

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS APLICADAS A EDUCAÇÃO
DEPARTAMENTO DE CIÊNCIAS EXATAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**USO DE MINERAÇÃO DE DADOS PARA DESCOBERTA
DE CONHECIMENTO: ESTUDO DE CASO NO TRIBUNAL
DE CONTAS DO ESTADO DA PARAÍBA**

FRANCISCO IHAGO ALENCAR FERREIRA
Orientadora: ADRIANA ZENAIDE CLERICUZI

RIO TINTO - PB
2013

FRANCISCO IHAGO ALENCAR FERREIRA

**USO DE MINERAÇÃO DE DADOS PARA DESCOBERTA
DE CONHECIMENTO: ESTUDO DE CASO NO TRIBUNAL
DE CONTAS DO ESTADO DA PARAÍBA**

Monografia apresentada para obtenção do título de Bacharel à banca examinadora no Curso de Bacharelado em Sistemas de Informação do Centro de Ciências Aplicadas e Educação (CCAEE), Campus IV da Universidade Federal da Paraíba.
Orientador: Prof. Dra. Adriana Zenaide Clericuzi.

RIO TINTO - PB
2013

F383u Ferreira, Francisco Ihago Alencar.

*Uso de mineração de dados para descoberta de conhecimento:
estudo de caso no*

*Tribunal de Contas do Estado da Paraíba / Francisco Ihago
Alencar Ferreira. –*

Rio Tinto: [s.n.], 2013.

55f.: il. –

Orientadora: Adriana Zenaide Clericuzi.

Monografia (Graduação) – UFPB/CCAEE.

*1. Informática. 2. Mineração de Dados. 3. Inteligência de
Negócio. I.Título.*

FRANCISCO IHAGO ALENCAR FERREIRA

**USO DE MINERAÇÃO DE DADOS PARA DESCOBERTA
DE CONHECIMENTO: ESTUDO DE CASO NO TRIBUNAL
DE CONTAS DO ESTADO DA PARAÍBA**

Trabalho de Conclusão de Curso submetido ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal da Paraíba, Campus IV, como parte dos requisitos necessários para obtenção do grau de BACHAREL EM SISTEMAS DE INFORMAÇÃO.

Assinatura do autor: _____

APROVADO POR:

Orientador: Prof^º. Dr^a. Adriana Zenaide Clericuzi
Universidade Federal da Paraíba – Campus IV

Prof. Dr. XXX
Universidade Federal da Paraíba – Campus IV

Prof. Dr. XXX
Universidade Federal da Paraíba – Campus IV

RIO TINTO - PB
2013

RESUMO

Este trabalho, baseado em uma revisão bibliográfica com o intuito de filtrar as informações para que se torne mais simples a compreensão, tem como objetivo explicar técnicas de descoberta de conhecimento utilizando mineração de dados, mostrando as fases e processos envolvidos na transformação de dados brutos em conhecimento, que podem ser valiosos para organizações no momento da tomada de decisão. São abordados também os conceitos sobre *Business Intelligence* e Descoberta de Conhecimento em Bases de Dados, além da análise das características e funcionalidades de ferramentas de *Data Mining* e de como pode ser usada a mineração de dados no Tribunal de Contas do Estado da Paraíba, para a identificação de possíveis fraudes ou irregularidades em processos licitatórios, de acordo com as associações descobertas entre os dados coletados e armazenados no banco de dados da referida instituição. Dessa forma o trabalho dos auditores pode ser agilizado, fazendo com que a fiscalização das contas públicas seja mais produtiva.

Palavras chave: Descoberta de Conhecimento em Bases de Dados, Mineração de Dados, Inteligência de Negócios, Tomada de Decisão.

ABSTRACT

This work, based on a literature review with the purpose of filtering the information to make it easier to understand, It has as goal to explain technical knowledge discovery using data mining, showing the stages and processes involved in transforming hard data into knowledge which can be valuable to organizations at the time of decision making. It also takes the concepts of Business Intelligence and Knowledge Discovery in Databases, in addition to the analysis of the features and functionality of Data Mining tools and how it can be used as data mining in the Account Court of Paraiba State , for identification of possible fraud or irregularities in the bidding process, according to the associations found in the data collected stored in the database of this institution. Thus the work of auditors can be streamlined, making the audit of public accounts becomes more productive.

Keywords: Knowledge Discovery in Databases, Data Mining, Business Intelligence, Decision Making.

LISTA DE FIGURAS

Figura 1: Processo de Descoberta do Conhecimento em Base de Dados.....	11
Figura 2: Refinamento dos dados na tomada de decisão.....	13
Figura 3: Fases do CRISP-DM <i>Process Model</i>	15
Figura 4: Diagrama de Contexto do SAGRES	28
Figura 5: Tabela de dados do módulo de Licitações utilizadas pelo SAGRES.....	31
Figura 6: Resultado do processamento dos dados minerados pelo RapidMiner:	34
Figura 7: Filtro realizado por proponente nas regras de associação geradas.....	36
Figura 8: Grafo gerado pelo RapidMiner com as regras de associação de um proponente.	36

LISTA DE TABELAS

Tabela 1: Comparativo entre ferramentas utilizadas para a mineração de dados.....	19
Tabela 2: Métodos de mineração de dados que podem ser aplicados em cada tarefa de KDD	19
Tabela 3: Base de dados da tarefa exemplo “Jogar tênis”	22
Tabela 4: Um exemplo de transações em cesta de compras.....	22
Tabela 5: Indícios de irregularidade nos processos licitatórios.....	30
Tabela 8: Data mart de Licitações. Fonte: SAGRES (2013).....	32
Tabela 9: Matriz Esparsa Licitações versus Proponentes.....	33
Tabela 10: Medidas de interesse objetivas geradas no exercício de 2010.....	35
Tabela 11: Medidas de interesse objetivas geradas no exercício de 2011.....	37
Tabela 12: Medidas de interesse objetivas geradas no exercício de 2012.....	38

LISTA DE SIGLAS

BI	<i>Business Intelligence</i>
CRISP-DM	<i>Cross Industry Standard Process for Data Mining</i>
DCBD	Descoberta de Conhecimento em Bancos de Dados
DM	<i>Data Mining</i>
KDD	<i>Knowledge Discovery in Databases</i>
LRF	Lei de Responsabilidade Fiscal
OLTP	<i>On-Line Transaction Processing</i>
PROMOEX	Programa de Modernização do Sistema de Controle Externo
SAGRES	O Sistema de Acompanhamento da Gestão dos Recursos da Sociedade
SI	Sistemas de Informação
SIE	Sistemas de Informação Estratégicos
SIG	Sistemas de Informação Gerenciais
TCE	Tribunal de Contas do Estado
TCEPE	Tribunal de Contas do Estado do Pernambuco

SUMÁRIO

RESUMO	V
ABSTRACT	VI
LISTA DE FIGURAS	VII
LISTA DE TABELAS.....	VIII
LISTA DE SIGLAS.....	IX
SUMÁRIO	X
1 INTRODUÇÃO	1
1.1 DEFINIÇÃO DO PROBLEMA DE PESQUISA.....	1
1.2 TRABALHOS RELACIONADOS	2
1.3 JUSTIFICATIVA	3
1.4 OBJETIVOS	4
1.4.1 <i>Objetivo geral</i>	4
1.4.2 <i>Objetivos específicos</i>	5
1.5 QUESTÕES DE PESQUISA.....	5
1.6 METODOLOGIA.....	5
1.7 RELEVÂNCIA DO ESTUDO.....	6
1.8 ORGANIZAÇÃO DO TRABALHO.....	6
1.8.1 <i>Resultados esperados</i>	6
2 FUNDAMENTAÇÃO TEÓRICA.....	8
2.1 A MINERAÇÃO DE DADOS	8
2.2 TIPOS DE PADRÕES	9
2.3 PASSOS PARA A MINERAÇÃO.....	11
2.4 DATA MINING NA ORGANIZAÇÃO.....	12
2.5 MOTIVOS PARA A MINERAÇÃO DE DADOS.....	13
2.6 FASES DA MINERAÇÃO DE DADOS.....	14
2.6.1 <i>Entendimento do Negócio</i>	16
2.6.2 <i>Seleção de Dados</i>	16
2.6.3 <i>Limpeza dos Dados</i>	16
2.6.4 <i>Modelagem de Dados</i>	16
2.6.5 <i>Avaliação do Processo</i>	17
2.6.6 <i>Execução</i>	17
2.7 PRINCIPAIS FERRAMENTAS DE MINERAÇÃO DE DADOS.....	18
2.8 ALGORITMOS DE MINERAÇÃO DE DADOS.....	19
2.9 APLICAÇÕES DE MINERAÇÃO DE DADOS	19
2.10 EXEMPLOS DE APLICAÇÕES DE MINERAÇÃO DE DADOS	21
2.11 CASES DE SUCESSO UTILIZANDO DATA MINING	23
3 OS PROCESSOS NO TRIBUNAL DE CONTAS.....	25

3.1	CONTROLE NO TRIBUNAL DE CONTAS.....	25
3.2	DATA MINING NO TRIBUNAL DE CONTAS	26
3.3	SISTEMA DE GESTÃO SAGRES.....	27
4	ASPECTOS METODOLÓGICOS.....	29
4.1	PRINCIPAIS INDÍCIOS DE FRAUDES OU IRREGULARIDADES EM LICITAÇÕES	29
4.2	DELIMITAÇÃO DO TEMA E SELEÇÃO DOS DADOS DO CASO.....	30
4.2.1	<i>Coleta dos dados.....</i>	<i>31</i>
4.3	ANÁLISE DOS DADOS	31
4.4	PROCEDIMENTO REALIZADO.....	33
5	ANÁLISE DOS RESULTADOS.....	33
6	CONCLUSÃO	39
	REFERÊNCIAS BIBLIOGRÁFICAS	41

1 INTRODUÇÃO

Neste capítulo será feita uma apresentação do problema de pesquisa, da justificativa do tema, dos objetivos gerais e específicos, da estrutura do trabalho e limitações da pesquisa.

1.1 DEFINIÇÃO DO PROBLEMA DE PESQUISA

A capacidade de armazenamento em banco de dados, assim como sua utilização, vem crescendo na mesma proporção dos avanços em novas tecnologias de informação e comunicação, isto segundo Quoniam, Tarapanoff, Araújo e Alvares (2001). Ainda de acordo com os autores a atividade de extrair informações relevantes, por conseguinte, está se tornando bastante complexa. Este processo de mineração de dados é chamado de *Knowledge Discovery in Databases* – KDD (Descoberta de Conhecimento em Base de Dados).

De acordo com Han e Kamber (2006), a mineração de dados tem atraído muita atenção na indústria da informação e na sociedade como um todo nos últimos anos, devido à grande disponibilidade de grandes quantidades de dados. E Freitas Jr. (2003) acrescenta que a Tecnologia da Informação torna-se extremamente útil ao processo decisório, pois possibilita coletar, processar e armazenar dados pela utilização de sistemas de informação. Assim a informação e o conhecimento adquirido podem ser usados para aplicações que vão desde análise de mercado, detecção de fraude e retenção de clientes, para controle de produção e exploração científica.

Neste âmbito o processo de tomada de decisão tem como base a informação e o conhecimento, que são de extrema importância para auxiliar os gerentes e apoiar as ações estratégicas da empresa. Mas para que esse apoio seja dado de forma concreta e confiável, são necessárias algumas adequações aos dados, informações e conhecimento.

Seguindo este ramo podem-se encontrar soluções de *Business Intelligence* (Inteligência de Negócios), que são usadas recorrentemente, visando oferecer maneiras necessárias para a transformação de dados em informações e assim dar suporte para o processo decisório. Business Intelligence pode ser definido como um conjunto de modelos matemáticos e metodologias de análise que exploram os dados disponíveis para gerar informações e conhecimentos úteis para os processos de tomada de decisões complexas (VERCELLIS, 2009).

Temos o *Data Mining* (Mineração de Dados) como uma das etapas do processo de BI. Segundo Han e Kamber (2006) a mineração de dados pode ser vista como resultado da evolução natural da tecnologia da informação. Avanços nas funcionalidades como a coleta de dados e criação de banco de dados, gerenciamento de dados e análise avançada dos dados têm sido testemunhada pelas indústrias de banco de dados.

De acordo com Cardoso e Machado (2008), o *data mining* usa ferramentas de análise estatísticas, técnicas de inteligência artificial, técnicas baseadas em regras e outros métodos inteligentes. Para eles este processo pode ser executado em um banco de dados operacional, ou em *data warehouse*, constituindo um sistema de suporte a decisão.

1.2 TRABALHOS RELACIONADOS

Alguns trabalhos encontrados na literatura dizem respeito à utilização de processos de mineração de dados na área da educação. Tais trabalhos estão descritos a seguir:

a) Carvalho, Sampaio e Mongiovi usaram técnicas de *data mining* para o reconhecimento de caracteres manuscritos.

b) Lemos (2003), desenvolveu um trabalho, usando técnicas de *data mining*, que tinha como objetivo principal gerar classificadores para que através dos mesmos pudesse fazer uma classificação de novas empresas como adimplentes ou inadimplentes.

c) Cardoso e Machado (2008) fizeram a utilização da mineração de dados para tratar dados disponíveis na plataforma Lattes, referentes à pesquisa científica na UFLA.

d) Almeida, Siqueira e Onusic (2005) utilizaram a mineração de dados em um estudo sobre o perfil dos candidatos que se inscreveram em um processo seletivo. Os resultados permitiram identificar diferentes públicos, relacionando regiões de origem, poder aquisitivo, meios de comunicação mais eficazes e motivos mais significativos. O objetivo era o auxílio a melhores ações de marketing voltadas a divulgação da instituição, especificamente do processo seletivo.

e) Batista (2010) fez a utilização de ferramentas de *data mining* para que pudesse reduzir, de forma mais eficiente do que apenas usando o conhecimento empírico do investigador, o número de atributos necessários para a formação do preço de habitação.

f) Saraiva e Masson (2003). No âmbito da UFRJ, Barroso e Falcão (2004) identificaram entre os alunos do Instituto de Física, os principais fatores que são obstáculos na conclusão dos cursos de graduação daquela unidade:

i) econômico - impossibilidade de permanecer no curso por questões sócio-econômicas;

ii) vocacional – o aluno não se identificou com o curso;

iii) institucional – abandono por fracasso nas disciplinas iniciais, deficiências prévias de conteúdos anteriores, inadequação aos métodos de estudo, dificuldades de relacionamento com colegas ou com membros da instituição.

g) Santos e Oliveira (2002) utilizaram técnicas de mineração de dados para descobrir e analisar as interações de usuários com um ambiente de ensino à distância analisando a navegação dos usuários. Com o resultado do trabalho o autor contribuiu com uma taxonomia de acessos para associar padrões na execução de atividades propostas durante o curso e análise de melhores *templates* de navegação.

