



A Mineração de Dados como ferramenta de apoio à tomada de decisão aplicada à área de gerenciamento de projetos de uma indústria

Carlos Adriano Arruda Teixeira

carlosadrianotex@hotmail.com

IFFluminense

Simone Vasconcelos Silva

simonevs@iff.edu.br

IFFluminense

Resumo: Durante o processo de gerenciamento de projetos um volume de dados considerável pode ser gerado, porém tais dados não têm valor se não forem estruturados e analisados apropriadamente, com intuito de gerar base de conhecimento para apoiar decisões futuras. O presente trabalho tem por objetivo apresentar um estudo de caso demonstrando a aplicação da mineração de dados através da técnica de classificação para análise de algumas características de projeto que podem gerar atraso na entrega. A análise possibilitou uma leitura sobre fatores que podem ter impacto na entrega de produtos, e gerou um entendimento sobre a relação destes fatores. Como resultado das análises foram evidenciadas algumas relações que não eram consideradas inicialmente e um consequente amadurecimento sobre os dados aplicáveis à tomada de decisão durante gerenciamento de projetos.

Palavras Chave: mineração de dados - projetos - tomada de decisão - classificação - conhecimento

1. INTRODUÇÃO

A grande quantidade de dados disponível no ambiente das organizações dificulta a obtenção da informação de forma rápida e eficiente sem o suporte de ferramentas para a realização de uma análise (AZARIAS *et al.*, 2009). Processar e interpretar as informações geradas pelas bases de dados atuais de forma correta está entre os requisitos essenciais para uma boa tomada de decisão (CARDOSO & MACHADO, 2008).

A partir do crescimento do volume de informações que as corporações manipulam, gera-se a necessidade urgente de técnicas e ferramentas que transformem dados em conhecimento útil de forma inteligente e automática. A solução para esta necessidade das organizações de obterem conhecimento de grandes volumes de dados está na utilização de técnicas de mineração de dados para extrair as informações implícitas existentes nos Bancos de Dados destas organizações (AGRAWAL *et al.*, 1993).

A mineração de dados (MD) é uma das alternativas mais eficazes para extrair conhecimento a partir de um grande volume de dados referentes (CARDOSO & MACHADO, 2008). MD é formado por um conjunto de ferramentas e técnicas que através do uso de algoritmos de aprendizagem ou classificação baseados em redes neurais e estatística, são capazes de explorar um conjunto de dados, extraíndo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento. Esse conhecimento pode ser apresentado por essas ferramentas de diversas formas: agrupamentos, hipóteses, regras, árvores de decisão e grafos (LAUDON & LAUDO, 2011).

É de extrema importância o sucesso de um projeto para criar mecanismos para identificar o conhecimento crítico, para que possam ser gerenciados. Desta forma, a extração de conhecimentos de base de dados de projetos torna-se evidente como um processo sistemático para transformar a informação e o conhecimento gerados durante os “projetos atuais” em valores para “projetos futuros”. Com a extração de conhecimento é possível a melhoria de processos, a redução de custos e o valor é agregado ao projeto (O’DELL ET HUBERT, 2011).

Com a premissa de que, para propiciar a gestão do conhecimento, é necessário que a organização adote práticas e ferramentas específicas, este artigo tem como objetivo a aplicação da mineração de dados como ferramenta para análise dos dados de projetos de uma organização do ramo industrial a fim de extrair conhecimento útil à tomada de decisão gerencial. Essa tomada de decisão refere-se aos conhecimentos necessários para subsidiar futuros projetos baseados nos conhecimentos relacionados à própria gestão atual dos projetos, principalmente ao que diz respeito ao tempo de entrega dos projetos e os possíveis atrasos.

A partir da introdução, esse artigo encontra-se dividido nas seções: Seção 2 aborda conceitos relacionados a mineração de dados, Seção 3 apresenta a metodologia proposta neste artigo, Seção 4 aborda os resultados obtidos e discussões, e Seção 5 traz as conclusões deste trabalho.

2. MINERAÇÃO DE DADOS (MD)

De acordo com Fayyad *et al.* (1996), a extração de dados que sejam relevantes aos interesses de um gestor no contexto de gerenciamento de projetos, tem se tornado cada vez mais complexa devido a quantidade de dados armazenados. Denomina-se Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* – KDD), a atividade de garimpar a informação contida nestes dados (ASSMANN *et al.*, 2012). Apesar de ser comum usar os termos KDD e Mineração de Dados com o mesmo significado, Fayyad *et al.* (1996) definem o KDD (Figura 1) como sendo o processo da extração de conhecimento dos

dados como um todo, e Mineração de Dados, como ferramenta de KDD, considerando que a extração de padrões dos dados é realizada através da aplicação de algoritmos específicos.

