Defining a data-driven maintenance policy: an application to an oil refinery plant

Sara Antomarioni and Maurizio Bevilacqua
Dipartimento di Ingegneria Industriale e Scienze Matematiche, Università Politecnica delle Marche, Ancona, Italy, and
Domenico Potena and Claudia Diamantini
Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, Ancona, Italy

Received 23 January 2018
Revised 27 February 2018
Accepted 6 March 2018

Abstract

Purpose – The purpose of this paper is developing a data-driven maintenance policy through the analysis of vast amount of data and its application to an oil refinery plant. The maintenance policy, analyzing data regarding sub-plant stoppages and components breakdowns within a defined time interval, supports the decision maker in determining whether it is better to perform predictive maintenance or corrective interventions on the basis of probability measurements.

Design/methodology/approach – The formalism applied to pursue this aim is association rules mining since it allows to discover the existence of relationships between sub-plant stoppages and components breakdowns.

Findings – The application of the maintenance policy to a three-year case highlighted that the extracted rules depend on both the kind of stoppage and the timeframe considered, hence different maintenance strategies are suggested.

Originality/value – This paper demonstrates that data mining (DM) tools, like association rules (AR), can provide a valuable support to maintenance processes. In particular, the described policy can be generalized and applied both to other refineries and to other continuous production systems.

Keywords Big data analytics, Predictive maintenance

Paper type Case study

1. Introduction

Competition on a global scale, fast-changing customer needs, shorter product life cycles require a high level of efficiency in all industrial environment (Gröger *et al.*, 2012). Indeed, maintenance represents a key factor for production reliability (Groba *et al.*, 2007): several different strategies are still applied, but a growing focus is put on predictive maintenance. Predictive maintenance can be considered as the attitude to "use the actual operating conditions of plant equipment and systems to optimize total plant operations" (Mobley, 2002). Future failures can be predicted through statistical data referring to the equipment, but a more accurate prevision is provided through the analysis of data related to the current production process (Groba *et al.*, 2007). Noteworthy, the monitoring of a production process is connected to several variables which generate a large amount of data. In this scenario, knowledge discovery in databases (KDD) techniques help in automatically analyzing data with the aim of bringing out valid, novel, potentially useful and ultimately understandable patterns (Fayyad *et al.*, 1996).

Defining the best maintenance policy represents a critical issue for all kind of production plants. Process industry is particularly affected by this aspect, as all the activities are sequentially connected. Many variables like temperature, flow rates, level and chemical characteristics of raw materials have to be measured (Seng and Srinivasan, 2009) and monitored. Therefore, a theoretical procedure will be developed with the aim of providing a research approach widely applicable. In order to validate the application of the generalizable

International Journal of Quality & Reliability Management
Vol. 36 No. 1, 2019
pp. 77-97

0265-671X
DOI 10.1108/IJQRM-01-2018-0012

research approach, an Italian oil refinery will be considered as an example of process industry: indeed, a high number of failures may occur, because of the normal wear of the components (Bevilacqua and Ciarapica, 2018; Fabiano and Currò, 2012) or due to unexpected event.

The refinery considered for the case study is a medium-sized plant, located on the eastern coast of central Italy. All information concerning components breakdowns and sub-plants stoppages were collected to thoroughly understand the entity of the problem at hand and to provide an effective maintenance roadmap.

In this paper, KDD is applied to derive AR describing components breakdowns after a stoppage of an oil refinery, where a stoppage of a plant is defined as an interruption of the mass flow.

Considering the stoppages of an oil refinery, the purpose of this work is to answer the following research questions:

RQ1. What are the components likely to break within a given time interval after a plant stoppage?

RQ2. Considering the probability level of breakdowns, is a predictive maintenance intervention preferable to a corrective one?

In particular, this work provides a general procedure supporting the user in deciding whether to apply a predictive maintenance strategy or, alternatively, acknowledge the risk of an actual breakdown and correct it. Moreover, an adaptation of the methodology to an Italian oil refinery is presented highlighting some interesting results.

The remaining of the paper is organized as follows: Section 2 is dedicated to a review of existing literature on data mining (DM) applications to maintenance and reliability, while Section 3 contains a depiction of the methodology developed in this study. Section 4 is dedicated to the presentation of the case study carried out to validate the research approach, while in Section 5 results are presented and discussed. Section 6 contains the conclusive remarks.

2. Literature review

2.1 DM applications for maintenance and reliability

KDD is an interdisciplinary field aiming at extracting important information and knowledge from a large amount of data (Diamantini *et al.*, 2013). DM represents a fundamental activity belonging to KDD field (Fayyad *et al.*, 1996). Specifically, DM involves the search for hidden information, patterns and tendencies in a large amount of data (Shahbaz *et al.*, 2006). Therefore, DM acts as a facilitator in discovering hidden information and knowledge, as well as patterns, trends and relationships from the large amount of data currently available (Seng and Srinivasan, 2009).

Data come from almost all processes of the organization, from the design to the production scheduling, control and maintenance (Harding *et al.*, 2005). According to Braha (2001), DM should be integrated into organizational processes, taking into consideration its objectives and potentialities together with goals and weaknesses of the manufacturing environments. For instance, the design process involves the setting and selection of parameters, actions and components (Kusiak, 1999) in order to make previsions, such as cost estimating: prior data are often applied to the cost estimation problem during the designing phases and several algorithms have been developed to this end (Kusiak *et al.*, 2000). Even quality analysis, such as the search for causes for deteriorating product quality, can be performed through DM techniques, as presented by Skormin *et al.* (2002) and Oh *et al.* (2001). Furthermore, the application of DM techniques can provide a valid support to process analysis: Maki and Teranishi (2001) developed an automated DM system to detect anomalies in a manufacturing process, in order to induce engineers to pinpoint and easily prevent their causes. Indeed, problematic issues can be found and resolved through DM technology as presented by Gardner and Bieker (2000) in an application to the semiconductor industry. Moreover, the application developed by Batanov *et al.* (1993) communicated to the user maintenance policy suggestions, together with machines diagnosis and maintenance scheduling for the analyzed devices. Romanowski and Nagi (2001), instead, managed to identify subsystems responsible for low equipment availability and, on the basis of these results, provided a preventive maintenance schedule. Bumblauskas *et al.* (2017) defined a smart maintenance decision support system integrating optimization algorithms and analytic decision models, in order to provide useful suggestions on maintenance execution, while Manco *et al.* (2017) performed an outlier-based fault prediction through the study of non-normal signals provided by sensors and validated their approach through an experimental case study. The review of the principal condition monitoring techniques applied to fault detection of offshore wind turbine presented by Kabir *et al.* (2015) is a further example of DM applications to maintenance policy.

Antony and Nasira (2014), instead, proposed a cluster analysis to perform a predictive analysis on board of train vehicles. In addition, Sammouri *et al.* (2014) presented a methodology aiming at the prediction of rare failures mining temporal data provided by sensors installed on commercial trains. Even Jin *et al.* (2015) dedicated their work to the railway field, developing a procedure for the predictive maintenance of railway point machines. Ming (2015) created a maintenance management system for urban transit rails: indeed, through the application of artificial neural networks and decision trees he was able to supervise the equipment and to mine valuable information.

In a paper by Sipos *et al.* (2014), a data-driven framework based on multiple-instance learning is applied for the prediction of equipment faults: in particular, through the mining of equipment event logs, they extracted the operational information useful for the predictive activity. Wang (2014) implemented a fault diagnostic and prognostic system, collecting data through DM techniques and exploiting artificial neural networks to train and validate the model.

Furthermore, Gröger *et al.* (2012) presented an innovative DM-driven methodology aiming at the optimization of the whole manufacturing process, describing both conceptual and practical cases, while Bevilacqua *et al.* (2017) developed an analytic model to carry out an IoT-based energy management, aiming at empowering the decision-making process through the integration of data provided by different smart devices.

Despite the wide application of DM techniques to manufacturing processes, this theme is not widely developed with reference to oil and gas refinery plants, even if it would represent a successful discriminant in this field (Saybani and Wah, 2010). Indeed, Köksal *et al.* (2011) found that only a 4 percent of the DM application to manufacturing industry regards coke or petroleum refineries. DM in refinery could be applied to analyze the influence of some variables on products quality, create rules to manage the manufacturing process, to predict price changes or requirement of different kinds of oil (Zhong and Wang, 2003): indeed, Zhong and Wang (2003) proposed a theoretical combination of computer integrated management systems and DM tools. Wang and Gao (2012), instead, developed an indicator to support maintenance decision-making process, exploiting the Internet of Things and applied it to an oil transfer station.

