

UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA AGRÍCOLA

**MODELO DE GESTÃO ESTRATÉGICA COM FOCO NO  
CLIENTE PARA COMERCIALIZAÇÃO DE PRODUTOS  
ORGÂNICOS**

**TATIANA DENOBILE**

CAMPINAS  
FEVEREIRO DE 2005

UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA AGRÍCOLA

**MODELO DE GESTÃO ESTRATÉGICA COM FOCO NO  
CLIENTE PARA COMERCIALIZAÇÃO DE PRODUTOS  
ORGÂNICOS**

Dissertação submetida à banca  
examinadora para obtenção do título  
de Mestre em Engenharia Agrícola  
na área de concentração em  
Planejamento e Desenvolvimento  
Rural Sustentável

**TATIANA DENOBILE**

**Orientador: Prof. Dr. Luiz Henrique Antunes Rodrigues**

CAMPINAS  
FEVEREIRO DE 2005

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

D43m	<p>Denobile, Tatiana</p> <p>Modelo de gestão estratégica com foco no cliente para comercialização de produtos orgânicos / Tatiana Denobile. --Campinas, SP: [s.n.], 2005.</p> <p>Orientador: Luiz Henrique Antunes Rodrigues. Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola.</p> <p>1. Agricultura orgânica. 2. Concorrência. 3. Gestão de negócios – Serviços ao cliente. 4. Marketing de banco de dados. 5. Marketing direto. 6. Produtos agrícolas – Comercialização. I. Rodrigues, Luiz Henrique Antunes. II. Universidade Estadual de Campinas. Faculdade de Engenharia Agrícola. III. Título.</p>
------	--

Titulo em Inglês: Strategic management model with focus on customer to commercialize organic products.

Palavras-chave em Inglês: Organic agriculture, Competitive intelligence, Customer relationship management (CRM), Data mining, Direcct marketing e Agricultural marketing

Área de concentração: Planejamento e Desenvolvimento Rural Sustentável

Titulação: Mestre em Engenharia Agrícola

Banca examinadora: Moacir Roberto Darolt, Orandi M. Falsarella e Sylvio Luis Honório

Data da defesa: 28/02/2005

*“A permanência é uma ilusão. Somente a mudança é real.*

*É impossível pisar duas vezes no mesmo rio.”*

*Heraclitus*

*Aos meus pais José Luiz e Ivanice,  
a minha irmã Michela e ao Luiz Fernando,  
pelo eterno apoio, carinho e compreensão.*

## **Agradecimentos**

Ao Prof. Dr. Luiz Henrique Antunes Rodrigues pela sábia e amiga orientação, sem a qual este trabalho não seria possível.

Ao Sítio Vila Yamaguishi pela confiança e disponibilidade no fornecimento dos dados.

Ao CNPq, agência financiadora deste projeto.

Ao grande amigo Marcelo Lima e ao querido companheiro Luiz Fernando Pinto pela imensa ajuda e compartilhamento de seus conhecimentos.

A todas as pessoas da FEAGRI que colaboraram para a concretização deste trabalho e a todos os outros que também de alguma maneira contribuíram para sua realização.

## SUMÁRIO

LISTA DE FIGURAS.....	VII
LISTA DE TABELAS.....	IX
RESUMO.....	XI
ABSTRACT.....	XII
1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	4
3. REVISÃO BIBLIOGRÁFICA.....	6
3.1. Agricultura Orgânica.....	7
3.2. IC – Inteligência Competitiva.....	16
3.3. CRM – Gestão do Relacionamento com Cliente.....	21
3.4. Mineração de Dados.....	31
4. MATERIAL E MÉTODOS.....	49
4.1. Material.....	49
4.2. Métodos.....	57
5. RESULTADOS E DISCUSSÃO.....	73
5.1. Resultados preparatórios para a construção do modelo.....	74
5.2. Apresentação do modelo.....	82
6. CONSIDERAÇÕES FINAIS.....	126
7. TRABALHOS FUTUROS.....	130
8. REFERÊNCIAS.....	132
9. ANEXOS.....	135

## Lista de Figuras

1. Fases do ciclo de IC (MILLER, 2002).....	18
2. Fases do processo de CRM (SWIFT, 2001).....	22
3. Processo de Descoberta de Conhecimento em Bancos de Dados (FAYYAD, PIATETSKY-SHAPIRO,SMYTH, 1996).....	33
4. Exemplo de árvore de decisão simples.....	43
5. Atividade social Yamaguishi Jaguariúna - Limpeza dos rios.....	50
6. Atividade social Yamaguishi Jaguariúna - Soltura de alevinos.....	50
7. Logomarca do Sítio Vila Yamaguishi Jaguariúna.....	52
8. Hortas do Sítio Vila Yamaguishi Jaguariúna.....	52
9. Gráfico da porcentagem da geração de receita no período de 2001 a 2004 por grupo de produto.....	53
10. Gráfico do número e porcentagem pertencentes a cada tipo de cliente no período de 2001 a 2004.....	54
11. Participação de cada um dos tipos de clientes na geração de receita no período de 2001 a 2004.....	54
12. Processo para criação do atributo Recência.....	61
13. Processo para criação do atributo Frequência Acumulada.....	62
14. Processo para criação do atributo Valor Acumulado.....	64
15. Exemplificação do processo de criação do arquivo de Histórico.....	67
16. Comparação da porcentagem do número de pedidos do Sabor Natural com todos os demais clientes parceiros, no período de 2001 a 2004 .....	74
17. Comparação da receita gerada pelo Sabor Natural e os demais clientes jurídicos e parceiros juntos, no período de 2001 a 2004.....	74
18. Histograma da curva da frequência acumulada - clientes físicos.....	76
19. Histograma da curva da frequência acumulada - clientes jurídicos e parceiros....	76
20. Histograma da curva do valor acumulado - clientes físicos.....	77
21. Histograma da curva do valor acumulado - clientes jurídicos e parceiros.....	78
22. Histograma da curva do número de compras - clientes físicos.....	80
23. Histograma da curva do número de compras - clientes jurídicos e parceiros.....	80



24. Histograma da curva da receita bruta - clientes físicos.....	81
25. Histograma da curva da receita bruta - clientes jurídicos e parceiros.....	81
26. Fases do Modelo de Gestão Estratégica com foco no cliente, baseado nas fases do modelo de IC proposto por MILLER (2002).....	82
27. Distribuição do percentual do número de clientes físicos em relação à receita obtida entre 2001 e 2004.....	85
28. Número e percentual de Pedidos de clientes físicos entre 2001 e 2004.....	86
29. Árvore de clientes físicos gerada pelo software de mineração.....	92
30. Representação gráfica da árvore de clientes físicos.....	94
31. Modelo para interpretação da tabela dos perfis de cliente.....	97
32. Árvore de clientes jurídicos e parceiros gerada pelo software de mineração.....	99
33. Representação gráfica da árvore dos clientes jurídicos e parceiros.....	99

## Lista de Tabelas

1. Diferenças entre o marketing tradicional e o CRM (PEPPERS and ROGERS GROUP, 2000).....	23
2. Exemplo de banco de dados fictício para ilustração de possíveis atributos.....	35
3. Banco de dados hipotético para exemplificação do cálculo do suporte e da confiança .....	41
4. Exemplo do cálculo do suporte.....	41
5. Exemplo do cálculo da confiança.....	41
6. Campos mantidos no processo de Mineração de Dados e para a criação de novos atributos.....	59
7. Atributos utilizados no arquivo das Transações para análise dos clientes físicos.....	66
8. Atributos utilizados no arquivo das Transações para análise dos clientes jurídicos e parceiros.....	66
9. Comparação entre os atributos presentes no arquivo das Transações e no arquivo de Histórico.....	67
10. Atributos e seus valores, utilizados no arquivo de Histórico.....	68
11. Modelo de diferenciação dos clientes – Perfis dos clientes físicos.....	98
12. Modelo de diferenciação dos clientes – Perfis dos clientes jurídicos e parceiros.....	100
13. Regras dos clientes físicos geradas por meio de análise de associação.....	105
14. Radiografia das Cidades - clientes físicos.....	106
15. Regras clientes jurídicos e parceiros - meta: cliente ativo.....	108
16. Relação das Cidades e seus respectivos clientes em relação à situação Ativo/Não-ativo, tendo como atributo-meta: cliente ativo.....	109
17. Regras clientes jurídicos e parceiros - meta : valor acumulado.....	110
18. Radiografia das Cidades - clientes jurídicos e parceiros.....	111
19. Regras geradas a partir de análise de associação por meio do arquivo geral de clientes físicos.....	113