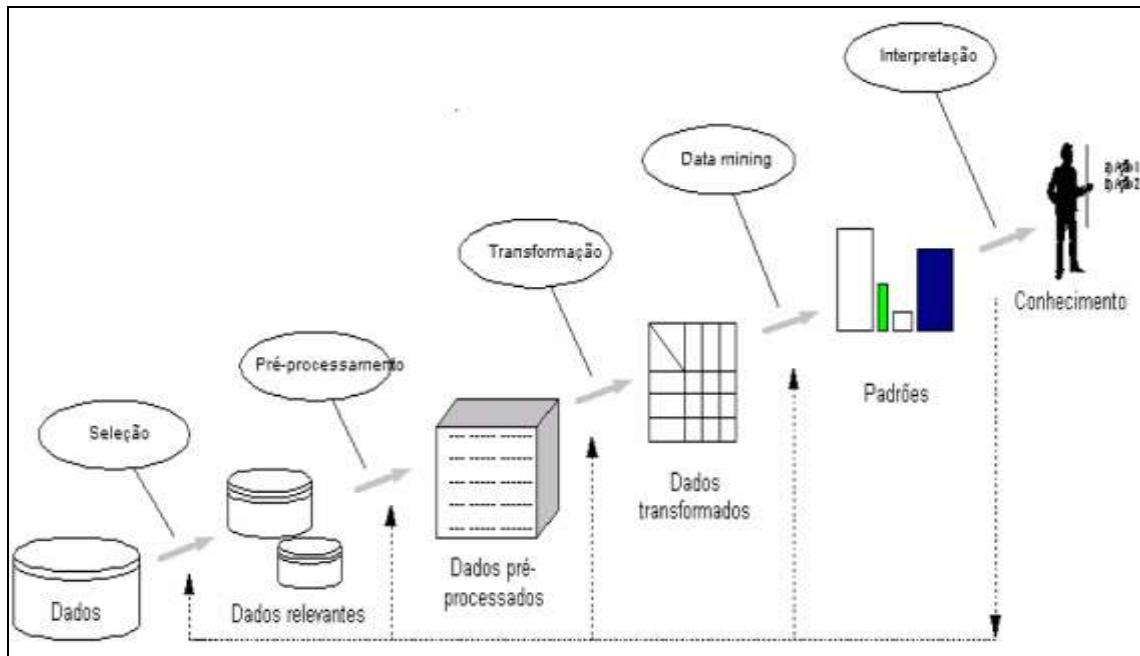


Figura 1: Processos do KDD

Fonte: Fayyad (1996)

De acordo com Pampa (2003), as etapas do KDD são:

- Seleção: nesta etapa destaca-se o conhecimento prévio dos objetivos de um usuário final e o conhecimento do domínio, e nesta etapa é preciso compreensão do domínio, conhecimento dos dados relevantes, definição do problema em termos de domínio, definição de objetivos da aplicação e metas específicas, e definição do modelo da solução desejada;
- Pré-Processamento: atividades que visam gerar um grupo de dados representativos convenientemente organizados e estruturados para ser minerados pelo algoritmo selecionado, e esta etapa é composta por integração de dados (múltiplas fontes de dados heterogêneos podem ser integradas em um único BD), seleção de dado (identifica um subconjunto de atributos onde será minerado, facilitando o trabalho dos algoritmos de mineração), e limpeza de dados (estratégia adequada de manipulação de dados ruidosos, errôneos, perdidos ou irrelevantes);
- Transformação de dados: os dados são transformados ou consolidados em formatos apropriados para minerar, e consiste em discretizar dados (os valores contínuos dos atributos são divididos numa lista de intervalos, convertendo valores contínuos em valores discretos), com isso obtém uma melhora da compreensão do conhecimento, redução do tempo de processamento, diminuição do espaço de busca, facilitação do algoritmo de tomada de decisões, agregação de dados (agrega dados existentes nas informações de modo que essas agregações contribuam no processo de descoberta de conhecimento), derivação de dados (adicionam-se novos dados derivados por uma

operação ou por séries de operações de dados existentes na tabela de dados), e redução de dados (reduz-se o número de variáveis a considerar);

- Processamento: os dados são analisados por um algoritmo e transformados em informações (resultados, padrões) úteis que serão avaliados no processo seguinte e consiste em escolha da função de mineração de dados, seleção de algoritmo de mineração de dados, e Data Mining (análise dos dados armazenados através de um programa computacional capaz de analisar os dados e encontrar padrões de interesse);
- Pós-Processamento: são feitas a avaliação e interpretação dos padrões, para serem representados em forma de conhecimento comprehensível e confiável ao usuário, e para serem incorporadas ao conhecimento anterior e consiste em interpretação e avaliação dos resultados (identifica os padrões interessantes que representam o conhecimento baseado em algumas medidas de interesse), apresentação e incorporação de conhecimento (são usadas técnicas de representação e visualização para apresentar o conhecimento de forma comprehensível ao usuário).