Friedemann *et al.* (2008) described some applications of prognostic and diagnostic systems in Energy and Rail fields, highlighting the need for adapting the operating plans to the information extracted. Moreover, they applied a similar management system to condition monitoring of subsea facilities in oil production processes, with the aim of improving availability and reliability of the infrastructure (Friedemann *et al.*, 2008). In addition, Li *et al.* (2006)

proposed a prototype system based on a rough set in order to deal with incomplete data in fault diagnosis and applied it to centrifugal pumps of a refinery plant.

In order to predict the deposition of scales in oil wells and to avoid the unavailability of the equipment, it was found that training a support vector machine can represent a valuable solution: indeed, an appropriate maintenance strategy can be defined according to the scaling values foreseen by the model (Moura *et al.*, 2012).

Hu, Zhang, Wang and Liang (2009) proposed a methodology for oil pumps fault detection, integrating multifractal theory and Mahalanobis–Taguchi system: in particular, they aimed at predicting failures through vibration signals monitoring. Moreover, they highlighted the necessity of storing real-time signals in an integrated diagnosis database, in order to enable the application of advanced faults detection techniques (Hu, Zhang, Liang and Wang, 2009).

Other methodologies have also been applied in order to provide useful maintenance procedures: for instance, Bertolini *et al.* (2009) developed a risk-based inspection and maintenance procedure that, through the integration of both risk analysis and reliability management methods, lead to an enhancement of the quality of maintenance performances. In addition, Bevilacqua and Braglia (2000) defined the best maintenance strategy for an Italian refinery using an approach based on analytic hierarchy process, while Bertolini and Bevilacqua (2006) developed a lexicographic goal programming method in order to settle the most appropriate maintenance procedure for centrifugal pumps in an oil refinery. A predictive control through simulation and 3D modeling was proposed by Witte and Ribeiro (2012) to prevent damages due to corrosion of structures exposed to critical weather conditions, such as refineries, port structures or offshore platforms. Moreover, Hanif and Agha (2012) studied the benefits deriving from the integration between predictive maintenance and statistical process control techniques, developing a case study in a Pakistani oil refinery. Even fuzzy logic has been applied to deploy a risk-based inspection planning for oil and gas pipelines (Singh and Markeset, 2009).

Interesting and non-trivial relations can be discovered implementing AR mining. For instance, AR has been applied to improve the design process and to define the most appropriate geometric dimensions (Shahbaz *et al.*, 2006). In addition, Chen (2003) applied AR to solve the problem of cell-formation, according to group technology requirements, evidencing the ability of this method in determining quality solutions. In particular, this method was applied to the binary version of the cell-formation problems and it resulted to be a satisfying procedure both for large and small-scale problems.

The application of AR mining in manufacturing field can be related to the purpose of enhancing overall performances. In order to pursue this aim, AR can be applied to frequent patterns extracted from industrial processes, as they represent a useful methodology in disclosing industrial failures (Kamsu-Foguem *et al.*, 2013; Martínez-de-Pisón *et al.*, 2012). Djatna and Alitu (2015) capitalized on the introduction of AR for the development of a total productive maintenance strategy, obtaining an increase in the effectiveness of maintenance response and efficiency considering time and costs. Bastos *et al.* (2012) studied a decentralized predictive maintenance system aiming at forecasting the possibility of a breakdown to increase the reliability of the system.

Furthermore, a manufacturing defect detective model was proposed by Chen *et al.* (2004): AR mining was applied in order to analyze the existing correlations between the combination of machines and defects and, integrating this procedure with a root cause machine identifier, the root cause could be detected. Wang *et al.* (2005) studied the generation of associative rules for manufacturing process planning to improve the performances previously obtained through a fuzzy decision technique and entropy-based analysis method: indeed, they combined the variable precision rough set and fuzzy clustering.

Another relevant application of rule mining algorithm to manufacturing processes is the one proposed by Agard and Kusiak (2004): indeed, their algorithm aimed at selecting subassemblies through the analysis of orders previously received from the customers. This application could ensure an improvement in performance in terms of delivery times to contractor. Moreover, AR can be applied to fault detection in assembly operations (Cunha *et al.*, 2006), achieving improvements in terms of quality of the assembly process thanks to the monitoring or to the avoidance of critical sequences (Choudhary *et al.*, 2009).

An applicative example to a drill production process was also found in literature: results showed that the approach based on AR mining was useful in providing important information on faults and related causes (Kamsu-Foguem *et al.*, 2013).

The development of an AR-based conceptual model was also found to be valuable in detecting the impact of human practices on risky situations in a refinery plant (Bevilacqua and Ciarapica, 2018).

A further application of AR in the refinery field has been analyzed by Li *et al.* (2009): indeed, they showed that basing AR on fuzzy systemic clustering could allow extracting useful advice on the production process, even if some of the rules extracted provided unnecessary or superfluous information.

3. Research approach

The current research approach aims at analyzing the refinery stoppages impact on components breakdowns: indeed, interpreting data related to previous stoppages and defining a timeframe of interest will allow decision makers to identify the need for predictive maintenance interventions on the components showing a high probability of breaking.

For the aims of this work, an oil refinery plant is considered as formed by a set of sub-plant, which, in turn, have several components on which a maintenance policy has to be defined. Furthermore, the stoppage of a sub-plant is defined as an interruption of the mass flow across the corresponding sub-plant.

AR mining represents a valuable methodology, as it aims at identifying relationships between knowledge contained in an information data set, revealing interesting attribute-value conditions frequently occurring together (Buddhakulsomsiri *et al.*, 2006).

Since AR mining results to be both easy to understand and to implement, it seems appropriate for being incorporated into the decision-making process (Chen *et al.*, 2005). Moreover, as observed by Chen (2003), AR can be applied to a lot of situations in manufacturing, such as resource consumption trends, malfunctioning recognition, performance analysis of supplier or of product mix. The application of these algorithms can lead to the improvement of the production processes, as stated by Martínez-de-Pisón *et al.* (2012).

Let $I = \{i_1, i_2, \ldots, i_n\}$ be the set of Boolean data called items and $D = \{t_1, t_2, \ldots, t_D\}$ be the set of all transactions, where each transaction t_i contains a subset of items (i.e. an itemset) chosen from I. In the current work, the item i_1 represents the breakdown of the component C_1, and a transaction is the set of components that break in a given timeframe after a stoppage.

An AR is an implication of the form $A \rightarrow B$, where A and B are itemsets such that $A, B \subseteq I$ and $A \cap B = \varnothing$. The itemset A is called body of the rule, while B is named head of the rule.

The most used metrics to evaluate the quality of a rule are support and confidence, as shown in the following equations, respectively:

$$Support = \frac{\#\{A \cup B\}}{\#\{D\}}, \tag{1}$$

$$Confidence = \frac{Support\ \{A \cup B\}}{Support\ \{A\}}, \quad (2)$$

where the function $\#\{x\}$ returns the number of transactions in D containing the item x: $\#\{D\}$ is the cardinalities of the database.

Confidence is the conditional probability of finding B under the hypothesis that the transaction contains A and can be considered as an indicator of the strength of the rule, of its certainty (Shahbaz *et al.*, 2006). On support provides an indication of the probability to find the transaction containing both A and B in the database D. Hence, support measures the statistical significance of the rule (Agrawal *et al.*, 1993).

In order to discover useful AR, first FP-growth algorithm is used to extract frequent itemsets (Tan *et al.*, 2006), which can be defined as sets of items that appear in a data set more frequently than a threshold imposed by the user (Han *et al.*, 2007). Then, rules are generated by frequent itemsets on the basis of a given criterion (i.e. minimum support or minimum confidence).

Next subsection is devoted to introducing a procedure that uses AR to define the best maintenance policy in a given timeframe is presented.

3.1 Defining the maintenance policy

The procedure to decide the most appropriate maintenance policy can be summarized as follows:

(1) Define the following parameters:
   - T: timeframe, is the time interval, starting from the sub-plant stoppage, during which the analysis is performed;
   - σ_{rule}: the minimum support requested to a rule to be considered;
   - σ_{rep}: the minimum value of the support for the execution of predictive maintenance on all the items composing the rule; and
   - σ_{conf}: the minimum value of the confidence for the execution of predictive maintenance on all the items composing the rule.