1.3 JUSTIFICATIVA

Segundo Carvalho (2002), a análise das contas de entes públicos brasileiros é realizado tanto pelo controle interno quanto pelo controle externo, de acordo com a Constituição Federal de 1988. O Poder Legislativo é responsável pelo controle além do ente público, responsabilidade esta que é dividida com os Tribunais de Contas, enquanto o próprio ente se encarrega do controle interno, apoiado por um sistema.

É citado no Manual de Organização, Competências e Atribuições dos Órgãos Auxiliares do TCEPE (2011) que duas das competências da Coordenadoria da Tecnologia da Informação são “Identificar oportunidades de aplicação de TI para otimização dos trabalhos do Tribunal” e “Promover intercâmbio com os demais Tribunais de contas e outras entidades na área de tecnologia da informação”.

No entanto, Moraes (2006), garante que existe uma defasagem gerencial e tecnológica nos Tribunais de Contas, no que se refere a tecnologia da informação, que reflete em um exaustivo trabalho manual e artesanal, produzindo pilhas de processos, lentidão e ineficácia.

Ainda de acordo com este autor, a TI é usada basicamente para automatizar etapas dos procedimentos como: receber informações dos jurisdicionados, na automação de escritórios e no acompanhamento de processos. De maneira que pouco altera a eficiência e eficácia dos mesmos. Revelando o modelo burocrático de trabalho que predomina no ambiente, tornando a tramitação de processos lenta.

Grilo Jr. e Clericuzi (2008) afirmam que uma redução desta defasagem vem acontecendo com o passar do tempo, devido ao baixo custo da aquisição de equipamentos de informática e a grande disponibilidade de ferramentas de software disponíveis no mercado.

Dessa forma os Tribunais de Contas passaram a ter um estágio de informatização mais avançado, porém ainda mantem vários dos seus processos com a mesma sistemática de antes, com fluxo manual, deixando de usar todo o poder tecnológico que tem em mãos.

Pesquisas associadas ao PROMOEX (Programa de Modernização do Sistema de Controle Externo dos Estados, Distrito Federal e Municípios Brasileiros) e realizadas pela FIA-USP e FGV (Fundação Getúlio Vargas), evidenciam que os Tribunais de Contas ainda operam com muitas deficiências (VIEIRA, 2008).

A informação coletada no Portal dos Tribunais de Contas do Brasil mostra que o objetivo geral do PROMOEX é “fortalecer o sistema de controle externo como instrumento de cidadania, incluindo o aperfeiçoamento das relações intergovernamentais e interinstitucionais, com vistas, inclusive, ao controle do cumprimento da Lei de Responsabilidade Fiscal (LC 101/2000)”.

Portanto visa a modernização do conjunto de Tribunais de Contas dos estados, Distrito Federal e Municípios, por meio da capacitação de pessoal, introdução de novos procedimentos e sistemas organizacionais e investimentos em informática, com a intensão de manter uma melhor integração e aumentar a eficiência e eficácia das ações de fiscalização e controle, ajudando a regular a aplicação dos recursos públicos dos governos.

Assim esta pesquisa é justificada pelo intuito de relacionar as possíveis evidências de fraudes em processos licitatórios, através da implantação de técnicas de *data mining* em uma organização governamental, neste caso no Tribunal de Contas do Estado da Paraíba (TCE-PB), analisando tanto os aspectos operacionais quanto o processo analítico das operações armazenadas de forma eletrônica, o acúmulo de informações e o seu uso para análise e tomada de decisão.

1.4 OBJETIVOS

1.4.1 Objetivo geral

Este trabalho tem como objetivo geral relacionar possíveis evidências de fraudes em processos licitatórios no Tribunal de Contas do Estado da Paraíba, fazendo associações e identificando alguns padrões que se repetem nos dados com alguma frequência, aplicando técnicas de *data mining*.

1.4.2 Objetivos específicos

Para chegarmos ao objetivo geral é importante que alguns objetivos específicos sejam alcançados como:

- Analisar os conceitos de técnicas de *data mining*,
- Verificar a aplicabilidade das técnicas de *data mining* para o processo de fiscalização nos órgãos de controle externo.
- Definir base de dados a ser utilizada para avaliação das ferramentas.

1.5 QUESTÕES DE PESQUISA

- QP01. Quais as principais características e funcionalidades das ferramentas de *data mining*?
- QP02. É viável a utilização de técnicas de *data mining* nos Tribunais de Contas?

1.6 METODOLOGIA

O mesmo trabalho foi dividido nas seguintes etapas:

- Etapa I – Revisão Bibliográfica: análise de artigos, livros, revistas científicas sobre *data mining*, ferramentas de mineração de dados, entre outros;
- Etapa II – Entender o domínio aonde será aplicada o processo de *data mining*;
- Etapa III – Coleta de dados: levantamento de dados e ferramentas necessárias para aplicar o processo de *data mining*;
- Etapa IV – Limpeza e Pré-processamento dos dados;
- Etapa V – Transformação dos dados;
- Etapa VI – Escolha do algoritmo de mineração de dados e o método de seleção a ser utilizado na busca pelos padrões;
- Etapa VII – Mineração de dados;
- Etapa VIII – Interpretação dos padrões.
- Etapa IX – Análise e apresentação dos resultados: analisar e apresentar os resultados da avaliação das ferramentas de *data mining*.

1.7 RELEVÂNCIA DO ESTUDO

Este trabalho é relevante na área acadêmica, pois estudantes podem ter uma fundamentação e entendimento na análise do processo de mineração de dados e onde o mesmo pode ser usado, com exemplos e definição das etapas dessa técnica de descoberta de conhecimento. De forma que terão ideia de como poderão utilizá-la em trabalhos futuros na vida profissional. Além da possibilidade de analisar e examinar os dados provenientes desta pesquisa.

1.8 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em 6 capítulos, que são expostos da seguinte forma:

O Capítulo 1 apresenta a definição do problema, alguns trabalhos relacionados, a justificativa, os objetivos geral e específico, as questões de pesquisa a metodologia e a relevância do trabalho.

O Capítulo 2 mostra fundamentação teórica do trabalho, fazendo uma explanação em temas como, a mineração de dados, seus padrões e passos, o seu uso na organização, os motivos para se usar a mineração de dados. Também são apresentadas as fases da mineração de dados, as principais ferramentas do mercado para fazer a mineração, os algoritmos, aplicações e por fim alguns exemplos e casos de sucesso utilizando esta técnica.

O Capítulo 3 trata de uma apresentação do Tribunal de Contas e algumas explicações, como, o controle no Tribunal de Contas, o uso de *Data Mining* no mesmo. Mostra também um pouco do sistema de gestão SAGRES.

O Capítulo 4 trás os aspectos metodológicos do trabalho, onde são mostrado alguns casos em que podem ser identificadas possíveis fraudes e também é exposto as delimitações da execução do processo.

No Capítulo 5 é feita uma análise dos resultados obtidos após o processamento dos dados coletados realizado pela ferramenta de mineração. E finalmente, o Capítulo 6 mostra a conclusão e as considerações que podem ser feitas em cima dos resultados.

1.8.1 Resultados esperados

O resultado esperado nesta pesquisa é gerar informação que sirvam de auxílio a estudantes e profissionais da área que desejam utilizar a mineração de dados para gerar conhecimento a partir de uma grande quantidade de dados. E ainda proporcionar apoio para

uma melhor análise e investigação das contas geradas pelos órgãos públicos, procurando diminuir o desvio de recursos públicos através de procedimentos irregulares.

2 FUNDAMENTAÇÃO TEÓRICA

Atualmente, o volume de dados gerados nos diferentes processos organizacionais define um cenário onde é possível observar dois aspectos básicos: a necessidade de definir abordagens eficientes para interpretação dos dados e a exigência crescente de criação de mecanismos que garantam a transparência das informações.

De acordo com Aló (2009), por exemplo, a transparência possibilita o acesso, a facilidade de uso, a qualidade do conteúdo e o entendimento das informações que são do interesse de um determinado público. A cada ano mais operações estão sendo computadorizadas, todas acumulando dados em operações, em atividades e em desempenho. Todos estes dados “prendem” valiosas informações, como tendências e padrões, que podem ser usados para melhorar ou aperfeiçoar decisões de negócio (Goebel 1999). As ferramentas e técnicas empregadas para análise automática e inteligente destes dados são os objetos tratados pelo campo emergente da Descoberta de Conhecimento em Bancos de Dados (DCBD), da expressão em inglês *Knowledge Discovery in Databases* (KDD) (Silva 2011).

2.1 A mineração de dados

Mineração de dados, ou *data mining*, é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis (AMORIM, 2007). Um padrão pode ser definido como sendo uma afirmação sobre uma distribuição probabilística. Estes padrões podem ser expressos principalmente na forma de regras, fórmulas e funções, entre outras.

O interesse por este tipo de informação se deve principalmente ao fato de que as empresas e organizações estão coletando e armazenando grandes quantidades de dados como consequência da queda dos preços de meios de armazenamento e computadores e do aumento da capacidade de ambos (COSTA, 2005). Ainda segundo Costa (2005), a popularização na utilização de armazém de dados, ou *data warehouses*, que são grandes bancos de dados criados para análise e suporte à decisão, tende a aumentar ainda mais a quantidade de informações disponível. Não é indicado utilizar os métodos tradicionais de análise de dados, como planilhas e consultas, pois os mesmos não são apropriados para

tais, já que eles criam relatórios informativos, mas não conseguem fazer uma análise do conteúdo destes relatórios.

Com a geração de informações e conhecimentos úteis para as empresas, os seus negócios podem dar mais lucratividade para as mesmas. Os recursos da Tecnologia da Informação, mais precisamente a capacidade do hardware e software disponíveis podem efetuar atividades em horas, o que tradicionalmente as pessoas levariam meses, desse ponto de vista efetivamente o DM cumpre o papel de descoberta de conhecimentos.

Os sistemas tradicionais de processamento de transações on-line (OLTP) das empresas são ferramentas capazes de manipular dados de forma rápida, segura e efetiva em bancos de dados, mas que apresentam restrições para gerar informações com análises significativas. Estas restrições são melhores trabalhadas quando as empresas se utilizam da tecnologia de DM aliada à outras tecnologias, tais como, *Knowledge Discovery in Databases*, *Database Marketing* e Inteligência de Negócios (*business intelligence*).

A tecnologia *Data Mining*, e seus recursos, podem ser aplicados no funcionamento dos modelos de Sistemas de Informação Executivos, principalmente nos Sistemas de Informação Gerencial (SIG) e Estratégicos (SIE).

2.2 Tipos de Padrões

As tarefas de mineração de dados são geralmente divididas em duas categorias: tarefas de previsão (preditivos) e tarefas descritivas (descritivos ou informativos). Isto de acordo com Tan, Steinbach e Kumar (2009). Para Costa (2005), os padrões preditivos são encontrados para resolver o problema de prever o valor futuro ou desconhecido de um ou mais atributos do banco de dados a partir do valor conhecido dos demais atributos. Por outro lado os padrões descritivos, ou informativos, têm por objetivo encontrar padrões interessantes, de forma interpretável pelo homem, que descrevam os dados.

Uma tem como objetivo prever o valor de algum atributo baseada em outros que já tem seu valor calculado, esta é a tarefa de previsão, já as tarefas descritivas objetivam derivar padrões (correlações, tendências, grupos, trajetórias e anomalias) que resumam os relacionamentos adjacentes dos dados. Para os autores existem quatro tarefas centrais da mineração de dados: modelagem de previsão, análise de associação, análise de agrupamentos e detecção de anomalias. A seguir são descritas cada uma destas tarefas:

A **modelagem de previsão** é a tarefa onde se constrói um modelo para a variável alvo como uma função das variáveis explicativas. Neste processo há dois tipos de tarefas: classificação, usada para variáveis alvo discretas, e regressão, que é usada para variáveis alvo contínuas. As duas tarefas tem como objetivo, criar um modelo para minimizar o erro que pode ser encontrado entre os valores previstos e os valores reais da variável alvo. Um exemplo de uso é a identificação de clientes que responderão a uma campanha de vendas ou avaliar se um paciente possui uma determinada doença baseado nos resultados de exames médicos.

A **análise de associação** é a tarefa que tem como objetivo descobrir padrões que possam descrever características que tem um alto nível de associação dentro dos dados, estes padrões são usados normalmente em forma de regras de implicação ou subconjunto de dados. Um exemplo de uso inclui a descoberta de genes que possuem funcionalidade associada, ou a análise de uma cesta de compras, analisando quais produtos foram comprados juntos.

A análise de **agrupamentos** ou *clustering* procura encontrar grupos de observações relacionadas que partilham padrões e tendências semelhantes, nesta tarefa não existe uma variável dependente, não são consideradas as peculiaridades específicas, são agrupados os registros semelhantes em segmentos ou clusters, de forma automática, podendo ser considerada uma aprendizagem não supervisionada. Segundo Fayyad *et al.* (1996), o agrupamento é uma tarefa onde se procura identificar um conjunto finito de categorias ou agrupamentos para descrever os dados. Um exemplo de uso é o agrupamento de documentos, outro exemplo é juntar conjuntos de clientes relacionados ou com as mesmas afinidades, tornando mais simples a oferta de produtos ou promoções.

A **detecção de anomalias** é a tarefa de descobrir grupos, utilizando a similaridade dos valores de seus atributos, onde as características apresentadas sejam significativamente diferentes dos demais dados. Segundo Tan, Steinbach e Kumar (2009), o objetivo de um algoritmo de detecção de anomalias é descobrir as anomalias verdadeiras e evitar a rotulação errônea de objetos normais como anômalos. Um exemplo de uso é a detecção de fraudes em cartões de crédito, padrões incomuns de doenças e perturbações no meio ambiente.

A importância relativa de ambos no processo de mineração pode sofrer variação, porém, no contexto da descoberta de conhecimento em bancos de dados, os padrões

descritivos tendem a ser mais importantes do que os preditivos. No entanto, pesquisadores afirmam que este tipo de padrão é mais difícil de avaliar, tendo em vista que seu valor verdadeiro não deixa claro se ele sugere alguma ação para o especialista do domínio e quanto efetiva esta ação seria. Isto se deve ao fato de que a predição normalmente é utilizada quando se tem um problema claro e bem especificado a ser resolvido, sendo que se busca, através da mineração, uma resposta para este problema. No caso da descrição, tem-se apenas um volume de dados como ponto de partida. Cabe ao analista perceber se algo pode ser feito com as informações extraídas. A seguir serão apresentados alguns tipos de padrões comuns, sendo três deles descritivos (agrupamento, regras de associação e padrões sequenciais) e dois preditivos (regressão e classificação).

