

UNIVERSIDADE DO VALE DO RIO DOS SINOS – UNISINOS
MBA EM ADMINISTRAÇÃO DE EMPRESAS DE TI

JOSIANE CAPAVERDE

UTILIZANDO MINERAÇÃO DE DADOS PARA
IDENTIFICAÇÃO DE PERFIS DE CLIENTES NA CADEIA DE
SUPRIMENTOS E DEMANDA: UM ESTUDO DE CASO

SÃO LEOPOLDO

2011

Josiane Capaverde

UTILIZANDO MINERAÇÃO DE DADOS PARA
IDENTIFICAÇÃO DE PERFIS DE CLIENTES NA CADEIA DE
SUPRIMENTOS E DEMANDA: UM ESTUDO DE CASO

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Especialista em Administração de Empresas de TI, pelo MBA em Administração de Empresas de TI, da Universidade do Vale do Rio dos Sinos.

Orientador: Prof. Dr. Sílvio Cesar Cazella

São Leopoldo

2011

RESUMO

Através da utilização de técnicas de *Data Mining* (Mineração de Dados) o presente trabalho tem por objetivo principal identificar de que forma essas técnicas podem auxiliar no processo de identificação de perfis de clientes para cada solução ofertada pela empresa XYZ. Uma vez escolhido o método de estudo de caso foram realizadas pesquisas com colaboradores da empresa XYZ para identificação de suas necessidades e também foram coletadas informações a partir de bases de dados, planilhas e outros documentos. Consolidando todas as informações adquiridas, foi possível constatar que a empresa XYZ tem alto potencial de vendas de suas soluções em empresas com perfis distintos e atividades variadas de negócio. A descoberta de conhecimento obtida com técnica de mineração de dados sobre os clientes na empresa XYZ trouxe a tona novos perfis de clientes com necessidades desconhecidas, conhecimento este que a empresa poderá abordar de maneira proativa.

Palavras-chave: mineração de dados, perfil de clientes, árvore de decisão.

LISTA DE FIGURAS

Figura 1: Processo de DCBD	11
Figura 2: Tela de Base no Formato Weka	31
Figura 3: Árvore de Decisão.....	32

LISTA DE TABELAS

Tabela 1: Tarefas de Mineração de Dados	15
Tabela 2: Técnicas de Mineração de Dados	19
Tabela 3: Base para Mineração	29

SUMÁRIO

1	INTRODUÇÃO	7
1.1	SITUAÇÃO PROBLEMÁTICA E PERGUNTA DE PESQUISA	7
1.2	OBJETIVOS	9
1.2.1	<i>Objetivo geral</i>	9
1.2.2	<i>Objetivos específicos</i>	9
1.3	JUSTIFICATIVA	9
2	FUNDAMENTAÇÃO TEÓRICA.....	11
2.1	DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS	11
2.2	PREPARAÇÃO DOS DADOS	12
2.3	MINERAÇÃO DE DADOS	13
2.4	TIPOS DE TAREFAS PARA MINERAÇÃO DE DADOS.....	13
2.4.1	<i>Classificação</i>	14
2.4.2	<i>Estimativa (ou Regressão)</i>	14
2.4.3	<i>Associação</i>	14
2.4.4	<i>Segmentação (ou Clustering)</i>	15
2.4.5	<i>Sumarização</i>	15
2.5	TÉCNICAS DE MINERAÇÃO DE DADOS	16
2.5.1	<i>Descoberta de Regras de Associação</i>	16
2.5.2	<i>Árvores de Decisão</i>	17
2.5.3	<i>Raciocínio Baseado em Casos</i>	17
2.5.4	<i>Algoritmos Genéticos</i>	18
2.5.5	<i>Redes Neurais Artificiais</i>	18
3	MÉTODOS E PROCEDIMENTOS	20
3.1	DELINEAMENTO DA PESQUISA	20
3.2	DEFINIÇÃO DA UNIDADE DE ANÁLISE	20
3.3	TÉCNICAS DE COLETA DE DADOS	20
3.4	TÉCNICAS DE ANÁLISE DE DADOS	21
3.4.1	<i>Entrevistas em profundidade</i>	21
3.4.2	<i>Índices e Relatórios</i>	22
3.4.3	<i>Textos e Documentos</i>	22
3.5	LIMITAÇÕES DO MÉTODO.....	22
4	APRESENTAÇÃO E ANÁLISE DOS DADOS.....	23
4.1	NECESSIDADE DA EMPRESA XYZ	23
4.1.1	<i>Primeiro Questionário</i>	24
4.2	DEFINIÇÃO DE TAREFA E TÉCNICA DE MINERAÇÃO DE DADOS.....	25
4.2.1	<i>Preparação de Dados</i>	26
4.2.2	<i>Atributos da Mineração</i>	27
4.2.3	<i>Tarefa e Algoritmo</i>	30
4.2.4	<i>Conhecimento Descoberto</i>	32
4.2.5	<i>Apresentação do Perfil de Clientes</i>	33
4.3	IDENTIFICAÇÃO DE OPORTUNIDADES DE NEGÓCIO	34
5	CONSIDERAÇÕES FINAIS.....	36

1 INTRODUÇÃO

No atual mundo dos negócios quem tem a informação tem o mercado em suas mãos. Saber quais são os principais clientes e serviços em potencial da área de atuação da empresa é um trunfo no momento da negociação e potencializa as chances de novos fechamentos de contratos e vendas.

Com a base de dados certa é possível utilizar técnicas de mineração de dados para descobrir uma informação antes desconhecida e que poderá trazer valor agregado à empresa e a seus clientes.

Como o número de possíveis relacionamentos existentes em um banco de dados é muito grande, a busca de relacionamentos corretos por meio da simples validação de cada um deles é praticamente impossível. Devido à grande quantidade de dados dos sistemas estes relacionamentos são desconhecidos. *Data Mining* (Mineração de Dados) é o processo de busca de relacionamentos e padrões globais existentes nas bases de dados (HAND, 2001).

O presente trabalho explicita de que forma técnicas de Mineração de Dados podem trazer conhecimento sobre os perfis de clientes de cada oportunidade de negócio de uma empresa de serviços de tecnologia para a cadeia de suprimentos e demanda.

1.1 SITUAÇÃO PROBLEMÁTICA E PERGUNTA DE PESQUISA

A descoberta de conhecimento através de técnicas de Data Mining é utilizada em várias áreas como o marketing, por exemplo. O presente trabalho identifica a técnica de Mineração de Dados para identificação de oportunidades de vendas e como essa técnica foi selecionada entre as tantas existentes.

A empresa da qual esse estudo de caso será baseado não será identificada por sua solicitação, sendo designada, portanto, como empresa XYZ.

A empresa XYZ, líder em soluções inovadoras para a cadeia de suprimentos e demanda, oferta diversos serviços para atender à grande demanda que esse ramo abrange. Seu portfólio de ofertas busca gerar resultados expressivos para seus clientes, alavancando

sua competitividade agilizando o processo de compras, melhorando o processo de vendas, automatizando reposições de mercadorias. Em sua base de clientes estão nove entre os dez maiores varejos do Brasil e sete entre os dez mundiais.

Com amplas bases de clientes e serviços a empresa XYZ encontra dificuldades em identificar as soluções ideais e ainda faltantes para determinados clientes. É importante que a empresa tenha de forma rápida e fácil acesso a informações sobre a localização destes clientes e baseada nos perfis de cada um possa indicar a solução ofertada ideal para determinado cliente.

Diretamente, a área comercial é a principal beneficiada com a descoberta do conhecimento de novas oportunidades de negócio de acordo com as características dos clientes, assim será possível realizar novos fechamentos e trazer novas receitas para a empresa XYZ. Indiretamente, toda a empresa é beneficiada, pois novos fechamentos trazem novos projetos a serem entregues, com novos desenvolvimentos, testes, faturamentos, suportes e atendimentos a um novo serviço antes não ofertado para aquele cliente.

