paolanti, M., Sturari, M., Mancini, A., Zingaretti, P., & Frontoni., E. (2017). Mobile robot for retail surveying and inventory using visual and textual analysis of monocular pictures based on deep learning. *2017 European Conference on Mobile Robots (ECMR)*. https://doi.org/10.1109/ECMR.2017.8098666

Pisacane, O., Potena, D., Antomarioni, S., Bevilacqua, M., Ciarapica, F. E., & Diamantini, C. (2021). Data-driven predictive maintenance policy based on multi-objective optimization approaches for the component repairing problem. *Engineering Optimization*, *53*, 1–20. https://doi.org/https://doi.org/10.1080/0305215x.2020.1823381

Sezer, E., David Romero, D., Guedeal, F., Macchi, M., & Emmanouilidis, C. (2018). An Industry 4.0-enabled low cost predictive maintenance approach for SMEs: A use case applied to a cnc turning centre. *24th International ICE-Conference on Engineering, Technology and Innovation*. https://doi.org/10.1109/ICE.2018.8436307

Thapa, S., Alsadoon, A., Prasad, P. W. ., Al-Dala'in, T., & Rashid, T. A. (2020). Software defect prediction using atomic rule mining and random forest. *5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)*, 1–8. https://doi.org/10.1109/CITISIA50690.2020.9371797

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, *21*(6), 2560–2574. https://doi.org/https://doi.org/10.1016/j.ymssp.2006.12.007

Yang, Q., Ji, G., & Zhou, W. (2017). The correlation analysis and prediction between mobile phone users complaints and telecom equipment failures under big data environments. *2nd International Conference on Advanced Robotics and Mechatronics (ICARM)*, 201–206. https://doi.org/10.1109/ICARM.2017.8273160

## Author Information

**Laura Lucantoni**
Universita Politecnica Delle Marche
Via Brecce Bianche, 12, 60131 Ancona AN
Contact e-mail: *l.lucantoni@pm.univpm.it*

**Filippo Emanuele Ciarapica**
Universita Politecnica Delle Marche
Via Brecce Bianche, 12, 60131 Ancona AN

**Maurizio Bevilacqua**
Universita Politecnica Delle Marche
Via Brecce Bianche, 12, 60131 Ancona AN

**To cite this article:**

Lucantoni, L., Ciarapica, F. E., & Bevilacqua, M. (2022). Learning-based algorithm for fault prediction combining different data mining techniques: A real case study. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 21,* 55-63.