

Descoberta de conhecimento aplicando regras de associação na ferramenta *WEKA*

Elio Ribeiro Faria Junior⁵⁸

Marcos Jesus de Oliveira Nielsen⁵⁹

RESUMO

Nos últimos anos vem aumentando o uso das técnicas de mineração de dados para descoberta de novos conhecimentos, devido à grande quantidade de registros armazenados em um ou mais bancos de dados. Torna-se complexa a utilização de técnicas tradicionais para produzir relacionamentos entre os elementos, que seja capaz de descobrir irregularidades, padrões e ligações para suportar a tomada de decisão. Sendo assim, este estudo vai utilizar a metodologia denominada KDD (Knowledge Discovery in Database – Descoberta de conhecimento em Banco de Dados). A base de dados é de uma avaliação institucional, a qual contém perguntas e respostas sobre um determinado curso. Para isso foi necessário utilizar o software *WEKA* (Waikato Enviroment for Knowledge Analysis), que permite o uso das regras de associação, a qual faz o uso do algoritmo *Apriori*. Dessa forma, o uso das regras de associação mostrou-se eficaz para descoberta de informações ocultas, as quais podem se tornar dados privilegiados para os gestores.

Palavras-chave: KDD. *WEKA*. Regra de associação. Informações ocultas. *Apriori*.

⁵⁸ Graduado em Sistemas de Informação pela Universidade do Contestado e Pós-Graduação em Engenharia de Software pela PUCPR. Atualmente atua como professor na área de informática no Centro Universitário de União da Vitória (UNIUV) e instrutor de informática no SENAC-PR. E-mail: prof.elio@uniuv.edu.br

⁵⁹ Graduado em Informática pelo Centro Universitário da Cidade de União da Vitória (2004). Atualmente é professor do Centro Universitário da Cidade de União da Vitória (UNIUV). Tem experiência na área de Ciência da Computação. E-mail: prof.marcos.nielsen@uniuv.edu.br

Discovery of knowledge applying association rules on the WEKA tool

Elio Ribeiro Faria Junior
Marcos Jesus de Oliveira Nielsen

ABSTRACT

In recent years the use of data mining techniques to discover new knowledge has increased due to the large number of records stored in one or more databases. It becomes complex to use traditional techniques to produce relationships between elements that are able to find out irregularities, patterns, and links to support decision making. Thus, this study will use the methodology called KDD (Knowledge Discovery in Database). The database is an institutional assessment that contains questions and answers about a particular course. It was necessary to use the WEKA software (Waikato Environment for Knowledge Analysis). This software allows the use of association rules, which make use of the *Apriori* algorithm. Thus, the use of association rules proved to be effective for the discovery of hidden information, which can become privileged data for managers.

Keywords: KDD. WEKA. Association Rules. Hidden information. *Apriori*.

1 INTRODUÇÃO

A grande quantidade de dados gerados pelas organizações torna-se cada vez maior, são muitos novos registros nas bases de dados, e, além de toda essa grande quantidade, existe a diversidade dos dados, que dificulta ainda mais o entendimento deles. Todo esse cenário motivou a necessidade de novos métodos e ferramentas para análise automatizada e inteligente em bases de dados (FAYYAD *et al.* 1996).

Em todas as áreas pode-se fazer o uso das técnicas. Pesquisadores de medicina, engenharia e ciência acumulam uma grande quantidade de informações, as quais são essenciais para descobertas importantes (TAN *et al.* 2009). Estudiosos determinados em transformar essa informação em conhecimento, depararam com o Knowledge Discovery in Database (KDD). Segundo FAYYAD e outros (1996), o KDD é um “processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. O KDD mapeia informações brutas em modelos mais reduzidos ou úteis que as informações originais, ou seja, ele realiza a seleção dos dados mais relevantes, que serão utilizados no decorrer do processo, assim, não sendo necessário utilizar toda a base de dados (MELO, 2010).

