Oil spills in European and North-American pipelines: an association rule-based analysis

Antomarioni S., Bevilacqua M., Ciarapica F. E., Mazzuto G.

Dipartimento di Ingegneria Industriale e Scienze Matematiche (DIISM) – Università Politecnica delle Marche, Via Brecce Bianche, 60131 - Ancona – Italy

(s.antomarioni@pm.univpm.it, m.bevilacqua@univpm.it, f.ciarapica@univpm.it, g.mazzuto@univpm.it)

Abstract: Oil is one of the principal energy sources used in worldwide applications hence its transportation should be treated carefully. Indeed, oil spills represent a critical issue form both the environmental and the economic point of view. Association rules mining can describe the frequent patterns characterising the dataset and relations explaining it can be extracted. Since this technique also provides probability measures of the rules extracted, it can be considered as valuable support in defining maintenance procedures and in identifying the existence of concurrent failures. The results of the analysis evidenced the characteristics of the plants that had experienced an oil spill in the period ranging from 1971 and 2014, as well as the most likely consequences deriving from such losses. For these reasons, they should be taken into consideration while defining both maintenance policies to limit the number of future oil spills or, eventually, to be able to face them and limit their impact.

Keywords: Association rules, Maintenance, Oil spills, Refining Industry, FMEA

1. Introduction

Production reliability represents a critical aspect of all industrial environment, especially in process industries, where all process phases are connected and constrained in sequence. In addition, several parameters, such as temperatures, flow rates, chemical properties, mass flows, have to be measured and monitored (Seng and Srinivasan, 2009). Deploying an effective maintenance strategy is fundamental to the success of a company. The vast availability of data represents an opportunity for companies to deepen the knowledge of their processes: introducing data mining (DM) techniques can provide valid support in this context. By definition, indeed, DM aims at discovering hidden information, relationships paths and trends in massive amounts of data (Shahbaz et al., 2006). This paper integrates the need for increasing process reliability and the chance of enhancing process quality capitalising on the application data mining techniques. In particular, association rules mining (ARM) is applied to analyse the characteristics of the oil pipelines that experienced at least an oil spill between 1971 and 2014. Oil pipelines failures, indeed, have vast importance both from an economic and an environmental point of view since oil represents one of the most common energy sources. Moreover, the risk of components failure is high, both because of the wear and because of unpredictable events (Bevilacqua and Ciarapica, 2018; Fabiano and Currò, 2012). Hence, this study aims at evidencing the technical and organisational criticisms that led to losses in oil pipelines using the ARM technique. Additionally, a methodology for the identification of potential failure risks capitalising on association rules performance measurement is proposed, highlighting the applicability of the study to maintenance fields.

The remaining of the paper is organised as follows: section 2 is dedicated to a brief literature review of data mining applications to reliability and maintenance, with a specific focus on the oil industry and ARM applications. Section 3 contains an accurate description of the ARM methodology, while section 4 explains the most important association rules extracted and proposes an integration between ARM and FMEA. Section 5 is dedicated to final remarks.

2. Literature review

The current data availability favours the application of DM techniques to all kind of processes since their potentialities and objectives can help in overcoming traditional methodologies weaknesses (Braha, 2001). For instance, Sipos et al. (2014) proposed a data-driven methodology for predicting machinery faults, extracting useful information to integrate the predictive maintenance policy. An innovative DM method was developed by Gröger et al. (2012), with the aim of optimising a manufacturing process and including applications both to theoretical and practical cases. Besides, an IoT-based energy management system was developed by Bevilacqua et al. (2017) and integrated the decision-making process with data collected and mapped out by smart devices.

However, the application of DM techniques to oil and gas plants is still limited, even though, as suggested by Saybani and Wah (2010) it could represent a source of competitive advantage in this field. Zhong and Wang, (2003) integrated DM tools and Computer Management Systems to predict oil prices and requirements, while Wang and Gao (2012) applied internet of things devices to an oil transfer station for supporting the decision-making in maintenance. Friedemann et al. (2008) also monitored subsea facilities in oil transportation aiming at the improvement of their reliability.