(2) Mine AR for the timeframe T having a support greater than or equal to σ_{rule}: For each component c_j of the sub-plant j, an item i_j in I is defined such that it assumes a true value if the component is broken because of a stoppage, false otherwise. A transaction t in D represents the set of components broken for the given time interval T and sub-plant j. FP-growth algorithm is applied over D to obtain the frequent item-set FI. From FI, AR in the form $r: A \rightarrow B$ are extracted, having support $(r) \geqslant \sigma_{rule}$.

(3) For each rule $r \in R$, if$support(r) \geqslant \sigma_{rep}$, then a predictive maintenance intervention is performed for all components included in (both head and body of) the rule r.

(4) Monitor and control the sub-plant for the whole timeframe T:
   - If the component A breaks, let R_A be the set of rules having A as body.
   - For each rule $r \in R_A$, If $confidence(r) \geqslant \sigma_{conf}$, then a predictive maintenance intervention is performed on all components in the head of r as well as a corrective intervention on A.

The definition of parameters (i.e. step 1) is of particular importance, as the analysis can be limited in two ways: first, according to the temporal dimension, in order to relate maintenance interventions only to stoppages in an interval relevant to maintenance policy purposes. The definition of σ_{rule} and σ_{rep} allows to limit the analysis only to critical components. Indeed, in this way, the predictive intervention is performed only for the components that present a statistically significant probability of breakdown. Since an oil refinery is composed of several components, σ_{rule} and σ_{rep} should be chosen to be enough low so that significant rules are not excluded, allowing at same time to exclude rarely occurring rules.

The monitoring activity recommended in the last step, instead, assures a control of the components not maintained prior to their breakdown. Since the confidence of the rule r: $A \rightarrow B$ represents the probability that the component B breaks if a work order for the component A is emitted, the choice of σ_{conf} affects the decision of predictively maintain or not the component B given the breakdown of A. The definition of the parameter σ_{conf} should take into consideration a two-fold aspect. Indeed, if it assumes a value too high, no predictive maintenance will be suggested.

On the other hand, if it is set too low, the maintenance policy will suggest performing predictive interventions even on components having a low breakdown probability. To derive this kind of trade-off a great experience, as well as the complete knowledge of the process, is necessary.

4. Case study

4.1 Oil refinery description

The refinery considered for the case study is located on the eastern coast of central Italy and it is characterized by a processing capacity of 3,900,000 ton/year of crude oil, that is about 85,000 barrel/day. The storage capacity of the refinery is more than 1,500,000 m^3 and the land-based shipping system is characterized by a total capacity of more than 12,000 tons per day. A fixed sea platform is located 16 km from the coast and it is able to accommodate tankers up to a tonnage of 400,000 tons. Furthermore, an island with a double mooring for ships up to 90,000 tons is located 4 km from the coast, while a pier for short-sea shipping is directly connected to the refinery and it is equipped with three mooring points. In Figure 1

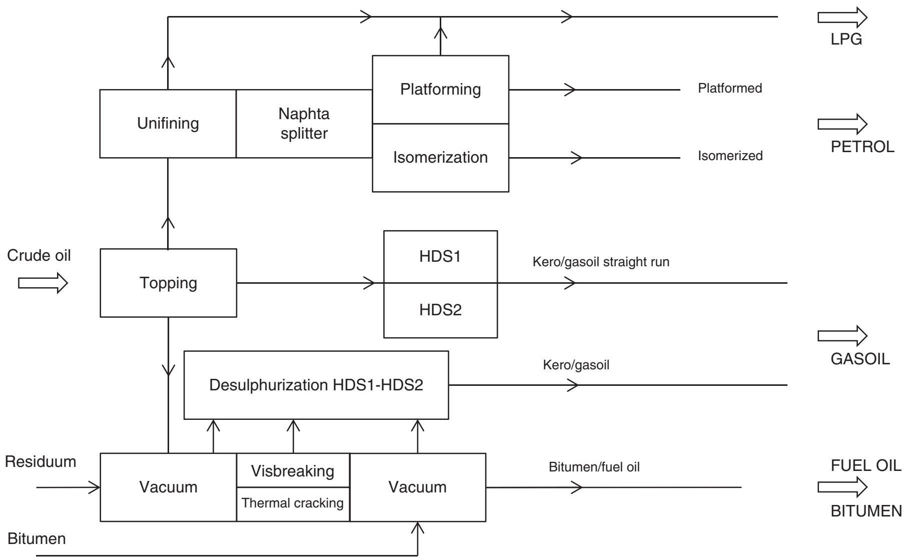

Figure 1. Representation of the oil refinery plant

the production process executed in the refinery is schematized: noteworthy, the plant is divided into 11 sub-plants dedicated to specific processes.

The refinery receives about 340,000 tons of crude oil per month and it goes in input in the topping sub-plant, that is the plant responsible for the primary distillation process. In general, distillates are classified into three categories, depending on their boiling point. As shown in Figure 1, three main processes branch off topping sub-plant. Light fractions, mainly characterized by petrol and liquefied petroleum gas (LPG), are processed through a unifining sub-plant; at this point, LPG can be extracted from the production process, while petrol is dispatched to isomerization or platforming processes. Medium distillate, instead, pass through hydro-desulfurization (HDS) in HDS1 or HDS2 plants. Heavy fractions and distillation residuum are subjected to thermal cracking and visbreaking.

4.2 Data and preliminary analysis

Data employed in this study belong to two different databases, containing information about production and maintenance, respectively.

From the former database, a view containing detailed information about the amount of product entered in the process cycle was extracted. In particular, the mass flow is provided considering hourly ranges for each of the sub-plants. The analyzed timeframe covers three years, from January 2001 to December 2003. Each instance of the view contains daily measurements. An excerpt of the view is reported in Table I, whose columns respectively contain information about the sub-plant, the code of the sensor used to evaluate the measurement (tag), the date of the measurement, the mean values of the mass flow per hour (v01, v02, …, v24), and their sum that represents the daily mass flow.

The considered database has some missing values in columns containing the mean values of the hourly mass flow, due to: sensors measurement errors or sub-plant stoppages. In order to distinguish between the two cases (a and b), the timeframe where the data were missing was compared with the table recording the stops of the sub-plants. If a stoppage was detected, then the missing value was replaced with zero (case b). Otherwise (case a), the mass flow of the previous timeframe has been used to replace the missing value.

Stoppages are grouped into three categories: shut down, if the daily mass flow is null; slow down, if the daily mass flow is lower than the 25 percent of the mean daily mass flow of the same year, but not null; non-significant (NS) stoppage, if the daily mass flow is greater than the 25 percent of the mean daily mass flow of the same year. On the basis of this classification, a further column reporting the kind of stoppage (see Table I) was added to the original view.

The maintenance database contains data related to work orders due to malfunctioning components[1] or breakdowns. A view with information about the malfunctioning

(a) Excerpt of the view extracted from production database

Sub-plant	Tag	Date	v01	v02	v03	…	v24	Daily mass flow
Vacuum	FC1401	April 30, 2002	0	0	0	…	0	0
Vacuum	FC1401	April 31, 2002	0	0	0	…	90.193	90.193
Vacuum	FC1401	June 1, 2002	90.193	18.039	0	…	79.740	187.972
Vacuum	FC1401	June 2, 2002	79.693	79.693	81.560	…	111.904	352.85

(b) Stoppages classification

stoppage
SHUT_DOWN
SLOW_DOWN
NS
–

Table I.
Excerpt of the view extracted from the production database (a) and further column indicating the stoppage category (b)

component, the identification code of the injured section of the component, the sub-plant where the component is located, and the date of the work order was extracted. An excerpt of the view is reported in Table II.

First, the two views have been integrated: when a stoppage of one of the sub-plants was detected, all work orders emitted in the following six months were considered. For example, looking at Table I, it can be noticed that a stoppage was detected in vacuum sub-plant on June 2, 2002; hence, all the work orders emitted for malfunctioning components – belonging to any of the sub-plants – from June 2, 2002 to December 2, 2002 were taken into consideration. Table III reports an excerpt of the output data.

As a preliminary analysis, the possible correlation between a stoppage in a sub-plant and work orders in the whole plant within different time slices has been evaluated. The different intervals considered were a week, a month, two months and six months. This analysis showed no significant correlations.

Hence, the analysis has been refined by considering only work orders emitted for the same sub-plant where a stoppage has been detected. In Table IV the output of the integration of the two views is reported.