O processo de KDD busca padrões que não são encontrados em pesquisas rotineiras em banco de dados, descobre informações que podem ser utilizadas por gestores na tomada de decisão. Um exemplo clássico na literatura é a venda de fraldas e cervejas, em que se identificou que homens casados, na faixa etária entre 25 e 30 anos, compravam fraldas e cervejas às sextas-feiras à tarde, no caminho do trabalho para casa. Com essa informação, o supermercado organizou as gôndolas, deixando as fraldas perto da cerveja, e isso fez com que houvesse um aumento no consumo.

Sendo assim, este artigo realiza uma análise da aplicação das regras de associação em uma base de dados de uma avaliação institucional. O software *WEKA* (Waikato Environment for Knowledge Analysis) foi utilizado para realização deste estudo. A base escolhida contém perguntas e respostas, com as quais, por meio das regras de associação, podem-se obter informações de grande relevância para os gestores.

O artigo está dividido da seguinte maneira: a seção 2 relata as fases do processo KDD, algoritmo utilizado e regras empregadas; a seção 3 apresenta os procedimentos metodológicos utilizados no trabalho; na seção 4 são evidenciados os resultados obtidos; e, na seção, 5 a conclusão.

2 KDD - KNOWLEDGE DISCOVERY IN DATABASES

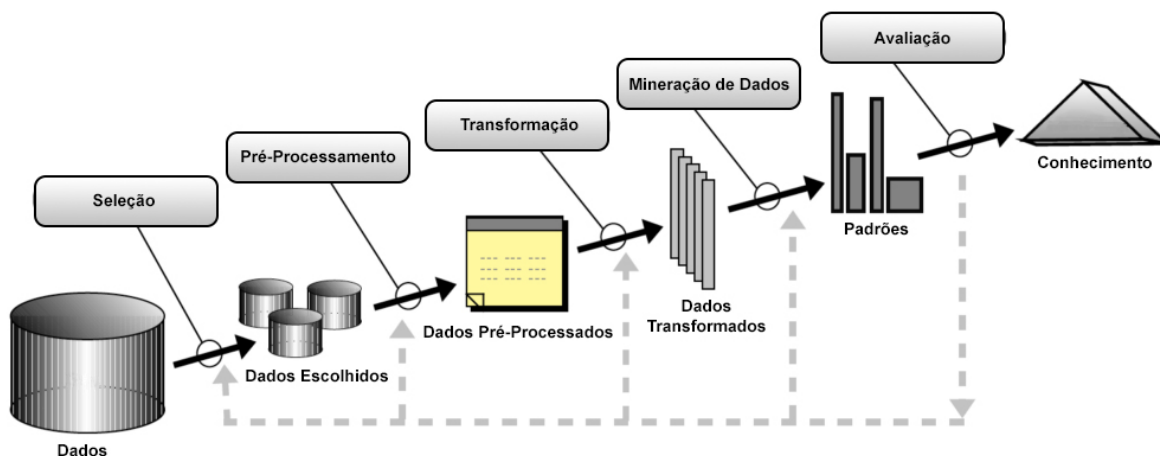
Na última década, houve um grande aumento na quantidade de informações armazenadas, e atribuímos esse aumento ao baixo custo para automatização dos processos de coleta de dados e outras tecnologias que facilitaram o armazenamento, por exemplo, a computação nas nuvens, que permite ao usuário armazenar toda sua informação em uma plataforma online.

Com toda essa informação armazenada, o processo para extrair informações relevantes torna-se cada dia mais difícil, com isso a metodologia Knowledge Discovery in Databases (KDD), que em português tem a denominação de “Descoberta de Conhecimento em Bases de Dados (DCBD)”, tem como objetivo extrair padrões de grandes bases de dados, ou seja, informações importantes que estavam ocultas devido à grande quantidade de informação. Segundo Fayyad e outros (1996), a tradicional forma de transformar os dados em conhecimento resulta em um processo manual realizado por especialistas, que produzem relatórios que devem ser avaliados. O grande volume de informações torna esse processamento manual inviável.

2.1 FASES DO KDD

O KDD faz o uso de técnicas de aprendizado de máquina, estatísticas e reconhecimento de padrões com o objetivo de encontrar informações que gerem conhecimento. De acordo com FAYYAD e outros (1996), o KDD é organizado em diversas fases, que vão da seleção dos dados até a avaliação. A Figura 1 apresenta o processo.

