Proceedings of the World Congress on Engineering 2012 Vol III
WCE 2012, July 4 - 6, 2012, London, U.K.

A Maintenance Prediction System using Data Mining Techniques

P. Bastos[1], I. Lopes[2], and L. Pires[3]

***Abstract* — In the last years we have assisted to several and deep changes in industrial companies, mainly due to market dynamics and the need to converge with a globalized and impatient world. These changes are transversal to the entire company also impacting on company maintenance function. In an attempt to eliminate faults and keep systems running without interruption, companies incorporated tools into their Information and Communication Technologies (ICT) systems. The benefits are clear in terms of resulting quality and in costs reduction, particularly those related with the data processing time and accuracy of the resulting knowledge.**

In their daily routine, companies produce and store endless and complex quantities of data of different nature, increasing the difficulty of use in real time. In this sense, considering the relevance of data collected on industrial plants, namely in its maintenance activities, it is intended with this paper to present a functional architecture of a predictive maintenance system, using data mining techniques on data gathered from manufacturing units globally dispersed. Data Mining will identify behavior patterns, allowing a more accurate early detection of faults in machines. The remote data collection is based on an intricate system of distributed agents, which, given its nature, will be responsible for remote data collection through the functional architecture.

***Keywords* — Agents, data mining, e-collaboration, maintenance, management.**

I. Introduction

There is a steady growing pressure on companies, urged by the worldwide competition, to streamline operations involving product and product related manufacturing system design, product manufacturing and system maintenance [1].

As markets become more dynamic, the need to introduce concepts of flexibility and agility, enabling companies to deliver customized products reacting promptly to fluctuating demands grows.

E-collaboration and collaborative systems have opened the possibility for geographically dispersed teams to work together by supporting coordination and cooperation [2]. In order to reach the competitive level of performances in terms of productivity, product quality and system availability, companies are motivated to join collaborative networks. Many companies have developed or are developing e-maintenance systems in order to provide a better service to customer's needs such as: i) service supplier, ii) service user, and iii) maintenance activities [3].

Maintenance activities are usually performed by integration of maintenance and process engineering functions at the phase of selection and application of machines and equipment; and also through proactive actions on those machines and equipments involving preventive and predictive maintenance [4].

In literature it is possible to find three generic types of maintenance [5, 6]:

- Corrective maintenance, consisting in repair actions when equipment or machine fails. The equipment is in action until the moment that it fails. At that moment it will be repaired or replaced. The main disadvantages of this approach include fluctuant and unpredictable production, high levels of non-conforming products and scraps as well as high levels of maintenance interventions motivated by catastrophic failures [7];
- Preventive maintenance which is characterized by periodic maintenance operations in order to avoid equipment failures or machinery breakdowns, determined through optimal preventive maintenance scheduling using a wide range of models describing the degrading process of equipment, cost structure, and admissible maintenance actions [8];
- Predictive maintenance uses some parameters measured in the equipment to "feel" when breakdown is eminent. This type of maintenance intends to make interventions on machinery before harmful events may occur.

The need to satisfy companies requirements leads to high pressure in factories maintenance systems. Maintenance, considered non-value-adding function, it is ever more requested to contribute for costs reduction, keeping the machines in excellent working condition [9].

Nowadays, the amount of data generated and stored during industrial activities exceeds the capacity to analyze them without the use of automated analysis techniques. As a consequence of that increase of information, data processing using traditional methods has become more difficult and complex [10].

Conventional tools of data analysis have limited capacity to detect patterns and discover the existent knowledge in data, because they only use statistical methods [11]. Hence there was a need for creating a new generation of computational tools and techniques in order to assist humans extracting useful information from data, in other words, knowledge. Thus, in the late 80's have emerged the area of Knowledge Discovery in Databases (KDD), using models and data mining techniques for extract useful knowledge, patterns and tendencies previously unknown, in a autonomous and semiautomatic way [12]. The application domain of data mining and its related techniques,

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.

[1]P. Bastos is with CIMO, Instituto Politécnico de Bragança, Campus de Stª Apolónia, 5301-855, Bragança, Portugal and with Centro Algoritmi, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal; e-mail: bastos@ipb.pt.

[2] I. S. Lopes is with Centro Algoritmi, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal; e-mail: ilopes@dps.uminho.pt.