5. Results and discussion

The current work aims at explaining the relationships between the work orders emitted for some components after a sub-plant stoppage and within a defined time interval. In order to qualitatively evaluate the effectiveness of the proposed approach, three academic experts and six members of the maintenance department of the refinery were interviewed.

Component	Item	Sub-plant	Date
filtro	P3304B	Desulphurization	January 1, 2001
tenuta	P2613B	Platforming	January 2, 2001
pilota	F3101	Desulphurization	January 2, 2001
filtro	P1846a	Visbreaking	January 3, 2001

Table II. Excerpt of the view extracted from maintenance database

Stopped sub-plant	Stoppage date	Kind of stoppage	Work-order sub-plant	Work-order date	Component	Item
Vacuum	June 1, 2002	NS	Vacuum	June 3, 2002	*valvola*	PSV1421B
Vacuum	June 1, 2002	NS	Desulphurization	June 3, 2002	*allarme*	ZL3328
Vacuum	June 1, 2002	NS	Unifining	June 3, 2002	*allarme*	PHH25165
Vacuum	June 1, 2002	NS	–	–	–	–
Vacuum	June 1, 2002	NS	Naphta-splitter	November 29, 2002	*controllore*	FC21005
Vacuum	June 1, 2002	NS	Desulphurization	November 30, 2002	–	U3200

Table III. Excerpt of the integration of the views presented in Tables I and II

Sub-plant	Stoppage date	Stoppage	Work-order date	Component	Item
Topping	January 7, 2001	NS	January 10, 2001	*cuscinetto*	P1004A
Topping	January 7, 2001	NS	January 10, 2001	*pilota*	U1000
Topping	January 7, 2001	NS	January 10, 2001	*analizzatore*	F1001
Topping	January 7, 2001	NS	January 11, 2001	*serbatoio*	P1002B
Topping	January 7, 2001	NS	January 11, 2001	*soffiatore*	F1101

Table IV. Excerpt of the integration of the views presented in Tables I and II, joining on the sub-plant

They considered the approach to be a valuable opportunity to anticipate the occurrence of components breakdown to improve the operational performances. Moreover, the six members of maintenance department have been involved in data preprocessing and in the parameter definition, to adapt the methodology to the specific requirements of the oil refinery analyzed. Next subsections are devoted to present and discuss some quantitative results based on the case study, in particular vacuum and topping sub-plants are considered.

5.1 Data processing

The data set presented in Table IV is employed as input for the AR extraction: indeed, through the analysis of the AR, this work aims at identifying the most appropriate maintenance strategy for the components that had frequently required maintenance interventions after sub-plants stoppages.

The first parameter to be set, according to step 1 of the procedure presented in Section 3, was the timeframe: for each stopped sub-plant, members of maintenance department required to analyze four time intervals: 1 day, 1 week, 2 weeks and 1 month. In particular, "1 day" explains that the work order is emitted within 24 h after the stoppage; "1 week" means that the indicated work order occurs within a week after the stoppage; "2 weeks" implies that the work order is emitted within fourteen days after the stoppage; "1 month" tells that work order is emitted within the 31st day.

Members of maintenance department also required to separately analyze the work orders depending on the classification of the stoppage (NS, slow down and shut down) and decided to set different thresholds of support and confidence for each sub-plant (see subsection 5.3 for example).

In order to extract the AR, all the components of the sub-plant that required maintenance within the given time interval were identified. An example is reported in Table V, where the first column identifies the stopped sub-plant which is the one requiring a maintenance intervention, while the second column provides the classification of the stoppage, followed by the date when it occurred and by the time interval. The following 349 columns represent the components constituting the various sub-plants: a value of "True" is assigned if a work order is emitted for the corresponding component during the indicated time interval. Otherwise, the assigned value is "False."

The tool chosen to deploy the analysis is RapidMiner, a widely used DM platform that allows the design of analysis processes by composing predefined tools. According to the structure of the record, the process in RapidMiner was structured as reported in Figure 2. The first operator, Read Excel, has the function of reading the data set in Microsoft Excel format exemplified in Table V. The second tool, filter time frame, is applied to

Sub-plant	Stoppage	Date of the stoppage	Time interval	*Accoppiamento*	*Allarme*	…	*Valvola*
Topping	NS	March 6, 2001	1 day	False	False	…	False
Topping	NS	March 6, 2001	1 month	False	False	…	True
Topping	NS	March 6, 2001	1 week	True	False	…	False
Topping	NS	March 6, 2001	2 weeks	False	False	…	True
Topping	SHUT_DOWN	April 8, 2001	1 day	False	False	…	False
Topping	SHUT_DOWN	April 8, 2001	1 month	False	False	…	False
Topping	SHUT_DOWN	April 8, 2001	1 week	False	False	…	False
Topping	SHUT_DOWN	April 8, 2001	2 weeks	False	False	…	False
Topping	SLOW_DOWN	June 11, 2001	1 day	False	True	…	True
Topping	SLOW_DOWN	June 11, 2001	1 month	False	True	…	False
Topping	SLOW_DOWN	June 11, 2001	1 week	True	False	…	True
Topping	SLOW_DOWN	June 11, 2001	2 weeks	False	True	…	True

Table V. Excerpt from the list of the components requiring maintenance interventions within a given time interval after a stoppage select only data referring to a specific timeframe. Then, a vertical selection is performed by Exclude Attributes, with the aim of excluding attributes like Time interval, that due to the previous filtering would not provide useful information in generating the AR. Nominal to Binominal and Numerical to Binominal tools transform the data set in a Boolean format. Hence, frequent itemsets are mined through FP-growth operator and AR are extracted (Create AR).

5.2 General remarks

When analyzing refinery processes and maintenance activities, it should be considered that some maintenance interventions can be executed without interrupting the production processes, while others require a complete stoppage of the sub-plant. In the former case, the procedure presented in Section 3 can be applied in order to decide whether is preferable to perform a predictive intervention or if it is better to wait for the actual breakdown of the component. In the latter case, instead, together with the cost of the component and of the maintenance intervention, the cost related to the production cost has to be added.

In particular, given the rule A→B, four cases can occur:

(1) Both A and B can be maintained continuing the production process: depending on the parameters set in the procedure and on the ones characterizing the rule, it is both possible to perform a predictive or a corrective intervention. Indeed, these interventions do not have an impact on the production process.

(2) A can be maintained during the production process, while B requires process interruption: in this case, it might be convenient to wait for the actual breakdown of component B, before intervening on it, while A can be replaced both predictively and correctively. If the probability associated with B's breakdown in the chosen timeframe, (i.e. ($\#\{B\}/\#\{D\}$ = confidence(B)) is high, the functioning of B should be monitored by means of proper KPIs. Moreover, its replacement can be planned and, based on the probability of breakdown, a new component can be purchased or ordered.

(3) A requires an interruption of the production process, while B can be replaced while the sub-plant is operating: in this case, waiting for the breakdown that imposes the interruption of the production process is more convenient. Component A should be monitored in order to promptly detect the breakdown; in addition, if the probability of the breakdown is high, the substitutive component should be purchased. When the breakdown of component A occurs, B's maintenance can be parallelized or it can be executed after a breakdown.

(4) Both A and B require an interruption of the production process to be maintained: in this case, waiting for the breakdown of one of the two components may result conveniently. A strict monitoring of the components could make fault detection more timely. Preventively purchasing two new components might also shorten the duration of the production interruption. In these conditions, the intervention could be performed on both A and B, possibly parallelizing the interventions or sequencing them.

Due to the complexity of cost estimation related to production losses, the methodology proposed in Section 3, as well as the following analysis, will take into consideration only the components that can be maintained without altering the production process, or, if necessary, recurring to the by-pass systems that assure the ordinary running of the process.

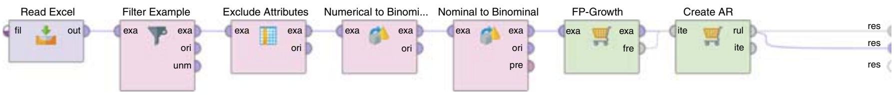

Figure 2.
Representation of the process implemented in RapidMiner

5.3 Methodology application to topping sub-plant

In this paragraph the policy explained in Section 3 will be exemplified through an application to the topping sub-plant. It is fundamental that topping sub-plant works efficiently, as it is situated at the beginning of the production process. The stoppage considered is a slow down. According to the suggestions provided by the technicians of the maintenance department, the parameters set for its analysis are the following:

(1) timeframe: 1 month;

(2) $\sigma_{rule} = 0.10$;

(3) $\sigma_{rep} = 0.50$;

(4) $\sigma_{conf} = 0.50$.

