
Kira: Uma Ferramenta Instrucional para Apoiar a Aplicação do Processo de Mineração de Dados

Eduardo Fernando Mendes, Marina Teresa Pires Vieira

Faculdade de Ciências Exatas e da Natureza – Universidade Metodista de Piracicaba
Piracicaba – SP – Brazil

eduardo@mococa.com, mtvieira@unimep.br

Abstract. *The available data mining tools require considerable knowledge of the user to perform the entire process of knowledge discovery in databases (KDD). Teaching how to prepare the data, how to choose the task of mining data appropriate to the purpose intended and analyse the generated rules are beyond the objectives of the current tools. This paper proposes several guides to abstract and facilitate the most important part of required knowledge to run the data mining process for mining association rules. Based on this guides it was implemented a tool named Kira that instructs the users to easily use the KDD process. This tool was validated through a case study applied in the company Metalúrgica Mococa S/A.*

Resumo. *As ferramentas de mineração de dados existentes não abstraem o conhecimento exigido para realizar o processo de mineração de dados. Ensinar como preparar os dados, como escolher a tarefa de mineração de dados adequada ao objetivo pretendido e como analisar as regras geradas transcendem os objetivos das ferramentas atuais. Este artigo propõe um conjunto de guias para abstrair as partes mais importantes do conhecimento necessário para executar os passos do processo de mineração de dados para a tarefa de regras de associação. Com base nesses guias foi implementada uma ferramenta chamada Kira, que instrui a realizar mais facilmente o processo de mineração de dados. A ferramenta foi validada através de um estudo de caso aplicado na empresa Metalúrgica Mococa S/A.*

1. Introdução

A mineração de dados tem atraído muito a atenção da indústria da informação devido à grande disponibilidade de dados e à crescente necessidade de transformar dados em conhecimentos úteis. Relatos de casos de sucesso têm comprovado a utilidade dos algoritmos de mineração de dados para encontrar padrões interessantes em conjuntos de dados. Porém o uso adequado dessa tecnologia exige um conhecimento abrangente do engenheiro de dados.

O processo envolvido na mineração de dados (referido como KDD) é complexo, sendo que vários conceitos e metodologias são utilizados, exigindo do usuário grande conhecimento do arsenal teórico que apoia o processo. Mesmo tendo disponíveis muitas ferramentas para diferentes tarefas de mineração, acima de 200 em 1999 (PIATETSKY-SHAPIRO, 2000), estas não eliminam a necessidade de conhecimento por parte do usuário.

A metodologia CRISP-DM (CHAPMAN, 2000) objetiva ser um processo padrão para auxiliar a definir, desenvolver e implementar um projeto de mineração de dados. Essa metodologia oferece detalhes para o desenvolvimento do processo de mineração de dados, porém é necessário que o usuário tenha conhecimento de aspectos envolvidos no processo.

Com a preocupação em auxiliar no aprendizado do processo de mineração de dados, foi desenvolvida a ferramenta Kira, que propõe um conjunto de guias para orientar o usuário na execução desse processo. Esses guias abstraem grande parte do conhecimento exigido do usuário para executar o processo de mineração de dados, centrado em regras de associação.

Para o desenvolvimento da Kira foram analisadas as ferramentas Oracle Data Mining, que tem como diferencial a metodologia Data-Centric Automated Data Mining (CAMPOS; STENGARD; MILENOVA, 2005), Weka (WITTEN; FRANK 2005), Dbminer (HAN et al, 1996), e a Bramining (GOLDSCHMIDT; PASSOS 2005), visando verificar o grau de conhecimento exigido do usuário para realizar o processo de mineração. Essas ferramentas tem como foco principal o tratamento e a visualização dos dados e não contemplam recursos instrucionais sobre a execução das etapas do processo de KDD.

O restante do artigo está organizado como segue: A seção 2 apresenta um breve resumo sobre o processo de KDD e mineração de dados com base em regras de associação. A seção 3 propõe um conjunto de guias para auxiliar o usuário a executar o processo de KDD e sua automatização na ferramenta Kira. A seção 4 apresenta um estudo de caso, utilizando a ferramenta Kira. Finalmente a seção 5 apresenta as conclusões.

2. Processo de KDD e Mineração de Dados

O processo de descoberta de conhecimento em bases de dados foi definido por Fayyad, Piatetsky-Shapiro, Smith (1996) como “o processo não trivial de identificar padrões válidos, novos, potencialmente úteis e compreensíveis, nos dados”. É composto de sete etapas: limpeza, integração, seleção e transformação dos dados, mineração dos dados, avaliação dos padrões e apresentação do conhecimento. As quatro primeiras etapas costumam ser chamadas de preparação dos dados. As etapas do processo de KDD envolvem técnicas apropriadas e diferentes abordagens dependendo do tipo de conhecimento que se busca nos dados. Para sua execução o usuário precisa ter muito conhecimento de como preparar os dados e sobre as técnicas de mineração. Este, talvez, seja o grande problema da mineração de dados não ter sido adotada em grande escala pelas empresas, devido à escassez, no mercado, de profissionais com tanto conhecimento acumulado.