É importante ressaltar também uma metodologia aplicada a mineração de dados conhecida como CRISP-DM, a qual é descrita em termos de um modelo de processo hierárquico, que consiste em conjuntos de tarefas descritas em quatro níveis de abstração, do geral para o específico: a fase, a tarefa genérica, tarefas especializadas e instância de processo (CRISP-DM, 2020). Conforme mostra a Figura 2.

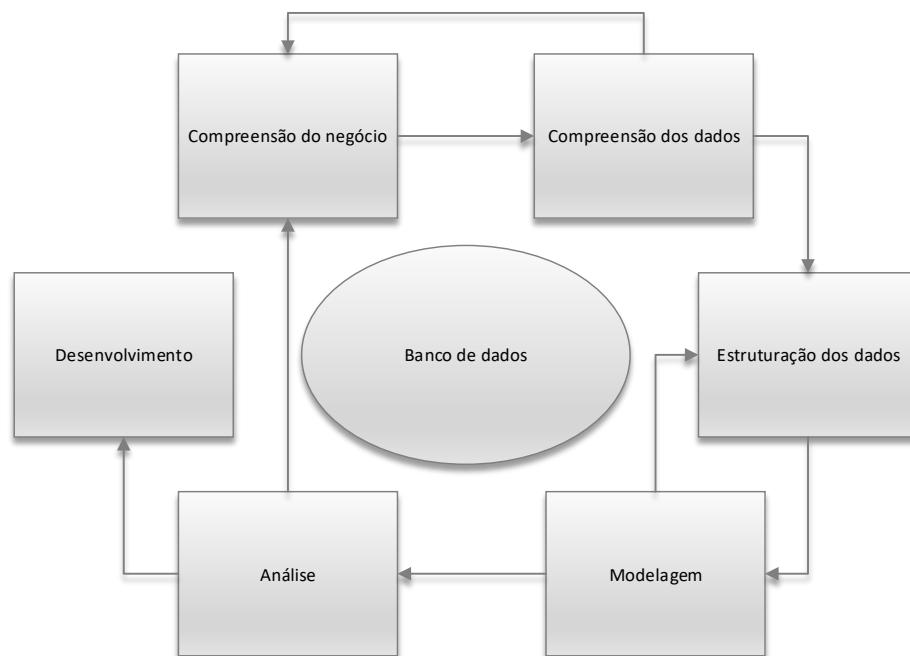


Figura 2: Fases do modelo de referência CRISP-DM

Fonte: Adaptado de CRISP-DM (2020).

De acordo com Pasta (2011), as fases apresentadas na Figura 2 possuem como tarefas: a elaboração de plano de distribuição, a criação de um plano de monitoramento e manutenção, a elaboração do relatório final e por último a revisão do projeto. CRISP-DM foi projetado para fornecer orientação para os iniciantes em mineração de dados e para fornecer um modelo de processo genérico que pode ser especializado de acordo com as necessidades de qualquer

ramo de atividade ou da organização. A mineração de dados pode ser desenvolvida de modo não-sistemático, sem que haja nenhum cuidado em seu desenvolvimento, o que não é recomendado, pois acarreta em resultados não esperados ou intuito de evitar este tipo de situação o uso de uma metodologia vem a garantir que o processo da MD seja desenvolvido de modo sistemático e padronizado, o que acarretará em resultados precisos e confiáveis.

Ainda de acordo com Pasta (2011), a metodologia CRISP-DM tem seu sucesso devido ao fato de ter sido desenvolvida à prática, não estar atrelada a nenhuma ferramenta específica de mineração de dados, mas sim a junção das melhores práticas que são utilizadas em um projeto de mineração de dados, aliada ao fato de atuar sobre todo o processo de mineração de dados.