Simple but powerful methodologies like ARM have already been applied to perform predictive analytics. In existing literature, some researchers deployed association rule (AR)-based applications to maintenance and reliability. For instance, Chen et al. (2004) analysed the relations between groups of machines and product defects through ARM, while Cunha et al. (2006) and Kamsu-Foguem et al. (2013) in fault detection. Integrating ARM and Total Productive Maintenance, Djatna and Alitu (2015) obtained a considerable improvement in maintenance both regarding time and costs.

3. Methodology: Association Rules Mining

Association rules mining is a powerful formalism since it allows to extract relationships between items belonging to the same dataset (Buddhakulsomsiri et al., 2006). This methodology is characterised by intuitiveness both regarding implementation and interpretation (Chen et al., 2005), aspect guaranteeing the possibility of applying it to many fields. For instance, in manufacturing, it has been widely employed to estimate resource consumption, supplier performance analysis and for diagnostic purposes (Chen, 2003). A formal definition is then provided: given a set of Boolean data $B=\{b_1,b_2,\dots, b_n\}$ called items, and a set of transactions $T=\{t_1, t_2, \dots, t_m\}$ chosen from B, such that each of the t_i contains a subset of items, the implication $X \rightarrow Y$ is called association rule. X is called "body" of the rule, while Y is the "head" of the rule. The head (Y) and the body (X) of the rule belong to B, and their intersection is an empty set. The key metrics applied in the evaluation of rules quality are:

$\text{Support}(X \rightarrow Y)=(\#\{X \cup Y\})/(\#\{T\})$, where $\#\{T\}$ represent the cardinality of the database (the transaction set) and $\#\{X \cup Y\}$ returns the number of transactions containing both X and Y. The support provides a measure of the statistical significance of the rule, as stated by Agrawal et al. (1993). Indeed, it represents the probability of having the transactions containing both X and Y in the database D.

$\text{Confidence}(X \rightarrow Y)=(\text{Support }\{X \cup Y\})/(\text{Support }\{X\})$: confidence indicates the strength of the relationship between X and Y. In particular, it represents the conditional probability of having Y in a transaction that already contains X, hence: $\text{Confidence}(X \cup Y)=P(Y|X)$.

The FP-growth algorithm is applied to extract the frequent item-sets (FI), namely sets of items appearing more frequently than a user-defined threshold (Han et al., 2007), included in the data-set. Then, through the definition of given criteria, such as support and confidence thresholds, rules are generated from the frequent patterns. According to Shahbaz et al. (2006), association rules mining is helpful in extracting interesting and non-trivial relationships, enhancing process design quality. For this reason, the analysis of the extracted AR and their interpretation represent the fundamental phases of the process.

4. Results

4.1 Dataset description

The dataset applied in the current work integrates information collected by the European Petroleum Refiners Association, Oil Spill Response Limited, Marine Spill Response Corporation, and Association of Control Spill America ranging from 1971 to 2014. Each instance of the dataset reports spillage data regarding safety and environmental impact of oil pipelines. Together with the year when the spills occurred, kind of failure happened is reported. Five different categories have been identified:

a. mechanical failures that are characterised by mechanical breakdowns of the facility;
b. corrosion issues;
c. third party activities;
d. operational failures, related to systems malfunctioning or human-related errors;
e. natural hazard-related problems, such as landslides, floods and earthquakes.

From 2006 on, additional information aiming at explaining the kind of failure has been added to the reports. In particular, in case of corrosion, it is specified whether the damage is internal or external to the pipeline. If the loss is attributed to mechanical issues, instead, it is specified whether it is related to construction errors or design inaccuracy. Third party activities are detailed into accidental, intentional or theft attempts. Operational and natural hazard-related have not been furtherly specified.

Additional information reported in the dataset detail the kind of facility damaged, the diameter of the pipes injured, the category of product spilt, the presence of wounded people, and fire. Furthermore, an estimate of the volume loss and the area contaminated. An excerpt of the dataset used in the analysis is reported in Figure 1.