Fayyad, Piatetsky-Shapiro, Smith (1996) definem a mineração de dados como uma etapa do processo de KDD, que consiste na aplicação de algoritmos particulares, sob limitações de eficiência computacional aceitável, produzindo um conjunto de padrões sobre os dados. Entre as principais tarefas de mineração de dados estão as Regras de Associação que buscam encontrar padrões frequentes, que são aqueles que aparecem frequentemente em um conjunto de dados. Apesar de ser uma etapa do processo de KDD, os termos KDD e mineração de dados têm sido usados como sinônimos (HAN; KAMBER, 2006).

O problema de minerar regras de associação foi abordado pela primeira vez em (AGRAWAL; IMIELINSKI; SWAMI, 1993), apresentando uma técnica para encontrar relacionamentos entre a ocorrência de itens em transações de uma base de dados. Formalmente, a mineração de regras de associação é definida como segue (AGRAWAL; SRIKANT, 1994). Seja $I = \{i_1, \dots, i_n\}$ um conjunto de literais, denominados itens. Um conjunto $X \subseteq I$ é chamado de **itemset**. Seja D um conjunto de transações T , tal que $T \subseteq I$. Diz-se que a transação T contém um **itemset** X se $X \subseteq T$. Uma regra de associação é uma expressão da forma $X \Rightarrow Y$, onde X e Y são itemsets de I e $X \cap Y = \emptyset$. Para um banco de dados que contém os itens de compras de clientes, essa associação pode representar que se um

cliente comprar X, ele também estará propenso a comprar Y. Para as regras encontradas, duas medidas de interesse são utilizadas, suporte e confiança. A medida de **suporte** da regra é a razão entre o número de transações em *D* que contém X e Y, e o número total de transações de *D*. A medida de **confiança** é a razão entre o número de transações que contém X e Y, e o número de transações que contém X. A meta das tarefas de associação é gerar todas as possíveis regras que obedecem a limites mínimos de suporte e confiança estabelecidos pelo usuário.

3. Técnica para Auxílio à Aplicação do Processo de Mineração de Dados

Nesta seção são apresentados os recursos oferecidos pela ferramenta Kira para auxiliar a execução do processo de mineração de dados, reduzindo o nível de complexidade envolvido e tornando essa tarefa mais intuitiva. Este objetivo é alcançado através do uso de guias para orientar o usuário em cada etapa do processo. Esses guias foram identificados após estudos do processo de KDD e da metodologia CRISP-DM e realização de experimentos práticos usando ferramentas de mineração de dados.

A arquitetura da ferramenta Kira (figura 1) é composta de três módulos principais: Apoio à Origem, Apoio à Preparação de Dados e Apoio à Análise. O Módulo de Apoio à Preparação de Dados é composto por módulos para orientar: a Integração, Limpeza, Seleção e Transformação dos dados. Cada módulo objetiva oferecer uma série de facilidades para auxiliar o usuário a preparar os dados, aplicar os algoritmos de mineração e avaliar os resultados obtidos.