[3] L. Pires is with CIMO, Instituto Politécnico de Bragança, Campus de Stª Apolónia, 5301-855, Bragança, Portugal; e-mail: luica@ipb.pt.

ISBN: 978-988-19252-2-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

methodologies and technologies have been greatly expanded in the last few years.

The development of automated data collection tools and the tremendous data explosion, the urgent need for interpretation and exploitation of massive data volumes, along with the existence of supporting tools, has resulted on the development and flourishing of sophisticated Data Mining methodologies.

Since Data Mining systems are comprised of a number of discrete, nevertheless dependent tasks, they can be thought of as collaboration networks, yet autonomous, units that regulate, control and organize all distributed activities involved in data cleaning, data transformation and reduction, algorithm application and results evaluation [13].

Research literature on intelligent agent system architectures has proven that such kind of problems that require the synergy of a number of distributed elements for their solution, can be efficiently implemented as a multi-agent system [14]. A multi-agent system consists of a group of intelligent agents that can take specific roles within an environment to cooperate with other agents [15].

The propose of this paper is to introduce a decentralized predictive maintenance system, based in the application of data mining techniques over maintenance data, generated by different machines in the same or in different production lines of industrial units globally dispersed and collected through a multi-agent system. The goal is to forecast a failure based on a pattern behavior matrix and generate a set of notifications for maintenance action scheduling. Predicting the possibility of breakdowns with bigger accuracy will increase systems reliability.

II. System Functionality

The aim of the system to be developed is to gather data about failures occurrence in similar machines from different factories, creating a system of distributed databases using intelligent agents, which allows through the use of data mining the prediction of failures, performing timely interventions in equipments and consequently increasing availability and productivity.

The system (Fig. 1) consists in three main modules:

i) Operational module;
ii) Modeling;
iii) Virtual platform.

Figure 1 shows that operational data is collected regarding corrective, preventive and predictive maintenance actions performed by the different project partners. These interventions are carried out following management maintenance strategies, incorporating all corrective maintenance actions performed, and also, all planned preventive and predictive interventions performed. The prediction prototype works over this data, applying data mining techniques in order to generate new knowledge and making it available thought a web platform.

The data collecting process is performed by a multi-agent system, which will be responsible for acquiring (agent 1), adapting and transforming the data (agent 2). Even when data on the factory floor is collected through maintenance operators (using a formalized internal registry, for example), this information has to be interpreted by a software agent, to adapt and transform the data structures into a semantically viable knowledge base (Fig. 2).

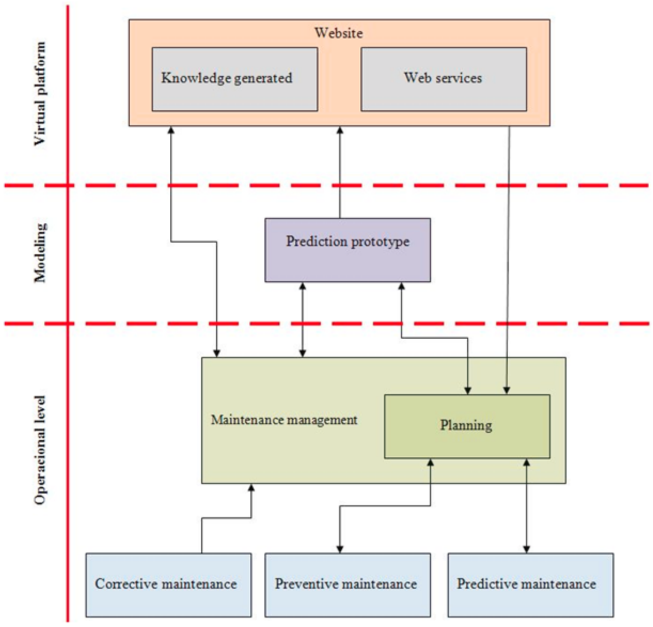

Fig. 1. Conceptual framework

All members that cooperate on the system must have mutual trust and a blind confidence on the infrastructure. The resultant information is critical. Therefore, it should be ensured that the information which flows from the agents to the database is a secure and trustful one. It is also important that the database server only accept information that come from authorized agents.