From the data set 459 frequent itemsets and 678 rules were extracted (see Appendix for an excerpt) with an execution time shorter than 1 s. In total, 120 of the extracted rules were excluded as they had a support lower than σ_{rule}. Considering the remaining 558 rules, it turned out that the support of 10 rules was higher than σ_{rep}. Hence, components of these rules (i.e. *accoppiamento, controllore, coibentazione* and *tenuta*) were immediately substituted.

For the 339 unique components contained in the remaining rules, the procedure suggests a monitoring policy.

During the monitoring phase, when a work order is emitted for a component A, maintenance attendants have to:

(1) perform a corrective intervention to replace A; and

(2) check the confidence of all rules having A as body. If the confidence is greater than the recommended threshold $\sigma_{conf} = 0.50$, then all components in the head of these rules have to be predictively replaced.

Table VI reports an example of rules having the component *indicatore* in the body. When a work order is emitted for component *indicatore*, components in the head of the 4 rules having a confidence greater than or equal to 0.50 are considered for maintenance too. In Table VI, the four rules are in bold and components to predictively repair are *presa campione, rilevatore, illuminazione* and *amperometro*.

Body	Head	Support	Confidence
indicatore	***presa campione***	**0.324324324**	**0.666666667**
indicatore	***rilevatore***	**0.324324324**	**0.666666667**
indicatore	***illuminazione***	**0.297297297**	**0.611111111**
indicatore	***amperometro***	**0.243243243**	**0.5**
indicatore	*allarme*	0.216216216	0.444444444
indicatore	*batteria*	0.189189189	0.388888889
indicatore	*tracciatura*	0.162162162	0.333333333
indicatore	*dreno*	0.162162162	0.333333333
indicatore	*trasmettitore*	0.162162162	0.333333333
indicatore	*strumentazione*	0.162162162	0.333333333
indicatore	*troppo pieno*	0.162162162	0.333333333
indicatore	*lubrificazione*	0.162162162	0.333333333
indicatore	*condensa*	0.162162162	0.333333333
indicatore	*refrigerante*	0.135135135	0.277777778
indicatore	*area*	0.108108108	0.222222222
indicatore	*baderna*	0.108108108	0.222222222

Table VI.
Excerpt of the rules extracted for topping sub-plant

If necessary, a corrective maintenance intervention will be performed on components presenting a low breakdown probability.

Maintenance department members will have to continually monitor activity until the chosen timeframe is not over and, in case of further work order, reiterate the procedure. The component could be preventively purchased to reduce the overall maintenance time evaluating component a breakdown probability, its cost and the procurement time.

5.4 Variation of the timeframe and stoppage category

In this paragraph, some rules extracted for vacuum and topping sub-plants will be analyzed at the varying of the timeframe considered. In particular, Table VII shows the variation of support and confidence, after a slow down of the vacuum sub-plant, of the rule *Tenuta* → *Valvola* when the time interval considered increases: the highest probability of occurrence of both components breakdown can be observed a week after the stoppage. Moreover, if the timeframe is enlarged to two weeks and to a month, the probability is lower and lower. On the contrary, confidence increases: if a work order for component *tenuta* is emitted in the day after the stoppage, only in 40 percent of cases this will cause a work order for component *valvola*. In all other cases, instead, a work order for *tenuta* will be followed by a work order for *valvola*.

Considering the same values of parameters σ_{rule}, σ_{conf} and σ_{conf} proposed for the topping sub-plant, the couple of components (*tenuta* and *valvola*) will not be predictively maintained, as all the support results lower than $\sigma_{rep} = 0.50$. In case of a work order for component *tenuta* after one day, no actions will be performed for *valvola*. For the other timeframes, instead, component *valvola* would be immediately replaced, as the confidence is higher than σ_{conf}.

Table VIII reports the performances of the rule *Tenuta* → *Valvola* when NS and shut-down stoppages are considered. It can be noted that, for NS stoppages, the rule is not significant; the two components never break together. Considering shut-down stoppages, support and confidence values grow when the timeframe increases: this means that the probability of occurring in a breakdown of both *tenuta* and *valvola* increases.

Tenuta → *Valvola*	Support	Confidence
1 day	0.182	0.400
1 week	0.286	1.000
2 weeks	0.125	1.000
1 month	0.118	1.000

Table VII. Support and confidence values of the rule *Tenuta* → *Valvola* for different timeframes

Tenuta → *Valvola*	Support	Confidence
1 day NS	–	–
1 day shut-down	–	–
1 week NS	–	–
1 week slow down	0.285714	1
1 week shut down	0.217391	0.625
2 weeks NS	–	–
2 weeks shut down	0.309091	0.68
1 month NS	–	–
1 month shut-down	0.375	0.91304348

Table VIII. Support and confidence values of the rule *Tenuta* → *Valvola* for different timeframes, in case of NS or shut-down stoppages

In Table IX, a comparison between the values of support and confidence for the rule *Controllore* → *Soffiatore* varying the kind of stoppage. The rule is significant for all cases related to 1 month time interval, while for a two-week timeframe it is only meaningful for slow-down stoppages. Furtherly reducing the time interval does not lead to relevant outcomes. Considering "1 month," the category of stoppage influences both support and confidence. In particular, in case of a shut-down, the rule results to be more likely than in case of slow-down or NS stoppages.

Hence, the monitoring of *controllore* and *soffiatore* components does not have to be particularly strict for "1 day" and "1 week" intervals, as well as for "2 weeks" in case of NS and shut-down stoppages. On the contrary, it should be intensified two weeks after a slow down and one month after all kind of stoppages – especially for shut-downs.

5.5 Discussion

When implementing the proposed maintenance policy, it is also important to schedule the updating of the extracted rules. Since the refinery continues processing, further stoppages may occur, and additional work orders could be emitted. A clever planning should foresee an updating interval proportional to the one chosen for the analysis (e.g. T in the current application). In this way, rules generated for the last timeframe would not be lost.

Moreover, during the update of the rules, maintenance department members can choose from two different strategies:

(1) consider a fixed interval – three years in the shown application – hence progressively exclude the oldest data; and

(2) progressively enlarge the time interval, adding new data to existing ones.

For instance, in case of modifications to refinery structure or components characteristics, the interval should be shortened. Otherwise, the second alternative should be preferred as it provides much more information and requires a limited processing time.

The definition of the parameters must be carried out by expert maintenance members since it represents the core of the maintenance policy. For instance, setting values of support (σ_{rule}, σ_{rep}) and confidence (σ_{conf}) too low would imply the execution of predictive interventions even on components with a low probability of breakdown. On the other hand, if these parameters were too high, a wide range of interventions would be performed after the occurrence of a breakdown, overcoming the aim of the predictive maintenance policy.

Noteworthy, σ_{rule}, σ_{rep}, σ_{conf}, as well as the timeframe T, can be adjusted and modified during the run-time, allowing the adaptation of the maintenance policy to refinery necessity. For example, if it is required to skimp on maintenance σ_{rule}, σ_{rep}, σ_{conf} – or at least some of them – can be scaled up, reducing the number of predictive interventions and maintaining the components only after the actual breakdown. Instead, if it is required to increase the safety of the process, support and confidence threshold could be lowered: in this case, the number of predictive actions would increase, eliminating the need for future corrective interventions.

Table IX. Support and confidence values of the rule *Controllore* → *Soffiatore* for different timeframes and different kind of stoppages

Controllore → *Soffiatore*	Support	Confidence
1 month	0.212	0.389
1 month NS	0.192	0.385
1 month slow down	0.176	0.250
1 month shut down	0.232	0.448
2 weeks	–	–
2 weeks NS	–	–
2 weeks slow down	0.125	0.333
2 weeks shut down	–	–

6. Conclusion

In this paper, a data-driven model for a maintenance policy definition is developed and applied to an oil refinery plant. Maintenance is a focal point in the production process of refinery plants since they should be run continuously. Hence, it is important to evaluate whether the occurrence of a stoppage in a sub-plant has an impact on the functioning of other components.

Basically, the maintenance policy developed in this work aims at supporting members of maintenance department in the definition of the best strategy to apply to components in case of a sub-plant stoppage. AR mining is the DM tool chosen to handle the big data extracted by the refinery databases, with the aim of detecting the relations between sub-plant stoppages and consequent components breakdowns occurring in a timeframe, defined in collaboration with members of maintenance department of the analyzed refinery.