Year	Failure	Facility	Line size	Product	Injuries	Net spill (m^3)	Area (m^2)
2013	THIRD PARTY ACTIVITY	Underground pipe	10''	Diesel	/	30	**3000**
2014	MECHANICAL	Underground pipe	10,75''	Crude oil	/	1	**0**
2014	THIRD PARTY ACTIVITY	Pump station	24''	Crude oil	/	3	**200**
2014	THIRD PARTY ACTIVITY	Underground pipe	6''	Gasoline	/	0	**100**

Figure 1 Excerpt of the dataset used.

4.2 AR extracted

The dataset has been analysed through RapidMiner, a popular data mining tool, which consents the development of analysis processes combining apposite operators. According to dataset structure and considering the aim of the work, RapidMiner process has been structured as reported in Figure 2. In particular, the first operator allows the reading of the dataset, while the second one is applied for the generation of ranges useful for data analysis. Through the third tool, "Exclude Attributes", a vertical selection is performed, in order to exclude the attributes that, due to the previous elaborations, would not provide useful information in generating the association rules. "Nominal to Binominal" and "Numerical to Binolminal" tools assure that data are analysed in a Boolean format, as required by the "FP-Growth" operator. At this point, FP-Growth algorithm is applied for calculating frequent patterns, while AR are mined through "Create AR" tool.

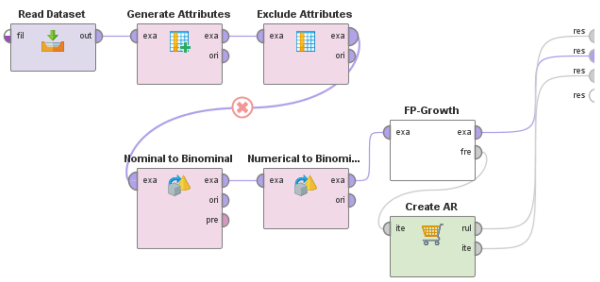

Figure 2 RapidMiner process developed for ARM.

Considering a maximum of 3 items per rule, 4193 AR were extracted. Only the most relevant in terms of support, confidence and general meaning will be discussed.

For instance, considering the five categories of failures as the body of the rules, the most interesting rules extracted – that are those with support higher than 0.1 - are reported in Table 1. The majority of the rules present "third party activities" as "body" of the rule: Table 1 indicates that in 38.5% (support(third party activity, pipe run)) of the cases analysed the failure is caused by third-party actions and the facility part injured is the pipe run. Besides, if the event is imputable to third party activities, in the 91.0% of cases the pipe run will be damaged.

Table 1 AR extracted having failure categories as "body" (Supp=support and Conf=confidence).

Body	Head	Supp	Conf
Third-party activity	Facility part: Pipe run	0.385	0.910
Third-party activity	Net oil spill (1 - 100)	0.220	0.520
Corrosion	Facility part: Pipe run	0.216	0.862
Third-party activity	Pipe diameter (6 -10)	0.212	0.502
Third-party activity	Facility part: Pipe run, Pipe diameter (6-10)	0.201	0.476
Third-party activity	Facility part: Pipe run, Net oil spill (1 - 100)	0.198	0.468
Third-party activity	Contamination (1 - 500)	0.149	0.352
Third-party activity	Facility part: Pipe run , Contamination (1 - 500)	0.136	0.322
Third-party activity	Net oil spill (1 - 100), Pipe diameter (6 -10)	0.123	0.292
Mechanical failure	Net oil spill (1 - 100)	0.122	0.508
Corrosion	Pipe diameter (6 -10)	0.112	0.449

In an analogue way, if the failure is caused by corrosion, the probability that the fault is on the pipe run – that is the confidence – is 0.862. Mechanical failure and an oil net spillage lower than 100 m^3 characterise the 12.2% of accidents, while the confidence of the rule is 50.8%. Corrosion failures and 6"-10" diameters, instead, have a support of 0.112 and a confidence of 44.9%.

Table 2 AR extracted having pipe diameters as "body" (Supp=support and Conf=confidence).