Uma das soluções ofertadas pela empresa XYZ chama-se EDI (*Electronic Data Interchange*), que é a troca eletrônica de documentos entre, por exemplo, varejos e sua malha de fornecedores. No EDI, grupos de empresas como indústrias, compradores e distribuidores são integrados eletronicamente pela empresa XYZ para que a troca de documentos ocorra sem divergências e de forma transparente para cada elo da cadeia. Ainda assim, podem ocorrer casos em que determinadas empresas não estejam integradas eletronicamente com o outro lado da cadeia, deixando o processo suscetível a erros de digitação, perda de dados impressos e sem controle gerencial das informações. Com a identificação das empresas faltantes na integração eletrônica para troca de informações a empresa XYZ oferece vantagens competitivas a seus clientes, agregando valor à operação de cada um e também aumentando o seu faturamento.

Sendo assim, a pergunta de pesquisa desse trabalho é: Como técnicas de Mineração de Dados podem auxiliar no processo de identificação de oportunidades de negócios para empresas de serviços da cadeia de suprimentos e demanda?

1.2 OBJETIVOS

1.2.1 Objetivo geral

Utilizar técnicas de Mineração de Dados visando adquirir conhecimento sobre características de clientes e soluções ofertadas pela empresa XYZ tendo em vista o aumento de faturamento desta empresa e agregação de valor a seus clientes da cadeia de suprimentos e demanda.

1.2.2 Objetivos específicos

- Entender a necessidade da empresa XYZ;
- Identificar tarefa e técnica de Mineração de Dados adequada para o objetivo;
- Avaliar os dados minerados a partir da base de dados da empresa XYZ;
- Apresentar conhecimento descoberto sobre os clientes da empresa XYZ;
- Apresentar perfis de clientes para as soluções da empresa XYZ;
- Confirmar se o processo de Mineração de Dados auxilia na identificação de oportunidades de negócio, a partir do perfil dos clientes.

1.3 JUSTIFICATIVA

A empresa XYZ se preocupa em oferecer a seus clientes, de forma proativa, as soluções que os auxiliarão a alavancar sua competitividade. Com a constante venda de soluções a novos clientes a malha de clientes da empresa XYZ cresceu consideravelmente nos últimos anos, tornando difícil a identificação de quais clientes ainda não tem implantados todos os serviços oferecidos e adequados a seu ramo de atividade.

A Mineração de Dados pode extrair conhecimento sobre os clientes e suas vendas, parceiros e serviços facilitando a identificação das condições sob as quais novos serviços são adquiridos, e a importância de cada serviço ofertado ao cliente, por exemplo.

O quanto antes esse tipo de mecanismo for implantado na empresa XYZ melhor será para a organização das vendas e também para o foco do resultado em faturamento e em novos fechamentos com clientes-chave para a cadeia.

Além do foco em novas oportunidades, a pesquisa realizada por esse trabalho possibilitará identificar outros conhecimentos que hoje não são considerados por não serem de fácil compreensão e acesso. Como, por exemplo, quais soluções são ofertadas com mais sucesso em determinada região do país, podendo evoluir para o mercado internacional. A continuidade da pesquisa utilizando técnicas de Mineração de Dados dependerá da utilização dos conhecimentos adquiridos na empresa XYZ e também da base utilizada para a mineração.

O presente trabalho pretende apresentar à empresa um padrão de conhecimento extraído das bases de clientes a fim de identificar conhecimentos para facilitar a iteração com clientes e abrir novos caminhos de negócios.

A estrutura do trabalho será a seguinte:

No segundo capítulo são abordados os temas teóricos que possibilitaram o embasamento para a criação da base para a mineração e o passo a passo deste processo.

No capítulo três são descritas as metodologias utilizadas na busca da informação na empresa XYZ utilizada no estudo de caso, bem como as técnicas de análise dos dados e as limitações da pesquisa.

No quarto capítulo são apresentados os tratamentos práticos dos dados coletados, como o processo de preparação dos dados para mineração, a mineração dos dados e os questionários realizados na empresa escolhida para o estudo de caso.

No último capítulo são apresentadas as considerações finais em relação ao trabalho, assim como um balanço final sobre os objetivos e expectativas para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

A Mineração de Dados consiste na exploração de bases de dados evidenciando padrões que auxiliem na descoberta de conhecimento utilizando técnicas de estatística, recuperação de informação, inteligência artificial e reconhecimento de padrões (ADRIAANS, 1996).

Com a utilização da Mineração de Dados é possível identificar padrões nas informações de bases de clientes, criando um padrão e tornando possível a identificação de novas oportunidades de negócios ainda não exploradas.

Nos próximos capítulos são detalhadas as etapas de preparação para a Mineração de Dados.

2.1 DESCOBERTA DE CONHECIMENTO EM BASE DE DADOS

O processo de Descoberta de Conhecimento em Base de Dados (DCBD) é dividido em duas fases principais: preparação de dados e mineração de dados. A figura 1 exibe todo o processo de preparação e de mineração em uma base de dados.

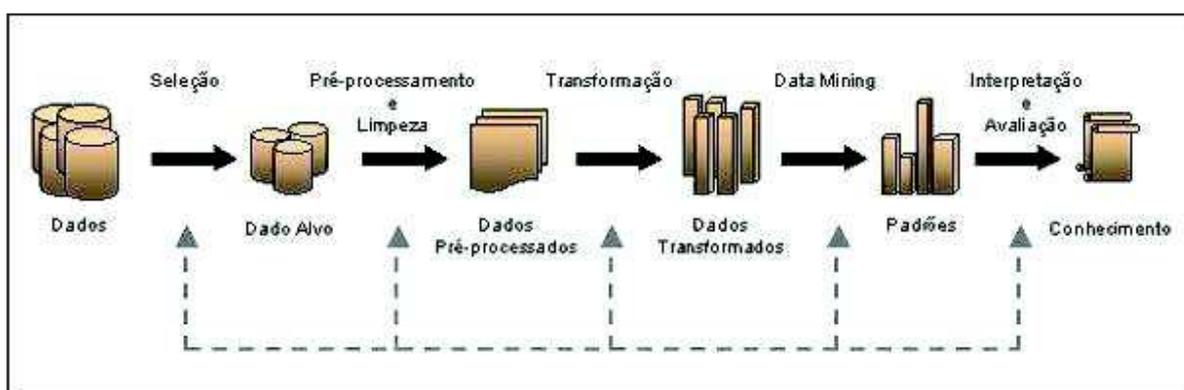


Figura 1: Processo de DCBD

2.2 PREPARAÇÃO DOS DADOS

Nessa etapa são identificados os dados relevantes para a solução do problema, pois apenas ter os dados não é suficiente, é necessário tê-los de forma correta e adequada para atender as características desejadas e para que os dados estejam disponíveis para o processamento e em condições de uso.

A falta de informação pode dificultar, ou até impedir, que o resultado obtido seja o esperado. Por isso, sempre que existir dúvida em relação à importância de um dado é aconselhável considerá-lo na análise dos dados (AMARAL, 2001).

A base a ser utilizada também pode ser um ponto de dúvidas no momento da mineração. É importante escolher a base que contenha dados coerentes com o objetivo da mineração.

Após a escolha da base são realizadas as etapas preparatórias para a mineração:

- **Transformação de Dados:** muitas vezes os dados a serem considerados na mineração de dados foram exportados de diferentes sistemas, documentações, planilhas ou até mesmo anotações feitas em um papel. Para que a mineração possa ser realizada é obrigatório que estes dados sejam transformados em um formato padrão.
- **Limpeza de Dados:** para o processo de mineração ser realizado com sucesso os dados devem ser tão puros e corretos quanto possível. Caso algum dado contenha informações incorretas pode provocar problemas na classificação dos resultados, como conclusões mais generalizadas do que as possíveis quando os dados estão corretos. É importante desconsiderar dados repetidos, por exemplo, e deve-se tomar cuidado para não confundir fenômenos interessantes com anomalias, pois essa pode se tornar a chave de pontos centrais.
- **Seleção de Dados:** para uma amostra considerável de dados pode ser realizada uma seleção dos dados a serem utilizados. Essa seleção deve ser realizada de forma a somente desconsiderar dados irrelevantes, pois uma análise incorreta pode acabar descartando dados necessários no futuro. Com uma quantidade de dados moderada, a ordenação de valores não é um processo complexo. Quando utilizada uma amostra maior, a ordenação dos dados torna-se uma tarefa difícil.