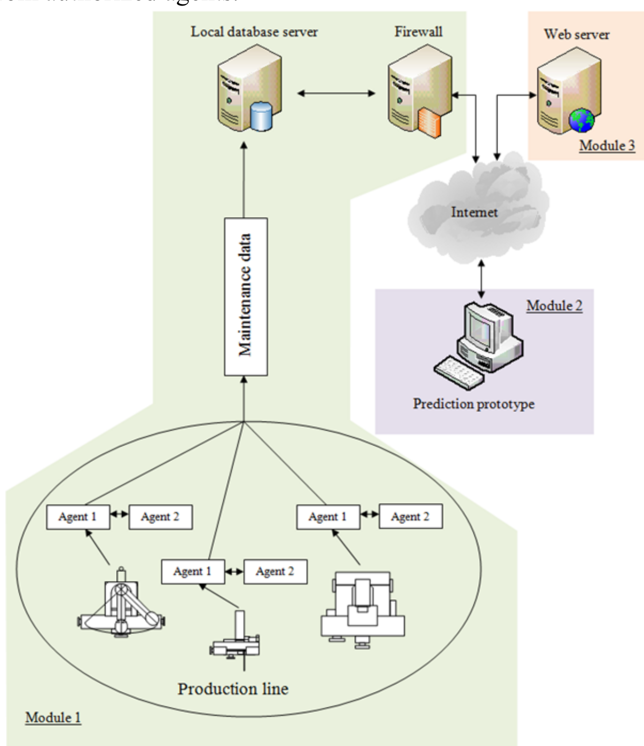

Fig. 2. System overview

Figure 2, which represents the system physically, shows the three main system modules:

1. The preprocessing unit (or operational module) which collects, normalizes and storages all the input data relating

ISBN: 978-988-19252-2-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

to corrective, preventive and preventive maintenance actions.

2. The prediction prototype (or modeling module), which is the core component, where Data Mining takes place. Data Mining techniques will focus on data to discover implicit and hidden knowledge in order to generate predict patterns of behavior and events. The possibility of events occurrence will be provided through a prediction system for each plant. Data related to intervention process, used material, the consequences of non-intervention and scenario generation will be provided over the form of decision support system module.

3. The web platform (or virtual platform module), which is the front-end of the data miner, provides the knowledge generated by the prediction prototype thought algorithm performance metrics, capabilities visualization and system alerts generation, ensuring a secure environment provided by a web portal that uses all the necessary web services (such as HTTPS and FTPS) and also controlled access by username and password.

The global system will be based on three main processes or activities, as shown in the Figure 3 using a IDEF0 representation: data management and communications (A1) which corresponds to the operational module, prediction knowledge system (A2) which corresponds to the modeling module and results and alert generation (A3) which corresponds to the virtual platform module.

The A1 activity will be responsible for data collecting. As output, this activity will produce a database based on normalization rules and the knowledge base required for study. The process of data gathering will be performed automatically by prospective agents using other mechanisms such databases and tools and communication channels. In this activity, after collecting all maintenance data, the first step to be performed consists in the identification of all data to be analyzed, producing a selected database. This database will passes through a phase of normalization resulting on a database that serves as input of the A2 activity.

The A2 activity is the main module of knowledge production and inference of knowledge related to equipment of a factory unit.

This activity will generate new knowledge that will feed of activity A3. A3 will use the resources of the A1 activity to send alerts that consists in proactive failures notifications. This output function aims to advise the maintenance responsible in order to act over the equipment before malfunctioning. The main activity (A2) consists in four sub activities (Fig. 4). The first sub activity is the data processing (A21) responsible for making all necessary changes in data so they can be used by different models of artificial intelligence. The data mining sub activity (A22) uses the generated database and KDD methodologies and applies different artificial intelligence techniques and algorithms.

In this sub activity, depending on the type of data to be analyzed, and the knowledge to be generated, the system will be able to apply different artificial intelligence algorithms such as:

• Decision trees, structures with a shape of a tree representing sets of decisions. These decisions create rules for classification of data sets. Decision trees can represent diverse types of data but the simplest and most familiar is numerical data. It is often desirable to organize nominal data as well. Nominal quantities are formally described by a discrete set of symbols;