2.3 Passos para a mineração

Os passos fundamentais de uma mineração bem sucedida a partir de fontes de dados (bancos de dados, relatórios, logs de acesso, transações) se iniciam em uma limpeza (consistência, preenchimento de informações, remoção de ruído e redundâncias). Disto nascem os repositórios organizados (*Data Marts* e *Data Warehouses*), que já são úteis de diversas maneiras. Veja na figura 01 o processo de descoberta do conhecimento a partir de uma aplicação de *Data Mining*.

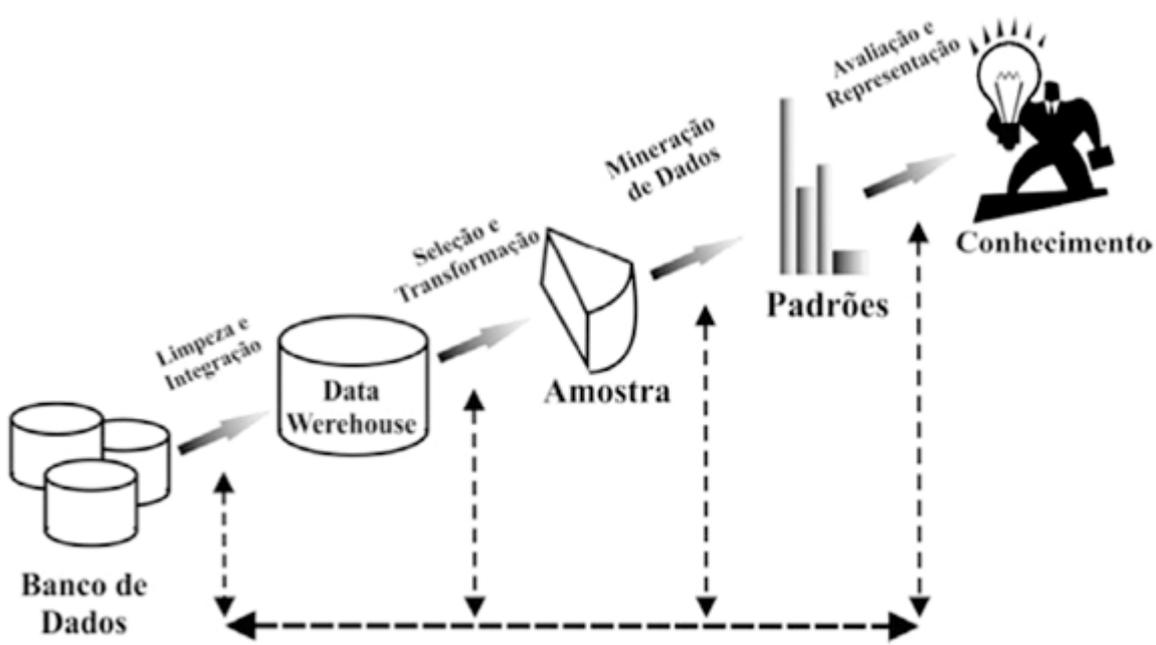


Figura 1: Processo de Descoberta do Conhecimento em Base de Dados , Site http://www.scielo.br/scielo.php?pid=S0102-77862009000400007&script=sci_arttext

De acordo com o professor Fabrício Varajão das Faculdades Integradas Campo-grandenses (2012), em geral o final de tudo não é aqui: o processo continua de forma interativa e frequentemente usando visualização gráfica, é necessário que o processo seja conduzido até que valiosos padrões apareçam. Observe que todo esse processo parece indicar uma hierarquia, algo que começa em instâncias elementares, com um grande número de dados, e terminam em um ponto relativamente concentrado, mas muito valioso.

O professor ainda defende que os dados brutos devem ser lapidados e sistematicamente simplificados de forma a desconsiderar aquilo que é específico e privilegiar aquilo que é genérico. Isto é feito, porque parece não se extrair muito conhecimento de eventos isolados, justamente por terem muitas características distintas que os isolam do restante. Uma loja de sua rede que tenha vendido a um cliente em particular uma quantidade impressionante de um determinado produto em uma única data pode apenas significar que esse cliente em particular procurava grande quantidade desse produto naquele exato momento. Mas isso provavelmente não indica nenhuma tendência de mercado.

Em outras palavras, esta informação isolada não nos trás conhecimento de mercado para que no futuro a empresa lucre mais. Apenas com conhecimento genérico é que isto pode ser obtido. Para que o processo dê certo, é necessário desprezar os eventos particulares para só manter aquilo que é genérico.

2.4 Data mining na organização

O fundamental a se perceber no diagrama abaixo é a sensível redução de volume que ocorre cada vez que subimos de nível. Essa redução de volume é uma natural consequência do processo de abstração.

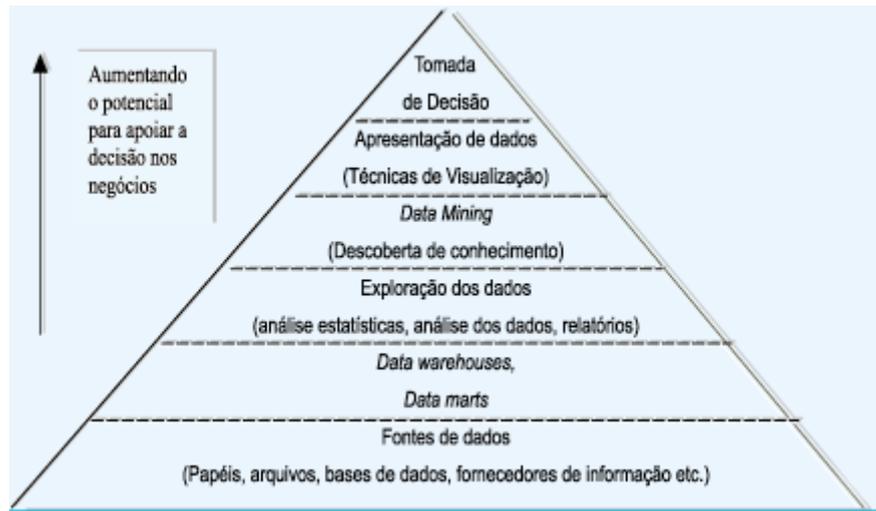


Figura 2: Refinamento dos dados na tomada de decisão, adaptado do site http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000200004

Abstrair, no sentido que usamos aqui, é representar uma informação através de correspondentes simbólicos e genéricos. Para ser genérico, é necessário "perder" um pouco dos dados, para só conservar a essência da informação (*et Al.* MARTINHAGO). O processo de *Data Mining* localiza padrões através da judiciosa aplicação de processos de generalização, algo que é conhecido como indução.

2.5 Motivos para a mineração de dados

As técnicas da Mineração de Dados passaram a ser usadas como exploração de dados, por vários motivos (Carvalho, 2005):

- O volume de dados disponível atualmente é enorme – Mineração de Dados é uma técnica que só se aplica a grandes massas de dados, pois necessita disto para calibrar seus algoritmos e extrair dos dados conclusões confiáveis. Empresas de telefonia, cartões de crédito, bancos, televisão por assinatura, comércio eletrônico, entre outras, vem gerando a cada dia uma grande quantidade de dados sobre seus serviços e clientes. Estes dados são passíveis de análise por mineração;
- Os dados estão sendo organizados, com a tecnologia *data warehouse* – Os dados de várias fontes estão sendo organizados e padronizados de forma a possibilitar sua organização dirigida para o auxílio à decisão. As técnicas de mineração de dados necessitam de bancos de dados limpos, padronizados e organizados;

- Os recursos computacionais estão cada vez mais potentes, a mineração de dados necessita de muitos recursos computacionais para operar seus algoritmos sobre grandes quantidades de dados – O aumento da potência computacional, devido ao avanço tecnológico e à queda dos preços dos computadores, facilita o uso da mineração de dados atualmente. O avanço da área de banco de dados, construindo bancos de dados distribuídos, também auxiliou em muito à mineração de dados;
- A competição empresarial exige técnicas mais modernas de decisão - As empresas da área de finanças, telecomunicações e seguro experimentam a cada dia mais competição. Como estas empresas sempre detiveram em seus bancos de dados uma enorme quantidade de informação, é natural que a mineração de dados tenha se iniciado dentro de seus limites. Atualmente, outras empresas buscam adquirir dados para analisar melhor seus caminhos futuros através dos sistemas de apoio à decisão.

2.6 Fases da mineração de dados

Para melhor contextualizar o ambiente de mineração de dados é interessante mostrar um pouco do projeto CRISP-DM. Em 1996, um conjunto de três empresas especializadas no então jovem e imaturo mercado de *data mining*, desenvolveram um modelo de processos genéricos, com o intuito de padronizar as etapas do processo de mineração de dados, dando início ao denominado projeto CRISP-DM (*Cross Industry Standard Process for Data Mining*) (*The CRISP-DM Consortium, 2000*). Este projeto desenvolveu um modelo de processo de mineração de dados industrial e livre de ferramenta. Começando pelos embrionários processos de descoberta de conhecimento usados nos primeiros projetos de mineração de dados e respondendo diretamente aos requerimentos do usuário, esse projeto definiu e validou um processo de mineração de dados que é aplicável em diversos setores da indústria.

Essa metodologia torna projetos de mineração de dados de larga escala mais rápidos, mais baratos, mais confiáveis e mais gerenciáveis. Até mesmo projetos de mineração de dados de pequena escala se beneficiam com o uso do CRISP-DM. O modelo CRISP, atualmente, é uma referência para que seja desenvolvido um plano de integração para a descoberta de conhecimento. O atual processo para mineração de dados propõe uma visão geral do ciclo de vida de um projeto de mineração de dados. Ele contém as fases

correspondentes de um projeto, suas respectivas tarefas e relacionamentos entre essas tarefas.

Na Figura 03 é mostrado o ciclo de vida de um projeto de mineração de dados, que é formado por 6 (seis) fases. A sequência de fases não é obrigatória, a transição pode acontecer para diferentes fases, dependendo do resultado de cada uma, e que etapa particular de cada fase precisa ser executada na sequência. As setas indicam as mais importantes e mais frequentes dependências entre as fases. O ciclo externo na figura simboliza o ciclo natural da mineração de dados. Um processo de mineração de dados continua após a solução ter sido desenvolvida. As lições aprendidas durante o processo podem provocar perguntas novas, frequentemente mais pertinentes ao negócio. Processos subsequentes se beneficiarão das experiências de processos anteriores.

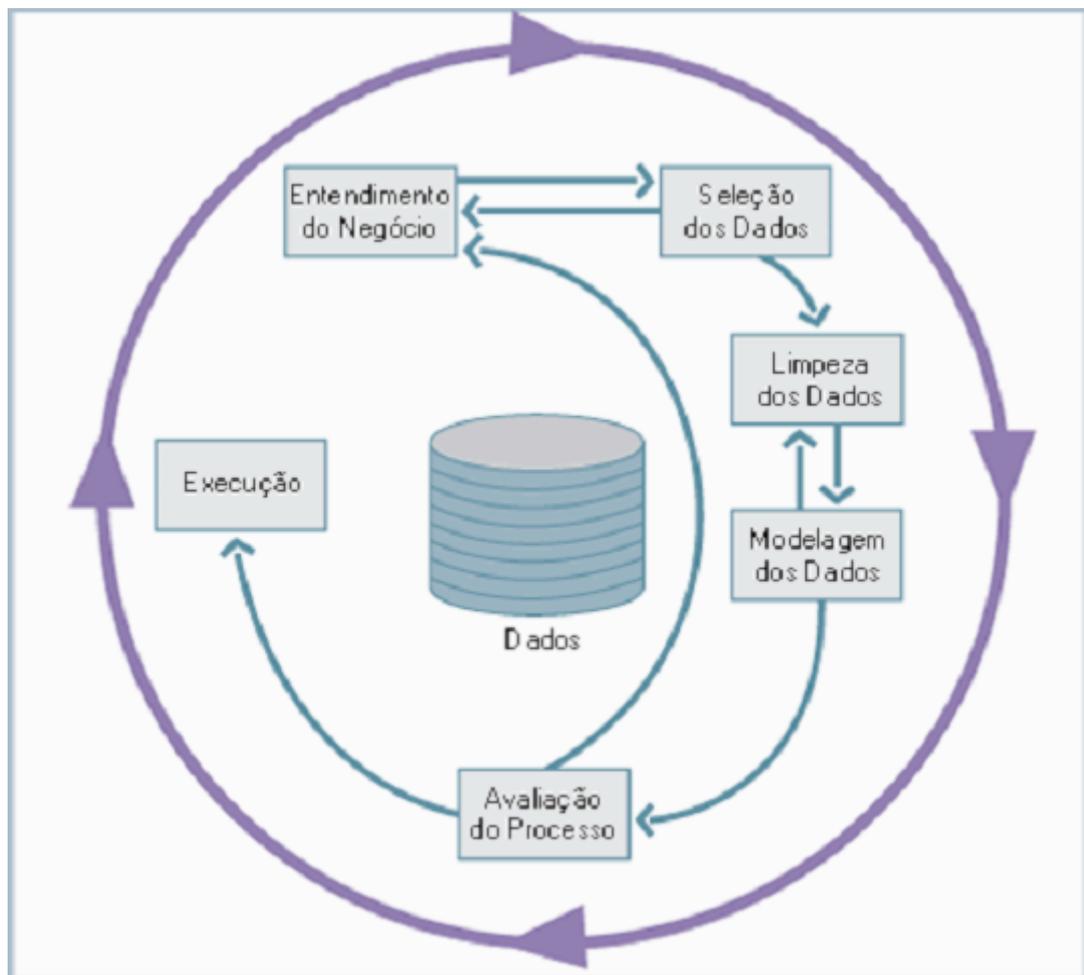


Figura 3: Fases do CRISP-DM *Process Model* (Baseado em The CRISP-DM Consortium, 2000).

Segue abaixo uma síntese das etapas pertencentes ao modelo CRISP (*The CRISP-DM Consortium, 2000*):

2.6.1 Entendimento do Negócio

Business Understanding, essa fase inicial tem o foco no entendimento do negócio que visa obter conhecimento sobre os objetivos do negócio e seus requisitos, e então converter esse conhecimento em uma definição de um problema de mineração de dados, e um plano preliminar designado para alcançar esses objetivos.