- Pré-Processamento de Dados: neste momento do processo atributos não existentes na base original podem ser criados, desde que os dados utilizados para gerá-los estejam presentes. Podem ser cálculos como, somatório de dados ou diferença entre dois valores.

2.3 MINERAÇÃO DE DADOS

Após as etapas preparatórias dos dados são realizadas as etapas de mineração dos dados para identificação de padrões existentes da base de dados.

- Tarefa de Mineração: deve ser determinada de acordo com o objetivo da mineração, por exemplo: classificação, regressão, agrupamento, agregação, modelagem de dependência e detecção de mudança e desvio.
- Técnicas de Mineração: estes serão os métodos utilizados para identificação de padrões de conhecimento. Alguns exemplos: métodos estatísticos, métodos lineares, método de árvores de decisão e regras, métodos de aprendizagem relacional, métodos de algoritmos genéticos e método de redes neurais.
- Padrões de Interesse: nesta etapa são identificados os padrões de interesse de forma representativa, como regras de classificação, regressão, agrupamento, árvores de decisão e outros.
- Interpretação: os dados descobertos são analisados e interpretados, se necessário podem ser realizados os passos anteriores novamente, de forma a identificar novos dados de outras formas.
- Consolidação: por último se tem a consolidação do conhecimento descoberto e as ações pertinentes, como documentação adequada e relatório aos interessados.

2.4 TIPOS DE TAREFAS PARA MINERAÇÃO DE DADOS

A seguir são descritos os tipos de tarefas que podem ser utilizadas para o processo de Mineração de Dados.

2.4.1 Classificação

A tarefa de classificação consiste em construir um modelo de algum tipo que possa ser aplicado a dados não classificados visando categorizá-los em classes. Um objeto é examinado e classificado de acordo com uma classe definida (HARRISON, 1998).

São exemplos de tarefas de classificação: classificar pedidos de créditos como de baixo, médio e alto risco, esclarecer pedidos de seguros fraudulentos, identificar a forma de tratamento na qual um paciente está mais propício a responder, baseando-se em classes de pacientes que respondem bem a determinado tipo de tratamento médico.

2.4.2 Estimativa (ou Regressão)

A estimativa é usada para definir um valor para alguma variável contínua desconhecida como, por exemplo, receita, altura ou saldo de cartão de crédito (HARRISON, 1998). Ela lida com resultados contínuos, enquanto que a classificação lida com resultados discretos. Ela pode ser usada para executar uma tarefa de classificação, convencendo-se que diferentes faixas (intervalos) de valores contínuos correspondem a diferentes classes.

Como exemplos de tarefas de estimativa têm-se (FAYYAD, 1996): estimar o número de filhos em uma família, estimar a renda total de uma família, estimar o valor em tempo de vida de um cliente, estimar a probabilidade de que um paciente morrerá baseando-se nos resultados de um conjunto de diagnósticos médicos, prever a demanda de um consumidor para um novo produto.

2.4.3 Associação

A tarefa de associação consiste em determinar quais itens tendem a serem adquiridos juntos em uma mesma transação. O exemplo clássico é determinar quais produtos costumam ser disponibilizados juntos em um carrinho de supermercado. As cadeias de varejo usam associação para planejar a disposição dos produtos nas prateleiras

das lojas ou em um catálogo, de modo que os itens geralmente adquiridos na mesma compra sejam vistos próximos entre si (HARRISON, 1998).

2.4.4 Segmentação (ou Clustering)

A segmentação é um processo de partição de uma população heterogênea em vários subgrupos ou clusters mais homogêneos (HARRISON, 1998). Na segmentação, não há classes predefinidas, os registros são agrupados de acordo com a semelhança, o que a diferencia da tarefa de classificação.

Exemplos de segmentação: agrupar os clientes por região do país, agrupar clientes com comportamento de compra similar, agrupar seções de usuários Web para prever comportamento futuro de usuário.

2.4.5 Sumarização

Segundo Fayyad (1996), a tarefa de sumarização envolve métodos para encontrar uma descrição compacta para um subconjunto de dados. Um simples exemplo desta tarefa poderia ser tabular o significado e desvios padrão para todos os itens de dados. Métodos mais sofisticados envolvem a derivação de regras de sumarização. As tarefas de mineração de dados, descritas acima, são apresentadas de forma resumida na Tabela 1.

Tabela 1: Tarefas de Mineração de Dados

Tarefa	Descrição	Exemplos
Classificação	Constrói um modelo de algum tipo que possa ser aplicado a dados não classificados a fim de categorizá-los em classes.	Classificar pedidos de crédito; Esclarecer pedidos de seguros fraudulentos; Identificar a melhor forma de tratamento de um paciente; Identificar mercado alvo;
Estimativa (ou Regressão)	Usada para definir um valor para alguma variável contínua desconhecida.	Estimar o número de filhos ou a renda total de uma família; Estimar o valor em tempo de vida de um cliente; Estimar a probabilidade de que um paciente morrerá baseando-se nos resultados de diagnósticos médicos; Prever a demanda de um consumidor para um novo

		produto;
Associação	Usada para determinar quais itens tendem a serem adquiridos juntos em uma mesma transação.	Determinar quais produtos costumam ser colocados juntos em um carrinho de supermercado;
Tarefa	Descrição	Exemplos
Segmentação	Processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos.	Agrupar clientes por região do país; Agrupar clientes com comportamento de compra similar; Agrupar seções de usuários Web para prever comportamento futuro de usuário;
Sumarização	Envolve métodos para encontrar uma descrição compacta para um subconjunto de dados.	Tabular o significado e desvios padrão para todos os itens de dados; Derivar regras de síntese;

2.5 TÉCNICAS DE MINERAÇÃO DE DADOS

Não há uma técnica que solucione todos os problemas de Mineração de Dados. É necessário utilizar diversas técnicas em conjunto com as tarefas para atingir diferentes objetivos, cada método oferece suas vantagens e suas desvantagens (HARRISON, 1996). A familiaridade com as técnicas é necessária para facilitar a escolha de uma delas de acordo com os problemas apresentados. A seguir são descritas as técnicas de mineração de dados normalmente usadas.

2.5.1 Descoberta de Regras de Associação

A técnica de descoberta de regras de associação estabelece uma correlação estatística entre certos itens de dados em um conjunto de dados (GOEBEL e GRUENWALD, 1999).

Um exemplo desta regra pode ser que 90% dos clientes que compram leite, também compram pão; o percentual de 90% é denominado a confiança da regra. O suporte da regra leite→pão é o número de ocorrências deste conjunto de itens na mesma transação.

A técnica de descoberta de regras de associação é apropriada à tarefa de associação. Como exemplos de algoritmos que implementam regras de associação tem-se: Apriori, AprioriTid, AprioriHybrid, AIS, SETM e DHP.

2.5.2 Árvores de Decisão

Uma árvore de decisão é uma árvore onde cada nó não terminal representa um teste ou decisão sobre o item de dado considerado (GOEBEL e GRUENWALD, 1999). O objetivo principal é separar as classes. Conjuntos de classes diferentes tendem a ser alocados em subconjuntos diferentes, cada um descrito por regra simples em um ou mais itens de dados. Essas regras podem ser expressas como declarações lógicas, em uma linguagem como SQL, de modo que possam ser aplicados diretamente a novos conjuntos. Uma das vantagens principais das árvores de decisão é o fato de que o modelo é intuitivo, uma vez que tem a forma de regras explícitas (HARRISON, 1998).

A técnica de árvore de decisão, em geral, é apropriada às seguintes tarefas: classificação e regressão.

Alguns exemplos de algoritmos de árvore de decisão são: CART, CHAID, C5.0, Quest, ID-3, SLIQ, J48 e SPRINT.