Figura 01 - Processo do KDD



Fonte: FAYYAD *et al*, 1996

Na fase de seleção são identificados os dados que serão trabalhados da base (GOLDSCHIMIDT; PASSOS, 2005, p.26). É necessária a identificação, para alcançar um determinado objetivo. Por exemplo, identificar possíveis compradores de um determinado produto, após a fase de seleção é realizado o pré-processamento, quando ocorre a eliminação de dados com inconsistência, redundância e ausência, por exemplo, a identificação do sexo masculino definida por “M” em um registro, e “H” em outro. Na fase de transformação é necessário exportar essa base tratada para um formato adequado para a mineração. No software *WEKA* o arquivo precisa ser convertido para ARFF (Attribute-Relation File Format), que é o formato aceito pelo programa, mas é possível encontrar na literatura outros softwares com diferentes formatos aceitos.

Após a fase de transformação, é realizada a extração do conhecimento mediante a mineração de dados. De acordo com Berry e Linoff (1997), “Mineração de dados é a exploração e a análise, por meio automático ou semiautomático, de grandes quantidades de dados, a fim de descobrir padrões e regras significativos”. Nessa fase pode-se fazer o uso de diversas técnicas e algoritmos, por exemplo, Classificação, Análise de Agrupamento, Modelo de Dependência, Regras de Associação e outras. E a última etapa do processo é a avaliação, quando o conhecimento obtido é interpretado por um especialista da área, e é identificado se os objetivos foram alcançados.

2.2 REGRAS DE ASSOCIAÇÃO

A tarefa de associação busca encontrar relacionamentos no conjunto de dados, ou seja, busca identificar padrões frequentes que ocorrem na base. Por exemplo, em uma base de dados, onde se registram as vendas por cliente, com o uso das regras de associação, poderia gerar a seguinte regra: {leite, café, pão} -> {margarina}, a qual sugere que o cliente que comprar leite, café e pão, pode vir comprar margarina também; o que vai determinar se ele vai levar ou não a margarina é o grau de certeza da regra. Para explicar o processo, será exibido na figura 02 um exemplo fictício de uma pequena base de dados, que contém compras efetuadas por clientes em um mercado.

Figura 02 - Registro de vendas

TID	Produtos Comprados
1	Café, Pão, Margarina, Leite
2	Leite, Margarina
3	Café, Leite, Bolacha
4	Pão, Café, Maionese
5	Leite, Margarina
6	Café, Maionese
7	Pão, Leite, Café, Margarina
8	Café, Pão, Margarina, Maionese
9	Leite, Maionese
10	Maionese, Bolacha

O grau de certeza da regra é definido por dois parâmetros, o primeiro é o suporte, que é a frequência de um determinado item em um conjunto de dados, por exemplo, a figura 02 possui um conjunto de dados com dez registros de vendas, o suporte é a quantidade de vezes que um determinado produto ou um conjunto de produtos aparece nesses registros. Ao utilizar a figura 02, observam-se as seguintes regras.

1-SE Café ENTÃO Pão

2-SE Pão ENTÃO Leite

3-SE Café, Pão ENTÃO Leite

Dessa forma, o suporte da primeira regra pode ser calculado pela seguinte fórmula:

$$\text{Suporte} = \frac{\text{Número de registro com x e y}}{\text{Número total de registros}}$$

O número de registros é a quantidade de vezes que o café e o pão saíram no mesmo registro de venda, ou seja, em quatro dos dez registros ocorreu a compra conjunta de café e pão, dessa maneira o suporte é medido com a utilização da fórmula descrita acima, cujo valor é de quarenta por cento (40%).

