

# Application of data mining in a maintenance system for failure prediction

P. Bastos

*Instituto Politécnico de Bragança, Bragança, Portugal*

I. Lopes

*Universidade do Minho, Escola de Engenharia, Guimarães, Portugal*

L. Pires

*Instituto Politécnico de Bragança, Bragança, Portugal*

**ABSTRACT:** In industrial environment, data generated during equipment maintenance and monitoring activities has become increasingly overwhelming. Data mining presents an opportunity to increase significantly the rate at which the volume of data can be turned into useful information. This paper presents an architecture designed to gather data generated in industrial units on their maintenance activities, and to forecast future failures based on data analysis. Rapid Miner is used to apply different data mining prediction algorithms to maintenance data and compare their accuracy in the discovery of patterns and predictions. The tool is integrated with an online system which collects data using automatic agents and presents all the results to the maintenance teams. The purpose of the prediction algorithms is to forecast future values based on present records, in order to estimate the possibility of a machine breakdown and therefore to support maintenance teams in planning appropriate maintenance interventions.

## 1 INTRODUCTION

In modern manufacturing environments, large amounts of data from different areas of activity are collected and stored in database systems, including data from product and process design, assembly, materials planning, quality control, scheduling, maintenance, etc. In these environments, the volume of data increases at an unprecedented rate. In addition to the vast quantity of data, the data stored in databases usually includes patterns, trends, associations and dependencies that makes complex its interpretation. Therefore, the analysis of accumulated data has been limited, which has led to the “rich data but poor information” problem (Wang & McGreavy 1998).

In an attempt to overcome this limitation, data mining techniques were introduced into the manufacturing field and have emerged as an important tool for knowledge acquisition from the manufacturing databases (Choudhary et al. 2009).

In industrial environment, maintenance activities are typically intended to reduce failures of machinery creating conditions for increasing equipment availability and consequently

increasing productivity. Manufacturing data collected in real-time contains valuable information and knowledge that could be integrated within prediction systems to improve decision making and enhance productivity (Elovici & Braha 2003).

The main objective of the on-going project is to develop a recording and prediction system to foresee machine failures in manufacturing units, globally dispersed. A crucial and core component of the functionality of this system lies in the ability to collect and interpret dispersed data, creating predictive models allowing maintenance teams to act before problems occur. In this paper a Rapid Miner prototype that complies with the conceptual specifications previously defined for the interpretative module on the global maintenance system, is presented. Into the system will flow real-time data gathered by intelligent agents from different machines in world scattered production facilities, allowing after data mining activities to produce warnings guiding future maintenance actions. Such a system could improve overall maintenance operations, reducing breakdowns and maintenance costs, assisting managers for a better planning of maintenance activities. An empirical study was conducted by using real data from production

lines of an industrial company in order to validate the proposed approach.

This paper is organized as follows. In section 2, the literature regarding maintenance and data mining is reviewed and the decision tree model that is applied in this work is described, presenting the reasons for its application. Section 3 starts with the presentation of the framework structure of the maintenance decision support as well the methodology developed for the data mining prototype creation. Then, the database used in the study is described as well as the data preparation process. In the same section, the data mining prototype created using Rapid Miner is also introduced, ending with the presentation of system outputs examples and results validation. Section 4 concludes the article with a brief summary of the main results.

## 2 LITERATURE REVIEW

### 2.1 *Maintenance*

The need to satisfy organizations requirements leads to high pressure in maintenance systems. The maintenance function, considered non-value-adding, is ever more asked to contribute for costs reduction, keeping the machines in excellent working condition (Bansal et al. 2006).

It is well known that 99% of machine failures are preceded by some indicators (Bloch & Geitner 2012). The earliest maintenance technique is basically breakdown maintenance (also called unplanned maintenance, or run-to-failure maintenance), which takes place only at breakdowns. A later maintenance technique is time-based preventive maintenance (also called planned maintenance), which sets a periodic interval to perform preventive maintenance regardless of the health status of a physical asset (Jardine et al. 2006). Breakdown maintenance is a primitive maintenance method and often with great economic loss. It is only suited to noncritical areas under the following preconditions: the consequences of failure are slight, there is no safety risk, the failure will be identified quickly, and the repair will be quick (Fu et al. 2004). Planned maintenance is centered on reducing the probability of failure (Lofsten 2000). This kind of maintenance is time-oriented, whereas predictive maintenance is actual condition and trend-oriented (Fu et al. 2004). Most of failures do not occur instantaneously, and usually there are some kinds of degradation process or symptoms from normal states to failures. Therefore, the actual conditions and their trends should be assessed and predicted during the degradation process, and appropriate maintenance actions should be taken before breakdown. This is the main target of predictive maintenance.