2.5.3 Raciocínio Baseado em Casos

Também conhecido como MBR (Memory-Based Reasoning – raciocínio baseado em memória), o raciocínio baseado em casos tem base no método do vizinho mais próximo. “O MBR procura os vizinhos mais próximos nos exemplos conhecidos e combina seus valores para atribuir valores de classificação ou de previsão” (HARRISON, 1998, p. 195). Tenta solucionar um dado problema fazendo uso direto de experiências e soluções passadas. A distância dos vizinhos dá uma medida da exatidão dos resultados.

Na aplicação do MBR, segundo Berry e Linoff (1997), existem quatro passos importantes:

- Escolher o conjunto de dados de treinamento
- Determinar a função de distância

- Escolher o número de vizinhos mais próximos
- Determinar a função de combinação

A técnica de raciocínio baseado em casos é apropriada às tarefas de classificação e segmentação.

Os seguintes algoritmos implementam a técnica de raciocínio baseado em casos: BIRCH, CLARANS e CLIQUE.

2.5.4 Algoritmos Genéticos

Os algoritmos genéticos são métodos generalizados de busca e otimização que simulam os processos naturais de evolução. Um algoritmo genético é um procedimento iterativo para evoluir uma população de organismos e é usado em mineração de dados para formular hipóteses sobre dependências entre variáveis, na forma de algum formalismo interno (GOEBEL e GRUENWALD, 1999).

Os algoritmos genéticos usam os operadores de seleção, cruzamento e mutação para desenvolver sucessivas gerações de soluções. Com a evolução do algoritmo, somente as soluções com maior poder de previsão sobrevivem, até os organismos convergirem em uma solução ideal (HARRISON, 1998).

A técnica de algoritmos genéticos é apropriada às tarefas de classificação e segmentação.

Exemplos de algoritmos genéticos: Algoritmo Genético Simples, Genitor e CHC, Algoritmo de Hillis, GA-Nuggets, GA-PVMINER.

2.5.5 Redes Neurais Artificiais

As redes neurais são uma classe especial de sistemas modelados seguindo analogia com o funcionamento do cérebro humano e são formadas de neurônios artificiais conectados de maneira similar aos neurônios do cérebro humano (GOEBEL e GRUENWALD, 1999).

Uma das principais vantagens das redes neurais é sua variedade de aplicação, mas os seus dados de entrada são difíceis de serem formados e os modelos produzidos por elas são difíceis de entender (HARRISON, 1998).

A técnica de redes neurais é apropriada às seguintes tarefas: classificação, estimativa e segmentação.

Exemplos de redes neurais: Perceptron, Rede MLP, Redes de Kohonen, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB

A Tabela 2 apresenta um resumo das técnicas de Mineração de Dados descritas.

Tabela 2: Técnicas de Mineração de Dados

Técnica	Descrição	Tarefas	Exemplos
Descoberta de Regras de Associação	Estabelece uma correlação estatística entre atributos de dados e conjuntos de dados.	Associação	Apriori, AprioriTid, AprioriHybrid, AIS, SETM e DHP.
Árvores de Decisão	Hierarquização dos dados, baseada em estágios de decisão (nós) e na separação de classes e subconjuntos.	Classificação Regressão	CART, CHAID, C5.0, Quest, ID-3, SLIQ e SPRINT.
Raciocínio Baseado em Casos ou MBR	Baseado no método do vizinho mais próximo combina e compara atributos para estabelecer hierarquia de semelhança.	Classificação Segmentação	BIRCH, CLARANS e CLIQUE.
Algoritmos Genéticos	Métodos gerais de busca e otimização, inspirados na Teoria da Evolução, onde a cada nova geração, soluções melhores têm mais chance de ter “descendentes”.	Classificação Segmentação	Algoritmo Genético Simples, Genitor, CHC, Algoritmo de Hillis, GA -Nuggets, GA – PVMINER.
Redes Neurais Artificiais	Modelos inspirados na fisiologia do cérebro, onde o conhecimento é fruto do mapa das conexões neuronais e dos pesos dessas conexões.	Classificação Segmentação	Perceptron, Rede MLP, Redes de Kohonen, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB

3 MÉTODOS E PROCEDIMENTOS

Neste capítulo serão apresentadas as formas definidas para a pesquisa e formulação do trabalho proposto.

3.1 DELINEAMENTO DA PESQUISA

Nesta pesquisa foi utilizado o método de estudo de caso, que é comumente utilizado quando o pesquisador tem pouco controle sobre os eventos e quando o foco se encontra em fenômenos contemporâneos, inseridos em algum contexto da vida real (YIN, 2001).

Como foram necessárias consultas a registros de projetos, negociações anteriores, contratos firmados, entrevistas com responsáveis pelas soluções ofertadas e acompanhamento de rotinas optou-se pelo estudo de caso para explicitar os conhecimentos do trabalho.

3.2 DEFINIÇÃO DA UNIDADE DE ANÁLISE

A pesquisa foi realizada junto à empresa XYZ e foram utilizadas suas bases de dados de clientes e serviços. A principal área de interesse é a área comercial e de novos negócios, devido ao seu alto interesse no faturamento da empresa a partir da identificação de clientes em potencial.

3.3 TÉCNICAS DE COLETA DE DADOS

A seguir são detalhadas as formas como as informações foram coletadas na empresa XYZ para a utilização neste trabalho.

Entrevistas: foram elaboradas duas entrevistas com as pessoas chave do processo de negócio da empresa XYZ com lista de tópicos a serem explicados pelos entrevistados. Foram entrevistados dois colaboradores da empresa XYZ, um responsável pelas vendas das

soluções ofertadas pela empresa e o outro pela criação e desenvolvimento de uma das soluções existentes. Foram realizadas duas entrevistas com cada colaborador. Nas primeiras foram identificadas as expectativas com o trabalho de Mineração de Dados e as informações a serem apresentadas. Após a mineração foram realizadas novas entrevistas, para apresentação dos resultados, verificação se as expectativas iniciais foram atendidas e programação para continuidade no processo de Mineração de Dados na empresa XYZ.

Índices e Relatórios/Textos e Documentos: E-mails, pautas de reuniões, relatórios de desempenho e especificações de serviços. A técnica de textos e documentos foi selecionada devido à confiabilidade de suas informações e também do processo elaborado para atualização periódica.

Além das entrevistas e da coleta de documentação foram utilizadas as bases de clientes que a empresa XYZ mantém em seu ambiente de testes. O ambiente tem informações reproduzidas do ambiente produtivo da empresa, no entanto, essa base foi utilizada a fim de preservar as informações dos clientes. Após apresentação dos resultados fica a cargo da empresa a decisão sobre a utilização da Mineração de Dados em sua base de clientes real.

3.4 TÉCNICAS DE ANÁLISE DE DADOS

Serão considerados e analisados todos os dados relevantes à busca de conhecimento e entendimento das funcionalidades e processos existentes na empresa para melhor detalhar as soluções que poderão ser sugeridas com as descobertas de conhecimentos através da mineração dos dados disponibilizados.

Detalhando por cada uma das técnicas:

3.4.1 Entrevistas em profundidade

Foram analisadas as entrevistas realizadas anteriormente com analistas de negócios e o gerente responsável pela solução, a fim de entender a real necessidade das áreas. As principais informações descobertas foram documentadas para consultas futuras.

3.4.2 Índices e Relatórios

Identificação de pontos de melhoria e foco para a pesquisa. As metas são importantes para o acompanhamento da evolução das melhorias que foram ser propostas.

3.4.3 Textos e Documentos

Informações de detalhamento das oportunidades e qual a prioridade de cada oferta para empresa XYZ, como, por exemplo, o percentual de faturamento.

3.5 LIMITAÇÕES DO MÉTODO

A pesquisa poderá ser afetada caso informações relevantes ao resultado do trabalho não sejam repassadas ao entrevistador. A forma de abordagem da pesquisa e a pré-elaboração da pesquisa é fundamental para que essa etapa seja realizada com sucesso.

O mesmo poderá ocorrer caso os índices e relatórios apresentados não estejam atualizados ou sua análise seja feita de forma incorreta.