20. Regras geradas a partir de análise de associação por meio do arquivo de pré-pós feriados de clientes jurídicos e parceiros.....	114
21. Regras de associação geradas por meio do arquivo de produtos de maior valor agregado de clientes físicos.....	115
22. Regras de associação geradas por meio do arquivo de produtos de maior valor agregado de clientes jurídicos e parceiros.....	116
23. Caso Valinhos A(180/60) – perfil 355.....	151
24. Caso Vinhedo NF3(17)/A(7) – perfil 355.....	151
25. Caso Cajamar N (140/3) – perfil 352.....	152
26. Caso São José dos Campos AF2(15/1) – perfil 242.....	152

## **Resumo**

DENOBILE, T. **Modelo de gestão estratégica com foco no cliente para comercialização de produtos orgânicos**, 2005. Dissertação (Mestrado). Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas. Campinas. 152p.

O objetivo deste trabalho foi o desenvolvimento de um Modelo de Gestão Estratégica com foco no cliente para a comercialização de produtos orgânicos por meio da venda direta. Este modelo foi desenvolvido por meio de um estudo de caso do Sítio Vila Yamaguishi de Jaguariúna – SP e se baseou nas teorias da Inteligência Competitiva e da Gestão do Relacionamento com Cliente (CRM), contando também com o auxílio da tecnologia de Mineração de Dados para elaboração de um modelo de diferenciação dos perfis dos clientes, bem como para a descoberta dos hábitos de compra. No contexto da venda direta é muito importante que se conheça os hábitos e perfis dos clientes, pois cada um se configura como uma transação comercial, portanto, sua satisfação e o atendimento de suas necessidades pode representar a continuidade do negócio. Assim, por meio do modelo de diferenciação dos clientes foi possível detectar os clientes ativos e os não ativos, além de destacar os de maior valor para o negócio e, ainda, os que costumam comprar regularmente. Deste modo, foi possível também sugerir ações de comercialização de acordo com os perfis e hábitos descobertos. Por último, este trabalho visa ilustrar como dados internos de um negócio, quando bem trabalhados, são transformados em conhecimento que, juntamente com ações comerciais adequadas, podem se tornar uma ferramenta valiosa para aquisição de vantagens comerciais e competitivas.

**Palavras-chave:** Agricultura orgânica, Inteligência competitiva, Gestão do relacionamento com cliente (CRM), Mineração de dados, Venda direta

## **Abstract**

DENOBILE, T. **Strategic management model with focus on customer to commercialize organic products**, 2005. Dissertação (Mestrado). Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas. Campinas. 152p.

The main objective of this research was to develop a Strategic Management Model with focus on customer to commercialize organic products by direct sales. It was based on Competitive Intelligence and Customer Relationship Management (CRM). Data Mining technology was used in order to elaborate a model of customer differentiation as well as to identify consumption habits. This research was developed by a study of case. It is very important to know the habits and the profile of each customer in the context of direct sales, once each client means one transaction. Therefore, their satisfaction and the response for their needs can mean the continuity of the business. This model allowed distinguishing which clients had been buying or not, which ones were more valuable to the business and also which of them had a regular frequency. Considering this scenery, it was possible to suggest commercialization actions according to the different profiles and habits discovered by the using of Data Mining technology. This study aimed to illustrate how internal data of a business can be transformed in knowledge by a planned way, which can become a valuable instrument in order to gain competitive and commercial advantages.

**Key words:** Organic agriculture, Competitive intelligence, Customer relationship management (CRM), Data mining, Direct sales

## 1. Introdução

Duas fortes tendências estão cada vez mais influenciando o consumo principalmente no mercado alimentício, fazendo com que as empresas desenvolvam novos produtos e direcionem suas estratégias de marketing para a satisfação das necessidades de clientes cada vez mais exigentes.

A primeira dessas tendências refere-se à busca crescente da qualidade de vida por meio de produtos e serviços que proporcionem este benefício e, aliada a esta busca, encontra-se a segunda, que diz respeito ao cuidado cada vez maior com a saúde, moldando, assim, o consumo por produtos que atendam a tais condições.

Deste modo, criou-se todo um segmento de mercado que nos Estados Unidos movimenta um grande montante de dólares ao ano. Uma estimativa realizada pelo Hartman-Group (citado por GUIVANT, 2003), mostrou que só a população americana gasta com produtos de saúde em torno de US\$ 66 bilhões de dólares ao ano.

Neste contexto, não é surpreendente o crescimento cada vez maior da agricultura orgânica, ainda mais após os fatos ocorridos na Europa, como o mal da vaca louca, e na Ásia com a gripe do frango.

Por apresentar um modo limpo de produção que é livre de aditivos agro-químicos sintéticos, a agricultura orgânica pode disponibilizar aos consumidores, produtos mais saudáveis que os da agricultura convencional em geral, pois são produzidos em equilíbrio com a natureza, evitando contaminações e desgaste do solo.

A venda de produtos orgânicos cresce a cada ano, tanto internacionalmente como nacionalmente, já que seu modo de produção proporciona um produto que responde fortemente aos anseios dos consumidores do bem-estar, ou seja, daqueles que buscam qualidade de vida.

Deste modo, este sistema de produção agrícola movimentou, no Brasil, quase 300 milhões de reais em 2002, ante 50 milhões de reais movimentados em 2000, fazendo o país subir para a 5ª posição em áreas cultivadas sob manejo orgânico (DINIZ e NAIDITCH, 2003).

Ainda assim, o mercado de alimentos orgânicos é considerado um nicho, já que suas vendas representam uma pequena parcela do mercado alimentício. Um exemplo deste fato é o mercado americano de produtos orgânicos, que apesar de apresentar crescimento nas vendas,

ele representa não mais do que 4% do total de alimentos vendidos, evidenciando que ainda há muito espaço para o seu desenvolvimento (O ESTADO DE SÃO PAULO, 2003).

Além disso, segundo a IFOAM (Federação Internacional de Movimentos da Agricultura Orgânica), o preço do alimento orgânico é, em média, 20% a 30% mais alto que o do produto convencional (VALOR ECONÔMICO, 2003).

No Brasil, a produção de orgânico é desenvolvida na maior parte por pequenos e médios produtores.

Os grandes produtores destacando-se pela produção de soja, cana-de-açúcar e café e geralmente destinam sua produção ao mercado externo.

Os pequenos e médios produtores são responsáveis pela produção de frutas, hortaliças e alguns produtos processados (como iogurtes, sucos, geléias) para o mercado interno, podendo estar organizados, ou não, em associações de comercialização, cooperativas ou empresas processadoras.

Diante de tal contexto, o foco deste estudo será os pequenos e médios produtores de orgânicos e a dinâmica de seu negócio, já que esta classe representa cerca de 90% dos produtores de orgânicos do país (ESCOBAR, 2002) e são responsáveis por 70% da produção orgânica nacional (DAROLT, 2002).

Geralmente, estes produtores enfrentam inúmeros problemas, que vão desde a falta de pesquisas na área para aumento da produtividade, até problemas de comercialização, aspecto no que este trabalho concentrará seus esforços.

Em relação à comercialização, um dos maiores problemas refere-se às exigências dos supermercados para a venda dos orgânicos, já que as grandes redes exigem quantidades, variedades e periodicidade muitas vezes não alcançadas pelos produtores, devido às limitações inerentes ao modo de produção orgânica.

Sendo assim, a comercialização também acaba sendo realizada por meio de feiras orgânicas, venda por meio de lojas (especializadas em produtos naturais) e pela venda direta, sendo esta última uma grande oportunidade para que o produtor possa garantir maiores retornos de seus investimentos, uma vez que não há intermediários, encurtando o caminho até o consumidor final.

