

## Identificação de Padrões de Falhas em Válvulas Submarinas Utilizando Mineração de Dados

### *Identification of Failure Patterns in Subsea Valves Using Data Mining*

Antonio Heverton Martins Silva<sup>1</sup>, Wallan Azevedo dos Santos<sup>2</sup>, Henrique Rego Monteiro da Hora<sup>3</sup>

#### RESUMO

Esse artigo se propõe a utilizar a descoberta de conhecimento em banco de dados para identificar pontos críticos de falhas em equipamentos em operação, especificamente de válvulas submarinas aplicadas na produção offshore de óleo e gás. Discute-se ainda sobre a possibilidade de aplicar o processo de mineração de dados na melhoria da qualidade de processos. Para tanto, foram aplicadas técnicas de mineração de dados de associação e de classificação através do software WEKA. A identificação de uma associação entre os parâmetros contidos neste banco de dados proporcionaria ganhos no setor de manutenção, aumento da confiabilidade e segurança dos componentes. Verificou-se que as falhas hidráulicas ocorrem de forma mais acentuada nas válvulas VGA. O mesmo ocorre em válvulas do tipo VGP. A ferramenta e metodologia que se mostraram adequadas para extrair conhecimento de maneira sistemática, que proporcionaram um teor científico aos resultados obtidos auxiliando na identificação de itens críticos e na proposição de maneiras de contorná-los. Obtiveram-se como produto final árvores de decisão consideradas tecnicamente relevantes para gestão da melhoria da qualidade de produtos.

**Palavras-chave:** Mineração de Dados, Válvulas Submarinas, Falha de Válvulas.

#### ABSTRACT

This paper uses the knowledge discovery in database to identify critical points of failure in equipment on operation, specifically submarine valves applied in offshore production of oil and gas. It is also discussed the possibility of applying the data mining process in quality process improvement. For that, we applied association and classification data mining techniques through WEKA software. Identifying an association between the parameters contained in this database would lead to gains in the maintenance process and increased component reliability and safety. The results found that hydraulic failures occur more in the annular gate valves. The same is true of production gate valves. The methodology proved adequate to extract knowledge in a systematic way, which provided a scientific content to the results obtained helping in the identification of critical items and in proposing ways to get around them. The final product was decision trees considered technically relevant for the equipment quality improvement.

**Keywords:** Data Mining, Subsea Valves, Valve Failure

<sup>1</sup> Mestre em Sistemas Aplicados à Engenharia e Gestão (IFF). Engenheiro de Equipamentos na Petrobras.

E-mail:

hevertonmartins@gmail.com

<sup>2</sup> Mestre em Sistemas Aplicados à Engenharia e Gestão.

<sup>3</sup> Doutor em Engenharia de Produção (UFF). Professor do Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFF).

## 1. INTRODUÇÃO

A indústria, de modo geral, está sempre perseguindo a redução dos seus custos operacionais. Nessa busca, um dos pilares fundamentais é manter alta a confiabilidade e a disponibilidade dos equipamentos de seus processos produtivos. Diante desse desafio, diversas instituições têm buscado na tecnologia os recursos e as ferramentas para suportar essas iniciativas (CHOPRA; MEINDL, 2013; VIEGAS, 2013).

Na área de exploração e produção de óleo e gás em campos marítimos, a confiabilidade e disponibilidades dos equipamentos são fundamentais. Haja vista que extrair e transportar o petróleo de reservatórios offshore em águas profundas e ultra-profundas exigem sistemas de produção submarinos. Logo, a tecnologia voltada para equipamentos submarinos utilizados na produção de óleo e gás é um campo de aplicação altamente especializado que coloca demandas específicas e pontuais para a engenharia (BAI; BAI, 2012).