3. METODOLOGIA

Com relação à metodologia, o método utilizado na pesquisa é classificado como estudo de caso (YIN, 2005). A metodologia proposta neste trabalho foi dividida nas quatro etapas seguintes:

- Etapa I – Revisão Bibliográfica sobre mineração de dados;
- Etapa II – Definição da base de dados para o Estudo de Caso: a base de dados selecionada para a aplicação da mineração de dados pertence a uma fábrica de dutos flexíveis voltada para indústria Off-Shore. Foram coletados dados de projetos que iniciaram a fase de fabricação em 2016 com previsão de entrega em 2017 com um total de 57 registros;
- Etapa III – Aplicação das etapas do KDD na base de dados do estudo de caso: para atingir os objetivos deste estudo, foram definidas atividades com base no processo KDD. A escolha das técnicas de Data Mining depende fundamentalmente do objetivo do processo de KDD e segundo Berry e Linoff (2000) as técnicas podem ser classificação, agrupamento ou associação. Neste trabalho o objetivo se atém na classificação, ou seja, distinguir exemplos de características das etapas do projeto onde houve atraso na entrega do produto em relação a data de entrega prevista em contrato. A classificação consiste na previsão de uma variável categórica, ou seja, para descobrir uma atividade que irá mapear um conjunto de registros em um conjunto de variáveis predefinidas chamadas classes. Esta atividade pode ser aplicada a novos registros, de modo a prever a classe em que esses registros se enquadra (ASSMANN *et al.*, 2012). Segundo Pasta (2011), para prever se um projeto tem potencial ou não de atrasar, algumas relações devem ser estabelecidas através de dados que já constam disponibilizados na base de dados da organização foco da análise. A partir desses dados, um algoritmo de classificação pode descobrir regras que preveem se um projeto pode ou não atrasar. Para a aplicação da mineração de dados foi utilizado a ferramenta Weka (WEKA, 2020);
- Etapa IV – Análise dos Resultados: aborda os resultados obtidos através da mineração de dados da base de dados dos projetos da fábrica através do uso da técnica de classificação. Os dados e gráficos foram gerados através da ferramenta Weka.

4. RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados obtidos na aplicação da metodologia proposta.

4.1. SELEÇÃO E MODELAGEM DOS DADOS

A primeira etapa do processo KDD consiste na seleção e modelagem dos dados que serão utilizados (ASSMANN *et al.*, 2012). Nesta etapa é essencial a participação dos *stakeholders* do negócio, que tenham conhecimento relevante acerca do problema abordado. Dentro deste contexto, foi apresentado ao departamento de projetos e de engenharia da fábrica as possíveis variáveis que seriam relevantes ao estudo, assim como foram explicados conceitos relacionados à mineração de dados. Nesta oportunidade foi possível identificar fatores de fabricação que estão disponíveis em base de dados e que podem contribuir para uma avaliação técnica com possibilidade de ganho de conhecimento em relação a fatores que impactam em prazos de entrega segundo as datas previstas em contrato de fornecimento. Desta forma, foi elaborada uma lista de variáveis que serão utilizadas na mineração de dados, conforme a Tabela 1.

Tabela 1: Lista de variáveis para a mineração de dados

Grau	Descrição
Quantidade de pedidos	Número de pedidos atrelado ao contrato
Tipo de contrato	Identificação do número de contrato atrelada aos projetos
Relevância	A relevância foi classificada de 1 à 9 onde 1 é o menos relevante e 9 o mais relevante em termos de prazo de entrega.
Aplicação	Na área de dutos flexíveis voltados para o mercado de óleo e gás, a aplicação foi dividida entre estática e dinâmica apenas
Volume (m)	O comprimento final do produto a ser entregue
Não conformidades	Se durante as etapas de manufatura de um projeto houveram ou não ocorrências de não conformidades
Tempo de fabricação (meses)	A duração da fabricação de cada produto especificamente
Ocorrência de atraso	Se houve atraso na entrega do produto para o cliente em relação a data de entrega contratual

Fonte: Adaptado de Assmann *et al.* (2012).

Para o desenvolvimento deste trabalho foram inicialmente consideradas 8 variáveis (Tabela 1) para cada um dos 57 casos abordados neste artigo, sendo que já se conhecia todo o histórico de entrega de pedidos até o momento, porém para efeitos de estudo, foi determinado que apenas uma amostra de dados seria suficiente para utilização como base de estudo. A variável “Relevância” é classificada de acordo com o grau de relevância em relação ao prazo de entrega, este grau varia de 1 (menos) a 9 (mais), conforme a Tabela 2.

Tabela 2: Descrição do grau de relevância

Grau	Descrição
1	Pouco relevante: prazo não é importante para o cliente, impacto muito baixo em custo de projeto baixo.
2	Pouco relevante: prazo não é importante, custo não é muito alto.
3	Relevante: Prazo é importante, custo médio.
4	Relevante: Prazo é importante, custo médio.
5	Relevância média: Prazo é importante, custo médio.
6	Relevância considerável: Prazo curto, custo considerável.
7	Muito relevante: Prazo urgente, alto custo.
8	Muito relevante: Prazo urgente, alto custo, alta demanda.
9	Extremamente relevante: Prazo urgente, altíssimo custo, alta demanda, aplicação de multa sobre atraso.