In these days of high technology and efficient data communication, preparation of effective planning, coordination, and scheduling of the maintenance tasks can be accomplished far more efficiently with the support of a Computerized Maintenance Management System (CMMS). A CMMS is much more than just a way to schedule maintenance actions. By using a CMMS, it is possible to create equipment logs to record events associated with equipment, create work orders automatically according to a schedule or manually from service requests, track scheduled services, create a maintenance history, record employee activity, downtime of a device, and much more. CMMS are used for all aspects of maintenance planning, management and control. CMMS must be flexible and adaptable, because every organization is considered unique (Mukattash et al. 2011). Creating a system using databases ensures a quick and efficient access to information, as well as their integrity. Thus, data can be easily analysed and results obtained can be provided more effectively to system users.

### 2.2 *Data mining*

Data mining is defined as the exploration and analysis, by automatic or semiautomatic means, of large quantities of data stored in databases. Its main focus is the discovery of useful knowledge, including meaningful patterns and rules, from raw and apparently unrelated data (Berry & Linoff 1993, Han & Kamber 2001). Data mining models are generally categorized in four main types: association rules, clustering, classification, and prediction (Fayyad & Piatetsky-Shapiro 1996). Association rules consists in discovering rules showing conditions for attribute values that usually occur together in a dataset. Clustering is the process of detaching a dataset into several different groups or clusters. Clustering objects goal consists in minimizing the inter-class similarity and maximizing the intra-class similarity. Classification approach develops a function or model that identifies the categorical class of an object based on its attributes. A classification model is generally constructed by analyzing the relationship between the attributes and the object classes in the training dataset followed by validation using testing dataset. Prediction involves a model that infers a continuous value or future data trends. These models must accurately recognize particular data patterns that indicate upcoming data behavior. Prediction models may be based on classification or regression. In classification, the predicted variable is a binary or categorical variable. Some popular classification methods include decision trees, logistic regression

and support vector machines. In regression, the predicted variable is a continuous variable. Some popular regression methods include linear regression, neural networks, and support vector machine regression (Kumar & Vijayalakshmi 2011).

The literature on machinery diagnostics and maintenance prognostics is huge and diverse primarily due to a wide variety of production systems. Hundreds of papers in this area, including theories and practical applications, appear every year in academic journals, conference proceedings and technical reports. Young et al. 2010 present a study about the implementation of data mining algorithms on an F-18 aircraft data as well as methods for properly structuring a database system to store this data. They develop data mining models for detection of problems before they become critical. Liao et al. 2012 present a work were a novel data-driven machinery prognostic approach for machine performance assessment and prediction is created. With this prognostic information, a predictive maintenance model is proposed for a repairable deteriorating machine. As machine performance can be assessed, once it reaches the maintenance threshold, a maintenance operation is performed to restore the machine. Pan et al. 2012 present a study based on the circumstance that, with the advancements in sensor and prognostic technologies, machine's condition can be monitored and assessed over time through conducting predictive maintenance. They propose a single machine-based scheduling model incorporating production scheduling and predictive maintenance. A machine's effective age and remaining maintenance life are introduced to describe machine degradation. A numerical example is given where the computational results show that the integrated scheduling model has better performance than the existing models, which proves its efficiency. Raheja et al. 2006 present a work proposing a combined data fusion/data mining-based architecture for Condition-Based Maintenance (CBM). In the proposed architecture, methods from both these domains (data fusion and data mining) analyze CBM data to determine the overall condition or health of a machine. The resultant information is then used by a predictive maintenance model to determine the best course of action for maintaining critical equipment.

### 2.3 Decision trees

Decision trees are powerful and popular tools for classification and prediction (Santos & Azevedo 2005). The attractiveness of decision trees, in contrast to other models such as neural networks, is based on the understandable form in which the

rules are represented. In some applications, the accuracy of a classification or prediction is the only thing that matters. In such cases, it does not matter how or why the model works. In other scenarios, the ability to explain the reason for a decision is crucial (Quinlan 1986). In maintenance scope, in order to allow the responsible for maintenance to establish a proper set of maintenance actions applicable and understandable, descriptions must be provided. Decision trees provide this information based on understandable representation. So, it fits well than ever on this scenario.