A venda direta demanda organização, tempo e atenção por parte do negociador, já que cada cliente passa a constituir uma transação comercial diferente, portanto, a manutenção

e satisfação de cada um destes clientes podem representar a fonte de lucro e a continuidade do negócio.

Deste modo, o produtor acumula outras tarefas com as quais deve se preocupar, que vão além da produção como, por exemplo, a venda individualizada da produção, a cobrança e o recebimento pelo que vendeu e, por fim, mas não menos trabalhoso, a entrega. Além disso, a venda direta apresenta uma outra desvantagem apontada por DAROLT (2002) que é a dificuldade de criar, manter e renovar clientes.

A venda direta é uma alternativa aos meios tradicionais de comercialização, já que proporciona melhores retornos ao produtor, por isso, é importante que haja estudos que foquem seus objetivos para amenizar e resolver os problemas pertinentes a este canal como, por exemplo, o problema da rotatividade de clientes.

Devido ao fato do produtor lidar diretamente com o consumidor final é importante que ele conheça os hábitos e perfis de cada um, de modo que tenha em mãos um processo de trabalho que o ajude não apenas a diferenciar seus clientes, mas também o ajude a identificar quais são seus melhores clientes e os mais importantes em termos estratégicos para o seu negócio.

Este trabalho visa ilustrar como dados internos de um negócio, quando bem trabalhados, podem ser transformados em conhecimento sobre o cliente que, juntamente com ações comerciais adequadas, podem se tornar uma ferramenta valiosa para aquisição de vantagens comerciais e competitivas.

Além disso, uma das grandes motivações deste trabalho foi mostrar que também é possível ao pequeno e médio produtor fazer uso de tecnologia de ponta (no caso a Mineração de Dados) para a gestão de seu negócio, uma vez que o software utilizado no desenvolvimento deste estudo é disponibilizado gratuitamente na Internet.



## **2. Objetivos**

### **2.1. Objetivo Geral**

Desenvolver um Modelo de Gestão Estratégica com foco no cliente para a comercialização de produtos orgânicos por meio da venda direta, baseado na teoria de Inteligência Competitiva (IC).

### **2.2. Objetivos Específicos**

- Desenvolver de um modelo de diferenciação de clientes, baseado na estratégia de Diferenciação da teoria de CRM, para descoberta dos perfis de clientes existentes, realizado por meio da Mineração de Dados com análises de classificação.
- Descoberta dos hábitos de consumo, realizada por meio da Mineração de Dados com análises de associação.
- Sugestão de ações de negócios para a comercialização de produtos orgânicos de acordo com os perfis de clientes descobertos pelo modelo de diferenciação.

Este Modelo de Gestão Estratégica com foco no cliente para a comercialização de produtos orgânicos, por meio da venda direta, representa um caminho para ajudar o pequeno e médio produtor a organizar e otimizar os processos de seu negócio, permitindo melhor conhecimento dos hábitos dos seus clientes, de modo que possa mantê-los e, portanto, quebrar o ciclo de rotatividade.

O trabalho foi realizado por meio de um estudo de caso, tendo como material de pesquisa os dados referentes ao período de três anos de registros de vendas do Sítio Vila Yamaguishi, produtor orgânico de Jaguariúna, São Paulo.

O método de trabalho utilizado para a atingir os objetivos propostos foi a Mineração de Dados, por meio de análises de associação e de classificação.

As hipóteses que se espera confirmar em relação aos hábitos de consumo é a influência dos diferentes atributos da dimensão tempo e das Cidades sobre o mesmo como, por exemplo:

- Em alguma estação do ano, há a predominância da venda de um determinado produto?
- Quando é pré ou pós feriado é vendido algum produto em especial?
- Há diferenças no consumo de determinados produtos em relação ao dia, quinzena do mês ou mês do ano?
- Em alguma cidade há a predominância da venda de determinado produto?
- Quais produtos têm sua venda associada à venda de outros?

Com a resposta destas questões, pode ser possível antecipar a venda de determinado produto e, assim, ter o planejamento de algumas ações comerciais como, por exemplo, a realização de promoções e descontos de acordo com determinada cidade e época do ano, mês, etc.

Em relação ao modelo de diferenciação dos clientes, espera-se mostrar quem são os clientes mais importantes para o negócio em termos estratégicos como, por exemplo, o melhor cliente é aquele que:

- Compra freqüentemente, mas gasta pouco em cada compra?
- Compra pouco, mas gasta muito em cada compra?
- O mais antigo?
- O mais recente?

Deste modo, o modelo de diferenciação pode permitir o planejamento de ações de comercialização adequadas aos perfis de clientes descobertos.

### **3. Revisão Bibliográfica**

Como este trabalho envolverá o estudo de quatro tópicos, a revisão bibliográfica será organizada pelos seguintes blocos de assunto:

1. Agricultura Orgânica
2. IC – Inteligência Competitiva
3. CRM (*Customer Relationship Management*) – Gestão do Relacionamento com cliente
4. Mineração de Dados

### **3.1. Agricultura Orgânica**

#### **3.1.1. Definição**

Segundo a lei 10.831 de 23 de dezembro de 2003, considera-se sistema orgânico de produção agropecuária todo aquele em que se adotam técnicas específicas, mediante a otimização do uso dos recursos naturais e socioeconômicos disponíveis e o respeito à integridade cultural das comunidades rurais, tendo por objetivo a sustentabilidade econômica e ecológica, a maximização dos benefícios sociais, a minimização da dependência de energia não-renovável, empregando, sempre que possível, métodos culturais, biológicos e mecânicos, em contraposição ao uso de materiais sintéticos, a eliminação do uso de organismos geneticamente modificados e radiações ionizantes, em qualquer fase do processo de produção, processamento, armazenamento, distribuição e comercialização, e a proteção do meio ambiente (INSTITUTO BIODINÂMICO, 2005).

Além disso, a agricultura orgânica baseia-se na melhora da fertilidade do solo por um processo biológico natural, pelo uso de matéria orgânica, o que é essencial à saúde das plantas, sendo contrária à utilização de adubos químicos solúveis (DAROLT, 2002).

Atualmente, para um produto ser considerado orgânico, deve receber o selo de uma das agências certificadoras do país. Estas agências atestam que a produção é feita de modo respeitoso aos fatores anteriormente citados e, também, certificam que os trabalhadores recebem remuneração digna e atuam em condições adequadas de trabalho, sem prejuízos à saúde.

### **3.1.2. Certificação**

A certificação é uma prática que surgiu da necessidade de se identificar a procedência e o processamento de um alimento orgânico, permitindo ao agricultor um produto diferenciado e mais valorizado, estabelecendo uma relação de confiança com o consumidor (INSTITUTO BIODINÂMICO, 2004).

É um procedimento que inspeciona e orienta a produção e o processamento de alimentos segundo as normas e práticas de produção orgânica, garantindo ao consumidor a sua procedência isenta de contaminação química, respeitando o meio ambiente e o trabalhador, e assegurando ao produtor um diferencial de mercado para os seus produtos (INSTITUTO BIODINÂMICO, 2004).

No Brasil, a Lei 10.831, de 23/12/2003, do Ministério da Agricultura, dispõe detalhadamente sobre as normas de produção, tipificação, processamento, envase, distribuição, identificação e certificação da qualidade para os produtos orgânicos de origem vegetal e animal. Nela, destacam-se os seguintes pontos (INSTITUTO BIODINÂMICO, 2005):

- a oferta de produtos saudáveis isentos de contaminadores intencionais; a preservação da diversidade biológica dos ecossistemas naturais e a recomposição ou incremento da diversidade biológica dos ecossistemas modificados em que se insere o sistema de produção;
- incrementar a atividade biológica do solo; promover um uso saudável do solo, da água e do ar, e reduzir ao mínimo todas as formas de contaminação desses elementos que possam resultar das práticas agrícolas;
- manter ou incrementar a fertilidade do solo a longo prazo; a reciclagem de resíduos de origem orgânica, reduzindo ao mínimo o emprego de recursos não-renováveis;
- basear-se em recursos renováveis e em sistemas agrícolas organizados localmente;
- incentivar a integração entre os diferentes segmentos da cadeia produtiva e de consumo de produtos orgânicos e a regionalização da produção e comércio desses produtos;

- manipular os produtos agrícolas com base no uso de métodos de elaboração cuidadosos, com o propósito de manter a integridade orgânica e as qualidades vitais do produto em todas as etapas.