O segundo parâmetro é a confiança, regras são criadas para determinar a confiança de um padrão na informação. Para descobrir o valor de confiança da primeira regra, utiliza-se a seguinte fórmula:

$$\text{Confiança} = \frac{\text{Número de registro com x e y}}{\text{Número de registros com x}}$$

É necessário verificar em quantos registros o café e o pão são vendidos no mesmo conjunto. Na figura 02, isso acontece em quatro registros, logo após é dividido esse valor pela quantidade de vezes que aparece somente o café nos registros, ou seja, seis vezes, cujo o valor é de sessenta e seis por cento (66%). A confiança da primeira regra é de 66%, isso quer dizer que em 66% das transações as pessoas que compraram café também levaram leite. O algoritmo utilizado nesse artigo para determinar as regras de associação é o *Apriori*, o qual vai ser abordado a seguir.

2.3 ALGORITMO *APRIORI*

O algoritmo *Apriori* busca informações que ocorrem com uma determinada frequência em uma base de dados, baseado em uma regra mínima determinada. De acordo com Miranda e outros (2003) o algoritmo *Apriori* é um clássico quando se utiliza extração de Regras de Associação. Ele faz o uso dos parâmetros de confiança e suporte para obter as melhores regras (CORRÊA, 2009).

No software *WEKA* é possível modificar alguns atributos do algoritmo, por exemplo, definir qual vai ser o suporte e a confiança mínima. O algoritmo faz recursivas buscas na base de dados, e os conjuntos frequentes que ele vai encontrar vai depender do suporte e confiança mínima que foi configurado antes de o algoritmo ser executado na base de dados.

Um ponto importante para se observar, é que se o suporte e a confiança forem muito baixos, o algoritmo vai gerar um grande número de regras, em que muitas delas não apresentarão nenhum ganho de informação, e se o suporte e a confiança forem muito altos, algumas regras importantes podem ficar de fora. A tabela 01 mostra os principais parâmetros que podem ser alterados no algoritmo *Apriori*.

Tabela 01 – Descrição dos parâmetros do algoritmo *Apriori*.

Parâmetro	Descrição
LowerBoundMinSupport	Alterar o limite inferior para o suporte mínimo.
MetricType	Definir o tipo de métrica que será utilizado para gerar as regras.
MinMetric	Determinar a métrica mínima.
NumRules	O número máximo de regras a serem encontradas.
UpperBoundMinSupport	Alterar o limite superior para o suporte mínimo.

2.4 BASE DE DADOS

A base de dados foi originada de uma avaliação institucional, a qual contém respostas de acadêmicos sobre um determinado curso. O conjunto possui 429 instâncias e 05 atributos. A tabela 02 apresenta as informações a respeito dos atributos e tipo de dados.

Tabela 02 – Descrição da base de dados

Atributo	Tipo de dado
Colaboram com um clima favorável ao desenvolvimento das aulas	Sempre; Às vezes; Não.
Conhecimento dos objetivos do curso	Bom conhecimento; Pouco conhecimento; Desconhece.
Cumprem adequadamente as atividades recomendadas	Sempre Cumpre; Cumpre às vezes; Não cumpre.
Desenvolvimento de pesquisas e leituras Adicionais	Sempre Desenvolve; Desenvolve às vezes; Não desenvolve.
Motivação para frequentar o curso	Pouca motivação; Nenhuma motivação.

No atributo Motivação para frequentar o curso, foi retirado o tipo de dado muita motivação, pois o objetivo é encontrar as melhores regras que determinem padrões em acadêmicos que têm pouco motivação ou nenhuma para frequentar o curso. A base original possuía pouco mais de mil instâncias.

3 DESENVOLVIMENTO

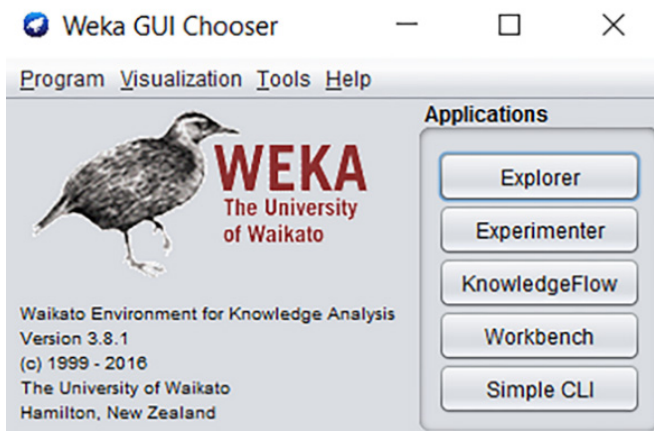
O avanço das técnicas de mineração de dados e da tecnologia tornou-se uma variante importante para as organizações. A informação é o bem mais precioso que uma organização possui, mas, devido ao seu grande volume, técnicas antigas de análise de dados não conseguem extrair informações que poderiam ser importantes e úteis para os gestores.