Body	Head	Supp	Conf
Pipe diameter (6 -10)	Net oil spill (1 – 100 m^3)	0.214	0.529
Pipe diameter (6 -10)	Facility part: Pipe run, Net oil spill (1 – 100 m^3)	0.176	0.435
Pipe diameter (11 -15)	Facility part: Pipe run	0.162	0.730
Pipe diameter (6 -10)	Contamination (1 – 500 m^2)	0.151	0.372
Pipe diameter (6 -10)	Facility part: Pipe run, Contamination (1 – 500 m^2)	0.127	0.314
Pipe diameter (16 -20)	Facility part: Pipe run	0.116	0.744
Pipe diameter (11 -15)	Net oil spill (1 – 100 m^3)	0.112	0.508

Table 2 reports the rules having pipe diameters as "body". It is noteworthy that the rules having Pipe diameter (6 - 10) and Pipe diameter (11 -15)as body both have Net oil spill (1 - 100) as head, highlighting that, in the given dataset, the dimension of the pipe does not have a significant impact on the net oil spills verified. Even in this set of data, the only facility part evidenced is pipe run, highlighting that its failures are the most common.

The contaminated area represents an interesting factor from an environmental point of view. In a 20.0% of accidents, the area contaminated (1 – 500) appears contextually to a net oil spill volume lower than 100 $m^{3.}$ However, the support reported for the other instances in Table 3, is very low, indicating that the item-sets composed by contaminated area larger than 500 m^2 and net oil spill rarely appear together: this result evidence that, despite the high number of oil spills occurred between 1971 and 2014, they were promptly noticed and repaired, avoiding a huge contamination of the area. Moreover, it means that measures preventing area contamination had also been applied.

Table 3 AR extracted having contaminated area as body (Supp=support and Conf=confidence).

Body	Head	Supp	Conf
Contamination (1 – 500 m^2)	Net oil spill (1 – 100 m^3)	0.200	0.622
Contamination (501 - 1000 m^2)	Net oil spill (1 - 100 m^3)	0.031	0.680
Contamination (1001 - 1500 m^2)	Net oil spill (1 - 100 m^3)	0.018	0.588
Contamination (9501 - 10000 m^2)	Net oil spill (1 - 100 m^3)	0.013	0.778
Contamination (5501 - 6000 m^2)	Net oil spill (101 - 300 m^3)	0.005	0.600
Contamination (7501 - 8000 m^2)	Net oil spill (101 - 300 m^3)	0.002	1.000
Contamination (6501 - 7000 m^2)	Net oil spill (301 - 500 m^3)	0.002	1.000
Contamination (6001 - 6500 m^2)	Net oil spill (301 - 500 m^3)	0.002	0.333
Contamination (6001 - 6500 m^2)	Net oil spill (3501 - 4000 m^3)	0.002	0.333

4.4 Methodology application: Integrating FMEA and ARM

The results highlighted in the previous section could be applied to conduct a Failure Modes Effects Analysis (FMEA). In particular, the performance measures extracted for the relevant rules could be used for Risk Priority Number (RPN) calculation.

The idea underlying FMEA methodology consists on anticipating all the potential failure modes, their causes and their potential consequences during the design phase of a product or process, in order to reduce inefficiencies in product or process development and to mitigate the risks (Bevilacqua et al., 2015). FMEA should be executed by cross-functional company members in order not to lose any potential issue (Carlson, 2014).

The conceptual phases describing FMEA procedure, as effectively summarised by Carlson (2014) are:

1. Identification of the potential failure modes causes and consequences on the systems and users;
2. Risk assessment of the identified items and prioritisation;
3. Identification of the corrective actions and planning of the maintenance strategy.

A ranking of the identified items is performed, and actions are undertaken on the most critical ones recursively, creating a continuous loop.

The general procedure proposed for this application can be summarised as presented in Figure 3. The first step regards the collection of data related to the process observed. As recommended by FMEA procedure, a list of all the potential failure modes is prepared, together with an investigation of the corresponding causes and effects.

The risk priority number is calculated to manage the risk level associated with each failure mode:

$RPN=P \times (w_s \times S) \times D$,

where S represents the severity of the failure, D is its detectability, and P is the probability of occurrence of the failure. w_s , instead, is the weight associated to S. Severity, Detectability and Probability are usually determined using data ranked on a given scale. The approach proposed in this work considers different paths for some parameters calculation: indeed, the evaluation of the probability of occurrence of each failure mode (P) and the weight w_s for the severity associated with each effect (S) are determined using the ARM.