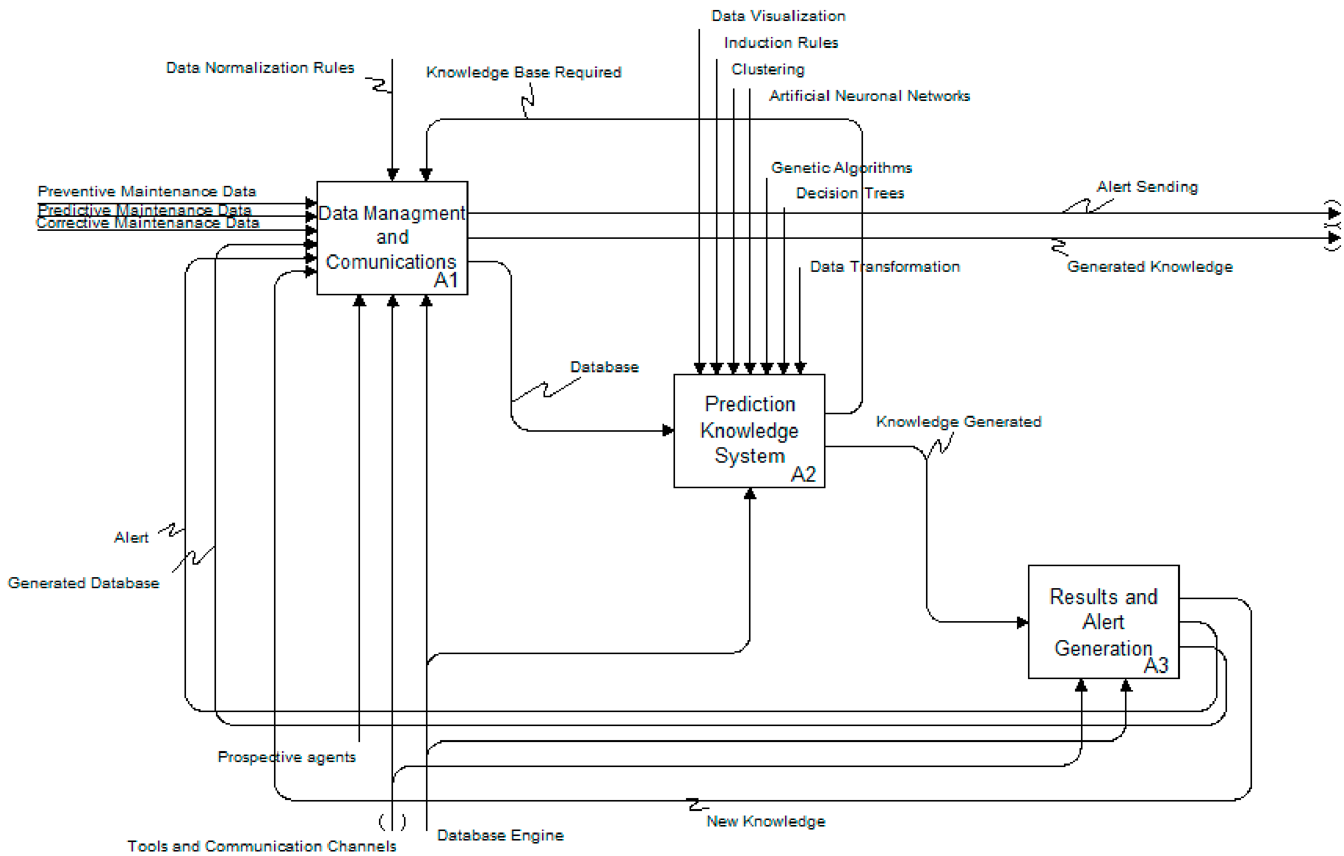

Fig. 3. Main System in IDEF0 format.