Algumas das principais certificadoras do Brasil são as seguintes (PLANETA ORGÂNICO, 2004):

- Associação de Agricultura Orgânica (AAO);
- Associação de Agricultores Biológicos (ABIO);
- Associação de Agricultura Natural de Campinas e região (ANC);
- Associação dos Produtores de Agricultura Natural (APAN);
- BCS Öko-Garantie (BCS);
- Associação de Certificação de Produtos Orgânicos do Espírito Santo (CHÃO VIVO);
- Fundação Mokiti Okada (CMO);
- Cooperativa Coolméia (COOLMÉIA);
- Ecocert Brasil (ECOCERT);
- Farm Verified Organic (FVO);
- Instituto Biodinâmico (IBD);
- Instituto de Mercado Ecológico (IMO);
- Organización Internacional Agropecuaria (OIA);
- Certificadora Sapucaí (SAPUCAI);
- Certificadora Holandesa Skal (SKAL).
- Rede EcoVida

Apesar da certificação ser um aspecto importante para a garantia de um produto de qualidade, adequado às normas de produção orgânica, ela também representa um dos grandes entraves para o crescimento do mercado orgânico pois, dependendo do tipo de certificação, ela pode influenciar na formação do custo do produto, segundo produtores (ORMOND *et al*, 2002).

As despesas referentes à certificação incluem taxa de filiação, o tamanho da área a ser certificada, despesas com inspeção, elaboração de relatórios, análise laboratorial do solo e da água, visitas de inspeção, acompanhamento e emissão do certificado (ORMOND *et al*, 2002).

### **3.1.3. Mercado**

#### **3.1.3.a. Consumidor: motivação e perfil**

Uma pesquisa realizada em feiras orgânicas na cidade de São Paulo, por CERVEIRA e CASTRO (1999), sobre os consumidores desses produtos, indicou que a principal motivação para seu consumo refere-se ao aspecto da saúde pessoal e familiar, seguido pela não utilização de agrotóxicos na produção, enquanto que a preocupação com o meio ambiente aparece apenas em 5º lugar. Além disto, tais consumidores afirmaram que pagariam até 30% mais caro por produtos orgânicos, em vista dos benefícios.

O primeiro fato acima citado também foi mostrado numa pesquisa realizada em Curitiba, em feiras orgânicas, a qual apontou a questão da saúde como principal razão para o consumo de produtos orgânicos (RUCHINSKI, 1999), (citado por DAROLT, 2002).

Um outro ponto importante levantado por ambas pesquisas é a falta de quantidade, diversidade e regularidades adequadas dos produtos, considerados pelos consumidores como algo que ainda há necessidade de ser melhorado.

Segundo DAROLT (2002), em relação às redes de supermercado, existe um grande desconhecimento do consumidor de orgânicos. Além disso, há também uma grande confusão gerada pela “onda” de produtos considerados naturais, dietéticos, *lighth*, integrais entre outros. Particularmente, entre as hortaliças orgânicas também há confusão, sobretudo, entre produtos orgânicos, hidropônicos e os processados minimamente (lavados e cortados, prontos para o consumo), evidenciando a necessidade de se desenvolver um trabalho para a conscientização dos consumidores deste canal de comercialização para maior valorização dos produtos orgânicos.



O perfil do consumidor orgânico apresentado por WILLER e YUSSEFI (2004), por meio de uma pesquisa com dados do mundo todo é o seguinte:

- Localização: moradores de áreas urbanas, geralmente grandes cidades;
- Comportamento de compra: fatores relacionados à qualidade, proveniência e métodos de produção influenciam na compra de comidas e bebidas;
- Demografia: geralmente possuem bom nível educacional e pertencem à classe média alta;
- Poder de compra: possuem renda familiar de médio a alto nível.

No Brasil o perfil do consumidor orgânico se aproxima com o do resto do mundo. Segundo a pesquisa de RUCHINSKI (1999), (citado por DAROLT, 2002), os consumidores de orgânico apresentaram o seguinte perfil:

- Normalmente profissional liberal;
- Maioria sexo feminino (66%);
- Idade entre 31 a 50 anos (62%);
- Renda individual superior a 12 salários mínimos (40%);
- Moradores da zona urbana;
- Hábito de praticar esporte com frequência (54,9%);
- Frequentador de parques e bosques (66,9%).

Segundo DAROLT (2002), esses dados refletem que o público das feiras orgânicas, tanto em termos de escolaridade quanto de renda, faz parte de um grupo de consumidores mais intelectualizados e de uma classe economicamente mais elevada. Assim, há um outro desafio que é levar o alimento orgânico para as outras camadas da população.

### **3.1.3.b. Preço**

Segundo DAROLT (2002) o preço final ao consumidor costuma variar conforme o canal de compra e o tipo de produto adquirido. Geralmente, não há grande diferença entre o preço do orgânico e do convencional, quando a compra é realizada por meio da venda direta ou feira livre, especialmente no caso de hortaliças. Já, em relação às redes de supermercado, os preços são, em média, 30% mais caros, podendo chegar a diferenças superiores a 100% como, por exemplo, nos casos do tomate e da batata.

O preço dos alimentos orgânicos pode ser considerado como um dos entraves para um rápido desenvolvimento da produção orgânica no Brasil.

No caso da comercialização via supermercado, há despesas extras com embalagens, pois é necessário diferenciar os produtos orgânicos dos demais. Além disso, os descartes dos supermercados são assumidos integralmente pelos distribuidores (associações e empresas), o que contribui para aumentar a margem de lucro dos supermercados e encarecer demasiadamente o produto final para o consumidor (DAROLT, 2002).

Segundo Darolt (2002), os principais motivos, que contribuem para que orgânico seja mais caro que o convencional, são os seguintes:

- Baixa escala de produção orgânica: implica maiores custos (mão-de-obra, insumos) por unidade de produto;
- Desorganização do sistema de produção e do processo de comercialização: devido ao fato de ser um sistema novo. A descontinuidade da produção é reflexo da falta de um bom planejamento que possa facilitar a tomada de decisão no momento da comercialização. Além disso, há falta de organização por parte dos pequenos produtores (responsáveis por 70% da produção nacional), falta de treinamento na área gerencial, desconhecimento técnico-agronômico, problemas de logística e distribuição;
- Pequeno número de pesquisas na área orgânica
- Custos extras devido ao processo de certificação e perdas econômicas, durante o processo de conversão, suportadas pelo produtor.

O autor destaca, ainda, que a concorrência econômica entre o sistema orgânico e o convencional é injusta, pois a agricultura convencional exclui dos cálculos da formação de preço a contabilidade ambiental.

### **3.1.3.c. Cenário Internacional**

Segundo dados de janeiro de 2002 do Centro Internacional de Comércio (ITC), órgão ligado à Organização Mundial de Comércio (OMC), o comércio mundial de alimentos orgânicos (considerando 16 países europeus, EUA e Japão) movimentou US\$ 16 bilhões no ano de 2000 e US\$ 19 bilhões em 2001 contra os US\$ 10 bilhões de 1997 (O ESTADO DE SÃO PAULO, 2003).

A agricultura orgânica é praticada aproximadamente em mais de cem países no mundo todo, representando mais de 24 milhões de hectares e movimentou um mercado estimado em US\$ 23 bilhões em 2002 (WILLER e YUSSEFI, 2004).

Já em 2003, o comércio de alimentos orgânicos movimentou cerca de US\$ 25 bilhões (ISTO É, 2004).

Estes números evidenciam a busca cada vez maior dos consumidores por produtos saudáveis, sem adição de hormônios ou outros produtos químicos e sintéticos, de modo que não agredam o meio ambiente.

Segundo Bernward Geier, diretor das relações internacionais da IFOAM (Federação Internacional de Movimentos da Agricultura Orgânica), o consumo mundial de alimentos orgânicos tem crescido, em média, 15% ao ano nas últimas duas décadas (VALOR ECONÔMICO, 2003).