A decision tree is a classifier in the form of a tree structure, where each node is either a leaf node that indicates the value of the target attribute (class), or a decision node that specifies some test to be carried out on a single attribute-value, with one branch and sub-tree for each possible outcome of the test. Usually, tree construction relies on the following concepts:

- Gini impurity, used by the CART algorithm (Classification and Regression Trees) (Equation 1). It is based on squared probabilities of membership for each target category in the node. It reaches its minimum (zero) when all cases in the node fall into a single target category. Suppose  $y$  takes on values in  $\{1, 2, \dots, m\}$  and let  $f(i, j)$  = frequency of value  $j$  in node  $i$ . That is,  $f(i, j)$  is the proportion of records assigned to node  $i$  for which  $y = j$  (Breiman et al. 1984).

$$I_G(i) = 1 - \sum_{j=1}^m f(i, j)^2 = \sum_{j \neq k} f(i, j)f(i, k) \quad (1)$$

- Information gain used by ID3, C4.5 and C5.0 tree generation algorithms (Equation 2). Information gain is based on the concept of entropy used in information theory (Quinlan 1998).

$$I_E(i) = -\sum_{j=1}^m f(i, j) \log_2 f(i, j) \quad (2)$$

Lim et al. 2000 compared the prediction accuracy, complexity and training time of 33 classification algorithms, indicating that decision tree provides good accuracy and data interpretation.

## 3 RESEARCH FRAMEWORK

### 3.1 Problem definition

There are different methods and techniques to assist in the implementation of predictive maintenance, i.e. to predict future equipment failures, or for predicting its performance degradation

starting from its current condition and its historical functional knowledge (Fornaro et al. 2004). The availability of databases containing relevant information provided by existing machines on different production lines and in different production environments may be a key element for any diagnostic tool. Figure 1 represents the logical structure for the maintenance decision support developed under the project. The decision support system is divided into three different levels:

- Level 1—Alarm. When a machine fails, data are collected through automatic agents and stored. At the same time an alarm is generated to the maintenance team or leader. This alarm also provides rough information about failure characteristics including failure nature and involved equipment or machine.
- Level 2—Maintenance actions. Automatic agents programmed for data collection installed in the machine collect data relative to the intervention performed and returns useful information regarding the origin of the problem and the resolution. Data related to the intervention is transmitted through the data infrastructure and stored in a historical database to be used at the next level.
- Level 3—Predictive maintenance. Algorithms and data mining techniques are applied based on stored maintenance historical data in combination with a monitoring architecture.

The scope of this work is not focused on nuclear maintenance activities specifications. It is intended to introduce an innovative approach for predicting equipment failures, incorporating to maintenance scope well proved techniques from other fields of science and technology.

The objective is to use techniques which can help to increase the efficiency of systems, leading to improvements in functionality that are translated into availability increase and consequently into higher productivity. These gains resulted from proactive changes to contingency and initial preventive plans derived from the analysis of the recorded data. Building a predictive maintenance

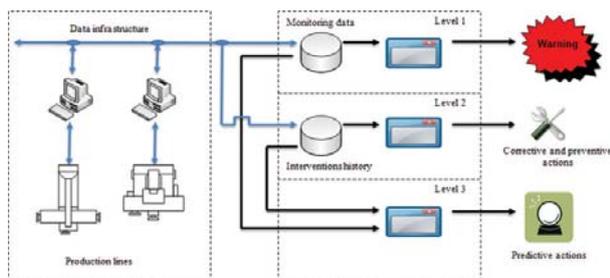


Figure 1. Scheme of the decision support system.

solution that alerts maintenance personnel of eminent breakdowns fits perfectly data mining nature. In order to work properly, data mining requires large quantities of data which in the proposed decision support system is achieved as shown in Figure 2.

At a local level, in each participant organization, data produced from maintenance activities and monitoring operations is collected at local databases by intelligent agents, being posteriorly transferred to a cloud system where prediction activities will be performed.

The cloud contains a prototype prediction system that uses the information produced in remote and sparsely plants, whose main aim is to discover implicit and hidden knowledge. Thus, it will be possible to generate models of behavior patterns for equipment, allowing the measurement of their functional deviations and therefore the prediction of the occurrence of breakdowns. Information will be resent through the cloud system to local facilities using alert warning on their local monitoring systems.