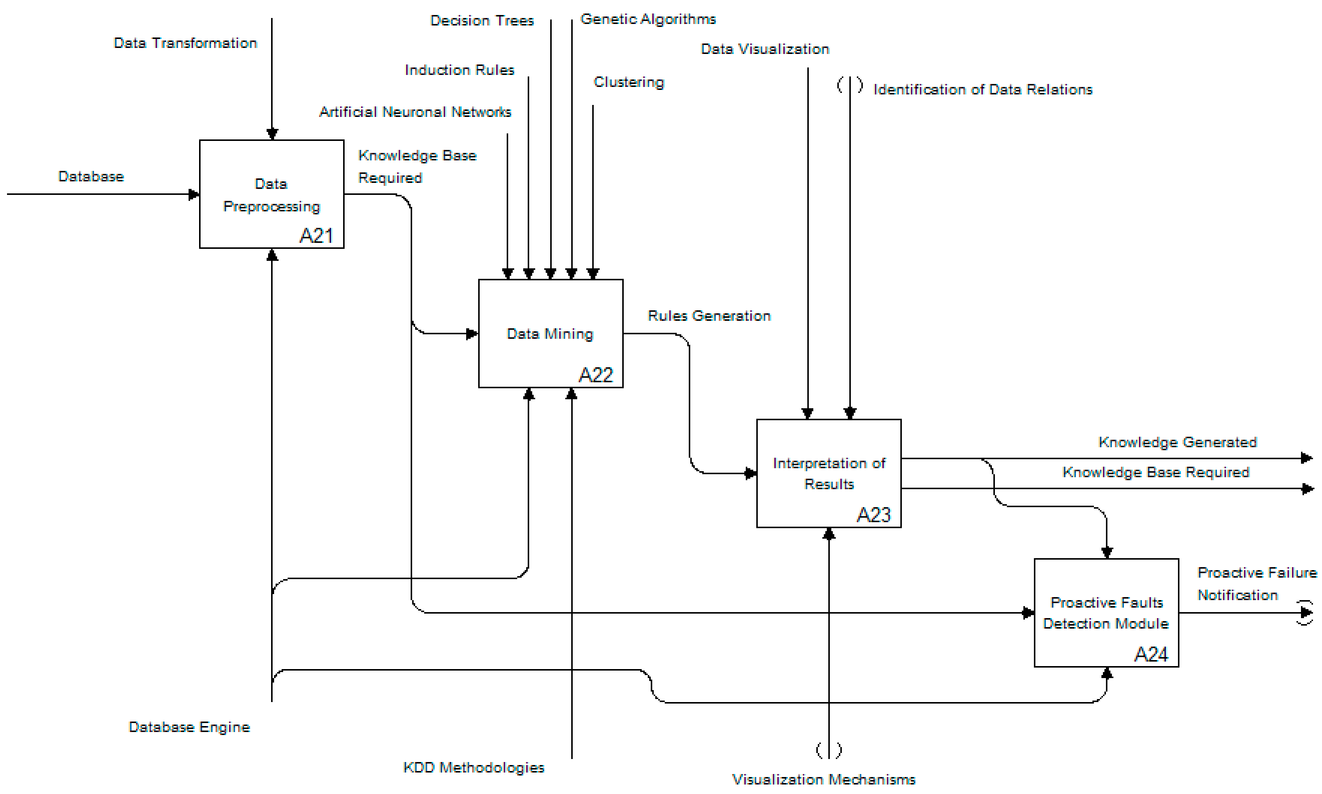

Fig. 4. A2 Sub-activities.

• Artificial neuronal network, which is one of the most known and used techniques in Data Mining [16]. It consists in a set of simple processing elements (nodes), with a large number of interconnections. The whole structure is based on recreations of human brain, more specifically on the ability to learn and self correct itself. Neuronal networks works over numeric or non-numeric data, with a remarkable ability to derive meaning from complicated or imprecise data. It can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and to answer "what if" questions;

• The induction rules that consist in the ability to detect trends and patterns in data groups, is one of the Data Mining techniques better known. This algorithm woks over numeric or non-numeric data and the main objective is to obtain dependencies between values or attributes. Generally the results are presented in the form of rules $X \rightarrow Y$ [17];

• Genetic algorithms works also with all kind of data. Is a technique generally used in optimization problems which is based on principles of natural laws of evolution proposed by Charles Darwin in 1859 [18];

• Clustering is a technique whose purpose is to detect the existence of different groups within all kind of dataset, and in case of existence, determine or characterize them. Cluster analysis is the organization of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based or similarity [19];

• Visualization techniques will be used in order to represent the data mining results through graphs or diagrams in way to provide a better representation of complex information.

Table 1 illustrates the available Data Mining techniques and the corresponding algorithms to be potentially applied.

TABLE I
DATA MINING TECHNIQUES AND ALGORITHMS

	Data Mining Technique		
	Classification	Association Rules	Clustering
Data Mining algorithm	C 5.0 Neuronal Network	Apriori	K-means

For classification, it is possible to apply a C5.0 decision tree algorithm based on genetic algorithms and Neuronal Net algorithm. C5.0 is a decision tree algorithm developed based on C4.5 by Quinlan [20]. This algorithm can produces a decision tree and induction rules for almost any kind of input. It includes all functionalities of C4.5 and apply a bunch of new technologies, among them the most important application is “boosting” technology for improving the accuracy rate of identification on samples.