A Árvore de Natal Molhada (ANM) é um equipamento instalado na cabeça do poço submarino e no Brasil, a maior operadora de campos de petróleo, já realizou a instalação de mais de 1.000 equipamentos desse tipo. A ANM é composta basicamente por válvulas, atuadores e sensores e tem por função controlar o fluxo de hidrocarbonetos e permitir a produção segura durante sua vida útil.

Por estar envolvida na segurança do meio ambiente e das pessoas (evitar ou mitigar vazamentos e fluxos descontrolados de óleo e gás), a confiabilidade e a análise de falhas da ANM e seus subcomponentes é uma área de estudo relevante. Usualmente, estudos de degradação em componentes mecânicos, tais como válvulas, são analisados em laboratório, sob condições controladas, cada componente individualmente.

Por outro lado, os dados de campo de equipamentos instalados podem também ser utilizados nesse tipo de análise. Entretanto, isto é raramente realizado porque o histórico das condições de operação e o grau de conhecimento destas condições dificultam sua utilização na modelagem da degradação dos componentes que resulte numa avaliação representativa (EUTHYMÍOU, 2013).

A melhoria da qualidade de processos e produtos industriais exigem a coleta e a análise de dados para identificar problemas relacionados à fabricação, montagem, operação, etc. Embora ferramentas de análise de dados tradicionais venham sendo utilizadas com sucesso na melhoria da qualidade, hoje existem melhores ferramentas para

minerar grandes massas de dados coletados através dos sistemas industriais computadorizados (KÖKSAL; BATMAZ; TESTIK, 2011).

Após a instalação da ANM, um dos componentes mais importantes na vida útil para a operação e produção do poço são as válvulas submarinas. Por esse motivo o conjunto válvula-atuador é objeto de estudo deste trabalho. Nesse sentido, o presente artigo se propõe a utilizar a descoberta de conhecimento em banco de dados para identificar pontos críticos de falha em equipamentos em operação e discutir sobre a possibilidade de aplicar essa ferramenta na melhoria da qualidade de produtos, especificamente de válvulas submarinas aplicadas na produção offshore de óleo e gás.

## 2. MATERIAIS E MÉTODOS

Essa pesquisa é classificada como quantitativa, descritiva e exploratória por utilizar ferramentas de mineração de dados com o objetivo de mensurar e identificar pontos de falhas em válvulas submarinas e suas correlações com os demais parâmetros analisados. É dita exploratória por utilizar um banco de dados de registros de falhas reais identificadas pelas equipes de operação e produção (SILVA; MENEZES, 2005).

Para o desenvolvimento do estudo utilizou-se um banco com 4.417 registros de falhas identificadas em campo dos componentes que integram o conjunto Árvore de Natal Molhada. Desse grupo foram extraídos os dados relativos as falhas do conjunto válvula-atuador que representam pouco mais de 57% dos dados registrados. Além do alto percentual de falha com relação aos demais componentes as válvulas são um dos componentes mais importantes desse tipo de equipamento por envolver questões operacionais de produção e de segurança. O banco de dados possui o registro de falhas de equipamentos relatadas formalmente nos relatórios técnicos desde 1996.

Os registros de falhas possuem diversos campos, são eles: identificação da ANM, fornecedor, tipo/projeto, componente falho, identificação do componente, meio de detecção, modo de falha, ação tomada e descrição da falha. Contudo, para a análise realizada foram eliminados alguns campos por conterem informações textuais que ficam a critério do técnico que faz o registro, como por exemplo, descrição da falha. Além desses campos, outros foram retirados após consultas a engenheiros de operação que trabalham especificamente com equipamentos submarinos de produção.

Dessa maneira, dos campos citados anteriormente, restaram fornecedor, tipo/projeto do equipamento, tipo de válvula e modo de falha que foram selecionados como atributos a serem processados pelo software. Cabe destacar que o atributo tipo de válvula foi pré-

processado a partir do campo identificação do componente presente do banco de registros de falha. Essa ação foi realizada para agrupar os componentes idênticos, por exemplo, duas válvulas semelhantes por projeto e instaladas no mesmo equipamento podem estar identificadas como V1 e V2.