Com essa capacidade de gerar novas informações para tomada de decisões, vários setores podem utilizar-se dessas técnicas, por exemplo, a realização de uma avaliação institucional em uma universidade pode gerar uma grande quantidade de informação. As avaliações buscam informações para confirmar hipóteses ou para descobrir motivos que atrapalham o processo de ensino-aprendizagem (RODRIGUES, 2010, no prelo). Com a utilização das técnicas de mineração de dados, além de confirmar hipóteses e descobrir motivos que atrapalham o processo de ensino, podem-se extrair informações que podem passar despercebidas, em métodos tradicionais.

A proposta deste artigo é descobrir padrões em informações por meio das técnicas de mineração de dados, especificadamente as regras de associação. O software utilizado para realizar a mineração de dados foi o *WEKA*, de acordo com Melo (2010), o software Waikato Environment for Knowledge Analysis (*WEKA*) possui técnicas de mineração de dados, ele pode ser utilizado em diferentes sistemas operacionais, é um software livre e está disponível em <http://www.cs.waikato.ac.nz/ml/weka>.

A figura 03 mostra a tela inicial do WEKA.

Figura 03 – Tela inicial do WEKA



3.1 ARQUIVO ARFF

No processo de KDD é necessário transformar os dados para realizar a mineração dos dados, o *WEKA* utiliza um formato chamado ARFF (Attribute-Relation File Format) para facilitar o processo de seus algoritmos. O arquivo utiliza algumas marcações, o nome dos conjuntos de dados é especificado com a marcação *@relation*, os atributos utilizam *@attribute* e os dados são definidos por *@data*.

Na figura 04 observa-se a base de dados que foi utilizada para o desenvolvimento deste artigo em formato ARFF.