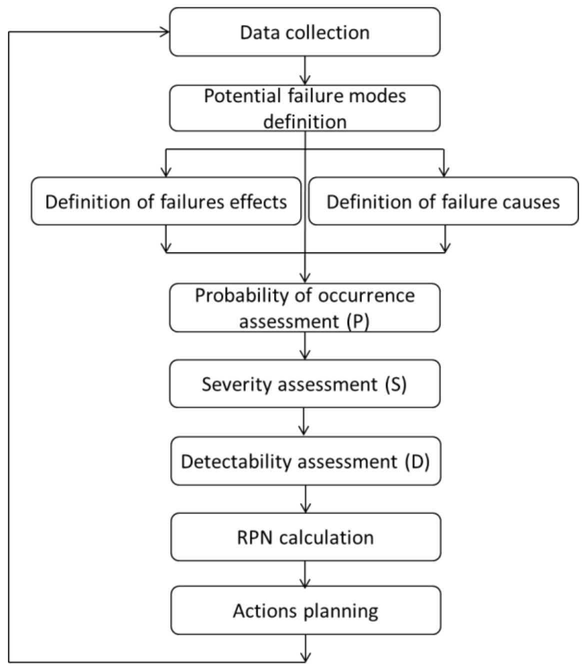

Figure 3 Procedure for the integration of ARM and FMEA.

A dataset of historical data recording information on failure modes and previous failures effects is required, and the procedure is developed as follows:

a. FI are extracted through FP-growth algorithm;
b. FI containing a single item are selected: if they contain one of the failure modes, then the probability P associated with the failure mode (FM) is indicated as P(FM)=sup(FM);
c. AR are extracted from the frequent patterns;
d. For each AR in the form X→Y, where X is the FM and Y is a failure effect, the weight w_s associated to the severity of the failure effect is w_s(Y)=Conf(X→Y);

The traditional ranking, instead, is applied for what concerns Detectability and Severity: Table 4 reports the most common ranking scales for D evaluation while Table

5 indicates the scale for Severity assessment. Both the qualitative assessments are performed by a cross-functional team of experts.

Table 4 Detectability scale.

Ranking	Degree	Description
10	Absolute uncertainty	It is not possible to detect the failure
9	Very remote	The failure is very difficult to detect
8	Remote	The failure is difficult to detect
7	Very low	The chance of detecting the failure is very low
6	Low	The chance of detecting the failure is low
5	Moderate	The chance of detecting the failure is moderate
4	Moderately high	The chance of detecting the failure is moderately high
3	High	The chance of detecting the failure is high
2	Very high	The chance of detecting the failure is very high
1	Almost certain	The chance of detecting the failure is certain

Table 5 Severity scale.

Ranking	Degree	Description
10-9	Catastrophic	Dangerous for safety
8-7	Critical	Ruptures and accidents may happen on pipelines
6-5	Significant	Structural safety and performance factors are compromised by the failure
4-3	Insignificant	Defects can be repaired
2-1	Very Insignificant	Performance effects are not influenced by the failure

The RPN calculated by multiplying the three contributions ranges between 0 and 100 since P can vary between 0 and 1, and S and D range from 1 and 10. The higher the RPN is, the more critical is the corresponding failure mode.

Considering the most critical issues, corrective actions can be applied, and a predictive maintenance strategy can be planned. The ARM is valuable to support the latter case since it provides information on the relationships existing between components’ concurrent failures.

As presented in Figure 3, the procedure is a continuous loop: after having implemented the corrective actions, data are newly gathered, and the new degree of risk is re-assessed.

The failure categories included in the dataset represent the possible failure modes to be analysed in FMEA application. In Table 6 the list of potential failure modes is reported, together with the causes identified.

Table 6 Failure modes, related causes and probabilities.

Failure modes	Causes	P
Mechanical failures	Construction and material defects	0.240
Corrosion failures	Chemical reactions between pipeline material and the fluid, external erosion	0.250
Third party activities	Excavations, falling of heavy materials, thefts	0.423
Operational failures	Human errors, system breakdowns	0.064
Natural hazard failures	Earthquakes, floods, landslides	0.023

All the failure modes identified include oil spills as an effect; moreover, the presence of injured people and fire represent relevant consequences, together with the volume of product lost and the dimensions of the area contaminated.