Para a mineração de dados, uma tabela foi construída com a informações extraídas de bases de dados internas do departamento de projetos da empresa alvo deste estudo de caso. Alguns dados foram tratados genericamente, porém as variações relacionadas a tempo de projeto correspondem a produções reais de linhas de dutos flexíveis. A Figura 3 ilustra a base de dados utilizada.

Quantidade de pedido	Tipo de contrato	Relevância	Aplicação	Volume (m)	NCR	Tempo de fabricação (meses)	Atrasado?
5	Contrato tipo G	5	Dinamica	875	1	10	Sim
5	Contrato tipo G	5	Dinamica	1075	1	10	Sim
5	Contrato tipo G	4	Estatica	1000	0	11	Sim
5	Contrato tipo G	2	Estatica	885	0	11	Sim
3	Contrato tipo B	1	Dinamica	200	0	8	Sim
5	Contrato tipo G	5	Dinamica	400	1	8	Sim
8	Contrato tipo J	6	Estatica	1387	0	9	Sim
8	Contrato tipo J	8	Estatica	740	0	8	Não
8	Contrato tipo J	1	Estatica	850	0	8	Sim
8	Contrato tipo J	6	Estatica	850	0	8	Sim
4	Contrato tipo E	7	Estatica	1200	0	12	Não
4	Contrato tipo E	9	Estatica	3256	0	16	Não
4	Contrato tipo E	6	Dinamica	1051	0	11	Sim
4	Contrato tipo E	6	Dinamica	1800	1	13	Sim

Figura 3: Banco de dados utilizado na modelagem do trabalho

4.2. PRÉ-PROCESSAMENTO E TRANSFORMAÇÃO DOS DADOS

Durante o desenvolvimento do processo KDD, por diversas vezes houve a necessidade de voltar para a etapa de pré-processamento. Os dados repetidos, por exemplo, só foram identificados na etapa de transformação dos dados, sendo necessário realizar este ajuste para continuar o processo.

Com os dados armazenados no modelo e já pré-processados, foi necessário desenvolver uma rotina para exportá-los para um formato que a ferramenta adotada interprete. Os arquivos de entrada da ferramenta Weka devem possuir um formato específico, contendo um cabeçalho onde são informados todos os atributos do processo de mineração, sendo que todos devem estar categorizados, ou seja, todos os seus valores possíveis devem estar explícitos no arquivo. Por exemplo, o atributo `encerrado_atrasado` pode assumir os seguintes valores: Sim ou Não, sendo que foram definidas as siglas S e N e estas foram adicionadas no cabeçalho do arquivo.

A extensão deste arquivo é ARFF (*Attribute Relation File Format*). Este pode ser criado em qualquer editor de textos, inclusive no bloco de notas. Entretanto, requer alguns cuidados em relação ao seu formato interno, pois o texto possui cabeçalho específico e uma área para os dados. Na Figura 4 o cabeçalho e uma parte dos dados podem ser visualizados.

Similarmente ao estudo executado por Pasta (2011), na primeira linha do arquivo código fonte que será utilizado como input para a ferramenta Weka, conforme representado pela Figura 4, e é referente a entidade que representa os dados. Em sequência são declarados os atributos da base de dados que serão abordados durante o processo de mineração. Um vetor é construído a partir das informações categorizadas para a aplicação das regras de classificação, do contrário o WEKA não permite executar o algoritmo.

```

@RELATION banco

@ATTRIBUTE 'Quantidade de pedidos' real
@ATTRIBUTE 'Tipo de contrato' { 'Contrato tipo A', 'Contrato tipo B', 'Contrato tipo C', 'Contrato tipo D', 'Contrato tipo E', 'Contrato tipo F', 'Contrato tipo G', 'Contrato tipo H', 'Contrato tipo I', 'Contrato tipo J', 'Contrato tipo K' }
@ATTRIBUTE Relevancia { 1, 2, 3, 4, 5, 6, 7, 8, 9 }
@ATTRIBUTE Aplicacao { 'Dinamica', 'Estatica' }
@ATTRIBUTE Volume real
@ATTRIBUTE NCR { 0, 1 }
@ATTRIBUTE 'Tempo de fabricacao' real
@ATTRIBUTE Atrasado { Sim, Nao }

@data
5,'Contrato tipo G',5,'Dinamica',875,1,10,'Sim'
5,'Contrato tipo G',5,'Dinamica',1075,1,10,'Sim'
5,'Contrato tipo G',4,'Estatica',1000,0,11,'Sim'
5,'Contrato tipo G',2,'Estatica',885,0,11,'Sim'
3,'Contrato tipo B',1,'Dinamica',200,0,8,'Sim'
5,'Contrato tipo G',5,'Dinamica',400,1,8,'Sim'
8,'Contrato tipo J',6,'Estatica',1387,0,9,'Sim'
8,'Contrato tipo J',8,'Estatica',740,0,8,'Nao'
8,'Contrato tipo J',1,'Estatica',850,0,8,'Sim'
8,'Contrato tipo J',6,'Estatica',850,0,8,'Sim'
4,'Contrato tipo E',7,'Estatica',1200,0,12,'Nao'
4,'Contrato tipo E',9,'Estatica',3256,0,16,'Nao'