Figura 04 – Arquivo em formato ARFF

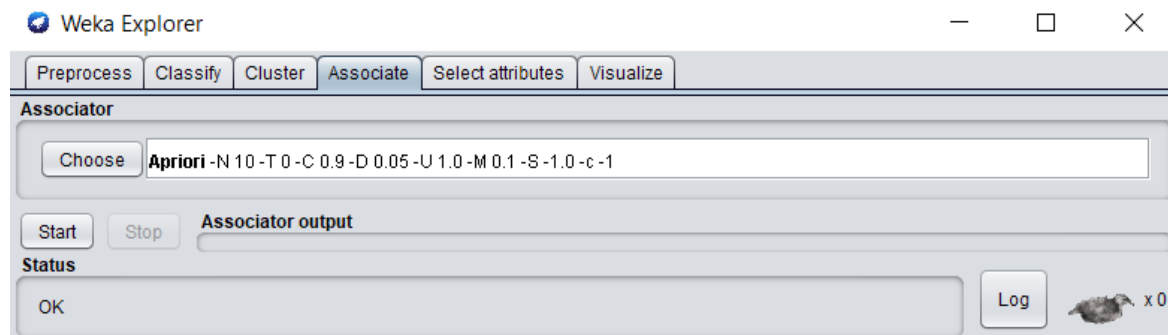
```

1 @relation SAI-Uniuv
2
3 @attribute ColaboramComUmaClimaFavoravelAoDesenvolvimentoDasAulas {
4   AsVezes, Sempre, Nao }
5 @attribute ConhecimentoDosObjetivosDoCurso {
6   PoucoConhecimento, BomConhecimento, Desconhece }
7 @attribute CumpremAdequadamenteAsAtividadesRecomendadas {
8   CumpreAsVezes, SempreCumpre, NaoCumpre }
9 @attribute DesenvolvimentoDepesquisasEleiturasAdicionais {
10  SempreDesenvolve, DesenvolveAsVezes, NaoDesenvolve }
11 @attribute MotivacaoaFrequentarOcurso { PoucaMotivacao, NenhumaMotivacao }
12
13 @data
14 AsVezes, PoucoConhecimento, CumpreAsVezes, SempreDesenvolve, PoucaMotivacao
15 Sempre, BomConhecimento, CumpreAsVezes, SempreDesenvolve, PoucaMotivacao
16 AsVezes, PoucoConhecimento, CumpreAsVezes, DesenvolveAsVezes, PoucaMotivacao
17 AsVezes, BomConhecimento, SempreCumpre, DesenvolveAsVezes, PoucaMotivacao
18 AsVezes, PoucoConhecimento, CumpreAsVezes, NaoDesenvolve, PoucaMotivacao
19 AsVezes, PoucoConhecimento, CumpreAsVezes, DesenvolveAsVezes, NenhumaMotivacao
20 AsVezes, PoucoConhecimento, CumpreAsVezes, DesenvolveAsVezes, PoucaMotivacao
21 AsVezes, BomConhecimento, CumpreAsVezes, NaoDesenvolve, PoucaMotivacao
22 Sempre, BomConhecimento, CumpreAsVezes, SempreDesenvolve, PoucaMotivacao
23 AsVezes, PoucoConhecimento, CumpreAsVezes, NaoDesenvolve, PoucaMotivacao
24 Sempre, BomConhecimento, CumpreAsVezes, DesenvolveAsVezes, PoucaMotivacao
25 Sempre, BomConhecimento, SempreCumpre, SempreDesenvolve, PoucaMotivacao
26 AsVezes, BomConhecimento, SempreCumpre, DesenvolveAsVezes, PoucaMotivacao
27 Sempre, BomConhecimento, SempreCumpre, DesenvolveAsVezes, PoucaMotivacao
28 AsVezes, PoucoConhecimento, CumpreAsVezes, SempreDesenvolve, PoucaMotivacao

```

Após a base de dados estar devidamente preparada no formato ARFF, ela pode ser executada pelo *software WEKA*. Como já mencionado anteriormente, vai ser utilizado o algoritmo *Apriori*. Dentro do *software WEKA* seleciona-se primeiramente a aba *Associate* e escolhe-se o pacote *weka.associations.Apriori*. A figura 05 mostra a tela *Associate* com a seleção do pacote *weka.associations.Apriori*.

Figura 05 – Tela *Associate* com a seleção do pacote *weka.associations.Apriori*



Como já mencionado, antes de executar o algoritmo *Apriori* pode-se alterar alguns parâmetros, por exemplo, a confiança e o suporte mínimo. A tabela 01 demonstra os principais parâmetros do algoritmo *Apriori*. A tabela 03 apresenta três parâmetros expressivos na aplicação do algoritmo *Apriori*.

Tabela 03 – Parâmetros do algoritmo *Apriori*

Opção	Função
Número de regras	Especifica quantas regras são desejadas. (Default = 12)
Confiança mínima	Especifica a confiança mínima exigida. (Default = 0.70)
Suporte mínimo	Especifica o suporte mínimo exigido. (Default = 0.20)

Com a configuração dos parâmetros dessa maneira, observa-se que, quanto mais próximo a confiança e o suporte for de 1.0 (100%), maior será a confiabilidade das regras geradas.

4 RESULTADOS OBTIDOS

Neste artigo foram utilizados os parâmetros definidos na tabela 03, ou seja, a confiança mínima é de 70%, suporte mínimo de 20% e o número de regras a ser gerado 12, aplicados na base de dados de uma avaliação institucional. Na tabela 04, são exibidas as regras geradas a partir do algoritmo *Apriori*.