Neural networks can be used to classify data. With this type of algorithm is possible to label each data point as belonging to one of *n* classes.

The Apriori algorithm is capable of mining for associations among items in a large database.

For Clustering, K-mean is one of the simplest unsupervised learning algorithms that solve clustering problems. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume *k* clusters) fixed a priori.

ISBN: 978-988-19252-2-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

III. EXPERIMENTAL RESULTS

Currently we have one project partner, an important international organization in the industry of electronic components for the car industry. The data collecting started on December of 2011 into a local database composed by the following fields:

i. Breakdown ID;
ii. Type of Breakdown;
iii. Breakdown Local;
iv. Breakdown Reason;
v. Date;
vi. Machine ID.

This data is related to maintenance activities (corrective and preventive) performed by maintenance teams and collected by intelligent agents placed in different machines.

An example of the output given by the prediction system is shown in the next figure (Fig. 5). This output consists in an example of Neuronal Network application (with several layers: back propagation algorithm), setting as studied target Breakdown reason field and as input all the other fields. This example reaches these results with an accuracy level of 72%. In general, when the Data Mining tasks involves prediction, it is important that the discovered knowledge holds a high predictive accuracy, even though in many cases the comprehensibility and interestingness of the discovered knowledge tends to be more important.

Figure 5 shows one of the results given by the algorithm, using visualization techniques given by the applied data mining tool.

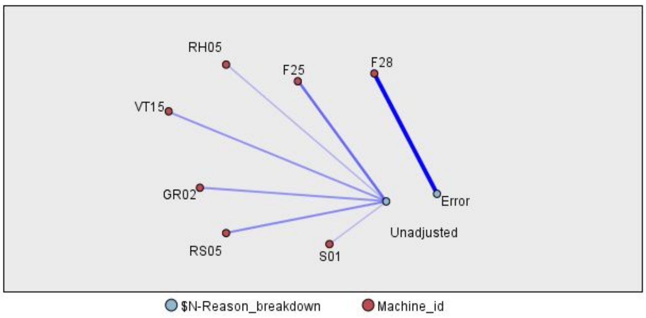

Fig. 5. Neuronal net results example.

With this kind of study we can identify strong relations between some machines and some breakdown reasons, so it is possible to predict a new intervention, of the same type, in same equipment. At this stage we can only predict the failure occurrence, it is not possible to predict the time window where it may happen. With the future addition of monitoring data into the system, the time of the malfunction episode and the evolution of monitoring parameters will be related. Thus, it will be possible to predict the failure occurrence within a proper time window.

Through the analysis of Figure 5, it is possible to predict, with accuracy, a future error breakdown occurrence in F28 machine. Concerned machines F25, RH05, VT15, GR02, RS05 and S01 it is possible to predict a future unadjusted breakdown occurrence.

In a next stage, it is intended to increase the database with some new fields, such as: i) equipment downtime; ii) products loss during breakdowns, and monitoring data about equipment (temperature, pressure, etc).

Having all of this data, the system will return more results and with more accuracy. It is intended that the system gives as output, alerts with a suitable time window, based in all the rules and models generated by the predictive prototype. With the information obtained through data mining, maintenance teams will be able to perform interventions on equipment before breakdowns.

With this kind of results and the future increase of database it will possible to built an effective system because it will not only detect potential system risk prior to failure and loss of that operating system, but will also have the ability to find adequate system inspection times, thus reducing maintenance effort and cost.

IV. CONCLUSIONS

This paper presents a new conceptual framework for collaborative prediction system to be used by industrial maintenance teams. Organization managers and maintenance leaders are more concerned with highlighting existing or potential maintenance problems in order to be able to improve performance and minimize the maintenance operational cost.

This system is characterized by autonomy in data collecting and the utilization 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” [21].

Applying data mining techniques on the available industrial maintenance data may help to discover useful rules that allow locating some critical issues that will have substantial impact on improving all maintenance processes. Before using rules to change operations, it is important to examine the rules. Unexpected rules that do not make sense may also signal other, more nefarious. The proposed system will help companies to collect, extract and create knowledge in a way that will allow companies to predict with more accuracy the moment to realize maintenance actions and thus to improve the productivity of manufacturing process.