A aplicação das técnicas de mineração de dados se deu através do software Waikato Environment for Knowledge Analysis (WEKA), no qual foram utilizados algoritmos de associação (Apriori e Predictive Apriori) e de classificação (J48) de forma interativa alterando-se os parâmetros iniciais, sendo geradas diversas regras de associação e árvores de decisão para serem avaliadas, extraíndo-se daí as mais relevantes tecnicamente.

## 2.1. Técnicas de Mineração de Dados

Mineração de Dados ou data mining (DM) é uma área de pesquisa multidisciplinar que inclui tecnologia de bancos de dados, inteligência artificial, aprendizado de máquina, redes neurais, estatística, reconhecimento de padrões, sistemas baseados em conhecimento, recuperação da informação, computação de alto desempenho e visualização de dados. O processo de data mining se baseia em extrair ou minerar conhecimento de grandes volumes de dados (REIS, 2015).

O processo de DM envolve tarefas, métodos e algoritmos para possibilitar a extração de novos conhecimentos e é geralmente classificada pela sua capacidade em realizar determinadas tarefas (GOLDSCHMIDT; PASSOS, 2005), dentre as quais estão:

□ Associação – A tarefa de associação consiste em identificar a relação existentes entre os atributos. Apresentam a forma comum: “SE atributo X ENTÃO atributo Y”. É considerada uma das tarefas mais conhecidas devido aos bons resultados obtidos, principalmente nas análises da “Cestas de Compras” (Market Basket), onde identificamos quais produtos são levados juntos pelos consumidores (CAMILO; SILVA, 2009).

□ Classificação – Esse tipo de tarefa visa identificar a qual classe um determinado registro pertence. Nesta, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à classe pertencente, com o objetivo de entender como classificar um novo registro, sendo classificado como um tipo de aprendizado supervisionado. Geralmente utilizada para prever valores de variáveis do tipo categóricas.

Existem alguns indicadores de análise devem ser considerados na utilização de DM. Os mais utilizados são: suporte, confiança, conviction, leverage e interest ou lift. A seguir a descrição de suporte e confiança que são os indicadores utilizados nas análises desenvolvidas no presente trabalho.

□ Suporte – o suporte de uma determinada regra mede a frequência com que esse padrão aparece na base de dados. Esse indicador é o número absoluto de ocorrências da regra obtida na amostra de dados utilizada.

□ Confiança – corresponde a um valor de correlação entre os itens que formam esse padrão. Esse indicador varia de 0 a 1, sendo que quanto mais próximo de 1 melhor é a qualidade do resultado.

Para evitar que sejam gerados um número de regras elevado, o que dificultaria a análise da relevância das regras, geralmente condicionam-se as regras para um suporte baixo e uma confiança alta.

### 2.1.1. Algoritmo de Classificação J48

O algoritmo J48 cria uma árvore de decisão binária simples para modelar o processo de classificação. Uma vez que a árvore é construída, ela é aplicada ao conjunto de valores inter-relacionados no banco de dados e resulta na classificação. As árvores de decisão são as abordagens mais poderosas na descoberta de conhecimento e na mineração de dados, sendo uma técnica muito importante, pois permite modelagem e extração de conhecimento a partir da maior parte dos dados disponíveis (BHARGAVA, 2013).

Ao construir uma árvore, o J48 ignora os valores faltantes, ou seja, o valor desse item pode ser previsto com base no que é conhecido sobre os valores dos atributos para os outros registros. A idéia básica é dividir os dados em um intervalo com base nos valores de atributo desse item que são encontrados na amostra de treinamento (PATIL, 2013).