According to the research approach explained, P and S were calculating capitalising on the extraction of AR from the collected data. Table 6 contains the probability measures associated with each failure mode, using the support of the single item-set, while Table 7 reports the severity weight of the effects calculated through the confidence of the AR. The last column of Table 7 contains the qualitative evaluation of S performed by company experts.

Table 7 Severity of the effects.

FM	Effects	w_S	S
Corrosion	Net volume loss (1 - 100 m³)	0.478	8
Corrosion	Contamination (1 - 500 m²)	0.312	9
Mechanical	Net volume loss (1 - 100 m³)	0.508	8
Mechanical	Contamination (1 - 500 m²)	0.326	9
Mechanical	Net volume loss (1 - 100 m³), Contamination (1 - 500 m²)	0.197	10
Natural Hazard	Net volume loss (1 - 100 m³)	0.692	8
Operational	Net volume loss (1 - 100 m³)	0.343	8
Operational	Contamination (1 - 500 m²)	0.257	9
Third party activities	Net volume loss (1 - 100 m³)	0.519	8
Third party activities	Contamination (1 - 500 m²)	0.352	9
Third party activities	Net volume loss (1 - 100 m³), Contamination (1 - 500 m²)	0.240	10

Company experts also evaluated the detectability of failure by current control activities. In particular, they affirmed that failure related to third-party actions were not detectable (10) as well as those associated with natural hazards, as any control was performed in this sense. Lower values were associated with operational, mechanical and corrosion issues. Table 8 reports all the indexes calculated and the final RPN for some of the relevant rules extracted.

Table 8 Results of RPN calculation

FM	Effects	P	w_S *x* S	D	RPN
Third party activities	Net volume loss (1 - 100 m³)	0.423	4.154	10	17.573
Third party activities	Contamination (1 - 500 m²)	0.423	3.167	10	13.398
Third party activities	Net volume loss (1 - 100 m³), Contamination (1 - 500 m²)	0.423	2.403	10	10.166
Mechanical	Net volume loss (1 - 100 m³)	0.240	4.060	3	2.924
Mechanical	Contamination (1 - 500 m²)	0.240	2.932	3	2.111
Mechanical	Net volume loss (1 - 100 m³), Contamination (1 - 500 m²)	0.240	1.970	3	1.418
Natural Hazard	Net volume loss (1 - 100 m³)	0.023	5.538	10	1.274
Corrosion	Net volume loss (1 - 100 m³)	0.250	3.826	1	0.956
Operational	Net volume loss (1 - 100 m³)	0.064	2.743	5	0.878
Operational	Contamination (1 - 500 m²)	0.064	2.314	5	0.741
Corrosion	Contamination (1 - 500 m²)	0.250	2.804	1	0.701

The most critical events were historically associated with third-party activities. Indeed, their probability of occurrence is higher than the other FM, as well as the difficulty in detecting them. Operational and corrosion issues represent the least critical FM: indeed, despite a value of severity comparable to those of other cases, the lower levels of P, w_s, and D favour them.

RPN related to third-party activities can be enhanced acting on detectability measures. Indeed, controlling on them can provide benefits for the whole company and, consequently, on environmental issues. Inserting periodical inspection could represent a valuable solution. In a second phase of the analysis, after having performed the re-calculation of the RPN, the controls executed on mechanical, corrosion, and operational failure modes could be improved.

5. Conclusion

Association rules mining is a powerful methodology for large and complex datasets. Hence its application should be favoured in all industrial field. In this paper, an analysis of the implementation of association rules mining to describe oil pipelines spills has been presented. Moreover, a procedure to reassess the risk priority number in FMEA methodology capitalising on the AR performance measures have been proposed, also presenting an example based on the AR mined in the current work. Both the AR extracted and the procedure presented could be applied by oil refineries or oil plants aiming at re-engineering their maintenance processes, in a predictive perspective.