2.6.2 Seleção de Dados

Data Understanding, consiste no entendimento dos dados, que visa à familiarização com o banco de dados pelo grupo de projeto, utilizando-se de conjuntos de dados "modelo". Uma vez definido o domínio sobre o qual se pretende executar o processo de descoberta, o próximo passo é selecionar e coletar o conjunto de dados ou variáveis necessárias. Essa fase se inicia com uma coleta inicial de dados, e com procedimentos e atividades visando a familiarização com os dados, para identificar possíveis problemas de qualidade, ou detectar subconjuntos interessantes para formar hipóteses.

2.6.3 Limpeza dos Dados

Data Preparation, a fase de preparação de dados consiste na preparação dos dados que visa a limpeza, transformação, integração e formatação dos dados da etapa anterior. É a atividade pela qual os ruídos, dados estranhos ou inconsistentes são tratados. Esta fase abrange todas as atividades para construir o conjunto de dados final (dados que serão alimentados nas ferramentas de mineração), a partir do conjunto de dados inicial. A utilização de *Data Warehouses* facilita em muito esta etapa do processo de mineração de dados, que costuma ser a fase que exige mais esforço, correspondendo geralmente a mais de 50% do trabalho. Por isso, é muito importante para uma organização, que ela possua em seus processos habituais boas práticas da administração de dados, como o *Data Cleansing*, que é uma parte fundamental da cadeia da administração da informação, responsável pelas etapas de detecção, validação e correção de erros em bases de dados (Chapman, 2005).

2.6.4 Modelagem de Dados

Modeling, fase que consiste na modelagem dos dados, a qual visa a aplicação de técnicas de modelagem sobre o conjunto de dados preparado na etapa anterior. Nessa fase,

várias técnicas de modelagem são selecionadas e aplicadas, e seus parâmetros são calibrados para se obter valores otimizados. Geralmente, existem várias técnicas para o mesmo tipo de problema de mineração. Algumas técnicas possuem requerimentos específicos na forma dos dados. Consequentemente, muitas vezes é necessário voltar para a etapa de preparação de dado. A maioria das técnicas de mineração de dados é baseada em conceitos de aprendizagem de máquina, reconhecimento de padrões, estatística, classificação e clusterização.

2.6.5 Avaliação do Processo

Evaluation, a avaliação do processo visa garantir que o modelo gerado atenda às expectativas da organização. Os resultados do processo de descoberta do conhecimento podem ser mostrados de diversas formas. Porém, estas formas devem possibilitar uma análise criteriosa para identificar a necessidade de retornar a qualquer um dos estágios anteriores do processo de mineração. Nesta etapa se construiu um modelo que parece de alta qualidade, de uma perspectiva da análise de dados. Antes de prosseguir, é importante avaliar mais detalhadamente o modelo, e rever as etapas executadas para construir o modelo, para se certificar de que ele conseguirá alcançar os objetivos de negócio. Deve se determinar se houve algum importante objetivo do negócio que não foi suficientemente alcançado. No fim desta fase, uma decisão sobre o uso dos resultados da mineração deve ser tomada.

2.6.6 Execução

Deployment, esta fase consiste na definição das fases de implantação do projeto de Mineração de Dados. A criação do modelo não é o fim do projeto. Mesmo se a finalidade do modelo for apenas aumentar o conhecimento dos dados, o conhecimento ganho necessitará ser organizado e apresentado em uma maneira que o cliente possa usar. Dependendo das exigências, a fase de execução pode ser tão simples quanto a geração de um relatório, ou tão complexo quanto executar processos de mineração de dados repetidamente. Em muitos casos será o cliente, não o analista dos dados, que realizará as etapas da execução. Entretanto, mesmo se o analista não se encarregar da execução é importante que ele faça o cliente compreender que medidas deverão ser tomadas a fim de empregar efetivamente os modelos criados.

2.7 Principais Ferramentas de mineração de dados

Na Tabela 1 temos relacionadas algumas ferramentas para mineração de dados, também são mostradas suas características e qual tarefa realizam dentro do processo de descoberta de conhecimento, além de informar alguns domínios onde as mesmas estão sendo utilizadas.

FERRAMENTA	CARACTERÍSTICAS	TAREFA DE KDD	DOMÍNIOS UTILIZADOS	FABRICANTE
SPSS/ Clementine	Permite o desenvolvimento rápido de modelos preditivos para as operações da corporação, melhorando a tomada de decisão.	Classificação, Regras de Associação, Clusterização, Sequência e Detecção de Desvios.	Associação Comercial de São Paulo, Credicard, CTBC Telecom, DirecTV, Globo.com, entre outras.	SPSS Inc. www.spss.com
PolyAnalyst		Classificação, Regressão, Regras de Associação, Clusterização, Sumarização e Detecção de Desvios.		Megaputer Intelligence www.megaputer.com
Intelligent Miner		Classificação, Regras de Associação, Clusterização e Sumarização.		IBM Corp. www.ibm.com
WizRule		Sumarização, Classificação e Detecção de Erros.		WizSoft Inc. www.wizsoft.com
SAS Enterprise Miner		Classificação, Regras de Associação, Clusterização, Agrupamento, .	Bank of America, Telefonica O2, Korea Customs Service, Australian Bureau of Statistics, entre outros.	SAS Corp. www.sas.com
Tamanduá		Associação, Agrupamento e Classificação	Auditoria-Geral do Estado de Minas Gerais; Sec.de Log.e TI do Min. do Planej., Orçamento e Gestão; Min.da Justiça; CGU.	Depart. Ciência da Comput. da UFMG http://tamandua.speed.decc.ufmg.br
Oracle Data Mining		Classificação, Regressão, Associação, Clusterização e Mineração de Textos.		Oracle www.oracle.com
WEKA	API e ambiente de testes com algoritmos de mineração de dados e aprendizado por computador.	Classificação, Regressão e Regras de Associação, Clusterização.		University of Waikato www.cs.waikato.ac.nz
RapidMiner (antigo YALE)	Derivado do WEKA, é um pacote mais completo de mineração de dados.	Classificação, Regressão e Regras de Associação, Clusterização.	Ford, Honda, Nokia, Miele, Philips, IBM, HP, Cisco, Bank of America, entre outras.	Rapid-I Rapid-i.com

Tabela 1: Comparativo entre ferramentas utilizadas para a mineração de dados, Fonte: Baseado em GOLDSCHMIDTH; PASSOS (2005); TAMANDUÁ (2010).

2.8 Algoritmos de mineração de dados

Na Tabela 2 são apresentados alguns dos principais métodos de mineração que podem ser usados nas diferentes tarefas da descoberta de conhecimento, não foram abortados todos os algoritmos que podem ser utilizados, apenas alguns métodos que são executados nas tarefas.

TAREFA DE KDD	MÉTODOS DE MINERAÇÃO DE DADOS
Descoberta de associações	Basic, Apriori, DHP, Partition, DIC, ASCX-2P
Descoberta de associações generalizadas	Basic, Apriori, DHP, Partition, DIC, ASCX-2P
Descoberta de sequências	GSP, MSDD, SPADE
Descoberta de sequências generalizadas	GSP, MSDD, SPADE
Classificação	Redes Neurais (Ex.: Back-Propagation, RBF), C4.5, Rough, Sets, Algoritmos genéricos (Ex.: Rule Evolver), CART, K-NN, Classificadores Bayesianos.
Regressão	Redes Neurais (Ex.: Back-Propagation), Lógica Nebulosa
Sumarização	C4.5, Algoritmos genéricos (Ex.: Rule Evolver)
Clusterização	K-Means, K-Modes, K-Prototypes, Fuzzy K-Means, Algoritmos genéricos, Redes Neurais (Ex.: Kohonen)
Previsão de Séries Temporais	Redes Neurais (Ex.: Back-Propagation), Lógica Nebulosa (Ex.: Wang-Mendel)

Tabela 2: Métodos de mineração de dados que podem ser aplicados em cada tarefa de KDD, Fonte: GOLDSCHMIDTH; PASSOS (2005).

Na mineração de dados, vários tipos de conhecimento podem ser extraídos, graças a grande diversidade de atividades relacionadas com o processo. Então é necessário especificar e decidir, no início do processo, quais tarefas devem ser executadas, qual o conhecimento que se quer extrair ou quais padrões intrínsecos aos dados poderão ser revelados. Segundo Fayyad *et al.* (1996), não há um método de mineração de dados universal e a escolha de um algoritmo particular é de certa forma uma arte.

2.9 Aplicações de mineração de dados

As aplicações contendo regras de associação podem ser usadas em várias áreas de aplicação, segundo Versellis (2009) ela é particularmente recomendada para:

Marketing relacional (*relational marketing*): as aplicações de mineração de dados nesta área deram grande contribuição para o aumento na popularidade dessas metodologias. Algumas aplicações relevantes dentro do marketing relacional são:

- A identificação de segmentos de clientes que são mais propensos a responder a campanhas de marketing, tais como *cross-selling* e *up-selling*;
- A identificação de segmentos de clientes alvo para campanhas de retenção;
- Previsão da taxa de respostas positivas às campanhas de marketing;
- Interpretação e compreensão do comportamento de compra dos clientes;
- Análise dos produtos adquiridos pelos clientes em conjunto, conhecido como análise de cesta de compras.

O *up-selling* é uma estratégia de venda que consiste em sugerir um adicional para o produto ou serviço que está sendo comprado, levando o cliente a comprar bens/serviços mais caros de forma a produzir uma venda mais rentável. Já o *cross-selling* é uma técnica de vendas onde são sugeridos produtos complementares.

Segundo Giudici (2003) a análise de cestas de compras é um estudo para compreensão das associações entre comportamentos de compra. Uma forma semelhante de análise pode ser aplicada a problemas em que o objetivo principal é *cross-selling* para aumentar o número de produtos que são comprados em uma determinada unidade comercial (supermercado, um banco, uma agência de viagens, ou de forma generalizada, uma empresa oferecendo mais de um produto ou serviços).

Detecção de fraude (*fraud detection*): a detecção de fraudes é um outro campo relevante nas aplicações de mineração dados. Algumas áreas como telefonia, seguros, cartões de crédito, cheques bancários, órgãos públicos, podem ser afetados por fraudes.

Avaliação de riscos (*risk evaluation*): o objetivo da análise de risco é estimar o risco ligado com as futuras decisões, que muitas vezes são bifurcadas. Por exemplo, um banco pode desenvolver um modelo preditivo para determinar se é conveniente conceder um empréstimo monetário ou um empréstimo à habitação, com base nas características do pretendente.

Mineração de texto (*text mining*): a mineração de dados pode ser aplicada a diferentes tipos de textos, a fim de classificar os artigos, livros, documentos, e-mails e páginas web. Temos como exemplo os motores de pesquisa na web ou a classificação

automática de comunicados de imprensa para fins de armazenamento. Outras aplicações de mineração de texto incluem a geração de filtros de mensagens de *e-mail* e *newsgroups*.

Reconhecimento de imagem (*image recognition*): o tratamento e classificação de imagens digitais, tanto estáticas quanto dinâmicas, é um assunto muito interessante, tanto para estudo como para um grande número de aplicações que as utilizam. É útil para reconhecer caracteres escritos, comparar e identificar rostos humanos, aplicar filtros de correção para equipamento fotográfico e detectar comportamentos suspeitos através de vigilância por câmeras de vídeo.

Web mining: aplicações de mineração na Web são destinadas à análise dos chamados *clickstreams* ou sequência de cliques, são as sequências de páginas visitadas e as escolhas feitas por um internauta. Eles podem ser úteis para a análise de sites de comércio eletrônico, ao oferecer a flexibilidade e páginas personalizadas para os seus usuários.

Diagnóstico médico (*medical diagnosis*): modelos de aprendizagem são ferramentas valiosas na área médica para a detecção precoce de doenças usando os resultados dos testes clínicos.

2.10 Exemplos de aplicações de mineração de dados

Serão apresentados a seguir alguns exemplos de aplicações de técnicas de mineração de dados para os casos de modelagem de previsão, análise de associação e agrupamento:

Modelagem de previsão → Para ilustrar a aplicação deste método, considere os dados constantes na Tabela 3:

Aparência	Temperatura	Umidade	Vento	Jogar ténis?
Ensolarado	Quente	Alta	Fraco	Não
Ensolarado	Quente	Alta	Forte	Não
Nublado	Quente	Alta	Fraco	Sim
Chuvoso	Moderado	Alta	Forte	Sim
Chuvoso	Fresco	Normal	Fraco	Sim
Chuvoso	Fresco	Normal	Forte	Não
Nublado	Fresco	Normal	Forte	Sim
Ensolarado	Moderado	Alta	Fraco	Não
Ensolarado	Fresco	Normal	Fraco	Sim
Chuvoso	Moderado	Normal	Fraco	Sim

Ensolarado	Moderado	Normal	Forte	Sim
Nublado	Moderado	Alta	Forte	Sim
Nublado	Quente	Normal	Fraco	Sim
Chuvoso	Moderado	Alta	Forte	Não

Tabela 3: Base de dados da tarefa exemplo “Jogar tênis” Fonte: Adaptado de Goldschmidth (2011, p. 61)

Nesta ilustração de Goldschmidth (2011), iremos considerar o atributo “Jogar tênis” como o objetivo da classificação, este problema tem duas classes: “jogar = sim” e “jogar = não”. Se desejarmos saber se devemos ou não jogar em determinadas circunstâncias basta fazer uma pergunta e inserir os dados no algoritmo para obter o resultado, por exemplo: devo jogar tênis em dia ensolarado, quente, de alta umidade e com vento fraco?

No exemplo podemos utilizar o Teorema de Bayes, relacionado ao cálculo de probabilidades condicionais. Pelo Teorema de Bayes podemos afirmar que a $P(A|B) = (P(B|A)*P(A))/P(B)$, assim, substituindo os valores da nossa problemática na fórmula temos:

$$P(\text{jogar} = \text{sim} \mid \text{ensolarado, quente, alta umidade, vento fraco}) = P(\text{ensolarado} \mid \text{jogar} = \text{sim}) * P(\text{quente} \mid \text{jogar} = \text{sim}) * P(\text{alta umidade} \mid \text{jogar} = \text{sim}) * P(\text{vento fraco} \mid \text{jogar} = \text{sim}) = 0,0071$$

$$P(\text{jogar} = \text{não} \mid \text{ensolarado, quente, alta umidade, vento fraco}) = P(\text{ensolarado} \mid \text{jogar} = \text{não}) * P(\text{quente} \mid \text{jogar} = \text{não}) * P(\text{alta umidade} \mid \text{jogar} = \text{não}) * P(\text{vento fraco} \mid \text{jogar} = \text{não}) = 0,0274$$

Portanto, a resposta do algoritmo seria **Jogar = Não**.