Tabela 04 – Regras geradas pelo algoritmo *Apriori*

Melhores Regras	
1	ColaboramComUmaClimaFavoravelAoDesenvolvimentoDasAulas=Sempre ConhecimentoDosObjetivosDoCurso=BomConhecimento 121 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 118 <conf:(0.98)>

Melhores Regras	
2	ConhecimentoDosObjetivosDoCurso=BomConhecimento CumpremAdequadamenteAsAtividadesRecomendadas=SempreCumpre 114 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 111 <conf:(0.97)>
3	ConhecimentoDosObjetivosDoCurso=BomConhecimento DesenvolvimentoDepesquisasEleiturasAdicionais=DesenvolveAsVeze 145 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 140 <conf:(0.97)>
4	ConhecimentoDosObjetivosDoCurso=BomConhecimento 234 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 223 <conf:(0.95)>
5	ColaboramComUmaClimaFavoravelAoDesenvolvimentoDasAulas=Sempre CumpremAdequadamenteAsAtividadesRecomendadas=SempreCumpre 116 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 110 <conf:(0.95)>
6	ConhecimentoDosObjetivosDoCurso=BomConhecimento CumpremAdequadamenteAsAtividadesRecomendadas=CumpreAsVeze 116 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 110 <conf:(0.95)>
7	ColaboramComUmaClimaFavoravelAoDesenvolvimentoDasAulas=Sempre 181 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 171 <conf:(0.94)>
8	CumpremAdequadamenteAsAtividadesRecomendadas=SempreCumpre DesenvolvimentoDepesquisasEleiturasAdicionais=DesenvolveAsVeze 120 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 113 <conf:(0.94)>
9	CumpremAdequadamenteAsAtividadesRecomendadas=SempreCumpre 190 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 175 <conf:(0.92)>
10	DesenvolvimentoDepesquisasEleiturasAdicionais=DesenvolveAsVeze 265 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 243 <conf:(0.92)>
11	ColaboramComUmaClimaFavoravelAoDesenvolvimentoDasAulas=AsVeze DesenvolvimentoDepesquisasEleiturasAdicionais=DesenvolveAsVeze 136 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 124 <conf:(0.91)>
12	ColaboramComUmaClimaFavoravelAoDesenvolvimentoDasAulas=AsVeze CumpremAdequadamenteAsAtividadesRecomendadas=CumpreAsVeze 134 ==> MotivacaoaFrequentarOcurso=PoucaMotivacao 122 <conf:(0.91)>

Após analisar a tabela 04, obtêm-se as seguintes informações sobre as 12 regras encontradas:

- Se o acadêmico sempre colabora com o clima favorável ao desenvolvimento das aulas e tem um bom conhecimento dos objetivos do curso, ele pode estar insatisfeito com os objetivos do curso, e isso acaba gerando pouca motivação para frequentar o curso. 98% de confiança.
- Se o acadêmico tem bom conhecimento dos objetivos do curso e sempre cumpre as atividades recomendadas, ele pode estar desmotivado com as atividades recomendadas pelo professor, gerando pouca motivação para frequentar o curso. 97% de confiança.
- Se o acadêmico tem bom conhecimento dos objetivos do curso e desenvolve às vezes as pesquisas e leituras adicionais, então ele pode ter pouca motivação para frequentar o curso. 97% de confiança.
- Se o acadêmico tem bom conhecimento dos objetivos do curso, então ele pode ter pouca motivação para frequentar o curso, devido à insatisfação dos objetivos. 95% de confiança.
- Se o acadêmico sempre colabora com o clima favorável ao desenvolvimento das aulas e sempre cumpre as atividades recomendadas, ele pode estar desmotivado com as atividades recomendadas, e isso acaba gerando pouca motivação para frequentar o curso. 95% de confiança.
- Se o acadêmico tem bom conhecimento dos objetivos do curso e cumpre às vezes as atividades recomendadas, então ele pode ter pouca motivação para frequentar o curso. 95% de confiança.
- Se o acadêmico sempre colabora com o clima favorável ao desenvolvimento das aulas, então ele pode estar insatisfeito com outros atributos que provocam pouca motivação para frequentar o curso. 94% de confiança.
- Se o acadêmico sempre cumpre as atividades recomendadas e desenvolve às vezes as pesquisas e leituras adicionais, então ele pode ter pouca motivação para frequentar o curso. 94% de confiança.
- Se o acadêmico sempre cumpre as atividades recomendadas, então ele pode ter pouca motivação para frequentar o curso, devido à insatisfação das atividades recomendadas. 92% de confiança.
- Se o acadêmico desenvolve às vezes as pesquisas e leituras adicionais, então ele pode ter pouca motivação para frequentar o curso. 92% de confiança.
- Se o acadêmico colabora às vezes com o clima favorável ao desenvolvimento das aulas e desenvolve às vezes as pesquisas e leituras adicionais, então ele pode ter pouca motivação para frequentar o curso. 91% de confiança.
- Se o acadêmico colabora às vezes com o clima favorável ao desenvolvimento das aulas e cumpre às vezes as atividades recomendadas, então ele pode ter pouca motivação para frequentar o curso. 91% de confiança.