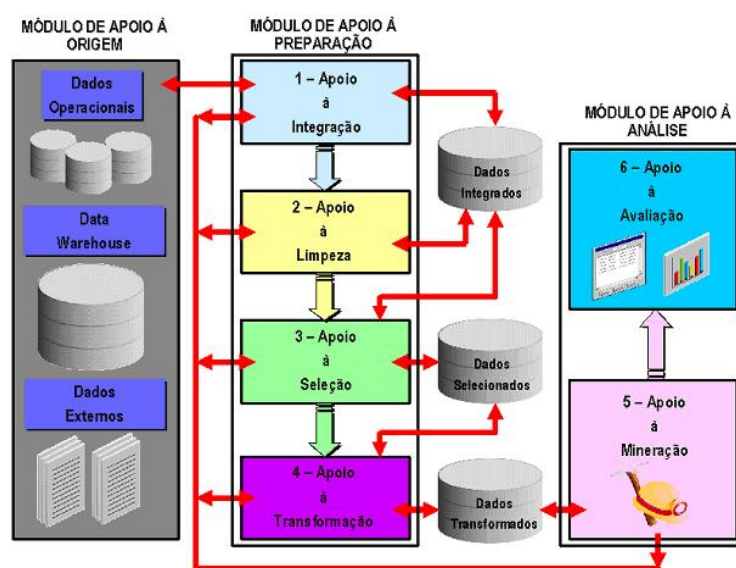


Figura 1 – Arquitetura da Ferramenta Kira

Ao entrar no ambiente da ferramenta Kira, inicialmente o usuário tem acesso às funcionalidades para parametrizar a fonte de dados que será utilizada para a mineração de dados (Módulo de Apoio à Origem). Após a parametrização dos dados o usuário é orientado gradativamente a fornecer determinadas informações, escolher e moldar os dados de interesse, escolher e aplicar um algoritmo de mineração de dados e analisar o resultado obtido, atividades estas embasadas em um conjunto de guias. Ao completar cada etapa, a ferramenta informa qual é a próxima etapa do processo, conduzindo o usuário de forma que ele consiga executar as tarefas, mesmo que não tenha conhecimento detalhado sobre as etapas do processo de KDD.

Nas próximas seções são apresentados os guias e algumas interfaces de usuário da ferramenta Kira, que implementam esses guias. Mais detalhes dos recursos oferecidos pela ferramenta podem ser encontrados em (MENDES, 2009). Para ilustrar as discussões é utilizada uma aplicação sobre inscrições para um congresso científico. Os dados dessa aplicação acompanham a ferramenta e servem de exemplo para aprendizagem do usuário.

3.1. Apoio à Preparação dos Dados

Para realizar a etapa de preparação dos dados é necessário inicialmente que seja identificada a tarefa de mineração de dados que será utilizada, uma vez que, para cada tarefa de mineração os dados devem ser preparados de uma forma específica. Antes, porém, é necessário que o usuário tenha identificado o problema e o objetivo da mineração, enfoque dos dois primeiros guias propostos.

Guia 01 – Induz o usuário a definir o *problema* de mineração de dados.

Definir o que se pretende fazer com a mineração de dados, com certeza, é o ponto de partida do processo de KDD. Para alcançar esse objetivo, o guia solicita que o usuário informe o problema a ser tratado pela mineração, para o seu ambiente de negócio, induzindo-o a pensar no problema. No exemplo considerado, o problema identificado foi: *A divulgação do congresso gera custo alto e nem sempre é eficiente* (parte 1 da interface da figura 2).

Guia 02: Induz o usuário a definir o *objetivo* que deseja cumprir para resolver o problema de mineração de dados.

A proposta desse guia é induzir o usuário a pensar, no início do processo, no objetivo que deve ser cumprido para a resolução do problema de mineração de dados e é mostrado na parte 2 da interface da figura 2, na qual o seguinte objetivo foi inserido: *Definir a melhor forma de divulgação do congresso conforme o perfil do congressista*. Com o exemplo, o usuário tem condições de avaliar o que é para ser informado.

Guia 03: Auxilia o usuário na identificação da *tarefa de mineração* que será usada para resolver o problema e a alcançar o objetivo de mineração.

Dentro do processo de KDD a etapa de identificação da tarefa de mineração de dados exige que o usuário saiba diferenciar as tarefas de mineração, o que não é trivial. É necessário que ele tenha conhecimento sobre as principais características de cada tarefa de mineração para identificar qual deverá ser aplicada para a solução do seu problema. A proposta do guia é auxiliar o usuário nesta etapa, disponibilizando perguntas para conduzi-lo à escolha correta da tarefa de mineração. As perguntas são frases genéricas, que são associadas a cada tarefa, conforme mostrado na parte 3 da figura 2. Nessa figura são mostradas as frases para escolha de uma das 3 principais tarefas de mineração de dados: Regras de Associação, Classificação e Agrupamento.

A etapa de integração dos dados é realizada quando os dados são provenientes de diversas fontes. A limpeza é realizada quando os dados a analisar não estão íntegros (por exemplo, estão incompletos, ou incorretos, entre outros). Alguns recursos referentes à integração e limpeza dos dados foram incluídos na ferramenta, porém não serão aqui tratados devido a limitações de espaço. Após essas etapas, procede-se à etapa de seleção dos dados, apresentada a seguir.

01 Qual o problema a ser resolvido pela mineração de dados?
A DIVULGAÇÃO DO CONGRESSO GERA CUSTO ALTO E NEM SEMPRE É EFICIENTE.

02 Que objetivo deve ser cumprido para resolver o problema?
DEFINIR A MELHOR FORMA DE DIVULGAÇÃO DO CONGRESSO CONFORME O PERFIL DO CONGRESSISTA

03 Dentre as frases abaixo, selecionar aquela que melhor representa a tarefa de mineração de dados que deseja executar para atingir o objetivo da mineração de dados.

Regras de Associação
☒ Deseja encontrar itens com uma grande probabilidade de ocorrerem juntos em uma mesma transação? Exemplo: Se um determinado cliente compra pão ele provavelmente comprará leite.