In order to better understand the methodology developed for the prediction prototype creation, the Figure 3 presents a graphical process representation through a flowchart schema.

### 3.2 Data preparation

Data that underpin the process of ongoing research are from a partner company of the project, an international semiconductor company. Data supplied was stored in two databases with distinct features. One contains equipment monitoring data (level 1) and the other saves data on maintenance corrective and preventive activities (level 2) performed by the maintenance team.

Although data mining find comfortable with large amounts of data, their quality is crucial for a fast and adequate feeding of the prediction system. During the activity of the plant, large quantities of data produced are not always considered properly

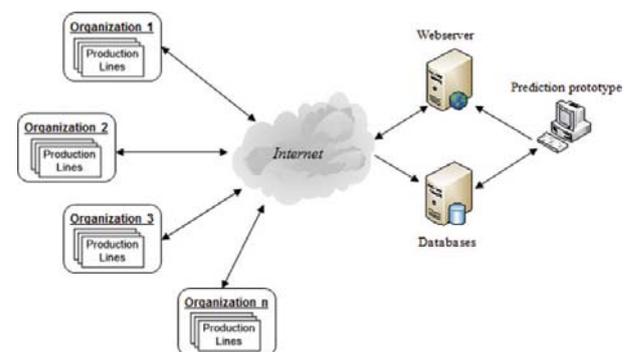


Figure 2. High level architecture.

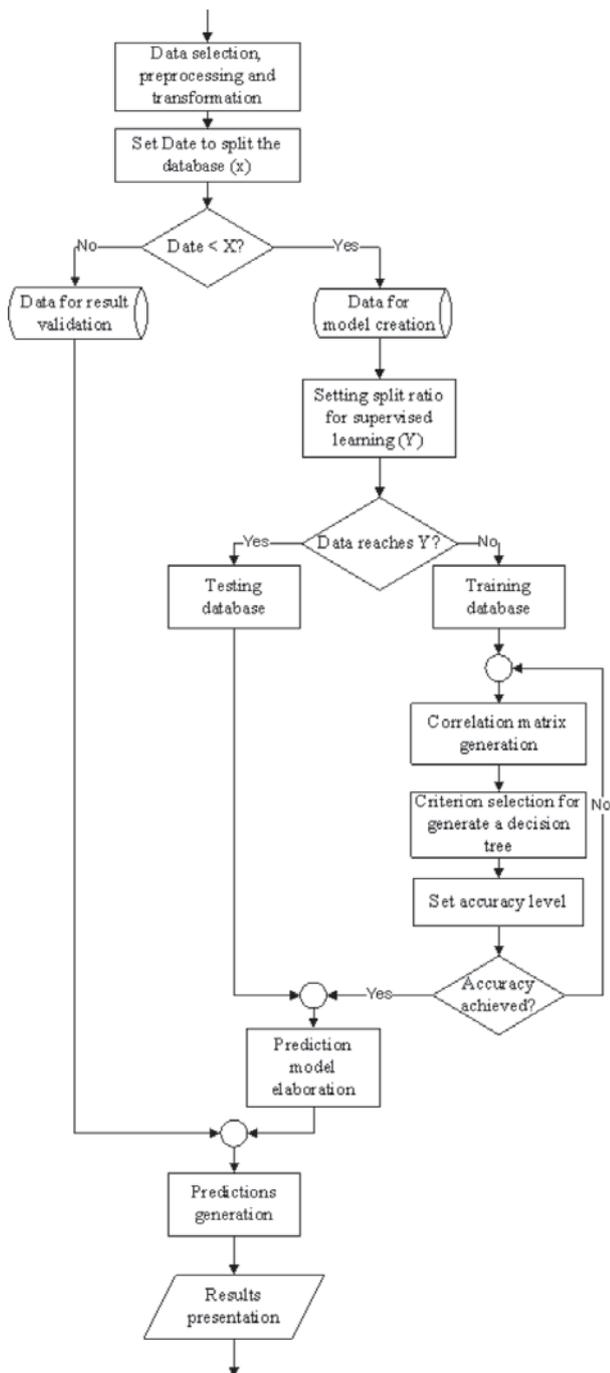


Figure 3. Methodology developed.