Análise de associação → Uma tarefa de associação busca por padrões que demonstrem o relacionamento entre conjuntos de itens, para ilustrar a aplicação deste método, considere os dados apresentados na Tabela 4.

Identificador	Item
100	Pão, leite, manteiga.
200	Pão, requeijão, leite.
300	Manteiga, farinha, leite.
400	Manteiga, pão, refrigerante.
500	Bolacha, leite, manteiga.

Tabela 4: Um exemplo de transações em cesta de compras. Fonte: Adaptado de Pizzi (2006).

De acordo com Pizzi (2006), analisando a regra *manteiga* \rightarrow *pão*, nota-se que dentre as 5 transações existentes, pão e manteiga ocorrem em 2 transações. Além disso, dentre as 4 transações em que manteiga ocorre, pão ocorre em 2 dessas transações. Com isso, pode-se dizer que a regra *manteiga* \rightarrow *pão* possui suporte de 40% e confiança de 50%, sendo que o fator de suporte determina a frequência na qual uma regra é aplicável a um determinado grupo de dados, o que demonstra que essa regra pode revelar um padrão de comportamento dos clientes: “clientes que compram manteiga tendem a comprar pão”.

Algoritmos de análise de associação possui um potencial de gerar uma variedade enorme de padrões com a combinação dos itens analisados, conforme são ajustados os limites de suporte e confiança.

2.11 Cases de sucesso utilizando data mining

Alguns *cases* de sucesso com o uso de técnicas de mineração de dados são facilmente encontrados na literatura e em consulta rápida na internet. A seguir serão listadas algumas aplicações de *data mining* e quais os benefícios oriundos com a análise dos seus resultados:

- Wal-Mart

Um dos casos mais divulgados é o da rede de supermercados americana *Wal-Mart*, que identificou um hábito curioso dos seus consumidores, segundo Murayama (2002), ao procurar relações entre o volume de vendas e os dias da semana, o *software* utilizado para exploração de dados apontou que, às sextas-feiras, as vendas de cerveja cresciam na mesma proporção que as de fraldas. Uma investigação mais detalhada revelou que, ao comprar fraldas para seus bebês, os pais aproveitavam para fazer o abastecimento de cervejas para o final de semana. A nova informação foi confirmada e uma ação válida foi o lançamento imediato da promoção e localização próxima de fraldas e cervejas. Este exemplo pode ser utilizado para exemplificar os recursos que uma ferramenta de *data mining* e *data warehouse* podem oferecer.

- Bank of America

Segundo Braz *et al.* (2009), um outro caso de sucesso é o do *Bank of America*, que utilizou técnicas de mineração de dados para selecionar dentre seus clientes, aqueles com menor possibilidade de não quitar os seus empréstimos. Com base nesses relatórios, foi feito um trabalho de marketing na oferta de crédito aos correntistas selecionados. Em apenas três anos, o *Bank of America* aumentou sua lucratividade em cerca de trinta milhões de dólares.

A detecção o de fraudes é uma das aplicações mais visadas pelos gerentes que procuram por soluções em *data mining*. Diversos bancos recorrem a esse recurso para avaliar a credibilidade de seus clientes.

- Outras aplicações

No setor público, a mineração de dados é utilizada, principalmente, como forma de descobrir fraudes. Ainda segundo Braz *et al.* (2009), os governos brasileiro e norte americano, por exemplo, monitoram transações financeiras procurando detectar padrões similares à lavagem de dinheiro do narcotráfico.

Ainda segundo Braz *et al.* (2009) podemos usar a mineração de dados na área de segurança pública de várias formas, por exemplo: determinar os locais com maior criminalidade, definir perfis de vítimas e criminosos, identificar a existência de quadrilhas e *serial killers*, detectar quais dias da semana em que ocorrem mais delitos e até mesmo as suas causas. Os benefícios em utilizar mineração de dados são muitos, por exemplo, descobrir os bairros que são mais perigosos pode induzir a uma melhor alocação dos policiais, bem como definir o perfil das vítimas pode ajudar a traçar campanhas preventivas focadas no público que se encaixa nesse perfil.

3 OS PROCESSOS NO TRIBUNAL DE CONTAS

3.1 Controle no Tribunal de Contas

O art. 70 da Constituição Federal de 1988 (BRASIL, 2009), determina que o exame das contas dos entes públicos brasileiros deve ser realizado pelo controle interno e externo. O controle externo fica sob responsabilidade do Poder Legislativo, com auxílio dos Tribunais de Contas, ao passo que o controle interno será levado a efeito por sistema do próprio ente público (CARVALHO, 2002).

O controle externo é exercido por um órgão que não faça parte da estrutura a qual pertencem os atos que devem ser controladas, como o controle que o Poder Judiciário pode vir a realizar sobre atos do Poder público, o controle direto ou indireto a cargo do Congresso Nacional e o realizado pelo Tribunal de Contas (NETO, 2004).

De acordo com a constituição os Tribunais de Contas, tem a responsabilidade de fiscalizar a parte contábil, financeira, operacional e patrimonial do uso e aplicação dos recursos públicos brasileiros, exercendo o papel da fiscalização da eficiência no uso do dinheiro público. Desta forma, é possível perceber a necessidade de instrumentos interessantes e eficazes para a detecção e correção da utilização de forma inapropriada dos recursos públicos, dado que as funções dos Tribunais de Contas têm dimensões significativas.

Segundo o TCE-PB (2009), os Tribunais de Contas tem como função, acompanhar a gestão, fiscalizar e analisar os processos de prestações de contas, atos de admissão de pessoal, aposentadorias e pensões, licitações, contratos e convênios. No Tribunal de Contas, os dados técnicos apresentados, ou seja, os demonstrativos contábeis deverão necessariamente ser tabulados, analisados e confrontados com a legislação regente e emitido parecer sobre sua aprovação ou desaprovação (MELO, 2005).

De acordo com Santana (2008), os Tribunais de Contas, atualmente, possuem as seguintes atribuições que estão estabelecidas na nossa Constituição Federal nos artigos 70 a 75:

- Apreciar as contas do chefe do Executivo;
- Julgar as contas dos Administradores Públicos e demais responsáveis por dinheiro, bens e valores públicos da administração direta e indireta;

- Appreciar, para fins de registro, a legalidade da contratação de pessoal, concessões de aposentadoria, reforma e revisão;
- Realizar auditoria e inspeções de natureza contábil, financeira, orçamentária, operacional e patrimonial nas entidades da administração direta e indireta, quanto à legalidade, legitimidade, economicidade, aplicação das subvenções e renúncia de receitas;
- Fiscalizar a aplicação de recursos públicos repassados;
- Aplicar sanções em caso de irregularidades;
- Determinar a adoção de providências para o exato cumprimento da Lei;
- Sustar ato impugnado se não atendido;
- Representar ao Poder competente sobre irregularidades ou abusos apurados.

Figueirêdo (2002) afirma que, o cumprimento da Lei de Responsabilidade Fiscal (LRF) deve ser verificado pelos Tribunais de Contas. A LRF está erguida sobre alguns pilares, entre eles está o da transparência. O objetivo mais nobre do princípio da transparência é permitir e estimular o exercício do controle social, que pode ser considerada a forma mais eficaz de vigilância da conduta do gestor público.

As Ouvidorias e o Ministério Público também fazem parte do controle externo. O grande diferencial é que o controle das Ouvidorias e do Ministério Público não depende da vontade política das autoridades a ser fiscalizadas, ou das autoridades fiscalizadoras quando atuam, mediante denúncia, representação ou o conhecimento de eventuais irregularidades, mas sim, cumprem determinações legais e não podem deixar de fazê-las, sob pena de prevaricação.

3.2 Data Mining no Tribunal de Contas

O processo de tomada de decisão em organizações sem fins lucrativos, segundo Slack *et al.* (2007), como no caso dos Tribunais de Contas, podem ser mais complexas pois envolvem uma maior diversidade de pontos de vista e de objetivos, por exemplo, objetivos políticos, econômicos, sociais ou ambientais, por causa disso, existe maior chance de que as decisões de operações ocorram em condições de objetivos conflitantes. Mesmo com

envolvendo esta complexidade, atualmente os Tribunais de Contas buscam ferramentas que auxiliem o processo de fiscalização, transparência e para a tomada da decisão.

As contas públicas são consideradas regulares, quando expressarem, de forma clara e objetiva, a exatidão dos demonstrativos contábeis, e atenderem aos princípios da legalidade, legitimidade, economicidade e razoabilidade dos atos de gestão do responsável (TCE-SE, 2009).

O processo de análise das contas públicas pode receber uma nova dinâmica com a adoção da tecnologia de BI, e conseqüentemente proporcionar um aumento na produtividade através de respostas mais rápidas de forma automática, reduzindo o tempo total de análise dos processos, envolvendo desde a fase de análise inicial do processo até o parecer final destas contas pelos Tribunais de Contas. Observemos os seguintes casos:

No Tribunal de Contas do Estado de Pernambuco (TCE-PE) foi realizado um estudo sobre *data mining* e o resultado deste, apresentou uma solução de mineração de dados, com a extração do conhecimento presente na própria base que compõe o estoque de processos do TCE-PE (SOUSA *et al.*, 2008), para a proposição de um Sistema de Apoio à Decisão (SAD) como instrumento de atuação efetiva na celeridade de execução dos processos daquela Corte de Contas.

É apresentado por Santana (2008) o Sistema de Acompanhamento da Gestão dos Recursos da Sociedade (SAGRES), criado e implantado pelo Tribunal de Contas do Estado da Paraíba (TCE-PB) em 2002, que possui uma ferramenta de consulta que utiliza conceitos básicos de mineração de dados e *data warehouse*. Para o autor, este sistema teve como objetivo dar suporte para que sejam realizadas as novas atribuições que a LRF impôs.

A LRF normatiza os princípios do equilíbrio - não gastar mais do que se arrecada; do planejamento - planejar para executar; da transparência - prestar contas e informar à sociedade; e da participação - governar com o cidadão.

3.3 Sistema de gestão SAGRES

Segundo Grilo Jr. (2010), o SAGRES é um sistema de informação que tem como objetivo, apoiar a atividade de controle externo, que tem como base a coleta mensal e a análise e disponibilização de dados sobre as ações orçamentárias e financeiras, sobre

licitações, contratos administrativos e sobre os gastos com pessoal das unidades gestoras municipais e estaduais.

O SAGRES possui diferentes módulos para atender as várias partes interessadas na fiscalização dos entes públicos são eles: um módulo de captura, um de armazenamento e outro de consultas dos dados dos seus jurisdicionados. O SAGRES também disponibiliza no web site denominado SAGRES ON-LINE alguns módulos deste sistema para que a sociedade possa exercer a fiscalização dos gastos públicos realizados pelas Prefeituras, Câmaras Municipais e Órgãos da Administração Indireta municipal, além das despesas realizadas pelo Governo do Estado (GRILO JR.; CLERICUZI, 2008). A Figura 4 mostra o contexto onde o *software* SAGRES está inserido para a coleta, análise e disponibilização de dados para a sociedade.

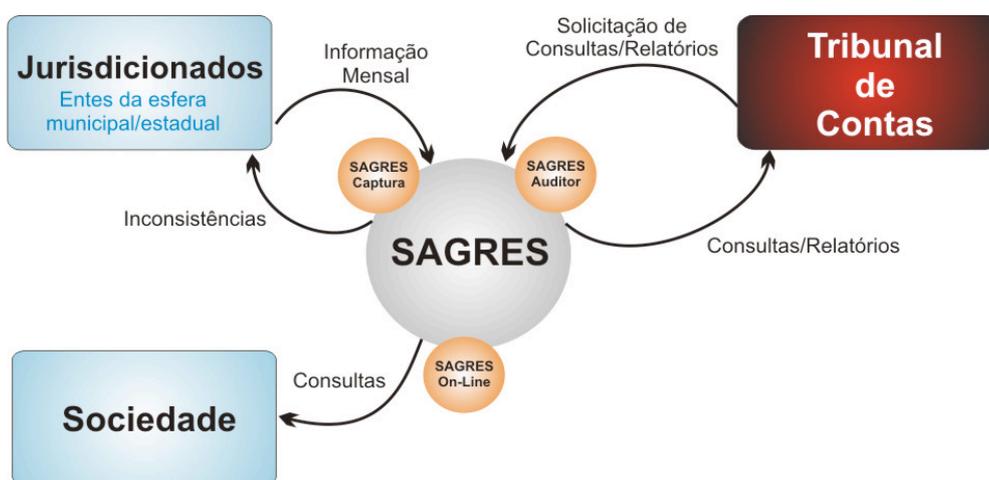


Figura 4: Diagrama de Contexto do SAGRES, Fonte: Adaptado de Santana (2008, p. 67)

Analisando a Figura 4, podemos observar a troca de dados, da execução orçamentária e financeira, com o módulo de captura do Sages, estes dados são enviados mensalmente ao TCE-PB por seus jurisdicionados, caso tenha alguma inconsistência nos mesmos, as suas unidades gestoras podem ter acesso a essas inconsistências para que seja tomada alguma ação. Então depois de tratados e armazenados, os dados são disponibilizados para consulta através do módulo Auditor, para os usuários internos, e pelo *web site* do Tribunal de Contas (Sages On-Line), para a consulta pública, para o cidadão e por órgãos fiscalizadores do controle externo.

O SAGRES possui dois mecanismos principais de acompanhamento da Gestão Pública: o cadastramento das informações relativas à Gestão, que constituem por si só a

formação de um banco de dados do TCE-PB, fonte de consulta ilimitada para toda a sociedade, e o acompanhamento da aplicação dos recursos públicos dentro dos ditames da legislação vigente (XAVIER, 2003).

O Tribunal de Contas do Estado do Piauí (TCE-PI) adotou o SAGRES em 2006 e fez um convênio de cooperação com o TCE-PB. Algum tempo depois, em maio de 2010, outro convênio foi acordado, desta vez com o Tribunal de Contas do Estado de Pernambuco (TCE-PB), que tinha como intuito a utilização do *software* para realizar a coleta dos dados jurisdicionados, já que o SAGRES é um sistema estável e de fácil operacionalização.

O objetivo do TCE-PI, ao utilizar o Sagres, é tornar a análise das contas anuais mais ágil, além de auxiliar o acompanhamento e a fiscalização das gestões orçamentária e financeira da administração municipal, com vistas ao recebimento e sistematização eletrônica da coleta e da remessa de dados necessários à composição da prestação de contas dos agentes públicos.