Considerando-se que a base de dados da empresa tenha um conteúdo extenso, será utilizada somente uma amostragem dessa base, trazendo como limitação que a amostragem selecionada corresponde somente à parte da realidade da empresa.

4 APRESENTAÇÃO E ANÁLISE DOS DADOS

Os fechamentos comerciais são negociados pela equipe comercial com clientes da base e também com novos clientes a partir de planilhas de clientes pré-estabelecidos, ranking de revistas, ou procuras por parte dos clientes, no qual a área comercial da empresa atende de forma passiva. Esses fechamentos são identificados de forma manual e podem ser solicitados também por atendentes do Suporte Técnico ou de demais áreas de atendimento a cliente, como por exemplo, a área de Implantação de Projetos. De qualquer maneira, a identificação das oportunidades de fechamento utilizada hoje é realizada somente com a intervenção humana.

4.1 NECESSIDADE DA EMPRESA XYZ

Para entender a necessidade da empresa XYZ foram necessárias leituras de documentações, análise de bancos de dados, mas principalmente a interação com colaboradores da empresa.

A seguir, seguem os dois questionários realizados na identificação da necessidade da empresa XYZ. A primeira entrevista foi realizada com um dos analistas comerciais da empresa, que trabalha diretamente em contato com os clientes e suas necessidades. Os analistas comerciais tem perfis de negociação e atendimento a clientes. O primeiro entrevistado tem mais de sete anos de experiência na função de analista comercial, sempre em contato com seus clientes para identificar qual solução atende às necessidades deste no momento. Na empresa XYZ os analistas de negócios são responsáveis pelos novos negócios fechados com os clientes, assim como a identificação de novos clientes em potencial. Já a segunda entrevista foi realizada com o coordenador de uma das soluções ofertadas pela empresa XYZ a seus clientes.

O analista de produto da empresa XYZ entrevistado trabalha há mais de oito anos na empresa e já passou por várias áreas e atividades, tendo criado ao longo destes anos um vasto conhecimento sobre os clientes, suas necessidades e as soluções que empresa XYZ pode ofertar a estes clientes. A responsabilidade de evolução de cada solução, como a

descrição do *roadmap* do produto, criação de novas soluções e aprimoramento das existentes é tratado diretamente com os coordenadores de soluções na empresa XYZ.

4.1.1 Primeiro Questionário

As questões feitas aos dois responsáveis foram as seguintes:

- 1- Qual a sua área dentro da empresa XYZ?
- 2- Como a Mineração de Dados pode auxiliar no seu processo?
- 3- Qual o tipo de informação que a Mineração de Dados pode descobrir?
- 4- Como acompanhar a informação descoberta?
- 5- Quem são os responsáveis/como será o processo para manter a descoberta do conhecimento?

A seguir serão apresentadas as respostas obtidas nas entrevistas. As respostas serão reportadas na íntegra.

Entrevistado 1: Perfil Comercial

- 1- Trabalho na área comercial, ofertando as soluções a nossos clientes de acordo com a necessidade de cada um.
- 2- Não conheço o processo de mineração, então posso te dizer o problema que eu tenho. Hoje tenho dúvidas sobre a solução ideal para cada tipo de cliente. Já aconteceu de eu perder uma oportunidade de venda por ofertar a solução incorreta, o cliente não conheceu as demais soluções que poderiam atendê-lo. A mineração pode me ajudar nisso?! Eu conheço as soluções que devo ofertar, mas é difícil pensar que uma empresa grande precisa de somente uma solução e não de outra, entende? É mais fácil com empresas pequenas.
- 3- O que eu gostaria mesmo era que a mineração informasse a solução a ser ofertada para determinada empresa.
- 4- Não faço ideia.
- 5- Como hoje cada analista comercial interpreta a necessidade dos seus clientes, acho que poderia continuar dessa forma.

Entrevistado 2: Perfil de Produto

- 1- Trabalho na área de produto. A partir das necessidades dos clientes verifico se ela pode se tornar uma nova solução a ser ofertada para outras empresas. Também é necessário verificar tendências, evolução na cadeia de suprimentos e no processo de cada empresa.
- 2- Identificando oportunidades, soluções, melhorias que possam ser ofertadas ou ainda desenvolvidas para os nossos clientes ou auxiliar na identificação de novos clientes.
- 3- Imagino que áreas onde certos clientes atuam e que a empresa XYZ ainda não domine, não que nós iríamos acatar, mas saber disso poderia nos ajudar de alguma forma. Clientes que ultrapassam seus limites de tráfego no EDI, por exemplo.
- 4- Hoje não existem estes indicadores. Mesmo com o auxílio de uma ferramenta que nos traga uma informação será necessário criar um processo para manter essa rotina de descobertas e acompanhamento. Como podemos ter certeza que foi a mineração que nos trouxe essa informação? Se utilizássemos as informações em planilhas excel, não poderíamos descobrir de outra forma?
- 5- Atualmente não temos responsáveis, seria necessário envolver a alta gerência para decidir esse ponto. Provavelmente cada produto ofertado seria responsável por manter essa busca, como um *roadmap* de produto.

As duas entrevistas trouxeram necessidades diferentes, mas com um ponto em comum, a identificação de soluções para determinadas empresas. Desta forma, para o presente trabalho optou-se por focar justamente no processo de identificação da melhor oferta para determinado cliente. Para isso foram buscadas bases de clientes e soluções para montarmos uma base única a fim de identificar perfis de empresas para determinadas soluções.

4.2 DEFINIÇÃO DE TAREFA E TÉCNICA DE MINERAÇÃO DE DADOS

A definição da técnica a ser utilizada na Mineração de Dados depende da tarefa definida e também dos dados disponibilizados para a mineração. Um ponto importante

nesta definição é o objetivo principal da mineração. Determinando-se o tipo de conhecimento a ser explorado, como por exemplo, regras características ou de classificação, é possível escolher a técnica necessária para a mineração.

Para determinação de tarefa e técnica a serem utilizados no presente do trabalho foram utilizadas as informações descritas nas tabelas 1 e 2. Em um dos exemplos de aplicação exibidos na primeira tabela é mencionado identificação de mercado-alvo, utilizando técnica de Classificação e técnica de Árvore de Decisão. Para a identificação dos perfis de clientes da empresa XYZ foram utilizadas essas duas formas, que comparada com os demais exemplos de aplicação foi escolhida como mais adequada ao objetivo principal.

Os detalhes do processo para mineração são detalhados nos próximos capítulos.

4.2.1 Preparação de Dados

- **Transformação de Dados:** Os dados utilizados para a mineração foram obtidos de documentação de projetos, bases de clientes e de soluções de formatos diferentes, alguns em documentos word outros em planilhas excel e até formato texto sem formatação, além de bases extraídas do banco de dados da empresa. Todos foram unificados em uma grande planilha contendo mais de 10.000 linhas. As colunas para identificação das informações também continham dados diversos, uma vez que foram unificadas informações de clientes (CNPJ, razão social, localização, número de funcionários, faturamento anual, ramo de atividade), de soluções (nome comercial da solução, características, valor inicial, valor recorrente) e também informações de bases que ligam uma informação a outra, como por exemplo, qual cliente utiliza determinada solução hoje.
- **Limpeza de Dados:** Para não descartar informações que poderiam ser interessantes no futuro, foram mantidas todas as linhas e colunas de informações no momento da limpeza. Somente foram desconsiderados dados incompletos, como clientes com dados faltantes e soluções que não estavam ligadas a nenhum cliente atualmente.

- Seleção de Dados: Como a base original era muito extensa e dificultava a mineração, informações irrelevantes para a identificação do perfil de cliente de cada solução foram descartadas. Ainda assim, a base a ser utilizada na mineração era muito grande e a fim de facilitar a mineração foi selecionada uma parte aleatória da base. A base ficou restringida a 35 linhas contemplando informações que ligam tipos de empresas a cada solução. Cada coluna restante para a mineração será detalhada no próximo capítulo.
- Pré-Processamento de Dados: Não foram criados novos atributos na base restante para a mineração.

4.2.2 Atributos da Mineração

A seguir são detalhados os atributos selecionados para a mineração de dados na empresa XYZ.