“cleaned” from practical and academic standpoint. Usually this data often include noisy, missing and inconsistent data. Data pre-processing can improve data quality and make easier efficient data mining tasks. Commonly, on this stage missing values are replaced or deleted, for instance new variables combining machine number with date are generated, i.e. selected data is transformed into appropriate formats allowing for further and suitable analysis.

Maintenance activities database start storing data in December 2011. This database is composed

by seven attributes: Ava\_id (breakdown id); Avaria\_tipo (type of breakdown); ID\_Locais (part of the machine); ID\_componente (machine component); Avarias\_motivos (reason of breakdown); Avarias\_accoes (action made in maintenance); Data\_registo (record date).

Monitoring data is captured and stored every second from all the factory machines, originating more than two hundred thousand of records, just from the last three years. This database is composed by seven attributes: Id (record id); Data\_registo (record date); Valor (value of the monitored indicator); Val\_low\_limit\_w (Lower limit warning); Val\_low\_limit\_cr (Critical lower limit warning); Val\_hi\_limit\_w (Upper limit warning); Val\_hi\_limit\_cr (Critical upper limit alert).

Thus, in order to create and test the prototype and not overload the system with a huge amount of data, which involves large periods of time testing different data mining models and subsequent validation of its performance, it was decided to work with one machine randomly selected and with monitoring data related to the reliability indicator given by the existent computer in this machine through his Hard disk drive S.M.A.R.T. capability (Self-Monitoring, Analysis and Reporting Technology). The monitored indicator, in this case, is the “Soft Read Error Rate”.

The prototype is running on an Intel Core Duo 2.9 GHz CPU 64 bits with 8 GB RAM, with Rapid-I’s Rapid Miner 5.3.008.

### 3.3 Data mining module

The development of the prediction prototype is based on the creation and application of two stages: (i) the first stage is responsible for creating the prediction model capable to predict the necessity to perform an intervention (Avaria\_tipo) based in the other maintenance and monitoring data fields (Fig. 4). (ii) The second stage relies on the use of

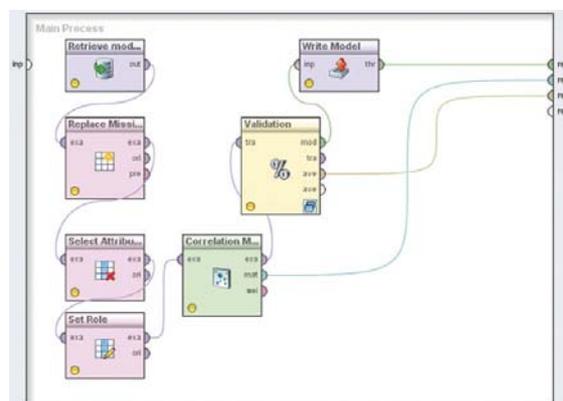


Figure 4. Stage of generating the prediction model with supervised learning using Rapid Miner.

the model generated in the first stage, applying new data inputs and generating a set of outputs, easily understandable, concerning the need and type of intervention that should be performed. Additionally it provides an easy and understandable way to show the relations between the system attributes.

Selected data to create the model concerns the period between 01/01/2012 and 31/08/2012, and results were validated recurring to data mining prototype outputs with data stored from 01/09/2012 until 31/12/2012. Data/hour field available in both databases were merged through comparison to a new database that includes all fields of both databases. At the first phase of the first stage, a selection of attributes was made filtering worthless attributes. The resultant database is composed by ten attributes and 529 records. The attributes are: Avaria\_tipo; ID\_Locais; ID\_componente; Avarias\_motivos; Avarias\_accoes; Valor; Val\_low\_limit\_w; Val\_low\_limit\_cr; Val\_hi\_limit\_w; Val\_hi\_limit\_cr.

Considering the designated dates for the database splitting, we obtain a database for the model creation with 330 records and a database for the results validation with 199 records.

As shown in Figure 4, during the first stage, a correlation matrix that determines correlation between all attributes was created. Based on these correlations, it is also possible to produce a weights vector. This matrix is used to show whether and how strongly pairs of attributes are related (Fig. 5). These weights are used by the system to create the decision tree.

Analyzing this matrix we can detect a high level of correlations between the monitoring attributes. The correlations level between the attributes “Valor” and “val\_low\_limit\_w” is 90%, and between the “val\_low\_limit\_w” and “val\_low\_limit\_cr” is 99%. At the next step a validation operator to model generation and to apply a supervised learning was used. With this operator the database was split, in a basis of 70% for model training and 30% for model testing.