A prestação de contas é exigida no parágrafo único do Artigo 70 da Constituição Federal que estabelece que, deverão prestar contas qualquer pessoa física ou jurídica, pública ou privada, que utilize, arrecade, guarde, gerencie ou administre dinheiros, bens e valores públicos ou pelos quais a União responda, ou que, em nome desta, assumam obrigações de natureza pecuniária (BRASIL, 2009).

4 ASPECTOS METODOLÓGICOS

4.1 Principais indícios de fraudes ou irregularidades em licitações

Algumas dúvidas podem ser levantadas, já que é falado em encontrar padrões que possam identificar possíveis fraudes, uma delas é: Quais os padrões que são procurados nos dados? Então a tabela a seguir (tabela 5) pode mostrar algumas situações que demonstram irregularidades em processos licitatórios de acordo com os procedimentos normatizados pela Lei nº 8666/93, que trata no Artigo 4º sobre a forma de proceder e fiscalizar as licitações realizadas. Dessa forma fica mais claro, quais padrões podemos relacionar e quais devemos ter um pouco mais de atenção.

Item	Padrões Identificados	Interpretação
1	Participação do mesmo grupo de fornecedores (proponentes) em outros processos licitatórios, alguma vez só alterando o nome do licitante vencedor.	<ul style="list-style-type: none"> · Inexistência de competitividade · Favorecimento de fornecedores · Direcionamento de resultados
2	Realização de aquisições por dispensa de licitação de um mesmo material várias vezes em um mesmo ano, e para um determinado proponente.	<ul style="list-style-type: none"> · Inexistência de competitividade · Favorecimento de fornecedores · Direcionamento de resultados · Fracionamento de Licitação · Não aplicação dos princípios relativos às compras governamentais
3	Realização de aquisições por dispensa de licitação para determinado proponente, ultrapassando o limite permitido estabelecido em lei.	<ul style="list-style-type: none"> · Inexistência de competitividade · Favorecimento de fornecedores · Direcionamento de resultados · Não aplicação dos princípios relativos às compras governamentais

Tabela 5: Indícios de irregularidade nos processos licitatórios. Fonte: Adaptado de Resende (2005).

É importante destacar que esta tabela mostra apenas algumas das irregularidades que podem ser detectadas no processo licitatório e que de forma alguma essas possibilidades limitam-se apenas às que foram destacadas.

4.2 Delimitação do tema e seleção dos dados do caso

O processo foi aplicado sobre os dados coletados pelo sistema SAGRES Captura na base de dados do TCE-PB, tendo como principal (mas não único) foco os dados presentes na tabela de Licitação. Tais dados foram enviados ao Tribunal de Contas pelos jurisdicionados dos municípios por meio do módulo de Captura do SAGRES, citado anteriormente.

Como o objetivo desse estudo está relacionado com os processos licitatórios, não poderiam deixar de fazer parte da seleção dos dados as tabelas: Participantes, Propostas, Fornecedores, Contratos e Aditivos, além dos dados da tabela de Licitações que já foram mencionados. Estas tabelas encontram-se no Sistema de Gerenciamento de Banco de Dados MS-SQL *Server* 2000, que integram a base de dados do SAGRES. A figura 5 mostra de maneira mais clara os relacionamentos que existem entre as tabelas ditas anteriormente, na forma de um diagrama de entidade relacional.

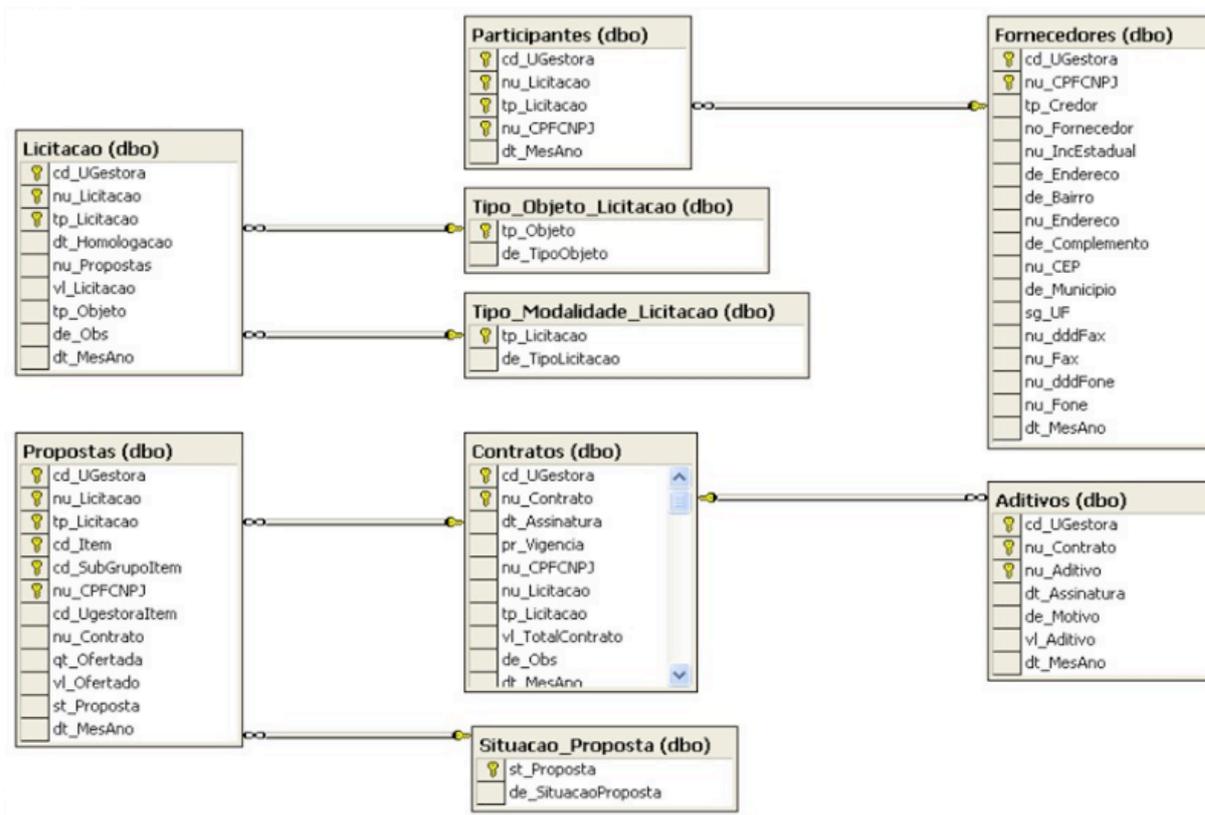


Figura 5: Tabela de dados do módulo de Licitações utilizadas pelo SAGRES, Fonte: TCE-PB (2010).

Os dados, que já foram mencionados, são enviados mensalmente ao Tribunal de Contas do Estado, onde são apreciados, os mesmos provêm das licitações realizadas pelos 233 municípios do estado da Paraíba, entre os anos de 2010 e 2012, nesse contexto são incluídas as Prefeituras Municipais, as Câmaras Municipais e os Órgãos da Administração Indireta Municipal.

4.2.1 Coleta dos dados

Como já foi falado, o trabalho se direciona principalmente à coleta dos dados relacionados ao módulo de licitações do TCE-PB, extraídos da base de dados do SAGRES. Pode ser observado na tabela a seguir um extrato da quantidade de processos licitatórios encontrados na base, organizado por tipo de modalidade de licitação, entre os períodos de 2010 e 2012.

4.3 Análise dos dados

A ferramenta utilizada para a execução do processo foi o *RapidMiner* na sua versão *Free 5.3.013*, sob licença *GPL (General Public License)* versão 3. É uma ferramenta com interface gráfica e que possui vários recursos para diferentes tarefas de descoberta de conhecimento e mineração de dados. Além de poder ser integrado com outros produtos por meio de uma *API (Application Programming Interface)*.

Como já visto anteriormente, vários algoritmos podem ser usados para encontrar padrões existentes nos bancos de dados, no caso deste trabalho, na base do SAGRES do Tribunal de Contas do Estado da Paraíba, com o intuito de encontrar padrões relevantes que possam representar indícios de irregularidades em processos de licitação. Com o objetivo de encontrar associações frequentes e válidas entre os dados, foram escolhidos então, nesta ocasião, os algoritmos para regras de associação.

Após a coleta dos dados, é preciso organizá-los e fazer uma seleção mais criteriosa dos mesmos, para que então possa ser feita sua análise e interpretação. Foi criada, para a fase de Extração, Transformação e Carga (ETL), um *data mart* com os dados dos processos licitatórios, contendo os proponentes e suas respectivas propostas. Ilustrado na figura 6.

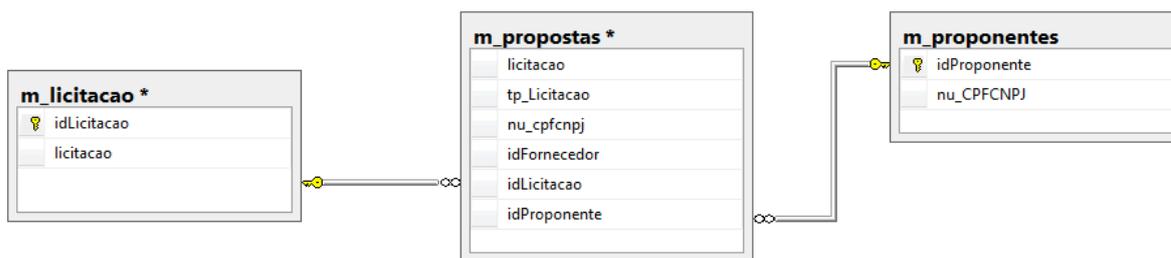


Tabela 6: Data mart de Licitações. Fonte: SAGRES (2013)

Os dados sofreram uma limpeza a fim de que os resultados pudessem ser mais concisos e que pudesse ser usado um computador com processamento não muito avançado, já que se fossem usados todos os dados que estivessem disponíveis, o resultado poderia sofrer algum desvio ou seria necessário uma máquina com um porte bem maior. Nas primeiras tentativas de execução dos algoritmos, com um número maior de dados, a máquina usada, com processador intel core i3 e 4 GB de RAM, chegou a apresentar erro de estouro de memória.

Do *data mart* que foi criado, os dados foram exportados para um arquivo CSV (*Comma-Separated Values*). Para que a manipulação dos dados se tornasse mais simples. Deste arquivo é gerada uma matriz “Licitação x Proponente” que é ilustrada na tabela 8:

Licitações	prop-001	prop-002	prop-003	prop-004	prop-00m
lic-001	1	1	1	0	...
lic-002	0	0	0	1	...
lic-00n

Tabela 7: Matriz Esparsa Licitações versus Proponentes

Para ficar melhor explicado: o valor “1” no campo $\{i; j\}$ significa que o prop-00*i* participou da licitação lic-00*j*, e o valor “0” no campo $\{i; j\}$ significa que o prop-00*i* não participou da licitação lic-00*j*.

4.4 Procedimento realizado

Aqui é apresentado o procedimento utilizado no *rapidminer*, para fazer a mineração de dados a partir do arquivo CSV, usando o algoritmo *FP-Growth*, que é um dos vários indicados para tarefas de associação.

As etapas para a execução da mineração na ferramenta foram: a leitura do conteúdo do arquivo CSV que contém os dados das licitações com seus respectivos proponentes; A transformação dos dados para a matriz esparsa; Conversão dos atributos numéricos para binários; Aplicação do algoritmo *FP-Growth*, que é um método eficiente para a mineração de padrões frequentes sejam eles curtos ou longos; Geração de regras de associação para um dado conjunto de itens frequentes, este processo é dividido em duas partes: primeiro a geração de itens frequentes e depois a geração de regras de associação a partir desses conjuntos.

O único parâmetro ajustável é o *min support*, que é o valor do suporte mínimo a ser considerado de acordo com as características do problema a ser resolvido.

5 ANÁLISE DOS RESULTADOS

É encontrado finalmente um conjunto de informações, que é resultado do processamento dos dados que foram citados no capítulo anterior, tais informações precisam de uma maior atenção para que seja realizada a interpretação e validação. Esta é a fase de interpretação e validação dos padrões que foram extraídos do processo de Descoberta de Conhecimento em Base de Dados.

Foram utilizados alguns filtros para que os recursos computacionais fossem mais bem aproveitados e também para a obtenção de resultados satisfatórios. Nos processos licitatórios podem existir de 1 (um) a infinito (∞) proponentes, dependendo do tipo de modalidade da licitação. Então um filtro foi aplicado onde foram selecionados apenas os registros que continham de 2 (dois) a 10 (dez) proponentes por processo.

Os processos que continha apenas 1 (um) proponente foram descartados por serem considerados falta de competitividade, ou mesmo um erro na informação enviada. Já os

registros que tinham 10 (dez) ou mais proponentes foram descartados pelos filtros por não representarem relevância para a execução do estudo, além da baixa chance de ocorrer algum tipo de fraude com essa quantidade de proponentes.

No experimento foi usado o suporte mínimo de 0.003 (0,03%) e confiança mínima de 0.60 (60%). O suporte corresponde à frequência com que A e B ocorrem em uma base de dados, e a confiança representa a frequência com que B ocorre, dentre os eventos que contém A .