### **3.1.3.d. Cenário Nacional**

No Brasil, a situação não é diferente segundo DINIZ e NAIDITCH (2003):

- A produção aumentou 50% ao ano desde 2000. Nesse período, o crescimento médio na Europa e nos EUA foi de 20% ao ano;
- O mercado de orgânicos no país movimentou quase 300 milhões de reais em 2002, ante 50 milhões de reais em 2000;
- O país subiu para a 5ª posição em áreas cultivadas (orgânicas) desde 2002;
- A produção brasileira é de 300 mil toneladas anuais. Cerca de 70% da produção orgânica brasileira é exportada;
- Pelo menos 30 tipos de produto já são produzidos no Brasil. Os principais são soja, hortaliças e café.

A produção sob manejo orgânico ocupa uma área de aproximadamente 840 mil hectares, sendo cerca de 14 mil produtores no país – dos quais 90% são pequenos produtores. O crescimento da produção orgânica nacional é calculado em torno de 30 a 50 por cento anuais (WILLER e YUSSEFI, 2004).

Segundo WILLER e YUSSEFI (2004), o mercado nacional brasileiro é, juntamente com a Argentina, o mais desenvolvido da América Latina, sendo que se divide da seguinte forma entre os canais de comercialização:

- 45% supermercado
- 26% feiras
- 16% lojas especializadas
- 13% outros canais

## **3.2. Inteligência Competitiva**

### **3.2.1. Definição**

Se no início do século passado havia uma situação de escassez de produtos, hoje se vive uma abundância de escolhas. Esta atual situação pode decretar o fim daquelas empresas que não se adaptarem e, além disso, não se anteciparem ao mercado que se apresenta cada vez mais exigente e mutante.

Neste contexto cresce a importância da Inteligência Competitiva (IC), pois seu processo de trabalho permite o monitoramento dos fatores externos à empresa que podem indicar ameaças ou oportunidades. Assim, ao contrário de somente reagir às mudanças externas, a empresa poderá assumir uma postura pró-ativa frente às mesmas.

Segundo Stephen H. Miller, da Sociedade dos Profissionais de Inteligência Competitiva (SCIP), “em um mercado global saturado de dados que percorrem o mundo em nanossegundos, a necessidade de transformar informações relativas aos concorrentes em inteligência prática nunca foi maior” (PRESCOTT e MILLER, 2002).

Deste modo, IC pode ser definida como:

“Um programa sistemático e ético para coleta, análise e gerenciamento de informações externas que podem afetar os planos, decisões e operações de uma empresa” (SCIP, 2004).

Segundo PRESCOTT e MILLER (2002) os profissionais de IC coletam, analisam e aplicam, legal e eticamente, informações relativas às capacidades, vulnerabilidades e intenções de seus concorrentes e monitoram acontecimentos do ambiente competitivo geral (como concorrentes antes desconhecidos ou novas tecnologias que podem alterar todo o paradigma competitivo), tendo como objetivo colocar a empresa na fronteira competitiva dos avanços.

Também pode ser entendida como um processo que usa meios legais e éticos para descobrir, desenvolver e entregar em tempo, a inteligência relevante para a tomada de decisão de modo que a empresa se torne mais competitiva. Essa inteligência pode ser usada para o desenvolvimento de produtos, fusões entre empresas e alianças, bem como iniciativas táticas como a antecipação dos movimentos dos clientes, dos competidores e órgãos reguladores que afetam o negócio (FARRELL, 2003).

Os dados necessários para a realização de uma boa análise de IC encontram-se disponíveis a todos, seja na Internet, contato com clientes e fornecedores, congressos, jornais e revistas, bastando apenas uma equipe bem treinada, que possa transformar tais dados em informação, de modo que esta seja entregue ao responsável pela tomada de decisão.

Além disso, é comumente sabido que 80% da IC de que uma empresa necessita está disponível nela própria, principalmente com seus funcionários. No entanto, o problema maior é como juntar e sintetizar todo esse conhecimento em um formato trabalhável o qual possa beneficiar o planejamento e operações da organização (COBB, 2003).

É neste contexto, que vêm crescendo a importância da Mineração de Dados como método de IC para a aquisição de conhecimento por meio de informações já arquivadas de modo que possam ser aplicadas a tarefas analíticas (PRESCOTT e MILLER, 2002).

### **3.2.2. Análise e Ciclo de IC**

Segundo BESSA (2001) uma boa análise de IC deve procurar responder às seguintes perguntas básicas:

- *Competidores:* Quem são? Quais são seus objetivos? Como estão organizados?
- *Clientes:* Quem são? Com quais recursos contam? Quais são seus padrões de reação? Onde e como serão suas próximas ações?
- *Própria organização:* Quais são as debilidades e fortalezas? Quais são as ameaças e oportunidades?

Segundo uma pesquisa conduzida por *The Futures Group* (1997), os entrevistados foram questionados para identificar em que áreas IC é mais necessária para a tomada de decisão. Tais áreas foram classificadas em ordem de importância (FARRELL, 2003):

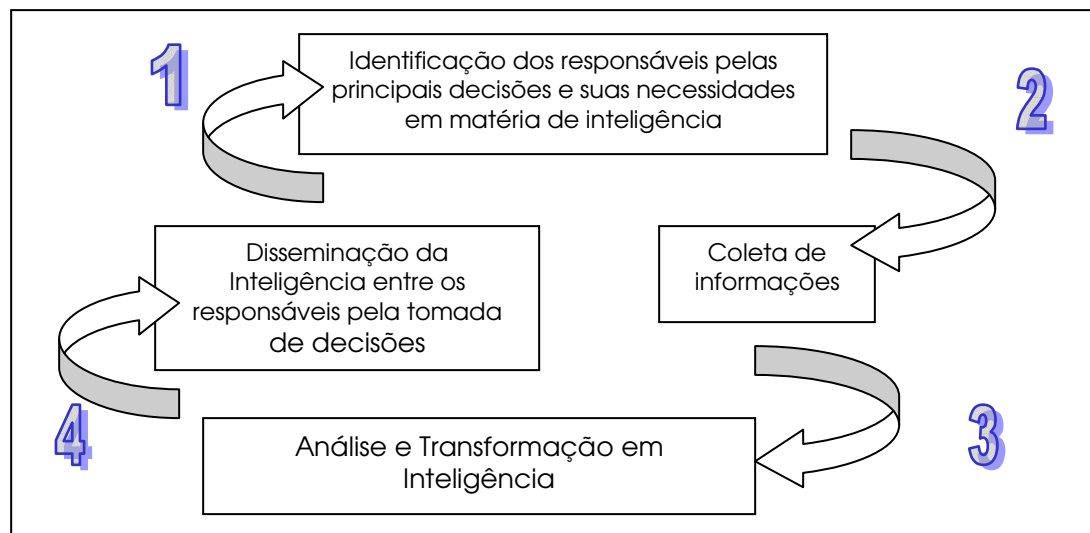
1. Atividades dos competidores
2. Mudanças no mercado e na estrutura industrial
3. Atividades dos clientes e fornecedores
4. Tecnologias emergentes

5. Condições da economia global
6. Ambiente regulador
7. Ambiente político

Segundo Benjamin Gilad (citado por FARRELL, 2003), a competitividade é baseada em aprender, ou seja, na habilidade de escutar: os clientes, os consumidores, os parceiros, os competidores, os *experts* da indústria e, mais importante, os próprios funcionários.

A essência desta filosofia é bastante simples. O ambiente competitivo manda mensagens o tempo todo: sinais sobre mudanças, tendências, ameaças, oportunidades. Tais sinais são, muitas vezes, fracos, ambíguos ou escondidos. Encontrá-los, entendê-los e, então, aprender, é uma arte que requer olhos, ouvidos e mente aberta.