In Figure 6 the operator configuration to generate the predictive model and its validation is shown.

In the stage of creating (training) the model, we created and tested two different trees: (i) The first one applying a gini index criterion for its creation and with a confidence level of 0.25 for

Attributes	ID_Locais	ID_Componentes	Avarias_motivos	Avarias_accoes	Valor	Val_low_limit_w	Val_low_limit_cr	Val_hi_limit_w	Val_hi_limit_cr
ID_Locais	1	0.055	-0.028	-0.007	0.084	0.009	0.094	0.040	0.024
ID_Componentes	0.055	1	0.133	0.120	0.037	0.009	0.006	0.053	0.057
Avarias_motivos	-0.028	0.133	1	0.029	0.006	-0.038	-0.041	-0.002	0.000
Avarias_accoes	-0.007	0.120	0.029	1	-0.091	-0.016	-0.011	0.005	0.010
Valor	0.084	0.037	0.006	-0.091	1	0.901	0.983	0.775	0.690
Val_low_limit_w	0.009	0.009	-0.038	-0.016	0.901	1	0.998	0.515	0.444
Val_low_limit_cr	0.094	0.006	-0.041	-0.011	0.983	0.998	1	0.481	0.411
Val_hi_limit_w	0.040	0.053	-0.002	0.005	0.775	0.515	0.481	1	0.905
Val_hi_limit_cr	0.024	0.057	0.000	0.010	0.690	0.444	0.411	0.905	1

Figure 5. Correlation matrix.

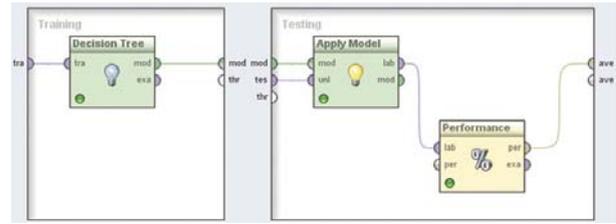


Figure 6. Configuration and implementation of the validation operator.

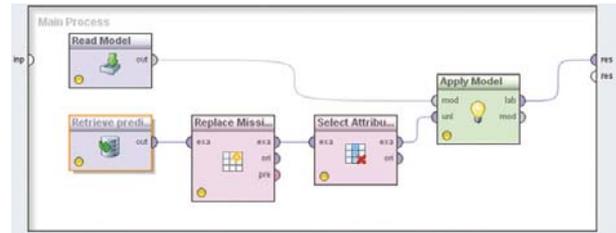


Figure 7. Prediction stage.

the confidence level used for the pessimistic error calculation of pruning. (ii) The second one applying an information gain criterion, so the entropy of all the attributes is calculated. The attribute with minimum entropy is selected for splitting. More accurate results were obtained applying the gini index criterion for the decision tree creation. A prediction accuracy level of 71.3% was achieved. In the testing phase, the generated model was applied and its performance was tested applying a “performance” operator. This operator delivers a list of performance criteria values of the classification task.

As showed in Figure 4, the model is written to a file (write model operator), allowing applying it to the database which was created to validate the model obtaining perdition outputs.

Figure 7 represents the second stage which is responsible for applying the generated model to the validation database (fields from 01/09/2012 to 31/12/2012), not including the field predicted by the model (i.e. Avaria\_tipo). Along with the new entries, the model generated is also provided as stage input, then the model is applied using the “apply model” operator in Rapid Miner.

Using the output of this stage, the team responsible for maintenance activities has at his disposal elements (graphs and tables) that indicate the need for preventive interventions, as well as the confidence level of this prediction.

### 3.4 Evaluation and interpretation

The output data presented in the preceding subsection are the result of applying the conceptual

Table 1. Occurrence number of both real and foreseen interventions.

	Preventive	Corrective	Total
Real	50	178	228
Foreseen	45	154	199

features described previously to a selected machine of the partner of this project, to test the prototype. Table 1 compares the number of interventions carried and the number of interventions that the system suggested. The real number of occurrences was calculated by counting the occurrences performed by the maintenance teams. The predicted one was calculated through the counting of the prediction table occurrences displayed in the main system output (Fig. 9). The values obtained are the result of clustering and subsequent accounting of all corrective interventions regardless of their nature (i.e. Electric, Mechanical and Setup), and the preventive ones are those that were labeled as “Manutenção”.