```

Figura 4: Exemplo do arquivo ARFF utilizado como input para a ferramenta Weka

4.3. MINERAÇÃO DE DADOS UTILIZANDO O WEKA

A próxima etapa consiste na aplicação dos algoritmos PART e JRIP para classificação e criação de regras. A Figura 5 apresenta exemplos de relações geradas pelo algoritmo PART que estão associadas ao atrasado de entrega e na Figura 6 são apresentados exemplos de relações geradas pelo algoritmo JRIP que estão associadas ao atrasado de entrega são apresentadas.

```

PART decision list
-----
Tipo de contrato = Contrato tipo K: Sim (9.0)

Tipo de contrato = Contrato tipo J: Sim (8.0/1.0)

Tipo de contrato = Contrato tipo H: Nao (6.0)

Relevancia = 6: Sim (6.0)

Relevancia = 4: Sim (4.0)

Aplicacao = Dinamica: Sim (14.0/3.0)

Quantidade de pedidos <= 4: Nao (6.0)

: Sim (4.0/1.0)

Number of Rules : 8

```

Figura 5: Exemplo de regras geradas pelo algoritmo PART na ferramenta Weka

```

JRIP rules:
=====
(Volume >= 1883) => Atrasado=Nao (11.0/2.0)
(Tipo de contrato = Contrato tipo H) => Atrasado=Nao (3.0/0.0)
=> Atrasado=Sim (43.0/5.0)

Number of Rules : 3

```

Figura 6: Exemplos de regras geradas pelo algoritmo JRIP na ferramenta Weka

4.4. ANÁLISE DOS RESULTADOS

Definida como última etapa do processo de mineração de dados e considerada de extrema importância, a análise dos resultados consiste em interpretar os dados gerados em conjunto com os especialistas do negócio. Esta etapa é essencial para validar se os objetivos foram alcançados, verificando se foi possível gerar conhecimento novo e útil para o processo de tomada de decisão.

Como resultado do processamento dos dados, alguns gráficos foram elaborados, onde é possível observar algumas relações entre as variáveis dos projetos. Através da Figura 7 é demonstrado que quanto maior a prioridade do projeto, ou seja, quanto maior a relevância do projeto para o cliente em relação ao prazo de entrega, menor é a chance de atrasos de acordo com o modelo estratégico atual da fábrica. A Figura 7 mostra a relação da relevância de acordo com o volume produzido.