Faturamento anual: Na base utilizada para mineração foram considerados faturamentos alto, médio e baixo, sendo que empresas com faturamento maior que R\$300.000.000,00 foram consideradas com faturamento alto, empresas com faturamento menor que R\$2.400.000,00 foram consideradas com faturamento baixo e empresas com o faturamento entre estes valores foram classificadas como faturamento médio.

Número de Empregados: Para manter a menor quantidade de dados numéricos possíveis na base a ser minerada, a informação de número de empregados também foi convertida para alto, médio e baixo. Empresas com mais de 500 empregados tem alto número de empregados, empresa com menos de 100 empregados tem um número baixo e empresas com número entre estes valores tem classificação média para o número de empregados.

Parceiros de Negócio: Como a empresa XYZ atende empresas da cadeia de suprimentos e demanda, existem soluções específicas para cada tipo de parceiro desta malha. Sendo assim, um dos atributos mantidos na base é o tipo de parceiro envolvido no processo: compradores, fornecedores, distribuidores e representantes,

Ramo de atuação: Este atributo determina o ramo de atuação de cada empresa, como: alimentos, bebidas, beleza, eletroeletrônicos e construção.

Sistema de Gestão Integrado: Continuando com os atributos de classificação das empresas é importante determinar o grau de tecnologia aplicado em cada uma. Uma regra básica é determinar aquelas que têm sistemas de gestão integrados para fins gerenciais em seus processos.

Média Mensal de Documentos: Na cadeia de suprimentos os documentos mais comumente trocados entre as empresas parceiras de negócio são pedidos e notas fiscais. Este atributo classifica a média mensal de troca destes documentos entre as empresas. Mantendo a mesma linha dos demais campos numéricos, foi realizada classificação de alto, médio e baixo. Alto tráfego mensal de documentos para empresas que trocam mais de 10.000 documentos, baixo tráfego para empresas que trocam menos de 5.000 documentos e médio tráfego para aquelas que trocam documentos entre 5.001 e 9.999 documentos.

Solução: O último atributo, classificador, determina a solução utilizada por cada uma das empresas utilizadas na base da mineração. As soluções utilizadas na base foram somente três das comercializadas pela empresa XYZ: EDI (sigla em português para Troca Eletrônica de Informações), Web e Portais Colaborativos.

A nomenclatura das soluções ofertadas pela empresa XYZ foi adaptada para facilitar sua interpretação neste trabalho. Os detalhes das soluções utilizadas na mineração foram adquiridos a partir dos documentos disponibilizados pela empresa e um resumo de cada uma segue abaixo:

A solução Web disponibiliza em um site as informações de documentos de pedidos e notas fiscais para visualização e impressão dessas informações. O download destes documentos somente é possível em formato PDF. Essa solução exige da empresa contratante somente acesso à internet, sem necessidade de investimentos em infraestrutura.

Já a solução de EDI disponibiliza as informações trocadas entre os parceiros de negócio diretamente no sistema de gestão de cada uma. O formato para recebimento das informações é estabelecido por cada empresa, de acordo com o formato importado pelo sistema escolhido. Nesta solução também estão contempladas soluções de conectividade que são responsáveis pelo envio e recebimento dos documentos em cada empresa. Esta solução é mais detalhada que a solução Web, pois exige tempo de implementação e também envolvimento técnico das empresas parceiras.

Portais Colaborativos são criados para processos mais detalhados e complexos. Somente a troca eletrônica das informações não é necessária para suprir a necessidade de uma empresa que tenha processos de negócio muito específicos e muitos parceiros envolvidos. O Portal é criado com a identidade visual da empresa contratante e suas funcionalidades são específicas para cada cliente.

A base já pronta e utilizada na Mineração de Dados ficou da forma exibida na tabela 2 a seguir. Na criação da base foram tomados os devidos cuidados para que os clientes da empresa XYZ não fossem identificados.

Tabela 3: Base para Mineração

Faturamento	Empregados	Parceiros	Atuação	ERP Integrado	Docs Mensal	Solução
Alto	Alto	Compradores	Alimentos	Sim	Alta	EDI
Alto	Alto	Compradores	Alimentos	Sim	Alta	EDI
Alto	Alto	Compradores	Alimentos	Sim	Alta	EDI
Alto	Alto	Compradores	Bebidas	Sim	Alta	EDI
Alto	Alto	Fornecedores	Bebidas	Sim	Alta	EDI
Alto	Alto	Compradores	Beleza	Sim	Alta	EDI
Alto	Alto	Compradores	Bebidas	Sim	Alta	EDI
Alto	Alto	Compradores	Beleza	Sim	Alta	EDI
Alto	Baixo	Compradores	Beleza	Sim	Alta	EDI
Alto	Alto	Fornecedores	Eletro	Sim	Alta	EDI
Alto	Alto	Fornecedores	Eletro	Sim	Alta	EDI
Alto	Alto	Fornecedores	Eletro	Sim	Alta	EDI
Alto	Alto	Compradores	Construção	Sim	Baixa	EDI
Alto	Alto	Compradores	Eletro	Sim	Alta	EDI
Alto	Alto	Compradores	Eletro	Sim	Baixa	EDI
Alto	Alto	Compradores	Eletro	Não	Baixa	WEB
Alto	Alto	Compradores	Alimentos	Não	Baixa	WEB
Alto	Alto	Distribuidores	Alimentos	Sim	Alta	Portal
Alto	Alto	Representantes	Alimentos	Sim	Alta	Portal
Alto	Alto	Representantes	Bebidas	Sim	Alta	Portal
Alto	Alto	Representantes	Beleza	Sim	Alta	Portal
Alto	Alto	Representantes	Bebidas	Sim	Alta	Portal
Alto	Alto	Representantes	Beleza	Sim	Alta	Portal
Alto	Médio	Representantes	Beleza	Sim	Alta	Portal
Baixo	Alto	Compradores	Alimentos	Sim	Baixa	EDI
Baixo	Baixo	Compradores	Alimentos	Não	Baixa	WEB
Médio	Alto	Compradores	Eletro	Sim	Baixa	EDI
Médio	Médio	Compradores	Alimentos	Sim	Média	EDI

Faturamento	Empregados	Parceiros	Atuação	ERP Integrado	Docs Mensal	Solução
Médio	Médio	Compradores	Bebidas	Sim	Baixa	EDI
Médio	Alto	Compradores	Construção	Sim	Baixa	EDI
Médio	Alto	Compradores	Eletro	Não	Baixa	WEB
Médio	Alto	Compradores	Eletro	Sim	Baixa	EDI
Médio	Alto	Compradores	Bebidas	Não	Baixa	WEB
Médio	Médio	Distribuidores	Bebidas	Sim	Baixa	EDI
Médio	Alto	Distribuidores	Construção	Sim	Baixa	EDI

4.2.3 Tarefa e Algoritmo

Como mencionado anteriormente, a técnica e algoritmo de Mineração de Dados utilizado neste trabalho foi Árvore de Decisão que representam a informação através de Classificação de regras ou árvores.

O principal objetivo da tarefa de classificação é descobrir algum tipo de relacionamento entre os atributos preditivos (faturamento, empregados, etc.) e o atributo objetivo (solução) (CAZELLA, 2010).

Já a Árvore de Decisão é uma forma simples de representação, classificando os exemplos de uma base finita de atributos.

O algoritmo J48 utilizado na mineração constrói um modelo de Árvore de Decisão baseado num conjunto de dados de treinamento, e usa esse modelo para classificar a exatidão do classificador num conjunto de teste (VIANNA, 2006).

A figura 2 mostra a base de dados formatada para mineração utilizando o software Weka.