Without taking into consideration that a maintenance action directly affects the possible need to make a new intervention, whether corrective or preventive, Table 1 shows, although in a simplistic way, the comparison of real interventions data and the prediction model outputs. The prediction provides a smaller number of interventions than reality. In total of all interventions the difference between the forecast and the reality is about 13% less. This difference could be a prediction error and/or can also signify an existence of a small interventions increase in the company over the previous period. These results have not yet been considered for redesign the maintenance plans, but with the predictions made preventive interventions can be carried when the system predict with a high degree of certainty the future occurrence of a fault.

Figure 8 shows a system output example capable to provide a clustering vision of the monitored value of the “Valor” attribute, relating it to the prediction made and the confidence level of each prediction. These relations are the source for the creation and exhibition of the main output system. Each cluster has an associated degree of confidence and this is visible in the prediction table (Fig. 9). For the same monitored value the system can provide different types of interventions, however one of them have greater confidence than another. This chart offers to the maintenance team specific information about the value monitored and the necessity of performing a kind of maintenance action over this machine.

The system shows good accuracy and confidence to predict interventions that can be used by maintenance teams. It is possible to see through

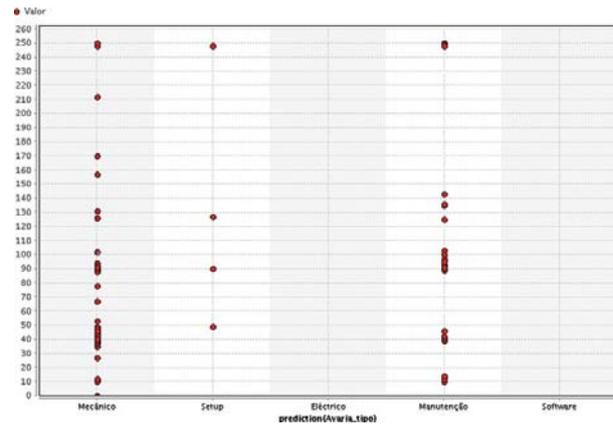


Figure 8. Relations between the value monitored of the “Valor” attribute and the prediction made.

prediction(Atributo_tipo)	confidence(Mecânico)	confidence(Setup)	confidence(Eléctrico)	confidence(Manutenção)	confidence(Software)	Valor
Mecânico	1	0	0	0	0	131
Mecânico	0.495	0.093	0.009	0.393	0.009	44
Manutenção	0.125	0.125	0	0.750	0	250
Mecânico	0.402	0.154	0	0.385	0	45
Mecânico	0.495	0.093	0.009	0.393	0.009	46
Manutenção	0.500	0	0	0.500	0	249
Mecânico	0.455	0.273	0.182	0.091	0	48
Mecânico	0.402	0.154	0	0.385	0	248
Mecânico	0.455	0.273	0.182	0.091	0	47
Mecânico	0.495	0.093	0.009	0.393	0.009	43
Mecânico	0.455	0.273	0.182	0.091	0	47
Mecânico	0.495	0.093	0.009	0.393	0.009	48
Mecânico	0.402	0.154	0	0.385	0	35
Mecânico	0.706	0	0.059	0.235	0	44
Mecânico	0.590	0.080	0	0.240	0	44
Mecânico	0.680	0.080	0	0.240	0	37
Mecânico	0.495	0.093	0.009	0.393	0.009	91
Mecânico	0.680	0.080	0	0.240	0	43
Mecânico	1	0	0	0	0	157

Figure 9. Prediction table.

the observation of Figure 9, the confidence of each prediction obtained based in the created model and the monitoring values obtained from the system, which will be studied in later stages of this project.

#### 4 CONCLUSIONS

This paper presents early stages of a prediction system based in monitoring data acquired during a manufacturing productive work and also with data related to historical records concerning different types of maintenance actions. Specifically, this paper focuses on the construction and validation of the prototype of the core of the entire system, which is the one that using data mining can direct the actions of the maintenance teams with high levels of accuracy and confidence.

The proposed approach is characterized by the application of data mined knowledge that offers the capability of deploying “Interesting (non-trivial, implicit, previously unknown, and potentially useful) information or patterns from data in large databases” (Fayyad & Piatetsky-Shapiro 1996).