Nesta figura é mostrado o resultado dos dados gerados pelo processamento do RapidMiner.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
1	10754.0	9781.0	0.003	0.607	0.998	-0.007	0.003	106.189	2.531
2	1964.0	1768.0	0.003	0.607	0.998	-0.007	0.003	122.526	2.533
3	3383.0	11076.0	0.008	0.632	0.995	-0.018	0.008	33.515	2.669
4	19656.0	16883.0	0.003	0.640	0.998	-0.006	0.003	69.960	2.752
5	11953.0, 14402.0	11772.0	0.003	0.640	0.998	-0.006	0.003	55.968	2.746
6	11953.0, 25.0	11772.0	0.003	0.640	0.998	-0.006	0.003	55.968	2.746
7	1768.0	1964.0	0.003	0.654	0.998	-0.007	0.003	122.526	2.873
8	19666.0	19109.0	0.004	0.655	0.998	-0.007	0.004	101.109	2.881
9	14402.0	11772.0	0.004	0.656	0.998	-0.008	0.004	57.389	2.876
10	17240.0	21071.0	0.003	0.667	0.998	-0.006	0.003	174.900	2.989
11	17240.0	4385.0, 21071.0	0.003	0.667	0.998	-0.006	0.003	174.900	2.989
12	4385.0, 17240.0	21071.0	0.003	0.667	0.998	-0.006	0.003	174.900	2.989
13	9176.0	1768.0	0.004	0.677	0.998	-0.008	0.004	136.708	3.085
14	442.0	6238.0	0.004	0.704	0.998	-0.007	0.004	160.536	3.360
15	21547.0	18750.0	0.004	0.710	0.998	-0.008	0.004	55.577	3.400
16	15113.0	16260.0	0.005	0.714	0.998	-0.009	0.005	45.706	3.445
17	4385.0	21071.0	0.004	0.714	0.998	-0.007	0.004	187.393	3.487
18	8466.0	5114.0	0.003	0.720	0.999	-0.006	0.003	78.705	3.539
19	19666.0	18750.0	0.004	0.724	0.998	-0.007	0.004	58.710	3.579
20	11140.0	12277.0	0.004	0.724	0.998	-0.007	0.004	82.599	3.593
21	11953.0	11772.0	0.010	0.754	0.997	-0.016	0.010	65.904	4.012
22	1289.0	21833.0	0.003	0.762	0.999	-0.005	0.003	72.886	4.156
23	11772.0, 14402.0	11953.0	0.003	0.762	0.999	-0.005	0.003	57.938	4.145
24	21345.0	5015.0, 12948.0	0.004	0.769	0.999	-0.006	0.004	183.462	4.315
25	9767.0	10331.0	0.003	0.773	0.999	-0.005	0.003	238.500	4.386
26	8176.0	1964.0	0.005	0.774	0.999	-0.007	0.005	145.078	4.405
27	14402.0	11953.0	0.005	0.781	0.999	-0.007	0.005	59.409	4.511

Figura 6: Resultado do processamento dos dados minerados pelo RapidMiner:

Os valores do suporte e da confiança devem ser analisados para apoiar a tomada de decisão, eles são exemplos de medidas de interesse objetivo, que são aquelas, que segundo Gonçalves (2005) empregam índices estatísticos para avaliar a força de uma regra. Pode-se ver na tabela 10 as regras com maior índice de confiança com os dados proveniente do processo executado no RapidMiner.

Outra medida de interesse importante a ser considerada é o *Lift*. Dada uma regra de associação $A \Rightarrow B$, esta medida indica o quanto mais frequente torna-se B quando A ocorrer. Esta medida é útil para regras com suporte baixo (como é este caso) e representa a dependência positiva e negativa entre o predecessor e sucessor da regra.

Regra	Premissas	Conclusões	Suporte	Confiança	Lift	Convicção
55	17240.0	4385.0	0.004	1.0	18.739	Infinito
56	21071.0	4385.0	0.003	1.0	18.739	Infinito

57	8193.0	2639.0	0.003	1.0	218.625	Infinito
58	10331.0	9767.0	0.003	1.0	238.5	Infinito
59	9365.0	18318.0	0.003	1.0	291.5	Infinito
60	18318.0	9365.0	0.003	1.0	291.5	Infinito
61	11772.0 e 25.0	11953.0	0.003	1.0	7.604	Infinito
62	17240.0 e 21071.0	4385.0	0.003	1.0	18.739	Infinito
53	7091.0	8009.0	0.003	0.95	11.357	2.082.389.937.1 06.910
54	21345.0 e 12948.0	5015.0	0.003	0.95	19.988	20.899.942.824. 471.100
52	5015.0	21345.0	0.004	0.92	18.566	12.438.059.843. 720.200
51	12948.0	5015.0	0.004	0.91	192.39	11.942.824.471. 126.300
50	5015.0 e 12948.0	21345.0	0.003	0.90	18.346	10.945.492.662. 473.700
49	21345.0	5015.0	0.004	0.88	18.566	8.625.373.229.1 46.810
48	5015.0	12948.0	0.004	0.88	192.39	829.521.631.40 8.424
47	12948.0	21345.0	0.004	0.875	17.658	7.960.358.299.9 80.940
45	2026.0	5114.0	0.003	0.86	9.505	7.596.531.351.2 48.330
46	21345.0 e 5015.0	12948.0	0.003	0.86	1.901	7.631.599.008.9 57.490

Tabela 8: Medidas de interesse objetivas geradas no exercício de 2010.

Fazendo uma análise desta tabela, é possível afirmar que em 100% dos casos em que os proponentes 17240 e 21071 estão juntos em uma licitação o proponente 11953 também esteve, assim como quando os 21345 e 12948 estão juntos, existe uma probabilidade de 95% do 5015 também estar presente e assim por diante.

Observe que para a maioria das regras geradas a partir do experimento, o *Lift* tem um valor muito alto. Isto significa que todos os proponentes relacionados na regra tem uma dependência grande.

A variável *Convicção* (*conviction*) tem como objetivo avaliar uma regra de associação como uma verdadeira implicação. Quando um valor muito alto aparece significa que $A \Rightarrow B$ é muito forte, mas $B \Rightarrow A$ não necessariamente será.

Na ferramenta é possível ainda analisar as regras em que um proponente faz parte, mesmo que ele não apareça nas premissas. Para ilustrar, na figura 7, foi escolhido como exemplo o proponente 17240.

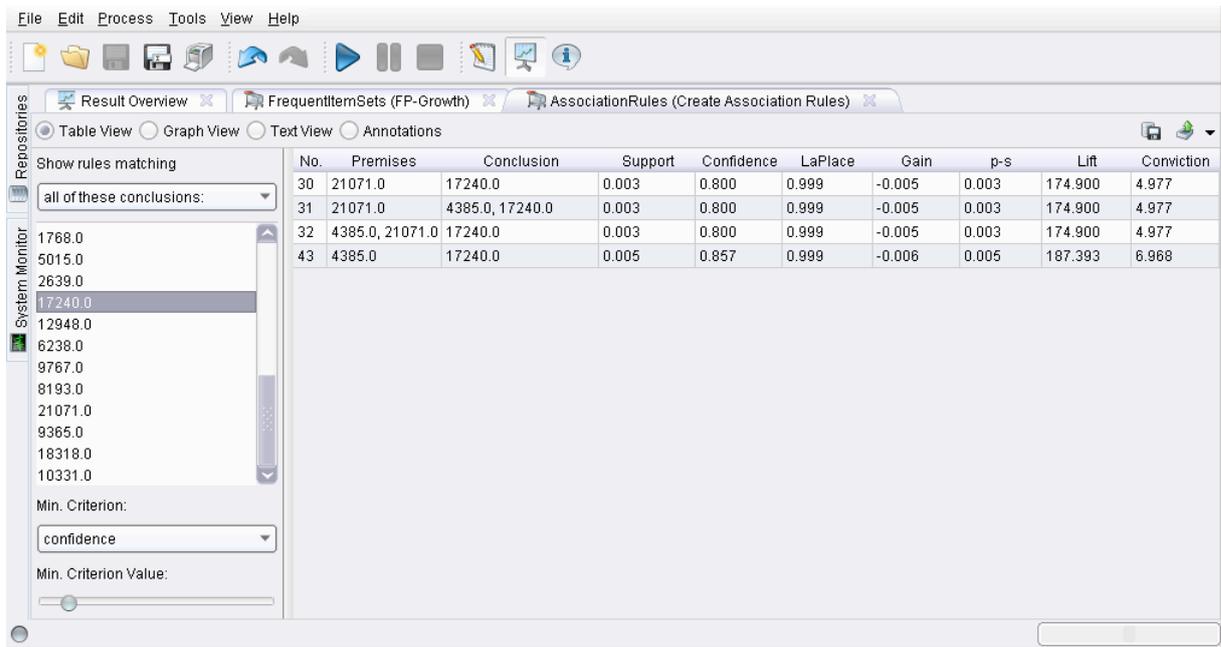


Figura 7: Filtro realizado por proponente nas regras de associação geradas.

Veja que são apresentadas variáveis, como conclusão, confiança, *lift*, que podem ser utilizadas para auxiliar o processo de decisão. É fácil a visualização dos registros e dos seus relacionamentos. A figura seguinte mostra os relacionamentos criados no momento da geração das regras, em forma de grafos. Assim fica mais fácil a observação e a análise das correlações.

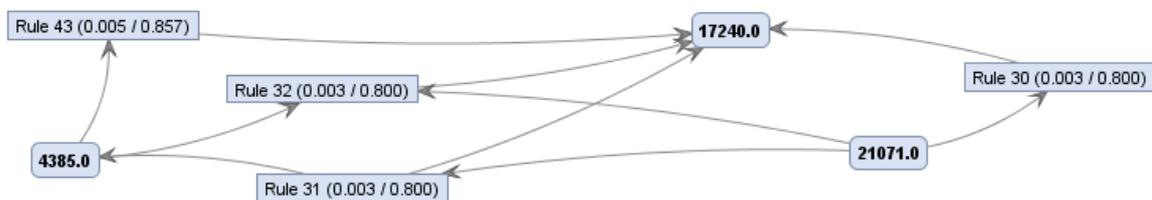


Figura 8: Grafo gerado pelo RapidMiner com as regras de associação de um proponente.

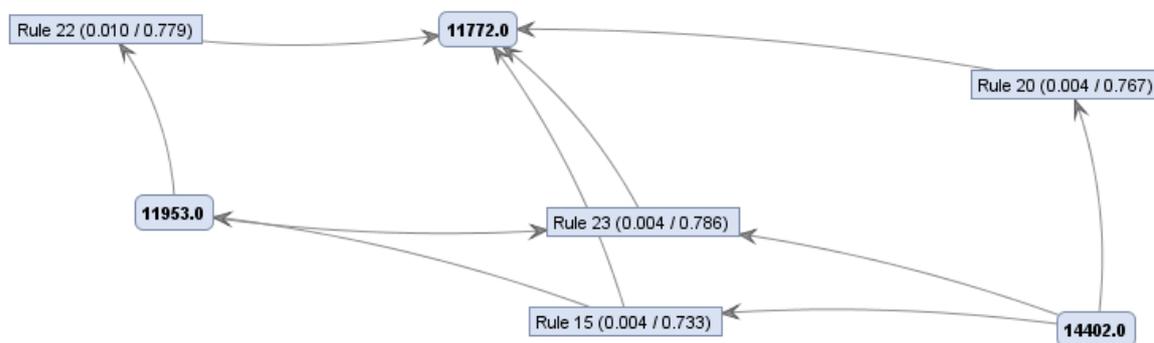
Alguns resultados dos experimentos realizados nos dados dos demais anos são descritos a seguir:

Regra	Premissas	Conclusões	Suporte	Confiança	Lift	Convicção
43	21071.0	9816.0	0.005	1.0	15.031	Infinity
44	21071.0	4385.0	0.005	1.0	15.473	Infinity
45	21071.0	9816.0, 4385.0	0.005	1.0	15.942	Infinity
46	9816.0, 21071.0	4385.0	0.005	1.0	15.473	Infinity
47	4385.0, 21071.0	9816.0	0.005	1.0	15.031	Infinity
42	4385.0	9816.0	0.006	0.97	14.589	#####
41	11772.0, 14402.0	11953.0	0.004	0.95	7.400.	#####
40	17997.0	10804.0	0.003	0.94	12.779	#####
39	7310.0	14867.0	0.003	0.94	2.484.	#####
38	9816.0	4385.0	0.006	0.94	14.589	#####

37	14402.0	11953.0	0.005	0.93	7.220	#####
36	11772.0	11953.0	0.01	0.92	7.193	#####
34	6238.0	14873.0	0.003	0.85	2.032	#####
35	14867.0	7310.0	0.003	0.85	2.484	#####
33	9816.0, 4385.0	21071.0	0.005	0.84	15.942	#####
32	14203.0	12993.0	0.003	0.84	11.973	#####
30	4385.0	21071.0	0.005	0.82	15.473	#####
31	4385.0	9816.0, 21071.0	0.005	0.82	15.473	#####
29	3015.0	20039.0	0.003	0.81	694.266	#####
28	4652.0	3282.0	0.004	0.81	13.828	#####
27	11140.0	12277.0	0.006	0.81	9.111	#####
26	1964.0	1768.0	0.004	0.80	11.785	#####
24	9816.0	21071.0	0.005	0.8	15.031	#####
25	9816.0	4385.0, 21071.0	0.005	0.8	15.031	#####
23	11953.0, 14402.0	11772.0	0.004	0.78	7.252	#####

Tabela 9: Medidas de interesse objetivas geradas no exercício de 2011.

Da mesma forma como foi visto na tabela 12, aqui se pode analisar e perceber que existem relações entre os proponentes, como por exemplo, sempre que o 21071 apareceu o 9816 também esteve presente.



Selecionando o registro 11772 é possível observar as regras e relações mostradas acima, onde aparecem os proponentes 11953 e 14402.

Regra	Premissas	Conclusões	Suporte	Confiança	Lift	Convicção
64	2844.0	14338.0	0.003	1.0	12.421	Infinity
65	8835.0	4385.0	0.005	1.0	1.528	Infinity
66	16691.0	11849.0	0.004	1.0	1.806	Infinity
67	7419.0	18904.0	0.003	1.0	198.75	Infinity
68	7310.0	14867.0	0.003	1.0	265.0	Infinity
69	7310.0	11730.0	0.003	1.0	265.0	Infinity
70	8565.0	16618.0	0.003	1.0	28.392	Infinity
71	14951.0	16618.0	0.003	1.0	28.392	Infinity
72	11076.0, 16260.0	11061.0	0.003	1.0	6.973	Infinity
73	8675.0, 8835.0	4385.0	0.004	1.0	1.528	Infinity
74	8675.0, 7419.0	18904.0	0.003	1.0	198.75	Infinity
75	7310.0	14867.0, 1730.0	0.003	1.0	28.392	Infinity
76	14867.0, 7310.0	11730.0	0.003	1.0	265.0	Infinity
77	11730.0, 7310.0	14867.0	0.003	1.0	265.0	Infinity
63	14402.0	11953.0	0.005	0.95	63.489	23.637
62	7091.0	8009.0	0.005	0.95	14.053	2.185
59	3740.0	11953.0	0.004	0.95	629.375	1.969
60	11953.0, 14296.0	11772.0	0.004	0.95	7.706	19.753

61	11772.0, 14296.0	11953.0	0.004	0.95	629.375	1.969
57	3282.0	4652.0	0.004	0.94	1.706	17.900
58	19033.0	3189.0	0.004	0.94	17.876	1.790
55	14867.0	11730.0	0.003	0.93	24.733	14.943
56	11730.0	14867.0	0.003	0.93	24.733	14.943
52	16618.0	8565.0	0.003	0.92	28.392	13.954
53	11140.0, 11606.0	12277.0	0.003	0.92	10.545	13.876
54	14867.0, 11730.0	7310.0	0.003	0.92	28.392	13.954

Tabela 10: Medidas de interesse objetivas geradas no exercício de 2012.