Classificação
☐ Deseja classificar um determinado objeto em uma entre várias possíveis classes? Exemplo: Todo o correntista de um determinado banco com idade entre 25-30 anos, casado e com renda mensal entre 2.000,00-3.000,00 é um bom cliente para receber uma proposta de cartão de crédito.

Agrupamento
☐ Deseja a organização de uma coleção de objetos em grupos de objetos similares? Exemplo: Agrupar clientes em função de altura e peso. Atributos devem ser numéricos.

Descrição da tarefa de mineração:
 Regras de Associação - (Projeto: CONGRESSO DE TECNOLOGIA)

Incluir Localizar Salvar Excluir Cancelar

Identificação da Tarefa de Mineração
 Tem por objetivo orientar o analista de dados com algumas questões para selecionar corretamente as tarefas de mineração de dados.
 Dentre as frases abaixo, selecionar aquela que melhor representa a tarefa de mineração de dados que deseja executar para atingir o objetivo da mineração de dados. Selecionar a frase que melhor representa a sua necessidade de mineração.
Restrições: Apenas uma frase pode ser selecionada.
Descrição da tarefa de mineração: Informar a descrição da tarefa de mineração de dados. Esta descrição será utilizada para diferenciar as diversas tarefas que podem ser cadastradas para um mesmo objetivo.
Incluir: Inclui uma nova tarefa de mineração de dados.
Localizar: Localiza todas as tarefas de mineração de dados associadas a um determinado objetivo.
Salvar: Salva os dados.
Excluir: Exclui a tarefa de mineração de dados.
Cancelar: Cancela as alterações realizadas.

Figura 2 – Identificação do Problema de Mineração de Dados

3.1.1. Apoio à Seleção dos Dados

Essa etapa tem como meta selecionar os dados que serão utilizados pelo algoritmo de mineração de dados. O guia 4 oferece subsídios para auxiliar nessa tarefa.

Guia 04: Auxilia o usuário a identificar e selecionar os dados relevantes ao problema.

Em essência, esta etapa objetiva identificar quais informações, dentre as bases de dados existentes, devem ser efetivamente consideradas para a mineração de dados. A proposta do guia é auxiliar o usuário nesta etapa, de forma a torná-la o mais simples possível, eliminar grande parte de conhecimento exigido e tentar evitar que erros possam ser cometidos. Uma facilidade oferecida é permitir uma visão geral das tabelas e colunas, para facilitar o trabalho do usuário, conforme mostrado na parte 1 e 2 da figura 3. Assim que uma determinada tabela é selecionada, a ferramenta automaticamente desabilita todas as tabelas que não possuem relacionamento com ela, para evitar que o usuário selecione dados de tabelas que não possuem relacionamento entre si (representado na parte 2 da figura 3).

A ferramenta informa se uma coluna não for apropriada para ser usada na tarefa de mineração de dados escolhida. A parte 3 da figura 3 mostra a situação em que o usuário selecionou a coluna *código* da tabela *Alunos*. Essa coluna é a chave primária da tabela *Alunos*, ou seja, todos os registros possuem valores diferentes. Em regras de associação, essa coluna não é apropriada, então o usuário é orientado por meio de uma mensagem.