### 2.1.2. Algoritmo de Associação

O algoritmo Apriori é usado para gerar conjuntos de itens frequentes no banco de dados. Esse algoritmo gera um conjunto de conjuntos de itens candidatos cujos comprimentos são  $(k + 1)$  dos conjuntos de itens  $k$ , e elimina os candidatos que não contém o subconjunto. Então, para os demais candidatos, apenas os candidatos que satisfaçam o limite de suporte mínimo (definido anteriormente pelo usuário) são considerados conjuntos de itens (AGGARWAL, 2013).

Notação utilizada no algoritmo Apriori:

- $k$  – qualquer conjunto de  $k$  itens.
- $C_k$  – Conjunto de candidatos de  $k$  itens.
- $L_k$  – Conjunto com maior conjunto de  $k$  itens. Esses conjuntos de itens são gerados a partir dos conjuntos candidatos de cada passo.

### 3. RESULTADOS E DISCUSSÃO

Depois do pré-processamento e seleção dos atributos mais relevantes, o banco de dados de registro de falha em válvulas submarinas utilizado apresenta 2622 instâncias e 4 atributos, sendo este último: fornecedor, características de projeto (relacionadas as requisições e premissas da época do projeto e ao ambiente de instalação do equipamento que está correlacionada na maioria dos equipamentos com o tempo de instalação e em operação, dessa forma estão divididos em três fases), tipo de válvula e modo de falha. Na Tabela 1 estão representados os atributos para facilitar o entendimento.

**Tabela 1.** Parâmetros avaliados.

Fornecedor	Projeto	Tipo de válvula	Modo de falha
Fornecedor 1	Fase 1 (Rasa)	Gaveta 2" – VGA	Indicador posição
Fornecedor 2	Fase 2 (Profunda)	Gaveta 4" – VGP	Estanqueidade
Fornecedor 3	Fase 3 (Ultra)	Gaveta 1" – VGH	Hidráulica
Fornecedor 4		Gaveta – VGM	Mecânica
Fornecedor 5		Direcional – VD	Outro
		Outras	