Quando o algoritmo de associação é executado são geradas algumas regras, como as que seguem:

Association Rules

```
[12277.0] --> [11606.0] (confidence: 0.600)
[1498.0] --> [18464.0] (confidence: 0.600)
[16739.0] --> [18923.0] (confidence: 0.600)
[21646.0] --> [11568.0] (confidence: 0.600)
[18904.0] --> [8675.0, 7419.0] (confidence: 0.600)
[10804.0] --> [20785.0] (confidence: 0.606)
[11606.0] --> [11140.0] (confidence: 0.609)
[20785.0, 7010.0] --> [10804.0] (confidence: 0.615)
[7010.0, 10804.0] --> [20785.0] (confidence: 0.615)
[7010.0] --> [20785.0] (confidence: 0.619)
[7010.0] --> [10804.0] (confidence: 0.619)
[12277.0, 11606.0] --> [11140.0] (confidence: 0.619)
[18904.0] --> [7419.0] (confidence: 0.650)
[1658.0] --> [8176.0] (confidence: 0.654)
[4385.0] --> [8675.0, 8835.0] (confidence: 0.654)
[12277.0] --> [11140.0] (confidence: 0.686)
[1498.0] --> [12551.0] (confidence: 0.700)
[10754.0] --> [9781.0] (confidence: 0.720)
[11849.0] --> [16691.0] (confidence: 0.727)
[11953.0] --> [11772.0] (confidence: 0.733)
[1964.0] --> [8176.0] (confidence: 0.750)
[8675.0, 18904.0] --> [7419.0] (confidence: 0.750)
[4385.0] --> [8675.0] (confidence: 0.769)
[8835.0] --> [8675.0] (confidence: 0.773)
[4652.0] --> [3282.0] (confidence: 0.773)
[8835.0] --> [8675.0, 4385.0] (confidence: 0.773)
[4385.0, 8835.0] --> [8675.0] (confidence: 0.773)
[8009.0] --> [7091.0] (confidence: 0.778)
[10804.0] --> [7010.0] (confidence: 0.788)
[18904.0] --> [8675.0] (confidence: 0.800)
[20785.0, 10804.0] --> [7010.0] (confidence: 0.800)
[3189.0] --> [19033.0] (confidence: 0.810)
[12062.0] --> [1476.0] (confidence: 0.812)
[14296.0] --> [11953.0, 11772.0] (confidence: 0.826)
```

Nestas regras podem ser observadas as frequências que estão associadas entre dois proponentes, como exemplo, pode ser destacada a primeira regra, que mostra a probabilidade de 60% do 11606 ocorrer quando o 12277 estiver presente.

Este experimento mostra apenas uma pequena parte do potencial, que pode ser explorado, das técnicas de *data mining*. É importante deixar claro que nem todos os processos

licitatórios que apresentaram participação recorrente de mútuos proponentes, como mostrados nas regras geradas, são necessariamente irregulares. Mas isto pode nortear e proporcionar uma melhor visão aos técnicos que analisam estes processos, dando mais confiança e um melhor desempenho ao trabalho realizado por eles.

6 CONCLUSÃO

Como pôde ser observado, o processo de descoberta de conhecimento em base de dados, tem dentro de si uma série de estágios ou etapas, que vão desde a coleta, preparação e limpeza dos dados, passa pela consolidação e extração dos padrões, até chegar ao momento de avaliação, onde é analisado o quanto os padrões ou regras, que foram identificadas, agregam valor sobre o conhecimento que o gestor já possui com relação ao problema em questão.

Diante deste trabalho, é possível compreender que o recurso da tecnologia da informação é essencial no apoio da fiscalização. Principalmente se a organização está imersa em um ambiente dinâmico, onde as mudanças estão presentes no dia a dia. Ainda é válido salientar a importância deste recurso no momento que o cumprimento da lei e a satisfação da sociedade entram em cena com o interesse na utilização de recursos de forma correta, tendo em vista a transparência, eficiência e eficácia proveniente do sistema de fiscalização da gestão pública. Mesmo que os sistemas gerem boas informações, estas podem ainda não ser adequadas para determinada necessidade, sendo necessárias informações complementares para que possam apoiar uma tomada de decisão.

Para chegarmos ao final deste estudo tivemos que alcançar alguns objetivos específicos como: analisar os conceitos de técnicas de *data mining*, verificar a aplicabilidade das técnicas de *data mining* para o processo de fiscalização nos órgãos de controle externo e definir base de dados a ser utilizada para avaliação das ferramentas. Onde existia a pretensão de investigar a contribuição da aplicação de técnicas de *data mining* na área de auditoria, com objetivo de apoiar a fiscalização para que fosse possível a identificação de possíveis fraudes nos processos licitatórios.

Respondendo as questões de pesquisa: “Quais as principais características e funcionalidades das ferramentas de *data mining*?” e “É viável a utilização de técnicas de *data mining* nos Tribunais de Contas?”, diante dos resultados do estudo, é possível afirmar que o processo de *data mining* tem um guia bem definido para que as informações desejadas sejam adquiridas, sabendo que algumas etapas são de grande importância para que o resultado esperado seja encontrado, como por exemplo, a seleção e a limpeza dos dados, o desenvolvimento dessas atividades é essencial para que o foco se mantenha no decorrer da

execução do processo e o objetivo não apresente muita discrepância. Sendo, dessa forma, possível a utilização de técnicas de *data mining* nos Tribunais de Contas, com o intuito de auxiliar os auditores na parte de fiscalização, dando mais celeridade aos processos analisados neste âmbito.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALÓ, C. C. (2009) **Uma abordagem para transparência em processos organizacionais utilizando aspectos**. Dissertação de Doutorado em Informática. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro.
- ALMEIDA, F.; SIQUEIRA, J.; ONUSIC, L. **Data mining no contexto de customer relationship management**. Caderno de Pesquisas em Administração, São Paulo, Brasil, 02, ago. 2005. Disponível em: <http://www.ead.fea.usp.br/cad-pesq/arquivos/v12n2art6_ult.pdf>. Acesso em: 04 jun. 2013.
- AMORIM, Thiago. **Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de bases de dados**. Recife, 2007. 50p. Monografia (Graduação em Ciência da Computação) – Universidade Federal de Pernambuco.
- BATISTA, Emerson de Oliveira. **Sistema de Informação: o uso consciente da tecnologia para o gerenciamento**. São Paulo: Saraiva, 2006.
- BRASIL. **Constituição de 1988. Constituição da República Federativa do Brasil**. Disponível em: <<http://www.planalto.gov.br>>. Acesso em: 28 mai. 2013.
- CARDOSO, Olinda Nogueira Paes; MACHADO, Rosa Teresa Moreira. **Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras**. Rev. Adm. Pública v.42 n.3 Rio de Janeiro maio/jun. 2008.
- CARVALHO, J.; SAMPAIO, M.; MOGIOVI, G. **Utilização de Técnicas de “Data Mining” para o Reconhecimento de Caracteres Manuscritos**. Universidade Federal da Paraíba.
- CLERICUZI, Adriana Zenaide. **Desenvolvimento adaptativo para Sistema de Apoio a Decisão específico**. 175 f. Tese (Doutorado) – Universidade Federal de Pernambuco, Recife, 2006.
- DECOMAIN, Pedro Roberto; **TRIBUNAIS DE CONTAS NO BRASIL: natureza jurídica e alcance das suas funções, à luz dos princípios constitucionais da Administração Pública**. 397 f. Dissertação (Mestrado). Universidade do Vale do Itajaí – UNIVALI. Itajaí. 2005.
- VARAJÃO, Fabrício. **Fundamentos de Sistemas de Informação**. Campo Grande, 2012. Apostila do Curso de Bacharelado em Sistemas de Informação e Licenciatura em Informática - FACULDADES INTEGRADAS CAMPO-GRANDENSES.
- FRANCO, Itamar. **Lei Federal nº 8.666, 1993, Lei de LICITAÇÕES**. Brasília: Congresso Nacional. Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/L8666cons.htm>. Acesso em: 1 jun. 2013.
- FAVARETTO, Fábio. **Melhoria da qualidade da informação no controle da produção: estudo exploratório utilizando Data Warehouse**. Revista Produção. Vol. 17, n. 2, p. 343-353, Maio/Ago, 2007.
- FIGUEIRÊDO, Carlos Maurício C. **Ética na gestão pública e exercício da cidadania: o papel dos tribunais de contas brasileiros como agências de accountability - O caso do**

Tribunal de Contas de Pernambuco. VII *Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública*, Lisboa, Portugal, 8-11 Oct. 2002.

FREITAS JR., O. G. **Um modelo de Sistema de Gestão do Conhecimento para Grupos de Pesquisa e Desenvolvimento.** 296 f. Tese (Doutorado em Engenharia) – Universidade Federal de Santa Catarina, Florianópolis, 2003.

GIUDICI, Paolo. **Applied Data Mining Statistical Methods for Business and Industry.** John Wiley & Sons Ltd, 2003.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data mining: um guia prático.** Rio de Janeiro: Elsevier, 2005, 3ª reimpressão.

GRILO JR., Tarcísio Ferreira; CLERICUZI, Adriana Zenaide. **Otimização do processo produtivo e de tomada de decisão através de *Business Intelligence* aplicado ao controle externo.** III SEPRONe – Juazeiro, BA. 2008.

GRILO JR., Tarcísio Ferreira; SILVA, José Gentil Bezerra da; CLERICUZI, A. Z. . **Business Intelligence: uma abordagem com ênfase no controle externo.** XV Simpósio de Engenharia de Produção - SIMPEP. Baurú/SP, 2008.

HAN, Jiawei; KAMBER, Micheline. **Data mining: concepts and techniques.** 2ª ed. San Francisco: Morgan Kaufmann Publishers, 2006.

LAUDON, Kenneth. LAUDON Jane – **Sistemas de informação gerenciais.** 7ª. edição São Paulo: Pearson Prentice Hall, 2007.

LEMOES, Elaine P. **Análise de Crédito Bancário com o uso de Data Mining: Redes Neurais e Árvores de Decisão.** Curitiba: Dissertação de Mestrado da Universidade Federal do Paraná, 2003.

MACEDO, Dayana. MATOS, Simone. **Extração de conhecimento através da mineração de dados.** v.2, n.2, ago. 2010.

MACHADO, Bruno de Abreu; JAMIL, George Leal. **Mineração de dados como ferramenta de tomada de decisão: estudo de caso de uma universidade.** 5º Congresso Internacional de Gestão da Tecnologia e Sistema de Informação - CONTECSI. São Paulo. 2008.

MARCHIORATO, Haroldo José. **Percepção de gestores de TI/SI e executivos sobre o papel da TI/SI e sua utilização estratégica: um estudo exploratório em indústrias de autopeças.** 114 f. Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção. Recife. 2006.

MARTINHAGO, Dariana. MAIA, Moisés. SEDIYAMA, Aline. CARVALHO, Nádia. **Data Mining: Definição, importância e aplicação na gestão organizacional.**

MORAES, Tiago Cacique. **O processo de modernização dos Tribunais de Contas no contexto da reforma do Estado no Brasil.** 109 f. Dissertação (mestrado) – Escola de Administração de Empresas de São Paulo. 2006.

O'BRIEN, James A. **Sistemas de informação e as decisões gerenciais na era da internet.** 2ª ed., São Paulo: Saraiva, 2004.

PETRINI, Maira. **Incorporando a Gestão da Sustentabilidade aos Sistemas de Inteligência de Negócio**. 157 f. Tese (doutorado) - Escola de Administração de Empresas de São Paulo. 2006.

PIZZI, Luciene Cristina. **Mineração multi-relacional: o algoritmo GFP-growth**. 107 p. Dissertação (Mestrado) - Universidade Federal de São Carlos. São Carlos, 2006.

PORTAL DOS TRIBUNAIS DE CONTAS DO BRASIL. **Programa de Modernização do Sistema de Controle Externo**. Disponível em: <http://www.controlepublico.org.br/index.php?option=com_content&view=article&id=11&Itemid=17>. Acesso em: 04 jun. 2013.

QUONIAM, L., TARAPANOFF, K., ARAÚJO JÚNIOR, R., ALVARES, L.. **Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil**. Ciência da Informação, Brasília, DF, Brasil, 30, out. 2001. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/183/162>. Acesso em: 04 Jun. 2013.

SAGRES. **Sistema de Acompanhamento da Gestão dos Recursos da Sociedade**. Disponível em: <<http://www.tce.pb.gov.br>>. Acesso em: 26 mai. 2013.

SANTANA, ED W. F. **Uma análise sobre a utilização da informação contábil como um instrumento de controle social dos municípios paraibanos com mais de 50.000 habitantes**. 116 f. Dissertação (Mestrado). Programa Multiinstitucional e Inter-Regional de Pós-Graduação em Ciências Contábeis da Universidade de Brasília, da Universidade Federal da Paraíba e da Universidade Federal do Rio Grande do Norte. 2008.

SANTOS, A.; OLIVEIRA, J. **Data mining aplicado ao ensino a distância**. Disponível em: <<http://perquirere.unipam.edu.br/documents/23456/33987/documento8.pdf>>. Acesso em: 29 Jul. 2013.

SARAIVA, S.; MASSON. M. **Evasão e Permanência em uma Instituição de Tradição: um estudo sobre o processo de evasão de estudantes em cursos de Engenharia na Escola Politécnica da UFRJ**, Relatório de Pesquisa.

TCE-CE. **Tribunal de Contas do Estado do Ceará**. Disponível em: <<http://www.tce.ce.gov.br/sitetce/arq/PROMOEX/Index.htm>>. Acesso em: 25 mai. 2013.

TCE-PB. **Tribunal de Contas do Estado da Paraíba**. Disponível em: <<http://www.tce.pb.gov.br>>. Acesso em: 25 mai. 2013.

TCE-PE. **Manual de Organização, Competências e Atribuições dos Órgãos Auxiliares**. Gerência de Desenvolvimento Organizacional. 2011.

TURBAN, Efrain; SHARDA, Ramesh; ARONSON, Jai E. & KING, David. **Business Intelligence: um enfoque gerencial para a inteligência do negócio**. Tradução Fabiano Bruno Gonçalves. Porto Alegre: Bookman, 2009.

VERCELLIS, Carlo. **Business Intelligence: data mining and optimization for decision making**. John Wiley & Sons Ltd, 2009.

Vieira, Luiz Sergio Gadelha. **O futuro dos Tribunais de Contas no Brasil**. Disponível em: <<http://www.controlepublico.org.br/index.php/publicacoes/artigos-tecnicos>>. Acesso em: 06 jul. 2013.