In particular, the maintenance policy first extracts the relationships between the work order emitted for some components after a sub-plant stoppage and within a given time interval. Then, it suggests whether, according to the probability measures provided by rules support and confidence, it is more convenient to perform a predictive maintenance intervention on or to wait for the actual breakdown of the components included in rules head and body.

The analysis has been carried out for each sub-plant of the refinery and it has been detailed considering:

(1) the three categories of stoppages typically characterized in the analyzed refinery (NS stoppages, slow-down stoppages and, shut-down stoppages); and

(2) four different time intervals during the analysis ("1 month," "2 weeks," "1 week" and "1 day").

The rules have been analyzed comparing the values of support and confidence with the thresholds (σ_{rule}, σ_{rep}, σ_{conf}) set by maintenance department members. Results highlighted that different rules can be extracted for the various sub-plants. Moreover, even the kind of stoppage and the timeframes considered resulted to be influent on support and confidence values associated with each rule. Thus, depending on these parameters, predictive interventions or corrective ones were recommended by the maintenance policy.

Future development may regard the definition of specific thresholds of support and confidence for each component belonging to a given sub-plant. In addition, an economic evaluation could be added to the maintenance policy to increase its relevance.

Note

1. The list of components is reported in Italian, the original language of the database. Anyway, this naming does not have any impact on the discussion of the case study and of the maintenance policy.

References

Agard, B. and Kusiak, A. (2004), "Data mining for subassembly selection", *Journal of Manufacturing Science and Engineering*, Vol. 126 No. 3, pp. 627-631, available at: https://doi.org/10.1115/1.1763182

Agrawal, R., Imieliński, T. and Swami, A. (1993), "Mining association rules between sets of items in large databases", *ACM Sigmod Record*, ACM, Vol. 22 No. 2, pp. 207-216.

Antony, J.V. and Nasira, G. (2014), "Enabling predictive maintenance strategy in rail sector: a clustering approach", *Wseas Transactions On Computers*, Vol. 13, pp. 118-128.

Bastos, P., Lopes, I. and Pires, L.C.M. (2012), "A maintenance prediction system using data mining techniques", *World Congress on Engineering*, Vol. 3, pp. 1448-1453.

Batanov, D., Nagarur, N. and Nitikhunkasem, P. (1993), "EXPERT-MM: a knowledge-based system for maintenance management. Artificial Intelligence in Engineering", Vol. 8 No. 4, pp. 283-291, available at: https://doi.org/10.1016/0954-1810(93)90012-5

Bertolini, M. and Bevilacqua, M. (2006), "A combined goal programming – AHP approach to maintenance selection problem", *Reliability Engineering & System Safety*, Vol. 91 No. 7, pp. 839-848, available at: https://doi.org/10.1016/j.ress.2005.08.006

Bertolini, M., Bevilacqua, M., Ciarapica, F.E. and Giacchetta, G. (2009), "Development of risk-based inspection and maintenance procedures for an oil refinery", *Journal of Loss Prevention in the Process Industries*, Vol. 22 No. 2, pp. 244-253, available at: https://doi.org/10.1016/j.jlp.2009.01.003

Bevilacqua, M. and Braglia, M. (2000), "The analytic hierarchy process applied to maintenance strategy selection", *Reliability Engineering & System Safety*, Vol. 70 No. 1, pp. 71-83, available at: https://doi.org/10.1016/S0951-8320(00)00047-8

Bevilacqua, M. and Ciarapica, F.E. (2018), "Human factor risk management in the process industry: a case study", *Reliability Engineering & System Safety*, Vol. 169 No. 1, pp. 149-159, available at: https://doi.org/10.1016/j.ress.2017.08.013

Bevilacqua, M., Ciarapica, F.E., Diamantini, C. and Potena, D. (2017), "Big data analytics methodologies applied at energy management in industrial sector: a case study", *International Journal of RF Technologies*, Vol. 8 No. 3, pp. 105-122, available at: https://doi.org/10.3233/RFT-171671

Braha, D. (Ed.) (2001), *Data mining for Design and Manufacturing: Methods and Applications, Massive Computing*, Kluwer Academic, Dordrecht and Boston, MA.

Buddhakulsomsiri, J., Siradeghyan, Y., Zakarian, A. and Li, X. (2006), "Association rule-generation algorithm for mining automotive warranty data", *International Journal of Production Research*, Vol. 44 No. 14, pp. 2749-2770, available at: https://doi.org/10.1080/00207540600564633

Bumblauskas, D., Gemmill, D., Igou, A. and Anzengruber, J. (2017), "Smart maintenance decision support systems (SMDSS) based on corporate big data analytics", *Expert Systems with Applications*, Vol. 90 No. 5, pp. 303-317, available at: https://doi.org/10.1016/j.eswa.2017.08.025

Chen, M.-C. (2003), "Configuration of cellular manufacturing systems using association rule induction", *International Journal of Production Research*, Vol. 41 No. 2, pp. 381-395, available at: https://doi.org/10.1080/0020754021000024184

Chen, M.-C., Huang, C.-L., Chen, K.-Y. and Wu, H.-P. (2005), "Aggregation of orders in distribution centers using data mining", *Expert Systems with Applications*, Vol. 28 No. 3, pp. 453-460, available at: https://doi.org/10.1016/j.eswa.2004.12.006

Chen, W.-C., Tseng, S.-S. and Wang, C.-Y. (2004), "A novel manufacturing defect detection method using data mining approach", *Innovations in Applied Artificial Intelligence, Lecture Notes in Computer Science. Presented at the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer, Berlin and Heidelberg*, pp. 77-86, available at: https://doi.org/10.1007/978-3-540-24677-0_9

Choudhary, A.K., Harding, J.A. and Tiwari, M.K. (2009), "Data mining in manufacturing: a review based on the kind of knowledge", *Journal of Intelligent Manufacturing*, Vol. 20 No. 5, pp. 501-521, available at: https://doi.org/10.1007/s10845-008-0145-x

Cunha, C.D., Agard, B. and Kusiak, A. (2006), "Data mining for improvement of product quality", *International Journal of Production Research*, Vol. 44 Nos 18-19, pp. 4027-4041, available at: https://doi.org/10.1080/00207540600678904

Diamantini, C., Potena, D. and Storti, E. (2013), "A virtual mart for knowledge discovery in databases", *Information Systems Frontiers*, Vol. 15 No. 3, pp. 447-463, available at: https://doi.org/10.1007/s10796-012-9399-0

Djatna, T. and Alitu, I.M. (2015), "An application of association rule mining in total productive maintenance strategy: an analysis and modelling in wooden door manufacturing industry", *Procedia Manufacturing*, Vol. 4, pp. 336-343, available at: https://doi.org/10.1016/j.promfg.2015.11.049

Fabiano, B. and Currò, F. (2012), "From a survey on accidents in the downstream oil industry to the development of a detailed near-miss reporting system", *Process Safety and Environmental Protection*, Special Issue Celebrating Trevor Kletz's 90th Birthday, Vol. 90 No. 5, pp. 357-367, available at: https://doi.org/10.1016/j.psep.2012.06.005

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996), "From data mining to knowledge discovery in databases", *AI Magazine*, Vol. 17 No. 3, pp. 37-54.

Friedemann, J.D., Varma, A., Bonissone, P. and Iyer, N. (2008), "Subsea condition monitoring: A path to increased availability and increased recovery", presented at the Society of Petroleum Engineers – Intelligent Energy Conference and Exhibition: Intelligent Energy, pp. 298-308.

Gardner, M. and Bieker, J. (2000), *Data Mining Solves Tough Semi Conductor Problems*, KDD, Boston, pp. 376-383.

Groba, C., Cech, S., Rosenthal, F. and Gossling, A. (2007), "Architecture of a predictive maintenance framework", presented at the 6th International Conference on Computer Information Systems and Industrial Management Applications, 2007, CISIM '07, pp. 59-64, available at: https://doi.org/10.1109/CISIM.2007.14

Gröger, C., Niederman, F. and Mitschang, B. (Eds) (2012), "The 2012 international conference of manufacturing engineering and engineering management", *The 2012 International Conference of Mechanical Engineering, World Congress on Engineering, WCE, IAENG, Hong Kong, London.*

Han, J., Cheng, H., Xin, D. and Yan, X. (2007), "Frequent pattern mining: current status and future directions", *Data Mining and Knowledge Discovery*, Vol. 15 No. 1, pp. 55-86, available at: https://doi.org/10.1007/s10618-006-0059-1

Hanif, A. and Agha, M.H. (2012), "Utilizing quality tools: a predictive maintenance perspective", *International Journal of Performability Engineering*, Vol. 8 No. 6, pp. 699-704.