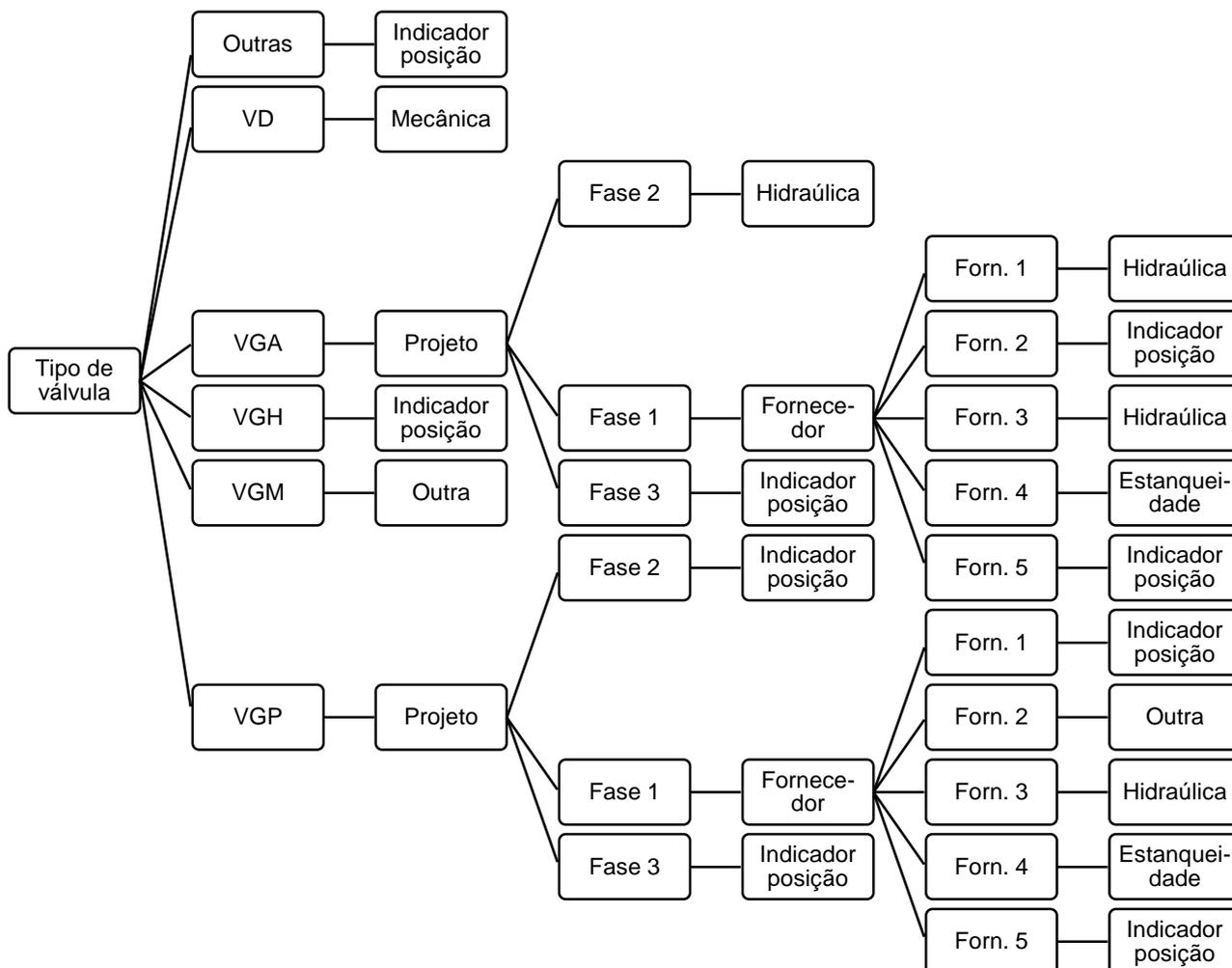
#### 3.1. Análise da Árvore de Decisão

A árvore de classificação resultante está ilustrada na Figura 1. Nela é possível observar que as falhas majoritariamente estão relacionadas ao sistema de indicação de posição. Nas válvulas VGH esse tipo de falha ocorre independente do fornecedor ou da fase do projeto do equipamento, que como dito anteriormente, está relacionada a profundidade de instalação em relação a lâmina d'água e ao tempo em operação. Ou seja, equipamento da Fase 1 estão instalados em profundidades menores e há mais tempo em operação, enquanto que o projeto da Fase 3 é o oposto. Entretanto, para as válvulas VD a falha mecânica independe de qualquer outro atributo, similar ao que ocorreu para VGH com relação a falha de indicação de posição da válvula.

A VGA apresentou uma ramificação da árvore que pode ser entendida como: em equipamentos projetados para operar em lâmina d'água profunda, as falhas no acionamento hidráulico, seja para abertura ou fechamento, ocorrem para qualquer fornecedor. Do mesmo modo, a falha de indicação de posição ocorre para qualquer fornecedor para equipamentos do projeto Fase 3. Entretanto, essa última afirmação pode

sofrer alteração futuramente porque esses equipamentos são os que estão em operação a menos tempo.

**Figura 1.** Árvore de decisão aplicando algoritmo de classificação J48.



Para o projeto de equipamento da Fase 1, dois fornecedores apresentaram falhas de indicação de posição, dois falhas de acionamento hidráulico e uma falha de estanqueidade da válvula. Esta última ocorrência pode ser atribuída ao tempo mais longo em operação deste equipamento quando comparado com os projetos Fase 1 e Fase 2. Apesar disso, os demais fornecedores não apresentaram essa falha com tanto destaque.

A VGP apresentou resultados bastante similares a VGA, isso pode ter acontecido devido as similaridades de projeto e operação dessas válvulas, sendo a principal diferença

entre elas o tamanho, 4” e 2”, respectivamente. O fluido que estão em contato também pode ter influenciado no tipo de falha apresentada (VGA – gás tratado e VGP – fluido produzido do reservatório).