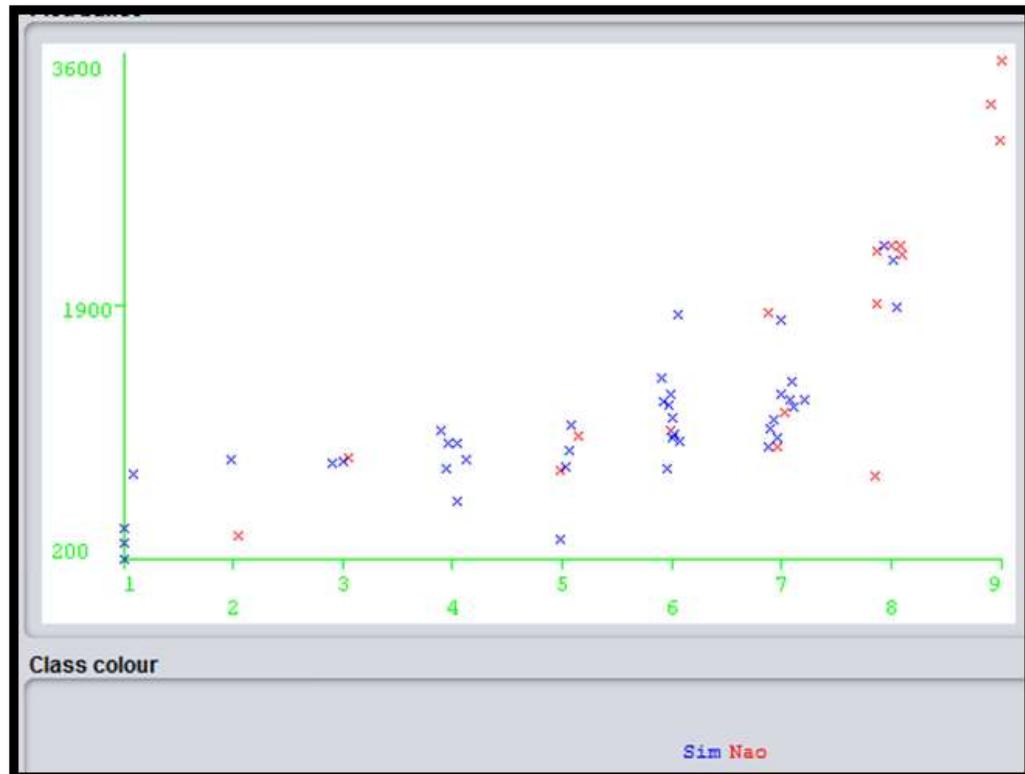


Figura 7: Relação gerada na ferramenta Weka, relevância (eixo x) X volume produzido (eixo y).

A Figura 8 mostra a relação da quantidade de pedidos de acordo com o volume produzido.

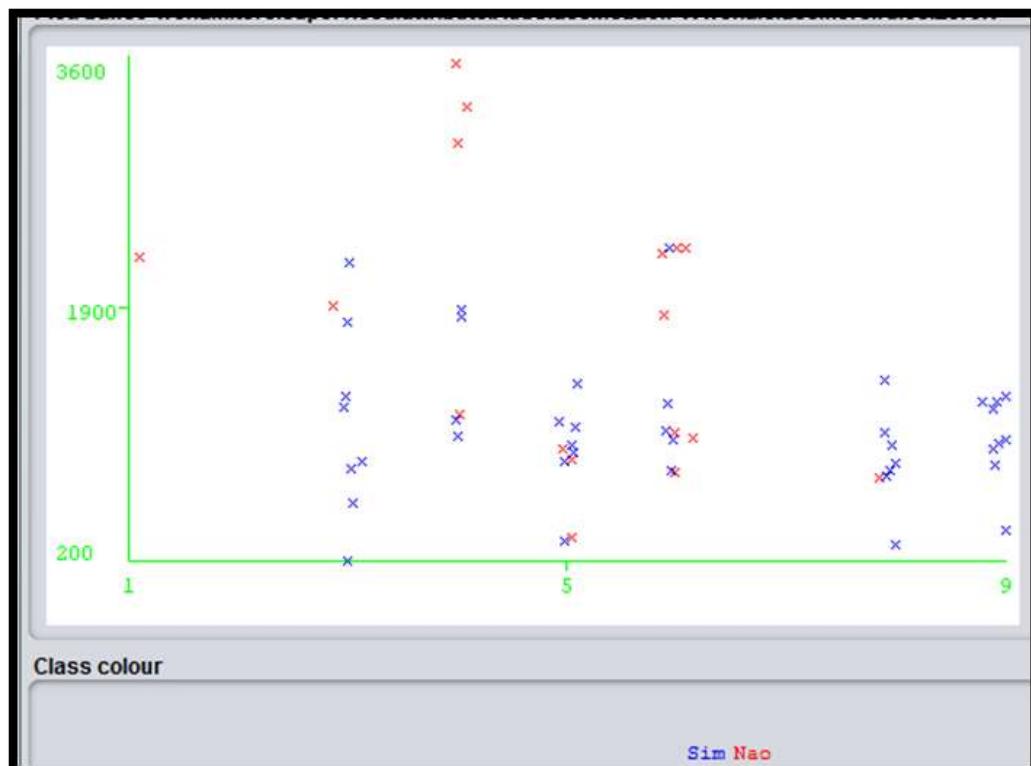


Figura 8. Relação gerada na ferramenta Weka, quantidade de pedidos (eixo x) X volume produzido (eixo y).