Harding, J.A., Shahbaz, M., Srinivas and Kusiak, A. (2005), "Data mining in manufacturing: a review", *Journal of Manufacturing Science and Engineering*, Vol. 128 No. 4, pp. 969-976, available at: https://doi.org/10.1115/1.2194554

Hu, J., Zhang, L., Liang, W. and Wang, Z. (2009), "Incipient mechanical fault detection based on multifractal and MTS methods", *Petroleum Science*, Vol. 6 No. 2, pp. 208-216, available at: https://doi.org/10.1007/s12182-009-0034-8

Hu, J., Zhang, L., Wang, Z. and Liang, W. (2009), "The application of integrated diagnosis database technology in safety management of oil pipeline and transferring pump units", *Journal of Loss Prevention in the Process Industries*, Vol. 22 No. 6, pp. 1025-1033, available at: https://doi.org/10.1016/j.jlp.2008.08.005

Jin, W., Shi, Z., Siegel, D., Dersin, P., Douziech, C., Pugnaloni, M., Cascia, P.L. and Lee, J. (2015), "Development and evaluation of health monitoring techniques for railway point machines", presented at the 2015 IEEE Conference on Prognostics and Health Management, pp. 1-11, available at: https://doi.org/10.1109/ICPHM.2015.7245016

Kabir, M.J., Oo, A.M.T. and Rabbani, M. (2015), "A brief review on offshore wind turbine fault detection and recent development in condition monitoring based maintenance system", *Australasian Universities Power Engineering Conference. Presented at the 2015 Australasian Universities Power Engineering Conference*, pp. 1-7, available at: https://doi.org/10.1109/AUPEC.2015.7324871

Kamsu-Foguem, B., Rigal, F. and Mauget, F. (2013), "Mining association rules for the quality improvement of the production process", *Expert Systems with Applications*, Vol. 40 No. 4, pp. 1034-1045, available at: https://doi.org/10.1016/j.eswa.2012.08.039

Köksal, G., Batmaz, İ. and Testik, M.C. (2011), "A review of data mining applications for quality improvement in manufacturing industry", *Expert Systems with Applications*, Vol. 38 No. 10, pp. 13448-13467, available at: https://doi.org/10.1016/j.eswa.2011.04.063

Kusiak, A. (1999), *Engineering Design: Products, Processes, and Systems*, 1st ed., Academic Press, Inc., Orlando, FL.

Kusiak, A., Kernstine, K.H., Kern, J.A., McLaughlin, K.A. and Tseng, T.L. (2000), "Data mining: medical and engineering case studies", *Industrial Engineering Research Conference*, pp. 1-7.

Li, J.R., Khoo, L.P. and Tor, S.B. (2006), "RMINE: a rough set based data mining prototype for the reasoning of incomplete data in condition-based fault diagnosis", *Journal of Intelligent Manufacturing*, Vol. 17 No. 1, pp. 163-176.

Li, R., Xu, L., Yang, X. and Yin, X. (2009), "The application of the association rules mining based on the Fuzzy systemic clustering in the hydrocracking refining device", presented at the 2009 First International Conference on Information Science and Engineering, pp. 994-997, available at: https://doi.org/10.1109/ICISE.2009.1179

Maki, H. and Teranishi, Y. (2001), "Development of automated data mining system for quality control in manufacturing", *International Conference on Data Warehousing and Knowledge Discovery*, Springer, Heidelberg, pp. 93-100.

Manco, G., Ritacco, E., Rullo, P., Gallucci, L., Astill, W., Kimber, D. and Antonelli, M. (2017), "Fault detection and explanation through big data analysis on sensor streams", *Expert Systems with Applications*, Vol. 87 No. 30, pp. 141-156, available at: https://doi.org/10.1016/j.eswa.2017.05.079

Martínez-de-Pisón, F.J., Sanz, A., Martínez-de-Pisón, E., Jiménez, E. and Conti, D. (2012), "Mining association rules from time series to explain failures in a hot-dip galvanizing steel line", *Computers & Industrial Engineering*, Vol. 63 No. 1, pp. 22-36, available at: https://doi.org/10.1016/j.cie.2012.01.013

Ming, Z. (2015), "Decision approach of maintenance for urban rail transit based on equipment supervision data mining", presented at the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 376-380, available at: https://doi.org/10.1109/IDAACS.2015.7340761

Mobley, R.K. (2002), *An Introduction to Predictive Maintenance*, Elsevier.

Moura, M.C., Lins, I.D., Ferreira, R.J. and Jacinto, C.M.C. (2012), "Predictive maintenance policy for oil well equipment in case of scaling through support vector machine", presented at the Advances in Safety, Reliability and Risk Management contains the papers presented at the 20th European Safety and Reliability (ESREL 2011), CRC Press, London, pp. 503-507.

Oh, S., Han, J. and Cho, H. (2001), "Intelligent process control system for quality improvement by data mining in the process industry", in Braha, D. (Ed.), *Data Mining for Design and Manufacturing, Massive Computing*, Springer, Boston, MA, pp. 289-309.

Romanowski, C.J. and Nagi, R. (2001), "Analyzing maintenance data using data mining methods", in Braha, D. (Ed.), *Data Mining for Design and Manufacturing, Massive Computing*, Springer, Boston, MA, pp. 235-254.

Sammouri, W., Côme, E., Oukhellou, L., Aknin, P. and Fonlladosa, C.E. (2014), "Pattern recognition approach for the prediction of infrequent target events in floating train data sequences within a predictive maintenance framework", presented at the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 918-923, available at: https://doi.org/10.1109/ITSC.2014.6957806

Saybani, M.R. and Wah, T.Y. (2010), *Data Mining and Data Gathering in a Refinery*, WSEAS Press, Wisconsin.

Seng, N.Y. and Srinivasan, R. (2009), "Data mining for the chemical process industry", *Encyclopedia of Data Warehousing and Mining*, 2nd ed., IGI Global, Springer, Berlin, pp. 458-464.

Shahbaz, M., Srinivas, M., Harding, J.A. and Turner, M. (2006), "Product design and manufacturing process improvement using association rules", *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 220 No. 2, pp. 243-254, available at: https://doi.org/10.1243/095440506X78183

Singh, M. and Markeset, T. (2009), "A methodology for risk-based inspection planning of oil and gas pipes based on fuzzy logic framework", *Engineering Failure Analysis*, Vol. 16 No. 7, pp. 2098-2113, available at: https://doi.org/10.1016/j.engfailanal.2009.02.003

Sipos, R., Fradkin, D., Moerchen, F. and Wang, Z. (2014), "Log-based predictive maintenance", *Presented at the Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1867-1876, available at: https://doi.org/10.1145/2623330.2623340

Skormin, V.A., Gorodetski, V.I. and Popyack, L.J. (2002), "Data mining technology for failure prognostic of avionics", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 38 No. 2, pp. 388-403, available at: https://doi.org/10.1109/TAES.2002.1008974

Tan, P.-N., Steinbach, M. and Kumar, V. (2006), *Introduction to data mining*, Addison-Wesley Longman Publishing, Boston, MA.