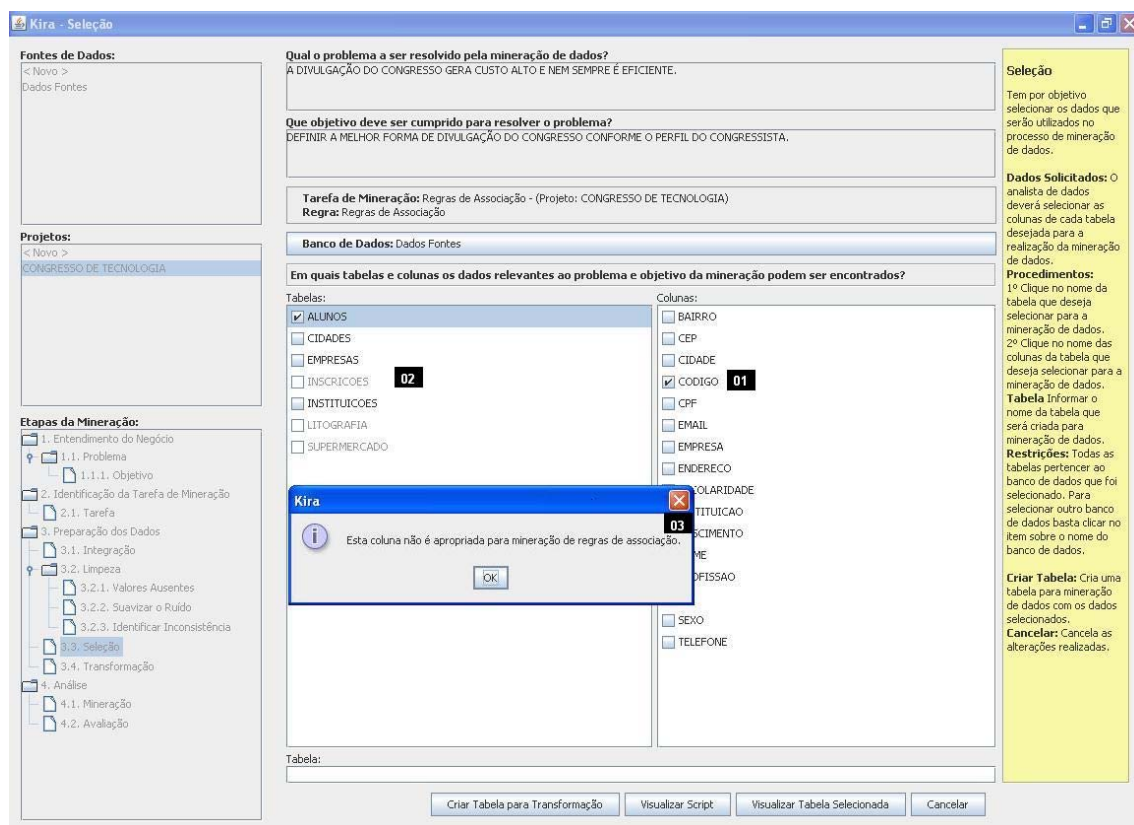


Figura 3 – Seleção dos Dados

3.1.2. Apoio à Transformação dos Dados

Esta etapa tem por objetivo orientar na transformação dos dados selecionados conforme as necessidades dos algoritmos de mineração de dados. O guia 5 auxilia o usuário a realizar a etapa de transformação.

Guia 05: Auxilia o usuário a avaliar individualmente cada coluna da tabela e sugere o que deve ser feito para deixar a coluna apta para a mineração.

Nesta etapa, os dados devem ser colocados em um formato que possam servir de entrada para um algoritmo de mineração. A proposta do guia é apoiar o usuário nesta etapa, de forma a orientar sobre o formato exigido e propor as alterações que devem ser realizadas para tornar os dados aptos à tarefa de mineração. Como exemplos de transformação de dados podem-se citar: a generalização de um atributo (por ex. *rua*) substituindo-o por conceitos de nível mais alto usando hierarquias de conceitos (ex: *cidade*, *país*); a substituição de um atributo que represente data de nascimento por faixas de idades.

3.2. Apoio à Análise

Este módulo objetiva auxiliar o usuário a executar as atividades que compreendem a mineração de dados e avaliação dos resultados obtidos.

3.2.1. Apoio à Mineração de Dados

A mineração de dados é a principal etapa do processo de KDD. Nesta etapa, ocorre a busca efetiva por conhecimento. Para alcançar este objetivo, é necessária a aplicação de

algoritmos sobre os dados selecionados. Ter conhecimento dos algoritmos disponíveis de cada tarefa de mineração, suas características e como parametrizar corretamente é fundamental para o sucesso da mineração de dados. A proposta do guia 6 é auxiliar o usuário nesta etapa, de forma a abstrair a necessidade de conhecimento detalhado sobre os algoritmos disponíveis e de detalhes sobre seus parâmetros de configuração.

Guia 06: Auxilia o usuário na parametrização do algoritmo que será usado para a execução da mineração de dados.

Para alcançar esse objetivo, o guia proposto deverá conduzir o usuário a selecionar o algoritmo mais adequado e, de forma intuitiva, o usuário irá parametrizar corretamente o algoritmo minerador.