Outra análise relevante foi em relação ao turno, pois a partir da sua análise foi possível dimensionar melhor os turnos dos técnicos e equipes de acordo com período que são mais demandados. O turno da noite por ser o menos requisitado, não requer mais técnicos disponíveis. A média de 1,4 técnicos por noite (normalmente 1 técnico e para duas noites são 2 técnicos) é suficiente em dias normais. Somente quando ocorre algum problema, mais técnicos alongam os seus turnos para resolvê-los. O turno de maior demanda de cada equipe é uma informação importante e que pode influenciar nas definições dos turnos das reuniões semanais, com o intuito de marcar reuniões no turno que abrem menos chamados que são de sua responsabilidade.

5. CONCLUSÃO

A contribuição deste trabalho foi apresentar um estudo de caso com base em situações similares a de uma situação real em uma fábrica. A aplicação da mineração de dados como técnica para suporte a tomada de decisão no departamento de projetos foi considerada relevante. Os resultados apresentados neste trabalho evidenciam que a utilização de uma ferramenta capaz de descobrir novas informações em banco de dados pode ser um diferencial para uma organização.

É importante destacar que a ferramenta de mineração de dados não substitui o papel do gestor e dos especialistas do negócio. Para ocorrerem descobertas de conhecimento a partir dos dados disponibilizados pela ferramenta Weka, são necessárias constantes análises e discussões. Desta forma, possibilita que a mineração de dados retorne bons resultados e a organização crie realmente um diferencial, por conseguir tratar suas informações e estas darem suporte à

tomada de decisão dos gestores. Uma organização que valoriza os conhecimentos gerados permite a construção de sua memória estratégica, pois os conhecimentos quando armazenados podem ser recuperados sempre que necessário.

Em relação à contribuição para a área de gestão de projetos, o nível de detalhamento adotado neste artigo, com base em exemplos reais, permite que outras pesquisas adotem as técnicas utilizadas para atender ao processo KDD, sendo o mesmo utilizado como um guia. Com os projetos em gestão do conhecimento as organizações alcançam a capacidade de transformar os dados disponíveis, mas de forma descentralizada, em informação e, a partir da aplicação de técnicas e análise de especialistas, em conhecimento de valor agregado, que dão apoio à tomada de decisão no negócio.

6. REFERÊNCIAS

- AGRAWAL R.; IMIELINSKI T. & SWAMI A.** Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, Washington, DC, p. 207-216. 1993.
- ASSMANN C. L.; FROZZA R. & KIPPER L. M.** Selection of attributes for use in Data Mining as support to decision-making in the information technology área. In: Proceedings of the IADIS European Conference Data Mining, Lisboa, p. 215-218. 2012.
- AZARIAS P.; MATOS S. N. & SCANDELARI L.** Aplicação da Mineração de Dados para a Geração do Conhecimento: um experimento prático. In: V Congresso Nacional de Excelência em Gestão – Gestão do conhecimento para a sustentabilidade, Niterói, p. 101-119. 2009.
- BERRY J. A. M. & LINOFF G.** Mastering Data Mining. Jonh Wiley & Sons, Inc., 2000.
- CARDOSO O. N. P. & MACHADO R. T. M.** Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. Revista de Administração Pública, Rio de Janeiro, p. 495-528. 2008.
- CRISP-DM.** Cross Industry Standart Process for Data Mining. 2020. Disponível em: <http://crisp-dm.eu>.
- FAYYAD, U.** Advances in knowledge discovery and data mining. Cambridge: MIT Press, 1996.
- FAYYAD, U; PIATETSKY-SHAPIRO, G. & SMYTH, P.** From Data Mining to Knowledge Discovery: An Overview. Menlo Park, CA: AAAI Press/The MIT Press, 1996.
- LAUDON. K. & LAUDO, J.** Sistemas de Informações Gerenciais: Fundamentos da inteligência de negócios/gestão da informação e de banco de dados. 9º ed. São Paulo, 2011.
- O'DELL, C. ET HUBERT, C.** The New Edge in Knowledge: How Knowledge Management Is Changing the Way We Do Business. Wiley & Sons, New Jersey, 2011.
- PAMPA N. R.** Técnicas e Ferramentas automáticas e inteligentes para a extração de conhecimento em bases de dados. Dissertação de tese de mestrado. DSIF, FEEC, UNICAMP. 2003.
- PASTA, A.** Aplicação da técnica de Data Mining na base de dados do ambiente de gestão educacional: um estudo de caso de uma instituição de ensino superior de Blumenau - SC. Dissertação de Mestrado, UNIVALI, São José, 2011.
- WEKA.** Ferramenta Weka. University of Waikato: Software. 2020. Disponível em: <http://www.cs.waikato.ac.nz/~ml/weka>.
- YIN, R.** Estudo de Casos: Planejamento e Métodos. Bookman, 3ª edição. 2005.