Dentre as semelhanças identificadas nesses dois tipos de válvulas estão as falhas para equipamentos do projeto Fase 1 terem a mesma relação com o fornecedor, ou seja, Fornecedor 1 e Fornecedor 5 apresentaram falhas do sistema de indicação, Fornecedor 3 falha no acionamento hidráulico, Fornecedor 4 falha de vedação (estanqueidade). Nesse ponto, apenas o Fornecedor 2 divergiu, como pode ser visto na Figura 1.

### 3.2. Análise da Árvore de Decisão

Assim como feito na análise de classificação, foram realizadas diversas interações com algoritmo de associação onde se verificou que a falha do sistema de indicação de posição se sobrepunha as demais. Então, foi realizado um pré-processamento nos dados de forma a retirar esse atributo da análise. Retirando o modo de falha indicador de posição, foi utilizado o algoritmo de associação Apriori para gerar um conjunto de 20 regras com confiança mínima de 0,4.

Das regras geradas, a Tabela 2 traz as onze regras de associação geradas com informações mais relevantes segundo a opinião de engenheiros especialistas e técnicos de operação. A partir delas é possível inferir que a falha de vedação ou estanqueidade, regras 1 a 6, é bastante acentuada no Fornecedor 4 para o projeto Fase 1. Isso pode estar relacionado com as características de projeto desse fornecedor. Além disso, equipamentos instalados que pertencem a esse projeto são mais antigos que os demais e essa falha, por sua característica, pode estar relacionada ao desgaste natural. Portanto, esse fornecedor pode ter melhorado algum aspecto de projeto ou construtivo nos equipamentos dos demais projetos ou ainda não houve tempo de vida útil suficiente para a manifestação dessa falha.

A falha hidráulica ocorre de forma mais acentuada nas válvulas VGA. Esse tipo de falha também é comum nas VGP, entretanto com uma confiança menor. Isso pode ser antagônico já que as válvulas VGP estariam, em tese, submetidas a condições operacionais mais adversas, tais como, o fluido do processo mais prejudicial e o número de acionamento mais alto. Pode-se inferir que o sistema de atuação da VGA necessita de um olhar crítico para identificar pontos de melhoria.

**Tabela 2.** Resultados da aplicação do algoritmo de associação.

	REGRA DE CONHECIMENTO	CONFIANÇA
1	Válvulas do Fornecedor 4 apresentaram 210 falhas de estanqueidade, sendo 195 destas no projeto da Fase 1.	0,93
2	Em 338 ocorrências de falhas de estanqueidade, 286 ocorreram em equipamentos do projeto da Fase 1 independente do fornecedor.	0,85
3	Nas 286 falhas de estanqueidade de válvulas do projeto Fase 1, 195 foram adquiridas do Fornecedor 4.	0,68
4	Dentre as falhas 338 falhas de estanqueidade de válvula, 210 estão relacionadas ao Fornecedor 4.	0,62
5	Na falha de estanqueidade, 195 delas envolveram equipamentos do projeto Fase 1 do Fornecedor 4.	0,58
6	Nos 360 registros de falhas para válvulas do Fornecedor 4 e do projeto Fase 1, 195 deles são falhas de estanqueidade.	0,54
7	Para as 508 falhas hidráulica do atuador, 308 foram nas válvulas VGA independentemente de qualquer outro critério.	0,52
8	A falha hidráulica do atuador se concentra nos equipamentos do projeto Fase 1, sendo responsável por 275 dos 588 registros dessa falha.	0,47
9	O modo de falha predominante na válvula VGA, 308 dos 726 registros, é a falha na atuação hidráulica, independentemente do projeto ou fornecedor.	0,42
10	Quando houve falha na atuação hidráulica (588 registros), 249 delas ocorreram em válvulas VGP.	0,42
11	Em 547 falhas ocorridas em válvulas do projeto Fase 2, 229 são relativas a atuação hidráulica.	0,42

#### 4. CONSIDERAÇÕES FINAIS

Os conceitos de descoberta de conhecimento em banco de dados foram discutidos e a aplicação de suas técnicas em um banco de dados real foram realizadas. Foram utilizados alguns algoritmos disponíveis na ferramenta WEKA para identificar padrões de falha em válvulas submarinas em operação. Dessa forma, o objetivo do trabalho foi atingido.