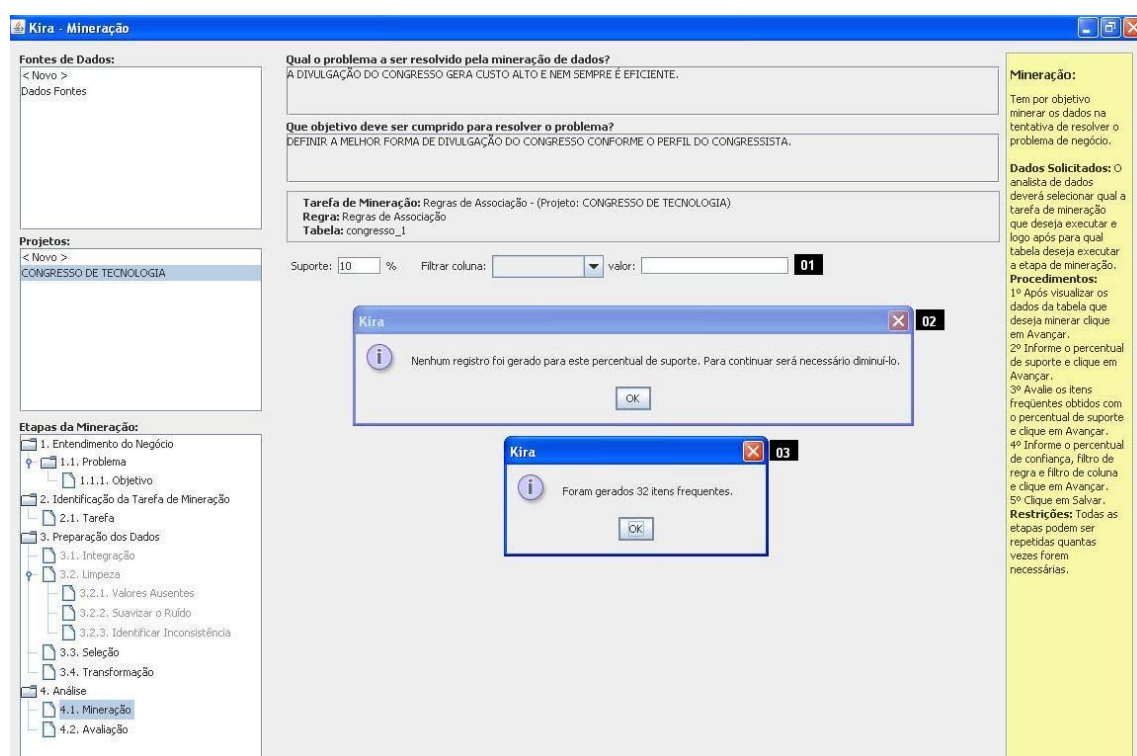


Figura 4 – Suporte Mínimo

A parte 1 da figura 4 permite ao usuário informar o valor de suporte mínimo (*minsup*) das regras e aplicar um filtro, se desejado. Um filtro tem por objetivo selecionar os dados que o usuário deseja para a geração dos itens frequentes; um exemplo seria: selecionar somente os registros de alunos que ficaram sabendo da realização do congresso por meio da escola.

Após informar o percentual de suporte mínimo pode-se executar o algoritmo minerador. A ferramenta retorna os itens frequentes encontrados (aqueles que atendem ao percentual informado), ou orienta a reduzir o percentual se não encontrar itens frequentes (parte 2 e 3 da figura 4). Em seguida o usuário informa o percentual de confiança mínima a utilizar e também o filtro de regra desejado. Esse filtro é válido apenas para as tabelas que já passaram pelo processo de mineração e cujas regras obtidas foram qualificadas quanto à qualidade (conforme mostrado na parte 1 da figura 5). Se não forem encontradas regras para o valor de confiança mínima estabelecido, o usuário é orientado a reduzir o percentual de confiança.

Qual o problema a ser resolvido pela mineração de dados?
A DIVULGAÇÃO DO CONGRESSO GERA CUSTO ALTO E NEM SEMPRE É EFICIENTE.

Que objetivo deve ser cumprido para resolver o problema?
DEFINIR A MELHOR FORMA DE DIVULGAÇÃO DO CONGRESSO CONFORME O PERFIL DO CONGRESSISTA.

Tarefa de Mineração: Regras de Associação - (Projeto: CONGRESSO DE TECNOLOGIA)
Regra: Regras de Associação
Tabela: congresso_1

Filtrar: Ruim

Seq.	Regra	Data	% de Suporte (m...	% de Confiança ...	% de Suporte	% de Confiança	Qualidade
1	SEXO=masculin...	14/05/2009	10.00	60.00	15.78	65.31	Ruim
2	ESCOLARIDADE...	14/05/2009	10.00	60.00	24.17	65.25	Ruim
3	ESCOLARIDADE...	14/05/2009	10.00	60.00	15.78	64.78	Regular
4	ESCOLARIDADE...	14/05/2009	10.00	60.00	15.04	70.28	Boa
5	SEXO=feminino ...	14/05/2009	10.00	60.00	24.56	63.93	Ótima
6	ESCOLARIDADE...	14/05/2009	10.00	60.00	11.05	62.22	
7	ESCOLARIDADE...	14/05/2009	10.00	60.00	13.56	64.86	
8	ESCOLARIDADE...	14/05/2009	10.00	60.00	24.37	65.78	

01

02

Regra
Dos 2027 registros analisados da tabela CONGRESSO_1 através de regras de associação, "SEXO=masculino; ESCOLARIDADE=Segundo Grau completo; FICOU_SABENDO=Escola" ocorrem juntos em 15.78% dos registros (ou seja, em aproximadamente 319 registros). Do total de registros em que "SEXO=masculino; ESCOLARIDADE=Segundo Grau completo" ocorre, "FICOU_SABENDO=Escola" ocorre em 65.31% desses registros.