A primeira linha @relation é como um título para as informações a serem mineradas. Em seguida são descritos cada atributo (@attribute) e por fim, a partir da linha @data, são descritas todas as informações a serem mineradas.

```

1 @relation solucoes
2
3 @attribute faturamento {alto, medio, baixo}
4 @attribute funcionarios {alto, medio, baixo}
5 @attribute parceiros {compradores, fornecedores, distribuidores, representantes}
6 @attribute atuacao {alimentos, bebidas, beleza, eletro, construcao}
7 @attribute erp {sim, nao}
8 @attribute mediadocsmes {alto, medio, baixo}
9 @attribute solucao {edi, web, portal}
10
11 @data
12 alto,alto,compradores,alimentos,sim,alto,edi
13 alto,alto,compradores,alimentos,sim,alto,edi
14 alto,alto,compradores,alimentos,sim,alto,edi
15 alto,alto,compradores,bebidas,sim,alto,edi
16 alto,alto,fornecedores,bebidas,sim,alto,edi
17 alto,alto,compradores,beleza,sim,alto,edi
18 alto,alto,compradores,bebidas,sim,alto,edi
19 alto,alto,compradores,beleza,sim,alto,edi
20 alto,baixo,compradores,beleza,sim,alto,edi
21 alto,alto,fornecedores,eletro,sim,alto,edi
22 alto,alto,fornecedores,eletro,sim,alto,edi
23 alto,alto,fornecedores,eletro,sim,alto,edi
24 alto,alto,compradores,construcao,sim,baixo,edi
25 alto,alto,compradores,eletro,sim,alto,edi
26 alto,alto,compradores,eletro,sim,baixo,edi
27 alto,alto,compradores,eletro,nao,baixo,web
28 alto,alto,compradores,alimentos,nao,baixo,web
29 alto,alto,distribuidores,alimentos,sim,alto,portal
30 alto,alto,representantes,alimentos,sim,alto,portal
31 alto,alto,representantes,bebidas,sim,alto,portal
32 alto,alto,representantes,beleza,sim,alto,portal
33 alto,alto,representantes,bebidas,sim,alto,portal
34 alto,alto,representantes,beleza,sim,alto,portal
35 alto,medio,representantes,beleza,sim,alto,portal
36 baixo,alto,compradores,alimentos,sim,baixo,edi

```

For Help, press F1 ln 13 col 45 46 00 PC ANSI

Figura 2: Tela de Base no Formato Weka

4.2.4 Conhecimento Descoberto

A árvore de decisão gerada após a mineração dos dados é exibida na figura 3.

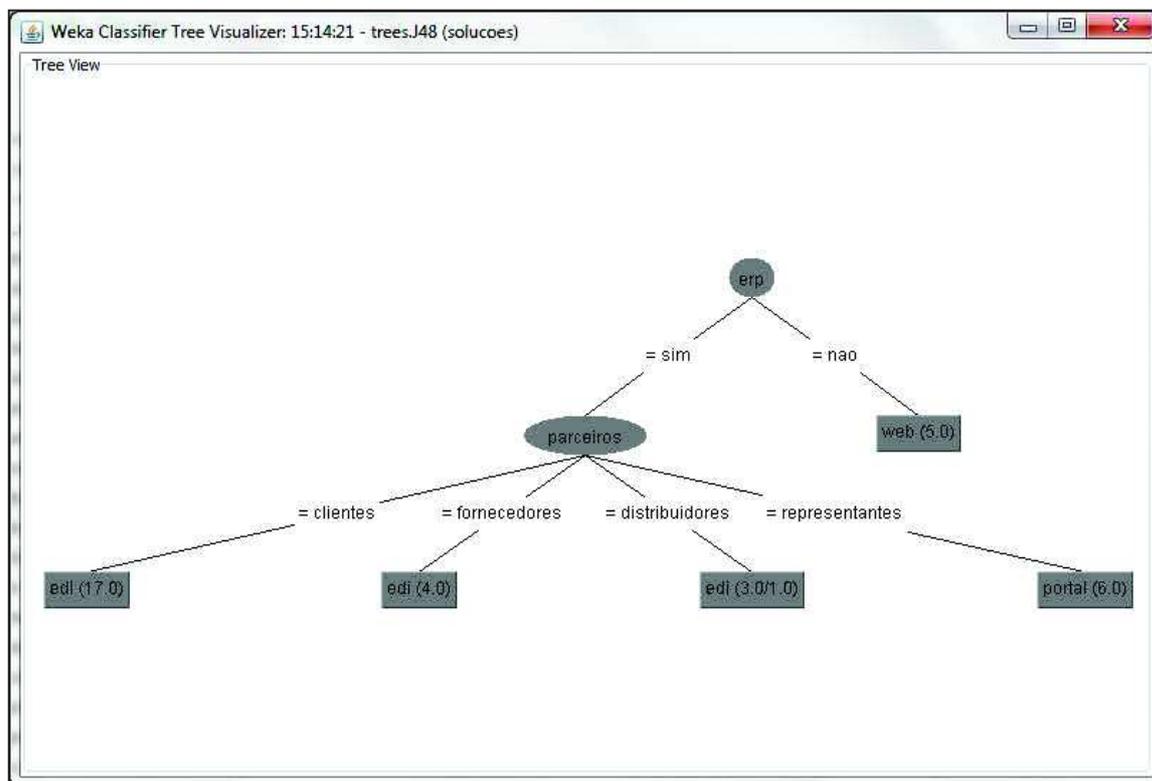


Figura 3: Árvore de Decisão

A partir da árvore de decisão gerada podemos interpretar algumas condições para determinadas soluções.

O atributo ERP exibido na tabela, equivalente ao Sistema de Gestão Integrado mencionado anteriormente, é determinante para a solução Web, na qual somente são exibidas informações para download ou acompanhamento visual no próprio site. Isto porque se uma empresa não tem um sistema de gestão integrado em sua empresa não será possível importar as informações recebidas em determinado formato, uma vez que não existirá destino para importação das informações. Como é possível identificar na árvore gerada, em cinco dos trinta e cinco dos casos minerados as empresas não continham sistema de gestão integrado e adquiriram a solução Web pela empresa XYZ.

As empresas da amostragem que têm o sistema de gestão integrado necessitam de uma interpretação dos parceiros de negócio para que então tenhamos uma definição da

solução. Para os parceiros como clientes, fornecedores e distribuidores a solução utilizada é EDI e para os parceiros que são representantes a solução são Portais Colaborativos.

Nota-se que os demais atributos utilizados na mineração não foram exibidos na Árvore de Decisão gerada. Isto devido às informações destes atributos não serem suficientemente informativos na amostra utilizada.

Antes da mineração acreditava-se que o porte da empresa era importante para a identificação do perfil das empresas para cada solução da empresa XYZ. Por este motivo, atributos como número de funcionários e faturamento anual foram mantidos na base para a mineração. Assim como a capacidade operacional da empresa em relação ao processo com seus parceiros, identificado na amostragem pelo atributo número de documentos trafegados.

O fato destes atributos não constarem na Árvore de Decisão gerada não significa que estes não são importantes no resultado. No entanto, a forma como estes atributos foram descritos na base não forneciam dados suficientemente importantes para a mineração.

4.2.5 Apresentação do Perfil de Clientes

Os resultados obtidos com a mineração dos dados foram apresentados aos colaboradores da empresa XYZ entrevistados antes da mineração. Com base na expectativa da mineração e visualização dos perfis de clientes obtidos como resultado novas entrevistas foram realizadas para interpretação dos colaboradores.

Nesta segunda etapa foram questionadas percepções dos responsáveis dos resultados obtidos e se as expectativas foram atingidas.

As questões foram as seguintes:

- 1- Os resultados obtidos estavam de acordo com a sua expectativa?
- 2- Como você acha que a Mineração de Dados possa ser realizada na empresa XYZ no futuro? Existem outros processos que ela poderia auxiliar?

Entrevistado 1: Perfil Comercial

- 1- Com os resultados do exemplo não me trouxe nenhuma informação nova. Embora me traga certeza de que a mineração funcionou, pois o resultado corresponde à realidade de nossos clientes. A informação pode não ser novidade

pra mim, mas pode ser para outras pessoas. Eu esperava o resultado justamente com os atributos que não apareceram no resultado.

- 2- Continuo com a mesma expectativa, que a mineração me auxilie na identificação da melhor solução para um cliente. Podemos evoluir com as demais soluções da empresa e adicionar outros atributos talvez. Agora que vi como funciona acho melhor que uma área seja responsável pelo controle dessas informações e então repasse aos analistas comerciais.