As informações geradas nessas regras podem ser de grande valia para os gestores da instituição, por exemplo, na primeira regra é possível analisar que mesmo o acadêmico colaborando com o clima favorável ao desenvolvimento das aulas e tendo um bom conhecimento dos objetivos do curso, ainda ele tem pouca motivação para frequentar. Isso pode significar que é necessário realizar uma alteração nos objetivos do curso, ou seja, os objetivos do curso não são satisfatórios para os acadêmicos, sendo necessário realizar mudanças.

5 CONCLUSÃO

A informação é o bem mais preciso de uma organização, saber analisar cria um diferencial e as tornam mais competitivas. Qualquer área pode fazer o uso da metodologia KDD, a qual apresentou ser eficaz para extração de informações ocultas em uma base de dados, gerando informações privilegiadas para futuras tomadas de decisões.

Sendo assim, com este artigo, foi possível visualizar as etapas do KDD, principalmente, a etapa de mineração de dados, utilizando regra de associação. Essas regras geradas buscam encontrar padrões frequentes na base de dados, algumas regras geradas não têm um valor significativo, mas outras acabam-se transformando em informações privilegiadas. É importante ressaltar que as regras encontradas tornam-se claras, quando o algoritmo é aplicado sobre um conjunto grande de elementos relacionados, e que, se uma regra originada pelo algoritmo não é bem explicada, então existe uma evidência de que talvez ela própria não tenha sido bem compreendida pelo analista.

Como trabalho futuro, pode-se estudar a aplicação de outros algoritmos para obter as regras de associação, utilizando-se a mesma base de dados, com o objetivo de comparar os resultados obtidos com o algoritmo *Apriori*, o qual foi utilizado para o desenvolvimento deste artigo.

6 REFERÊNCIAS

BERRY, M. J. A.; LINOFF, G. **Data mining techniques: for marketing, sales, and customer support.** New York: Wiley Computer Publishing, 1997.

CORRÊA, K. S. **Processo de mineração de dados no estudo de fenômenos solares e geomagnéticos.** 28 f. Trabalho Acadêmico (mestre) – INEP, São José dos Campos, 2009.

FAYYAD, U; SHAPIRO, G. P; SMYTH, P; UTHURUSAMY, R. **Advances in knowledge discovery and data mining.** MenloPark: Mit Press, 1996.

GOLDSCHIMIDT, R.; PASSOS, E. **Datamining: um guia prático.** Rio de Janeiro: Campus, 2005.

MELO, M.D. Introdução à Mineração de Dados usando o Weka. **V CONNEPI-2010.**

MIRANDA, D.; SABORÊDO, A. P., *et al.* **“Iniciação Científica – Data Mining.”** AEDB Associação Educacional Dom Bosco. Resende - Rio de Janeiro, 2003.

RODRIGUES, A. J.; SOUZA, E. V. de; REIS, M. S.; NUNES, K. de C. S. **Avaliação institucional e as contribuições no processo de ensino-aprendizagem escolar.** Araguaína, 2010. No prelo.

TAN, P-N.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining - mineração de dados.** Rio de Janeiro: Ciência Moderna, 2009.