Para que tais sinais possam não apenas ser encontrados, mas também serem monitorados em tempo hábil para a tomada de decisão, MILLER (2002) propõe o ciclo de quatro fases da IC ilustrado pela figura 1:



**Figura 1: Fases do ciclo de IC (MILLER, 2002)**

1. *A identificação dos responsáveis pelas principais decisões e suas necessidades em matéria de inteligência:*

Nesta fase é importante levar em consideração que as decisões, em geral, ocorrem também em níveis inferiores ao nível executivo, portanto é necessário mapear todos os atores envolvidos no processo, de modo que a informação possa alcançar todos os níveis hierárquicos da organização interessados em determinado assunto.

Esta fase também pode ser entendida como a fase de planejamento de todo o ciclo, ou seja, aqui é determinado todos os pontos-chaves para a execução das demais, bem como metas e prazos a serem alcançados.

2. *Coleta da informação*

É nesta fase que se determina os processos de coleta de informação mais adequados, além dos modelos analíticos que serão usados.

A obtenção de informações pode ser por meio de fontes primárias e secundárias. As fontes primárias referem-se: aos *experts* de setores determinados, consumidores, fornecedores e aos próprios empregados, estas fontes proporcionam maior vantagem competitiva, uma vez que são exclusivas à organização.

Já as fontes secundárias, que podem ser tanto impressas quanto eletrônicas, não são exclusivas e podem ser facilmente acessadas por todos os que tenham interesse no processo. As principais fontes deste tipo: banco de dados comerciais, publicações periódicas (como relatórios de analistas), publicações governamentais, relatórios setoriais, discursos de executivos, relatórios técnicos e de patentes.

As fontes secundárias proporcionam as informações de bastidores capazes de sustentar as sugestões obtidas a partir das fontes primárias.

O trabalho de coleta de informações é, na verdade, a montagem de quebra-cabeças, onde pedras isoladas não fazem sentido algum, mas quando organizadas corretamente podem revelar nitidamente o cenário, que antes se apresenta desconexo.



### 3. *A análise da informação e sua transformação em inteligência*

Tendo reunido as informações necessárias, os profissionais da inteligência devem identificar padrões e tendências significativas, buscando *insights* exclusivos e relações até então não detectadas entre os dados.

A fase de análise pode exigir uma característica de pesquisa científica: formular hipóteses e determinar a validade das suposições, bem como a probabilidade dos conseqüentes impactos.

O analista usa, neste processo, softwares estatísticos e/ou variáveis estatísticas de modelagem. No decorrer deste processo, seu executor pode concluir que precisa de mais dados. Portanto, a coleta e análise não são necessariamente fases sequenciais.

Esta fase exige persistência e criatividade por parte do pessoal de inteligência, além da capacidade de saber quando encerrar a análise.

### 4. *A disseminação da inteligência entre os responsáveis pelas decisões*

Na quarta fase do ciclo é necessário que o resultado seja disseminado eficazmente pela organização. Portanto, é preciso que sejam esclarecidos os meios e formatos (Intranet, reuniões de equipes, memorandos, etc) mais adequados para que as informações sejam apresentadas aos responsáveis pelas decisões, de modo que se aumente a integridade e a futura utilização do que foi relatado e para que, assim, o ciclo da IC possa ser realimentado.

Alguns autores afirmam existir uma quinta fase do ciclo de IC que corresponderia à avaliação dos processos e produtos de inteligência (HERRING, 2002). Segundo MILLER (2005), esta fase levaria em conta a resposta dos tomadores de decisão e suas outras necessidades em termos de inteligência frente ao trabalho já desenvolvido.

Cabe ressaltar, que a quantidade de fases do ciclo de IC não é uma unanimidade, há ainda outros autores que afirmam existir de seis a sete fases. No entanto, boa parte deles adota um ciclo de IC, geralmente, de quatro a cinco fases.

Para este trabalho, foi adotado o ciclo de IC de quatro fases definido por MILLER (2002) extraído do livro *O Milênio da Inteligência Competitiva*.

### **3.3. CRM – Gestão do Relacionamento com o Cliente**

#### **3.3.1. Definição**

A Gestão do Relacionamento com Cliente ou CRM, também recebe outros nomes, como *One to One Marketing* ou Marketing de Relacionamento, mas independentemente do modo como é chamado, o mais importante é a idéia de desenvolver e gerenciar relações individuais com clientes individuais. Tal atividade vem se tornando cada vez mais acessível, graças à diminuição dos custos referentes aos recursos computacionais (PEPPERS and ROGERS GROUP, 2000).

O CRM nada mais é do que um nome novo para expressar a importância de se voltar toda a empresa para a excelência do atendimento ao consumidor de uma maneira rentável (SWIFT, 2001).

Do ponto de vista tecnológico, CRM envolve capturar os dados do cliente ao longo de toda a empresa, consolidar todos os dados capturados interna e externamente em um banco de dados central, analisar os dados consolidados, distribuir os resultados dessa análise aos vários pontos de contato com o cliente e usar essa informação ao interagir com o cliente através de qualquer ponto de contato com a empresa (PEPPERS and ROGERS GROUP, 2000).

Na visão de BRETZKE (2000) o CRM pode ser entendido como uma estratégia de negócio que busca ações voltadas para a satisfação e retenção do cliente, suportadas pela tecnologia da informação, consistindo na integração desta tecnologia com o marketing. Seu objetivo é prover a empresa de meios eficazes e integrados para atender e reconhecer o cliente, em tempo real, transformando dados em informação que, disseminadas pela organização, permitam que o cliente seja conhecido e cuidado por todos.

Ou ainda, segundo SWITF (2001), o CRM é uma abordagem empresarial que visa entender e influenciar o comportamento do consumidor através de comunicações significativas com o objetivo último de se aumentar a aquisição, retenção, fidelidade e rentabilidade do cliente.

Deste modo, a Gestão do Relacionamento com o Cliente também pode ser entendida como uma estratégia de negócio voltada ao entendimento e antecipação das necessidades dos clientes atuais e potenciais de uma empresa.

O CRM apóia-se em três pontos: Base de Dados, Interatividade e Personalização em massa. A complexidade de cada um desses fatores dependerá do tamanho e complexidade da empresa. Assim, entende-se como:

- **Base de Dados:** banco de dados com informações referentes ao cliente, informações cadastrais, gostos e preferências, reclamações, sugestões, etc.
- **Interatividade:** estabelecimento de um diálogo entre empresa e cliente ou vice-versa;
- **Personalização em Massa:** consiste na disponibilização de um produto ou serviço básico cercado de opções, ou seja, é a adequação de produtos e serviços ou a forma como esses produtos e serviços são entregues e pagos de acordo com as necessidades de cada cliente. (PEPPERS and ROGERS GROUP, 2000).

Tais aspectos fornecem os subsídios necessários para que se estabeleça o que é chamado de Relação de Aprendizado ou Processo de gerenciamento da relação com cliente ilustrado na figura 2.

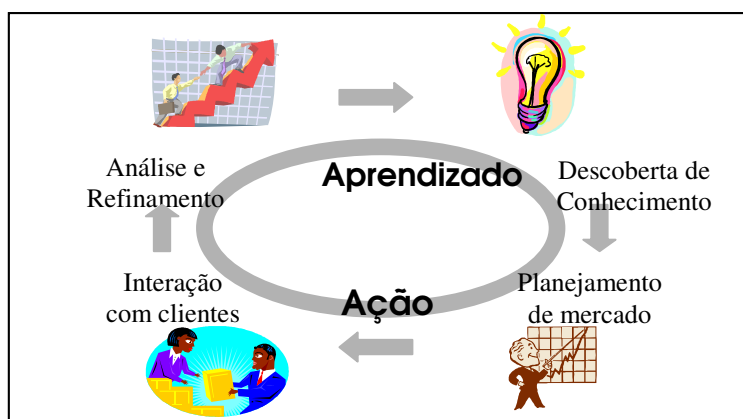


Figura 2: Fases do processo de CRM (SWIFT, 2001)

Isto significa que por meio da participação do cliente (fornecendo à empresa sugestões e reclamações) é possível desenvolver ações conduzidas de acordo com suas necessidades. Portanto, o relacionamento se torna mais inteligente a cada nova interação, pois o que se espera é, que cada vez mais, os produtos se tornem adequados às necessidades do cliente, satisfazendo-o e criando um sentimento de conveniência, de modo que o cliente não abandone o negócio com a empresa.