Data da última mineração: 14/05/2009
Suporte: 10.00%
Confiança: 60.00%
Número de Regras: 8

Salvar Relatário Visualizar Tabela Excluir Regras Selecionadas

Avaliação:
Tem por objetivo avaliar a qualidade das regras geradas.

Dados Solicitados: O analista de dados deverá selecionar qual a tarefa de mineração que deseja avaliar e logo após para qual tabela deseja executar a etapa de avaliação.

Procedimentos:
1º Analise a regra gerada.
2º Clique na coluna Qualidade e atribua a qualidade desejada a regra analisada.
3º Clique em Salvar.

Figura 5 – Avaliação das Regras

3.2.2. Apoio à Avaliação

Nesta etapa o usuário pode analisar cada regra gerada qualifica-la como ruim, regular, boa ou ótima, com base em seus valores de suporte e confiança. Esta etapa é apoiada pelo guia 07.

Guia 07: Auxilia o usuário a ler e interpretar as regras geradas.

A ferramenta auxilia a ler e interpretar corretamente as regras geradas, considerando os valores obtidos de suporte e confiança. É apresentado ao usuário um texto interpretativo da regra, conforme mostra a parte 2 da figura 5, para a regra gerada: *sexo=masculino, escolaridade=Segundo Grau completo => Ficou_sabendo=Escola (suporte=15,78%) (confiança= 65,31%)*.

4. Análise da Eficácia da Ferramenta Kira em Ambiente Empresarial

Para medir as dificuldades da aplicação da mineração de dados para diversos perfis de usuários e avaliar o potencial dos guias propostos na ferramenta Kira para auxiliar essa tarefa, foi desenvolvido um estudo de caso na empresa Metalúrgica Mococa S/A (<http://www.metalurgicamococa.com.br/>). Essa empresa é dedicada à produção de latas brancas e litografadas e produz embalagens metálicas para a maioria dos fabricantes nacionais de produtos alimentícios em conserva.

Procurou-se medir a eficácia da ferramenta com base na facilidade de ensino e aprendizado do processo de mineração de dados que ela propicia. Para o estudo de caso foram selecionados 20 funcionários da empresa, com diferentes níveis de conhecimento sobre

o processo de KDD e mineração de dados. Estes foram divididos em 4 grupos, cada um com 5 funcionários. Os dados utilizados são provenientes do setor de Litografia. Os operadores das máquinas digitam os dados relativos à produção realizada assim que a mesma é encerrada. As atividades de cada grupo foram: instalar a ferramenta Kira, selecionar os dados, executar a mineração de dados e analisar as regras geradas. O problema de mineração de dados identificado foi: *Reduzir as perdas do processo de litografia*. O objetivo da mineração: *Encontrar todos os itens que estão associados à perda no processo de litografia*.

O **primeiro grupo** foi formado por funcionários da área de informática (analistas de sistemas e desenvolvedores) com conhecimentos básicos sobre o processo de KDD e de mineração de dados. Esse grupo realizou as atividades em menos de 1 hora, sem apresentar dificuldade para a realização das tarefas e sem ler as instruções para a execução das mesmas.

O **segundo grupo** também foi formado por analistas de sistemas e desenvolvedores, porém sem qualquer conhecimento sobre o processo de KDD e mineração de dados. Esse grupo realizou as atividades em menos de 3 horas. Para cada etapa executada o grupo procurou ler as instruções relacionadas a cada tarefa.

O **terceiro grupo** foi formado por funcionários do setor administrativo com conhecimentos básicos de informática (usuários operacionais dos sistemas de gestão da empresa), sem conhecimento do processo de KDD e de mineração de dados. Os funcionários desse grupo tiveram dificuldade para executar as atividades; precisaram várias vezes de auxílio técnico. Inicialmente tentaram pular etapas, mas perceberam que tinham que executar passo a passo. A cada passo seguiram as instruções da ferramenta. Tiveram maior dificuldade na etapa de seleção dos dados: não entendiam o conceito de tabelas e colunas. Após 2 horas de uso conseguiram gerar as primeiras regras. Quando solicitados a executar novamente todas as etapas tiveram dificuldade, mas conseguiram gerar as regras. A ferramenta Kira serviu para que entendessem o processo de KDD.