As mentioned, the prototype was built based on a single machine and in one monitored indicator,

thus the next developments will pass through the integration of other monitoring indicators obtained from other similar machines in the prototype to strengthen it. After exceeded this step, system will be expanding to the whole universe of machines and indicators of the manufacturing unit and later the full integration into the cloud for universal use. As a first step of this ambitious proposal, the proposed approach was validated and its practical feasibility was showed based on empirical results.

## ACKNOWLEDGEMENTS

This work was financed with FEDER Funds by Programa Operacional Fatores de Competitividade—COMPETE and by National Funds by FCT—Fundação para a Ciência e Tecnologia, Project: FCOMP-01-0124-FEDER.

## REFERENCES

- Bansal, D., D. J. Evans & B. Jones 2006. Bjest: A reverse algorithm for the real time predictive maintenance system. *International Journal of Machine tools and manufacture* 46: 1068–1078.
- Berry, M. J. A. & G. Linoff, 1993. *Mastering data mining: The art and science of customer relationship management*. USA, John Wiley & Sons, Inc.
- Bloch, H. P. & F. K. Geitner, 2012. *Machinery failure analysis and troubleshooting* (fourth edition), Elsevier Inc.
- Breiman, L., J. H. Friedman, R. A. Olshen & C. J. Stone, 1984. *Classification and regression trees*, CRC Press.
- Choudhary, A. K., J. A. Harding & M. K. Tiwari 2009. Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing* 20: 501–521.
- Elovici, Y. & D. Braha 2003. A decision-theoretic approach to data mining. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans* 33(1): 42–51.
- Fayyad, U. & G. Piatetsky-Shapiro, 1996. *Knowledge discovery and data mining: Towards a unifying framework*. Second International Conference on Knowledge Discovery and Data Mining, Oregon.
- Fornaro, O., M. C. Magro & P. Pinceti, 2004. *Diagnostics for measure transmitters (predictive maintenance)*. 21st IEEE Instrumentation and Measurement Technology Conference, Como, Italy.
- Fu, C., L. Ye, Y. Liu, R. Yu, B. Iung, Y. Cheng & Y. Zeng 2004. Predictive maintenance in intelligent-control-maintenance-management system for hydroelectric generating unit. *IEEE Transactions on Energy Conversion* 19 n°1: 179–186.
- Han, J. & M. Kamber, 2001. *Data mining: Concepts and techniques*. USA, Morgan Kaufmann Publishers.
- Jardine, A. K. S., D. Lin & D. Banjevic 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20: 1483–1510.
- Kumar, S. A. & M. N. Vijayalakshmi 2011. Efficiency of decision trees in predicting student's academic performance. *Computer Science & Information Technology* 02: 335–343.
- Liao, W., Y. Wang & E. Pan 2012. Single-machine-based predictive maintenance model considering intelligent machinery prognostics. *The International Journal of Advanced Manufacturing Technology* 63(1–4): 51–63.
- Lim, T.-S., W.-Y. Loh & Y.-S. Shih 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine Learning* 40: 203–229.
- Lofsten, H. 2000. Measuring maintenance performance—in search for a maintenance productivity index. *Int. J. Production Economics* 63: 47–58.
- Mukattash, A., R. H. Fouad, H. Kitan & M. Samhoury 2011. Computer-aided maintenance planning system for industrial companies. *Jordan Journal of Mechanical and Industrial Engineering* 5(3): 227–234.
- Pan, E., W. Liao & L. Xi 2012. A joint model of production scheduling and predictive maintenance for minimizing job tardiness. *The International Journal of Advanced Manufacturing Technology* 60(9–12): 1049–1061.
- Quinlan, J. R. 1986. Induction of decision trees. *Machine Learning* 1: 81–106.
- Quinlan, J. R. 1998. *C4.5 programs for machine learning*, Morgan Kaufmann Publishers, Inc.
- Raheja, D., J. Llinas, R. Nagi & C. Romanowski 2006. Data fusion/data mining-based architecture for condition-based maintenance. *International Journal of Production Research* 44(14): 2869–2887.
- Santos, M. F. & C. Azevedo, 2005. *Data mining—descoberta de conhecimento de bases de dados*, FCA—Editora de Informática.
- Wang, X. Z. & C. McGreavy 1998. Automatic classification for mining process operational data. *Industrial & Engineering Chemistry Research* 37(6): 2215–2222.
- Young, T., M. Fehskens, P. Pujara, M. Burger & G. Edwards, 2010. Utilizing data mining to influence maintenance actions AUTOTESTCON, 2010 IEEE.