Segundo SWIFT (2001), o gerenciamento bem sucedido do relacionamento com os clientes significa aprender seus hábitos e suas necessidades, antecipando futuros padrões de compra e descobrindo novas oportunidades para fazer negócio.

### 3.3.2. Diferenças entre Marketing Tradicional e CRM

A tabela 1 evidencia os principais pontos de diferença entre o Marketing tradicional e o CRM.

**Tabela 1: Diferenças entre o marketing tradicional e o CRM (PEPPERS and ROGERS GROUP, 2000)**

<b>Marketing Tradicional</b>	<b>CRM</b>
Sucesso é medido pelo aumento de participação no mercado ( <i>market share</i> ) e por categoria de produto. Concentra-se em atender uma necessidade específica de um grupo muito grande de clientes.	Sucesso é medido em termos de participação no cliente ( <i>share of customer</i> ). O objetivo é identificar, nos clientes de maior valor e de potencial valor, o maior número de necessidades para que se possa fornecer outros produtos e serviços, aumentando a participação no cliente – ou seja, o <i>share of customer</i> .
Procura mais clientes para produtos específicos	Procura mais produtos e serviços para cada um dos clientes
Empresa tradicional gerencia produtos e funcionários são remunerados por vendas	Gerencia clientes e funcionários, os quais são remunerados pelo desenvolvimento dos clientes ao longo do tempo
Envolve um departamento	Envolve toda a empresa para o monitoramento de todas as transações e interações de cada cliente. Necessidade de grande interação e integração de dados e processos em todos os níveis da empresa

Essas diferenças evidenciam que o foco no cliente deixa de ser somente a preocupação de um departamento, passando a ser a estratégia de todo um negócio, envolvendo todas as partes de uma empresa.

Atualmente, vive-se num mundo de concorrência global, não somente pela abrangência geográfica, mas também porque as empresas não sabem mais de onde vem a concorrência, já que tecnologia e qualidade não são mais diferenciais entre produtos devido ao grande avanço alcançado e a variedade de opções existentes.

Por isso, saber identificar a ocorrência de eventos que determinam mudanças nas necessidades dos clientes, bem como conhecê-las, é fundamental como forma de obtenção de vantagem competitiva, já que essa vantagem é a informação que a empresa possui sobre os clientes e que a concorrência não tem (PEPPERS and ROGERS GROUP, 2000).

Segundo Frank Welch, CEO (*Chief Executive Officer*) da GE, a respeito da estratégia competitiva para o próximo século: “Nós temos somente duas fontes de vantagem competitiva: a capacidade de aprender mais sobre nossos clientes, mais rápido que nossos concorrentes e a capacidade de transformar esse conhecimento em ações, mais rapidamente que nossos concorrentes” (PEPPERS and ROGERS GROUP, 2000).

Assim, o conhecimento sobre mercados e clientes é um recurso, pelo menos, tão valioso quanto os produtos ou serviços fornecidos por uma empresa (SWIFT, 2001).

### **3.3.3. As quatro estratégias da Gestão do Relacionamento com Cliente**

O processo de implementação de um programa de CRM envolve quatro estratégias básicas, na seguinte ordem (PEPPERS and ROGERS GROUP, 2000):

- 1. Identificação dos clientes:** não é possível estabelecer nenhum tipo de relacionamento com cliente sem antes conhecê-lo.

É necessário que se conheça a identidade do cliente, forma preferida de contato com a empresa, as transações e interações com a empresa, todas as reclamações feitas, bem como as providências tomadas, além das sugestões. Além disso, é importante que o cliente seja identificado independente do canal escolhido por ele.

É importante que a identificação de cada cliente contenha: histórico das transações, receita, lucratividade, reclamações, canal de comunicação preferido, momento da vida, valor real, potencial e estratégico, potencial de crescimento e risco.

- 2. Diferenciação dos clientes:** O objetivo da diferenciação é encontrar os clientes de maior valor (CMV) e os clientes de maior potencial (CMP), mas também podem ser diferenciados pelas necessidades que possuem em relação aos produtos e serviços.

Na diferenciação, devem-se priorizar esforços e aprimorar o relacionamento com os clientes de maior valor. Além disso, é importante que se personalize o comportamento da empresa, baseado nas necessidades individuais dos clientes. Isto implica em se estabelecer algum critério de estratificação como, por exemplo: modelo de lucratividade, medida de valor dos clientes ou por necessidade.

A diferenciação por valor considera três tipos possíveis de valor. O primeiro deles é o Valor Vitalício ou *Lifetime Value* (LTV): é o valor que cada cliente tem em toda sua história transacional com a empresa e deve ser expresso em termos de lucratividade. Há, também, o Valor Potencial que é mais difícil de ser medido, uma vez que envolve dados externos à empresa, para sua mensuração é necessário saber quantas vezes o cliente comprou ou utilizou produtos ou serviços da concorrência. O Valor Estratégico é o menos tangível, não pode ser medido, mas sim interpretado, pois ele representa os clientes que colaboram com a empresa, fornecendo sugestões e reclamações, esses clientes tem o seu valor, pois não são passivos.

O segundo tipo de diferenciação é a diferenciação por necessidade. Ela mede o número de necessidades conhecidas para cada cliente, pois quanto maior for o número de necessidades conhecidas, maior será a oportunidade de participar no cliente. A empresa deve começar a diferenciar pelas necessidades inferidas para depois alcançar o nível individual, tido como ideal. Assim, quanto mais se

entende e responde às necessidades, os produtos e serviços são percebidos de modo diferenciado pelo cliente.

- 3. Interação:** O objetivo da interação é buscar a forma mais útil em termos de produção de informação, que possa ajudar a fortalecer a relação da empresa com o cliente.

A cada nova interação com o cliente, é importante que ela seja estabelecida dentro do contexto de todas as outras interações já ocorridas com este mesmo cliente, de modo que não haja necessidade de ser informado novamente os dados, sugestões ou reclamações já fornecidas.

Além disso, é fundamental que exista permissão do cliente, quando a interação parte da empresa e que, também, seja realizada pelo canal preferido do cliente em horário conveniente. A mensagem deve ter algum tipo de valor para o mesmo. Assim, é possível que se monitore mudança de suas necessidades.

- 4. Personalização:** o objetivo da personalização é incentivar que os clientes mantenham uma relação de aprendizado, de modo que a empresa seja capaz de fornecer produtos e serviços adequados às necessidades particulares de cada um deles, de acordo com o que já foi dito à empresa durante alguma interação.

Portanto, quanto mais se personaliza, mais valor é entregue ao cliente. Assim, cria-se um ciclo de personalização e feedback de modo que a empresa, a cada novo contato, possa atender mais e mais o cliente.

Estas quatro estratégias permitem à empresa não somente diferenciar os clientes, até que se chegue ao nível da personalização, mas principalmente redirecionar os esforços e investimentos para os clientes mais lucrativos ou de maior potencial (PEPPERS and ROGERS GROUP, 2000).

### **3.3.4. CRM Operacional, Colaborativo e Analítico**

A Gestão do Relacionamento com Cliente apresenta várias vantagens competitivas em relação à gestão tradicional com foco no produto, uma vez que permite conhecer melhor cada cliente e atender suas necessidades individuais, contribuindo para que ele continue a fazer negócio com a empresa e, assim, para conseqüente crescimento dos lucros.

Por isso, não é ao acaso que, a cada dia, cresce o número de empresas que adotam a gestão do relacionamento com cliente para a condução de seus negócios. Porém, juntamente com esse número cresce, também, o número de projetos de CRM que fracassaram, minando as expectativas de aumento do lucro.

A causa do insucesso destes projetos está marcada pela concepção errônea do que é CRM, a maioria deles confunde este conceito (já definido no item 3.3.1) com a idéia de solução tecnológica, de um software capaz de resolver todos os problemas de relacionamento com cliente e que contribui de maneira mágica para o aumento da lucratividade da empresa.

Na visão de PEPPERS and ROGERS GROUP (2000), CRM é por definição um projeto corporativo que envolve todas as áreas de uma empresa, refere-se a uma nova forma de ver e tratar o cliente e, portanto, tem mais a ver com mudança de cultura do que com implementação de um novo sistema.