It has to be noticed that a lack of historical data would make the analysis inapplicable, as well as a substantial modification of the operational condition of the plant in comparison to the one when data were collected. Moreover, since the procedure is based on well-known processes, it could be necessary to readapt it if it has to be applied to innovative methods.

References

Agrawal, R., Imieliński, T., Swami, A., (1993). Mining association rules between sets of items in large databases, in: *Acm Sigmod Record. ACM*, pp. 207–216.

Bevilacqua, M., Ciarapica, F.E., (2018). Human factor risk management in the process industry: A case study. *Reliab. Eng. Syst. Saf.* 169, 149–159.

Bevilacqua, M., Ciarapica, F.E., Diamantini, C., Potena, D., (2017). Big data analytics methodologies applied at energy management in industrial sector: A case study. *Int. J. RF Technol.* 8, 105–122.

Bevilacqua, M., Mazzuto, G., Paciarotti, C., (2015). A combined IDEF0 and FMEA approach to healthcare management reengineering. *Int. J. Procure. Manag.* 8, 25.

Braha, D. (Ed.), (2001). *Data mining for design and manufacturing: methods and applications*, Massive computing. Kluwer Academic, Dordrecht ; Boston.

Buddhakulsomsiri, J., Siradeghyan, Y., Zakarian, A., Li, X., (2006). Association rule-generation algorithm for mining automotive warranty data. *Int. J. Prod. Res.* 44, 2749–2770.

Carlson, C.S. (2014). *Understanding and applying the fundamentals of FMEAs.* Ariz. USA IEEE Inst. Electr. Electron. Eng.

Chen, M.-C. (2003). Configuration of cellular manufacturing systems using association rule induction. *Int. J. Prod. Res.* 41, 381–395.

Chen, M.-C., Huang, C.-L., Chen, K.-Y., Wu, H.-P., (2005). Aggregation of orders in distribution centers using data mining. *Expert Syst. Appl.* 28, 453–460.

Chen, W.-C., Tseng, S.-S., Wang, C.-Y., 2004. A Novel Manufacturing Defect Detection Method Using Data Mining Approach, in: *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, Heidelberg, pp. 77–86.

Cunha, C.D., Agard, B., Kusiak, A., (2006). Data mining for improvement of product quality. *Int. J. Prod. Res.* 44, 4027–4041.

Djatna, T., 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 Manuf.* 4, 336–343.

Fabiano, B., Currò, F., (2012). From a survey on accidents in the downstream oil industry to the development of a detailed near-miss reporting system. *Process Saf. Environ. Prot.*, Special Issue Celebrating Trevor Kletz's 90th Birthday 90, 357–367.

Friedemann, J.D., Varma, A., Bonissone, P., 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 2008*, pp. 298–308.

Gröger, C., Niedermann, F., & Mitschang, B. (2012). Data mining-driven manufacturing process optimization. In *Proceedings of the world congress on engineering.* 3, pp. 4-6.

Han, J., Cheng, H., Xin, D., Yan, X., 2007. Frequent pattern mining: current status and future directions. *Data Min. Knowl. Discov.* 15, 55–86.

Kamsu-Foguem, B., Rigal, F., Mauget, F., 2013. Mining association rules for the quality improvement of the production process. *Expert Syst. Appl.* 40, 1034–1045.

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

Seng, N.Y., Srinivasan, R., 2009. Data mining for the chemical process industry, in: *Encyclopedia of Data Warehousing and Mining, Second Edition*. IGI Global, pp. 458–464.

Shahbaz, M., Srinivas, M., Harding, J.A., Turner, M., 2006. Product Design and Manufacturing Process Improvement Using Association Rules. *Proc. Inst. Mech. Eng.* Part B J. Eng. Manuf. 220, 243–254.

Sipos, R., Fradkin, D., Moerchen, F., 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.

Wang, Q., Gao, J., 2012. Research and application of risk and condition based maintenance task optimization technology in an oil transfer station. *J. Loss Prev. Process Ind.* 25, 1018–1027.

Zhong, N.-N., Wang, X., 2003. Application of data mining in refinery CIMS, in: *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693).* Presented at the Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693), 1, p. 118–122.