Entrevistado 2: Perfil de Produto

- 1- De acordo com a base utilizada para a mineração sim, mas minha expectativa como consultor de produto é mais para identificação de melhorias das nossas soluções. O perfil de cliente atende à nossa área comercial, e é importante para toda a organização que ela tenha esse conhecimento. Mas para produto, os atributos e mineração deveriam ser feitos com outro foco.
- 2- Acho que a mineração possa ser utilizada em outras áreas sim, como no suporte. Poderíamos criar uma ferramenta interna para estes casos, poderíamos fazer uma ligação com a ferramenta utilizada pelo suporte para o registro dos casos e então ter uma base de chamados e soluções e a partir dessas informações realizar a mineração.

4.3 IDENTIFICAÇÃO DE OPORTUNIDADES DE NEGÓCIO

Considerando-se as entrevistas e os resultados obtidos na mineração é possível afirmar que o processo de mineração de dados pode auxiliar a empresa XYZ na percepção de suas soluções através dos perfis de seus clientes, podendo assim, gerar novas oportunidades de negócios para empresas com perfis similares aos clientes de sua malha.

A descoberta deste conhecimento para a empresa XYZ traz vantagens no momento de negociação com seus clientes, pois possibilitará o foco nos clientes em potencial para

determinada solução, potencializando seu poder de barganha com preços de soluções para clientes estratégicos.

5 CONSIDERAÇÕES FINAIS

Baseando-se nas informações dos atributos e na Árvore de Decisões gerada comprovou-se que o processo de DCBD através da etapa de Mineração de Dados pode ser utilizado de forma a identificar padrões nas características dos clientes da empresa XYZ e auxiliá-la na identificação de oportunidades de fechamentos comerciais de acordo com o perfil de cada um. Os resultados obtidos com a Mineração de Dados podem ser usados no gerenciamento de informação, processamento de pedidos de informação, tomada de decisão, controle de processo e muitas outras aplicações dentro da empresa XYZ.

Voltando à pergunta central do trabalho: “Como técnicas de Mineração de Dados podem auxiliar no processo de identificação de oportunidades de negócios para empresas de serviços da cadeia de suprimentos e demanda?” podemos destacar que no exemplo da empresa XYZ o processo de Mineração de Dados é bastante útil e necessário, uma vez que a empresa tem informações distribuídas em várias bases de conhecimento e que hoje não são utilizadas de maneira a trazer todo o resultado que poderiam à empresa. A maioria das informações extraídas dessas bases de forma manual hoje é aproveitada pela empresa em áreas focadas, impossibilitando que as demais áreas que se beneficiariam com o conhecimento tomem parte da informação.

Além disso, o processo de Mineração de Dados se faz necessário para uniformizar os dados armazenados pela empresa XYZ. Durante o processo de limpeza realizado antes da mineração foram encontradas muitas dificuldades e também muito tempo foi empregado no processo para que a base pudesse ser utilizada da maneira correta. Os dados armazenados são provenientes de várias bases diferentes e a importação destes dados é realizada por diversas áreas da empresa. Não há controle de registro padrão para o cadastramento dessas informações de forma generalizada para todos. Cada área realiza o registro das informações da forma que achar adequada ao seu dia-a-dia, o que nem sempre espelha o dia-a-dia de outras pessoas na empresa.

A partir da utilização do processo de Mineração de Dados a empresa terá vantagens como a padronização dos seus dados armazenados, e a partir disso realizar o processo para a descoberta de conhecimento, aprimorando o processo de negociação com novos e antigos clientes.

Como mencionado anteriormente, a empresa XYZ tem bases de dados de áreas diversas e pelo acompanhamento realizado durante a pesquisa e baseando-se também nas documentações consultadas é possível afirmar que a mineração pode ser explorada na busca de conhecimento para aprimoramento das operações em várias áreas da empresa, tais como marketing, suporte técnico, desenvolvimento de soluções entre outras áreas.

Acredita-se que como trabalhos futuros possam ser desenvolvidos novos processos de mineração na empresa XYZ com objetivos específicos para cada área, como os exemplos a seguir:

- Marketing: técnicas de mineração de dados podem ser aplicadas para descobrir preferências de cada cliente e padrões de compra de soluções. Especificamente para o Marketing o trabalho apresentado pode ser continuado, baseando-se no perfil do consumidor já conhecido, outros dados podem ser adicionados à base para novas minerações.
- Financeiro: podem ser criados modelos que predizem quem será um bom cliente ou aquele que poderá se tornar inadimplente em seus pagamentos com base no histórico e perfil de pagadores e inadimplentes.
- Qualidade: podem ser encontrados padrões de condições para entregas de projetos de desenvolvimento e implantações das soluções.
- Comercial: em conjunto com a área de Marketing a área comercial poderá filtrar clientes com características que possam exceder suas franquias contratadas, antes que estes se tornem clientes inadimplentes.
- Suporte: determinar as escalas de atendimento de acordo com o perfil cliente juntamente com o perfil de cada analista, analisar padrões de atendimento, prever períodos críticos para cada solução.

Com a finalização do trabalho e a confirmação da pergunta principal do mesmo, além da apresentação dos exemplos de continuidade sobre o tema, espera-se ter contribuído para o aperfeiçoamento do processo comercial da empresa XYZ, a partir da utilização de uma técnica até então desconhecida pelos responsáveis por essa decisão, trazendo uma nova vantagem competitiva frente a seus clientes e também concorrentes, podendo trazer resultados expressivos de faturamento direto e indireto com a utilização em demais áreas da empresa.

REFERÊNCIAS

ADRIAANS, Pieter. **Data Mining**. Harlow: 1996. 158 páginas.

AGRAWAL, R. et al. **Fast algorithms for mining association rules in large databases**. in: International Conference on Very Large Data Bases, VLDB, 20, Santiago, Proceedings Hove: Morgan Kaufmann, p. 478-499, 1994.

AMARAL, Fernanda. C. N. **Data Mining: Técnicas e Aplicações para o Marketing Direto**. São Paulo: Berkeley do Brasil, 2001. 110 páginas.

BERRY, Michael J. A., LINOFF, Gordon. **Mineração de Dados: techniques for marketing, sales and customer support**. USA: Wiley Computer Publishing, 1997.

CAZELLA, Sílvio. **Data Mining**. Notas de Aula: MBA. São Leopoldo, 2010.

CHEN, Ming S., PARK, Jong S., YU, Philip S. **Efficient Data Mining for Path Traversal Patterns - Knowledge and Data Engineering**. Vol. 10(2), 209-221, Mars 1998.

FAYYAD, U. M., Piatetsky Shapiro, G., Smyth, P. & Uthurusamy, R. **Advances in Knowledge Discovery and Data Mining**. 1996.

GOEBEL, M., GRUENWALD, L. **A survey of data mining and knowledge discovery software tools**. SIGKDD Explorations, v. 1, p. 20-33, 1999.

HAND, David. **Principles of Data Mining**. Cambridge: Massachusetts Institute of Technology, 2001. 546 páginas.

HARRISON, Thomas H. **Intranet Data Warehouse**. São Paulo: Bekerly Brasil, 1998.

MEHTA, M., AGRAWAL R., RISSANEN, J. **SLIQ: A Fast Scalable Classifier for Data Mining**. Avignon, France, March 1996.

COOLEY R., MOBASHER B., SRIVASTAVA J.: **Web Mining: Information and Pattern Discovery on the World Wide Web**. In Proceedings of the Nineth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), pages 558--567, November 1997.

VIANNA, Glacyene. **Aplicação da Técnica de Árvore de Decisão Utilizando Algoritmo J48 para Analisar Ocorrência de Sinistralidade em uma Operadora de Seguro Saúde**. Bahia, jan. 2006. Disponível

em: < <http://pt.scribd.com/doc/55184267/5/Arvore-de-decisao-usando-o-algoritmo-J48>>. Acesso em: 01 jun. 2011.

YIN, Robert. **Estudo de caso: planejamento e métodos**. Porto Alegre: Bookman, 2003.