O **quarto grupo**, formado por funcionários da área de produção, com conhecimentos básicos de informática (usuários operacionais dos sistemas de gestão da empresa), não possuíam conhecimento do processo de KDD e de mineração de dados. Tiveram um acompanhamento técnico desde o início. Primeiro foram ensinados alguns conceitos básicos de mineração de dados, sem entrar em detalhes técnicos. Durante o uso da ferramenta cada etapa foi executada com auxílio técnico que explicava seu objetivo.

Apesar das dificuldades iniciais dos 2 últimos grupos, a ferramenta facilitou o ensino do processo de KDD. Os integrantes dos 3 primeiros grupos não tinham conhecimento sobre o processo de litografia, o que dificultou a identificação das regras válidas para a solução do problema de mineração. Somente os funcionários do grupo 4 conseguiram ler e interpretar as regras geradas, por conhecerem os dados usados no experimento e o problema em análise. Atualmente esse grupo usa a ferramenta Kira no seu dia-a-dia, para análise de perdas na litografia, mostrando sua utilidade para a descoberta de padrões interessantes nos dados.

Esse experimento permitiu verificar que com a Kira, os usuários, no geral, aprendem o processo de mineração de dados de forma mais autônoma. Em experiências de ensino de mineração de dados em sala de aula para alunos de Ciência da Computação, comparando o uso da ferramenta Kira com a Weka (disponível em <http://www.cs.waikato.ac.nz/ml/weka>), constatou-se que é muito mais fácil o ensino de mineração de dados por meio da Kira. Em

geral, o uso dessa ferramenta em sala de aula, por alunos de graduação e pós-graduação, tem comprovado seu potencial para o processo de ensino-aprendizagem de mineração de dados usando regras de associação.

5. Conclusões

Este artigo apresentou a ferramenta Kira e a técnica adotada para apoiar o ensino/aprendizado do processo de mineração de dados, usando um conjunto de guias para orientar o usuário na execução desse processo, centrado em regras de associação. Foi apresentado um estudo de caso utilizando a ferramenta Kira. Esse estudo de caso permitiu verificar que o processo de mineração de dados tornou-se intuitivo e parte expressiva do conhecimento exigido do usuário foi abstraída. Com o estudo de caso verificou-se a eficácia da ferramenta para auxiliar o ensino do processo de KDD a usuários sem conhecimento de mineração. Apesar do enfoque inicial do trabalho ter sido voltado a usuários da área de informática, o uso da ferramenta na prática tem mostrado que mesmo usuários que não sejam dessa área conseguem usá-la com certa facilidade. Isso está sendo comprovado na empresa Metalúrgica Mococa S/A, cujos gerentes industriais têm usado a Kira.

Pretende-se refinar a ferramenta visando aprimorar as instruções para torná-la ainda mais eficiente no apoio ao ensino de mineração de dados e desenvolver recursos para outras tarefas de mineração de dados, tais como classificação e agrupamento.

6. Referências Bibliográficas

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. *Mining association rules between sets of items in large databases*. In: Proc. of the ACM SIGMOD Int'l Conf. on Management of Data. Washington, D.C., USA, 1993. p. 207-216.
- AGRAWAL, R. ; SRIKANT, R. *Fast algorithms for mining association rules*. In: Proc. of the Int'l Conf. on Very Large Databases, 1994. Santiago de Chile, Chile, 1994.
- CAMPOS, M. M.; STENGARD, P.J.; MILENOVA, B. L. *Data-Centric Automated Data Mining*. Oracle Data Min. Technol., Redwood Shores, CA, USA 2005.
- CHAPMAN, P. CRISP-DM 1.0: *Step-By-Step Data Mining Guide*. [S.I.]: 2000. Disponível em: <<http://www.crisp-dm.org/download.htm>>. Acesso em: 28 jan. 2008.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMITH, P. *From Data Mining to Knowledge Discovery: An Overview*. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press/ The MIT Press, MIT, Cambridge, Massachusetts, England, 1996.
- GOLDSCHMIDT, R.; PASSOS, E., *Data Mining: Um Guia Prático*, Rio de Janeiro, Elsevier, 2005.
- HAN, J.; FU, Y.; WANG, W.; CHIANG, J. *DBMiner: A System for Mining Knowledge in Large Relational Databases*. In: Proc. Intern. Conf. Knowledge Discovery in Databases, Portland, 1996.
- HAN, J.; KAMBER, M. *Data Mining - Concepts and Techniques*. 2a edição. Nova York: Morgan Kaufmann, 2006.
- MENDES, E. F. *Automatização do Processo de Descoberta de Conhecimento em Base de Dados Auxiliada por Guias*, Universidade Metodista de Piracicaba, Fev, 2009.
- PIATETSKY-SHAPIRO, G. *Knowledge Discovery in Data Warehouse: 10 years after*. SIGKDD Explorations, Vol.1, Issues 2, 2000.
- WITTEN, I.; FRANK, E. *Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 2005.