O banco de dados utilizado não foi pensado na sua origem para ser utilizado na ferramenta WEKA ou outra ferramenta de DM. Dessa forma, foi necessário um tempo razoável para prepara-lo e transforma-lo no formato adequado. Além disso, das diversas informações coletadas, apenas quatro foram, ao final das análises, utilizadas. Portanto, poderia se pensar em outras informações a serem coletadas no campo trazendo mais atributos para serem processados computacionalmente.

A análise identificou que o sistema de indicação de posição é a falha que ocorre em maior proporção. Dependendo do tipo de válvula essa falha se manifesta independente de

fatores como fornecedor e tipo de projeto. A falha de perda de estanqueidade apresentou alta confiança, relativamente, para o fornecedor 4, provavelmente por serem os equipamentos mais tempo em operação.

Os resultados obtidos e apresentados confirmam o que as equipes de operação percebem no dia a dia sobre as falhas mais comuns. Desse modo, esse trabalho trouxe uma ferramenta e metodologia que se mostraram adequadas para extrair conhecimento de maneira sistemática que poderá auxiliar na identificação de itens críticos e na proposição de maneiras de contorná-los. Obtiveram-se como produto final árvores de decisão consideradas tecnicamente relevantes para gestão da melhoria dos equipamentos.

Para os atributos utilizados do banco de dados, o algoritmo de associação Predictive Apriori não gerou regras relevantes. Sendo assim, uma continuação desse trabalho seria utilizar outros algoritmos e atributos disponíveis no banco de dados e avaliar a relevância dos resultados.

## REFERÊNCIAS

AGGARWAL, S.; KAUR, R. Comparative study of various improved versions of apriori algorithm. **International Journal of Engineering Trends and Technology (IJETT)-Volume4Issue4-April**, 2013.

BAI, Y.; BAI, Q. Subsea engineering handbook. Waltham, MA: Gulf Professional Pub, 2012.

BHARGAVA, N. et al. Decision tree analysis on j48 algorithm for data mining. **Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering**, v. 3, n. 6, 2013.

CAMILO, C. O.; SILVA, J. C. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. 2009.

CHOPRA, S.; MENDEL, P. Gerenciamento da cadeia de suprimentos: estratégia, planejamento e operação. S. Paulo: Pearson Prentice Hall, 2011.

EUTHYMÍO, E. Metodologia para modelagem da evolução da probabilidade de falha decorrente da degradação funcional de válvulas. Rio de Janeiro: UFRJ, 2013.

GOLDSCHMIDT, R.; PASSOS, E. Data mining: um guia prático, conceitos, técnicas, ferramentas, orientações e aplicações. São Paulo: Elsevier; 2005.

KÖKSAL, G.; BATMAZ, İ.; TESTIK, M. C. A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications*, v. 38, n. 10, p. 13448–13467, set. 2011.

PATIL, T. R.; SHEREKAR, S. S. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. **International Journal of Computer Science and Applications**, v. 6, n. 2, p. 256-261, 2013.

REIS, R. X. DOS. A utilização do fenômeno Big Data na Administração Pública - a experiência do PENSA na Prefeitura Municipal do Rio de Janeiro. Trabalho de conclusão de curso—Rio de Janeiro: Escola de Contas e Gestão, 2015.

SILVA, E. L.; MENEZES, E.. A pesquisa e suas classificações. **Metodologia da pesquisa e elaboração de dissertação. Cap**, v. 2, n. 4, p. 19-25, 2005.