At this stage different algorithms are being tested over the existing database in order to verify which ones produce the best results and which outputs is more comprehensible in order to generate the best system output.

The innovative point of this system is the capability of collecting and treats data dispersed in different facilities originated by maintenance interventions in different environments. The existent huge amount of data from maintenance actions are not fully used to increase the efficiency of maintenance prediction. Data mining seems to be the step forward that may change the actual state.

As future work, it is necessary to increase the database with all possible monitoring data in order to better understand the reason of all malfunction occurrences, so we can predict new failures based on studies of a larger number of parameters, resulting in more valuable and accurate outputs.

ISBN: 978-988-19252-2-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

REFERENCES

[1] Sluga A., Butala P., and Peklenik J., A Conceptual Framework for Collaborative Design and Operations of Manufacturing Work Systems. CIRP Annals Manufacturing Technology, 2005. **54**(1): p. 437-440.
[2] Wiendahl H.P. and Lutz S., Production in networks. in *CIRP*. 2002.
[3] Persona A., Regattieri A., Pham H., and Battini D., Remote control and maintenance outsourcing networks and its applications in supply chain management. Journal of Operations Management, 2007. **25**: p. 1275–1291.
[4] Palmer R., Maintenance planning and scheduling handbook. 1999, McGraw Hill: New Jersey.
[5] Chu C., Proth J.M., and Wolff P., Predictive maintenance:The one-unit replacement model. Int. J. Production Economics, 1998: p. 285-295.
[6] Pinjala S.K., Pintelona L., and Vereecke A., An empirical investigation on the relationship between business and maintenance strategies. Int. J. Production Economics, 2006: p. 214-229.
[7] Swanson L., Linking maintenance strategies to performance. Int. J. Production Economics, 2001: p. 237-244.
[8] Yao X., Fernandez-Gaucherand E., Fu M.C., and Marcus S.I., Optimal preventive maintenance scheduling in semiconductor manufacturing IEEE Transactions on Semiconductor Manufacturing, 2004. **17**(3): p. 345 - 356
[9] Bansal D., Evans D.J., and Jones B., BJEST: A reverse algorithm for the real time predictive maintenance system. International Journal of Machine tools and manufacture, 2006. **46**: p. 1068-1078.
[10] Goebel M. and Gruenwald L., A survey of Data Mining and Knowledge Discovery Software Tools, in *SIGKDD*. 1999. p. 20-33.
[11] Michalsky R.S., Bratko I., and Miroslav K., Machine Learning and Data Mining Methods and Aplications. 1998, England: John Wiley & Sons, Inc.
[12] Apte C., Liu B., Pednault E.P.D., and Smyth P., Business Applications of Data Mining. ACM, 2002. **45**(8): p. 49-53.
[13] Symeonidis A.L., Chatzidimitriou K.C., Athanasiadis I.N., and Mitkas P.A., Data mining for agent reasoning: A synergy for training intelligent agents. Engineering Applications of Artificial Intelligence, 2007. **20**: p. 1097–1111.
[14] Ferber J., Multi-agent systems: An introduction to distributed artificial intelligence. 1999, London: Addison-Wesley.
[15] Koulinitch A.S. and Sheremetov L.B., Coordination and communication issues in multi-agent expert system: concurrent configuration design advisor. Expert Systems with Applications, 1998. **15**: p. 295–307.
[16] Berson A., Smith S., and Thearling K., Building Data Mining Aplications for CRM. 2000, USA: McGraw-Hill.
[17] Toloo M., Sohrabi B., and Nalchigar S., A new method for ranking discovered rules from data mining by DEA. Expert Systems with Applications, 2009. **36**: p. 8503 - 8508.
[18] Beasley D., Bull D.R., and Martin R.R., An Overview of Genetic Algorithms, Part1, Fundamentals. University Computing, 1993. **15**(2): p. 58-69.
[19] Jain A.K., Murty M.N., and Flynn P.J., Data Clustering: A Review. ACM Computing Surveys, 1999. **31**(3): p. 264-323.
[20] Pang S. and Gong J., C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks. Systems Engineering — Theory & Practice, 2009. **29**(12): p. 94-104.
[21] Fayyad U. and Piatetsky-Shapiro G., Knowledge Discovery and Data Mining: Towards a Unifying Framework. in *Second International Conference on Knowledge Discovery and Data Mining*. 1996. Oregon.

ISBN: 978-988-19252-2-0
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)