Wang, K.S. (2014), "Key techniques in intelligent predictive maintenance (IPdM) – a framework of intelligent faults diagnosis and prognosis system (IFDaPS)", *Advanced Materials Research*, Vol. 1039, pp. 490-505, available at: https://doi.org/10.4028/www.scientific.net/AMR.1039.490

Wang, Q. and Gao, J. (2012), "Research and application of risk and condition based maintenance task optimization technology in an oil transfer station", *Journal of Loss Prevention in the Process Industries*, Vol. 25 No. 6, pp. 1018-1027, available at: https://doi.org/10.1016/j.jlp.2012.06.002

Wang, Z., Shao, X., Zhang, G. and Zhu, H. (2005), "Integration of variable precision rough set and fuzzy clustering: an application to knowledge acquisition for manufacturing process planning", Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Lecture Notes in Computer Science, presented at the International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, Springer, Berlin and Heidelberg, pp. 585-593, available at: https://doi.org/10.1007/11548706_62

Witte, C.C. and Ribeiro, M. (2012), "Structural integrity management: painting predictive control", *Presented at the SPE International Conference & Workshop on Oilfield Corrosion, Society of Petroleum Engineers*, Aberdeen, May 28-29, available at: https://doi.org/10.2118/155857-MS

Zhong, N.-N. and Wang, X. (2003), "Application of data mining in refinery CIMS", presented at the Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693), Vol. 1, pp. 118-122, available at: https://doi.org/10.1109/ICMLC.2003.1264454

Appendix

Sub-plant	Time interval	Stoppage	Premises	Conclusion	Support	Confidence
Topping	1 month	SLOW_DOWN	*accoppiamento*	*controllore*	0.837838	1
Topping	1 month	SLOW_DOWN	*controllore*	*accoppiamento*	0.837838	0.885714
Topping	1 month	SLOW_DOWN	*coibentazione*	*controllore*	0.621622	0.958333
Topping	1 month	SLOW_DOWN	*tenuta*	*controllore*	0.621622	0.92
Topping	1 month	SLOW_DOWN	*controllore*	*tenuta*	0.621622	0.657143
Topping	1 month	SLOW_DOWN	*controllore*	*coibentazione*	0.621622	0.657143
Topping	1 month	SLOW_DOWN	*coibentazione*	*accoppiamento*	0.567568	0.875
Topping	1 month	SLOW_DOWN	*tenuta*	*accoppiamento*	0.567568	0.84
Topping	1 month	SLOW_DOWN	*accoppiamento*	*tenuta*	0.567568	0.677419
Topping	1 month	SLOW_DOWN	*accoppiamento*	*coibentazione*	0.567568	0.677419
Topping	1 month	SLOW_DOWN	*indicatore*	*controllore*	0.486486	1
Topping	1 month	SLOW_DOWN	*indicatore*	*coibentazione*	0.486486	1
Topping	1 month	SLOW_DOWN	*coibentazione*	*indicatore*	0.486486	0.75
Topping	1 month	SLOW_DOWN	*tracciatura*	*controllore*	0.432432	1
Topping	1 month	SLOW_DOWN	*illuminazione*	*controllore*	0.432432	1
Topping	1 month	SLOW_DOWN	*tracciatura*	*accoppiamento*	0.432432	1
Topping	1 month	SLOW_DOWN	*illuminazione*	*coibentazione*	0.432432	1
Topping	1 month	SLOW_DOWN	*indicatore*	*accoppiamento*	0.432432	0.888889
Topping	1 month	SLOW_DOWN	*coibentazione*	*illuminazione*	0.432432	0.666667
Topping	1 month	SLOW_DOWN	*presa campione*	*controllore*	0.405405	0.9375
Topping	1 month	SLOW_DOWN	*illuminazione*	*accoppiamento*	0.405405	0.9375
Topping	1 month	SLOW_DOWN	*presa campione*	*coibentazione*	0.405405	0.9375
Topping	1 month	SLOW_DOWN	*coibentazione*	*presa campione*	0.405405	0.625
Topping	1 month	SLOW_DOWN	*amperometro*	*controllore*	0.378378	1
Topping	1 month	SLOW_DOWN	*allarme*	*controllore*	0.378378	1
Topping	1 month	SLOW_DOWN	–	*controllore*	0.378378	1
Topping	1 month	SLOW_DOWN	*amperometro*	*coibentazione*	0.378378	1
Topping	1 month	SLOW_DOWN	*rilevatore*	*controllore*	0.351351	1
Topping	1 month	SLOW_DOWN	*rilevatore*	*coibentazione*	0.351351	1
Topping	1 month	SLOW_DOWN	*rilevatore*	*presa campione*	0.351351	1
Topping	1 month	SLOW_DOWN	*amperometro*	*accoppiamento*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	*amperometro*	*tenuta*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	*allarme*	*tenuta*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	*allarme*	*coibentazione*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	–	*coibentazione*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	*amperometro*	*allarme*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	*allarme*	*amperometro*	0.351351	0.928571
Topping	1 month	SLOW_DOWN	*presa campione*	*rilevatore*	0.351351	0.8125
Topping	1 month	SLOW_DOWN	*area*	*controllore*	0.324324	1
Topping	1 month	SLOW_DOWN	*dreno*	*coibentazione*	0.324324	1
Topping	1 month	SLOW_DOWN	*rilevatore*	*indicatore*	0.324324	0.923077
Topping	1 month	SLOW_DOWN	*allarme*	*accoppiamento*	0.324324	0.857143
Topping	1 month	SLOW_DOWN	*presa campione*	*accoppiamento*	0.324324	0.75
Topping	1 month	SLOW_DOWN	*presa campione*	*tenuta*	0.324324	0.75
Topping	1 month	SLOW_DOWN	*presa campione*	*indicatore*	0.324324	0.75
Topping	1 month	SLOW_DOWN	*indicatore*	*presa campione*	0.324324	0.666667
Topping	1 month	SLOW_DOWN	*indicatore*	*rilevatore*	0.324324	0.666667
Topping	1 month	SLOW_DOWN	*dreno*	*controllore*	0.297297	0.916667
Topping	1 month	SLOW_DOWN	*dreno*	–	0.297297	0.916667
Topping	1 month	SLOW_DOWN	*rilevatore*	*accoppiamento*	0.297297	0.846154

(*continued*)

Table AI.
Excerpt of the rules extracted for topping sub-plant, considering a one-month timeframe and a slow-down stoppage

Sub-plant	Time interval	Stoppage	Premises	Conclusion	Support	Confidence
Topping	1 month	SLOW_DOWN	–	*accoppiamento*	0.297297	0.785714
Topping	1 month	SLOW_DOWN	*amperometro*	*presa campione*	0.297297	0.785714
Topping	1 month	SLOW_DOWN	*allarme*	*presa campione*	0.297297	0.785714
Topping	1 month	SLOW_DOWN	–	*presa campione*	0.297297	0.785714
Topping	1 month	SLOW_DOWN	–	*dreno*	0.297297	0.785714
Topping	1 month	SLOW_DOWN	*illuminazione*	*indicatore*	0.297297	0.6875
Topping	1 month	SLOW_DOWN	*presa campione*	*amperometro*	0.297297	0.6875
Topping	1 month	SLOW_DOWN	*presa campione*	*allarme*	0.297297	0.6875
Topping	1 month	SLOW_DOWN	*presa campione*	–	0.297297	0.6875
Topping	1 month	SLOW_DOWN	*indicatore*	*illuminazione*	0.297297	0.611111
Topping	1 month	SLOW_DOWN	*ausiliare*	*controllore*	0.27027	1
Topping	1 month	SLOW_DOWN	*ausiliare*	*accoppiamento*	0.27027	1
Topping	1 month	SLOW_DOWN	ausiliare	*tracciatura*	0.27027	1
Topping	1 month	SLOW_DOWN	*ausiliare*	*livello*	0.27027	1
Topping	1 month	SLOW_DOWN	*livello*	*controllore*	0.27027	0.833333
Topping	1 month	SLOW_DOWN	*livello*	*accoppiamento*	0.27027	0.833333
Topping	1 month	SLOW_DOWN	*area*	*accoppiamento*	0.27027	0.833333
Topping	1 month	SLOW_DOWN	*livello*	*tenuta*	0.27027	0.833333
Topping	1 month	SLOW_DOWN	*livello*	*tracciatura*	0.27027	0.833333
Topping	1 month	SLOW_DOWN	*livello*	*ausiliare*	0.27027	0.833333
Topping	1 month	SLOW_DOWN	*rilevatore*	*amperometro*	0.27027	0.769231
Topping	1 month	SLOW_DOWN	–	*tenuta*	0.27027	0.714286
Topping	1 month	SLOW_DOWN	*amperometro*	–	0.27027	0.714286
Topping	1 month	SLOW_DOWN	–	*amperometro*	0.27027	0.714286
Topping	1 month	SLOW_DOWN	*amperometro*	*rilevatore*	0.27027	0.714286
Topping	1 month	SLOW_DOWN	*allarme*	–	0.27027	0.714286
Topping	1 month	SLOW_DOWN	–	*allarme*	0.27027	0.714286
Topping	1 month	SLOW_DOWN	*tracciatura*	*livello*	0.27027	0.625
Topping	1 month	SLOW_DOWN	*tracciatura*	*ausiliare*	0.27027	0.625

Table AI.

Corresponding author
Sara Antomarioni can be contacted at: s.antomarioni@pm.univpm.it