Assim, para melhor compreensão da abrangência deste novo modo de ver e tratar o cliente é importante que se conheça seus três níveis, que são: CRM operacional, CRM colaborativo, CRM analítico.

O CRM operacional lida com os aspectos operacionais da relação com o cliente como, por exemplo, o funcionamento e automatização de um *call-center*, (local em uma empresa onde se concentram as ligações telefônicas de clientes) registrando todas as ligações e ocorrências de um determinado cliente, facilitando as próximas interações do mesmo com a empresa. (PEPPERS and ROGERS GROUP, 2000).

Ou seja, é propriamente a aplicação da tecnologia de informação para melhorar a eficiência do relacionamento entre clientes e empresa.

O CRM colaborativo engloba todos os pontos de contato (canais) onde há interação entre cliente empresa. Esses canais: internet, cartas/fax, telefone, interação direta, devem ser



preparados não apenas para permitir a interação, mas também para assegurar o fluxo adequado dos dados resultantes dela para toda a organização (PEPPERS and ROGERS GROUP, 2000).

Já o CRM analítico é a fonte de toda a inteligência do processo, serve para o ajuste das estratégias de diferenciação de clientes, bem como para o acompanhamento de seus hábitos, com o objetivo de identificar suas necessidades e mudanças no comportamento de consumo (PEPPERS and ROGERS GROUP, 2000).

É com ele que se monitora quem são os clientes mais rentáveis e os potenciais, permitindo que sejam traçadas ações de negócio de acordo com cada perfil descoberto.

É no contexto do CRM analítico que cresce a importância e a utilização da mineração de dados como ferramenta e tecnologia para a transformação de dados sobre clientes em fonte de informação para a busca de vantagem competitiva.

### **3.3.5. Ações de Negócio**

O conhecimento obtido por meio do CRM analítico, com auxílio da mineração de dados, é uma ferramenta valiosa para a formulação de ações de negócio.

O texto a seguir foi desenvolvido baseado nos exemplos programas de marketing de relacionamento da INFO SOCIETY (2004). Ele ilustra as ações que podem ser traçadas.

#### **3.3.5.a. Aquisição**

Refere-se à sondagem de clientes potenciais para determinado produto ou serviço, ou seja, vislumbra-se a aquisição de novos clientes. Geralmente pode estar alinhado a promoções ou descontos oferecidos a clientes que se encaixam no perfil de consumo do produto/serviço oferecido, tal perfil é traçado com base em características dos atuais clientes da empresa.

#### **3.3.5.b. Ativação**

Ao contrário da aquisição, na ação de ativação, a empresa não está à procura de novos clientes, mas sim fazer com que clientes pouco ativos passem a usar mais seus produtos ou serviços. Um exemplo desta ação pode ser um banco que oferece descontos nas compras pagas com cartão de crédito para clientes que utilizam pouco este produto, mas que apresentam um perfil de potencial de uso mais freqüente.

#### **3.3.5.c. Retenção**

Neste caso, a empresa está preste a perder um cliente lucrativo no qual é importante investir. Um exemplo é uma empresa de telefonia celular que detecta uma crescente queda na utilização do serviço por determinado cliente, assim, para que este cliente não migre para concorrência podem ser oferecidos serviços gratuitos, redução de tarifas, para que ele permaneça com esta empresa. Este é um exemplo de uma ação de retenção.

#### **3.3.5.d. Fidelização**

Ações de fidelização são realizadas para reforçar o sentimento de conveniência que o cliente tem por determinada empresa. São ações que reconhecem quem são os melhores clientes e oferecem-lhes vantagens por isso. Alguns exemplos são os programas de fidelidade das companhias aéreas que oferecem salas de espera especiais, passagens aéreas gratuitas com o acúmulo de milhas. Geralmente, clientes fiéis não são suscetíveis a promoções de preço por parte da concorrência, pois reconhecem que é melhor pagar mais pela conveniência e comodidade obtidas.

#### **3.3.5.e. Recuperação**

São ações que destinam recuperar clientes que deixaram de negociar com a empresa, estas ações podem oferecer alguns benefícios para a retomada do relacionamento como o oferecimento de descontos na compra em produtos e serviços em lançamento.

#### **3.3.5.f. Vendas Cruzadas (*cross selling*)**

São ações que visam a venda de produtos e serviços relacionados entre si para um mesmo cliente com o objetivo de aumentar a participação no cliente (*share of customer*). Um exemplo é uma casa de materiais de construção e decoração que comunica a nova linha de produtos para casa de praia para seus clientes que possuem esse tipo de residência, mas que em outra ocasião não necessariamente compraram produtos desta seção.

#### **3.3.5.g. Rentabilização (*up selling*)**

São ações que visam a venda de atualizações, complementos ou aperfeiçoamentos para um determinado produto ou serviço. Um exemplo clássico é uma empresa de informática que mantém registro das compras de todos os seus clientes, comunicando-lhes o lançamento de novos softwares, podendo oferecer descontos na aquisição de uma versão mais atual.

### **3.4. Mineração de Dados**

Desde os primórdios da humanidade, o homem procura não apenas entender, mas também explicar o mundo à sua volta. Essa inquietação pela busca do conhecimento permitiu, ao longo dos séculos, o acúmulo e desenvolvimento de tecnologias. Essa evolução tecnológica o auxiliou em sua trajetória para a conquista do mundo e em conseqüente geração de mais conhecimento.

Por exemplo, se antes o homem atribuía explicações místicas à ocorrência de chuvas, hoje, ele tem a capacidade de monitorar o comportamento dos fatores climáticos e determinar os padrões que levam a esta ocorrência.

Assim, a humanidade desenvolveu uma grande capacidade para explicar de modo lógico e científico não apenas fenômenos da natureza, mas também outras áreas de conhecimento, permitindo a evolução da história. No entanto, na medida em que mais conhecimento é gerado, sua complexidade e volume aumentam numa velocidade assombrosa, tornando difícil sua compreensão e apreensão.

Portanto, não é ao acaso que o homem se vê cercado de dados e informações por todos os lados, seja na Internet, nos jornais, na televisão, nos livros. Assim, um dos grandes problemas da modernidade se tornou exatamente o excesso de dados disponíveis, já que nem sempre é possível filtrar o que realmente há necessidade de se conhecer, perdendo-se aquilo que é importante.

Este problema pode ser ilustrado pela quantidade de dados armazenados eletronicamente, estima-se que essa quantidade dobra de tamanho a cada vinte meses (WITTEN e FRANK, 2000).

Este fenômeno de excesso de dados se faz presente em todas as áreas da vida, principalmente no campo empresarial.

Atualmente os bancos de dados das empresas podem alcançar um tamanho em terabytes – mais de 1.000.000.000.000 bytes de dados. Diante dessa imensidão de dados, há informações escondidas de importância estratégica. É neste contexto, que a mineração de dados tem sido usada tanto para incrementar os lucros, como para reduzir custos. Seu potencial de retorno é enorme. Empresas transnacionais já fazem uso da mineração de dados para localizar e atrair clientes de maior valor, para reconfigurar suas ofertas de produtos e

aumentar suas vendas e, também, para minimizar perdas e fraudes (TWO CROWS CORPORATION, 1999).

Muitas empresas têm usado a mineração de dados para encontrar informações escondidas, ajudando seus executivos na tomada de decisões para condução dos negócios. Bancos e instituições financeiras são as pioneiras no uso da mineração de dados para a tomada de melhores decisões em suas operações diárias. Uma comum aplicação do método de mineração de dados é a identificação de consumidores e produtos apropriados para vendas cruzadas (LIU e YAP, 2001).

Além dos bancos e instituições financeiras, também se destacam as empresas de telecomunicação na utilização desta tecnologia. Estas empresas se destacam pois, devido à natureza de seus negócios, já há muito tempo armazenam dados sobre seus clientes para a condução de suas atividades diárias.

No entanto, em outro extremo, estão as empresas varejistas, já que são poucas as que possuem dados sobre os clientes ou que sabem fazer uso daquilo que possuem.

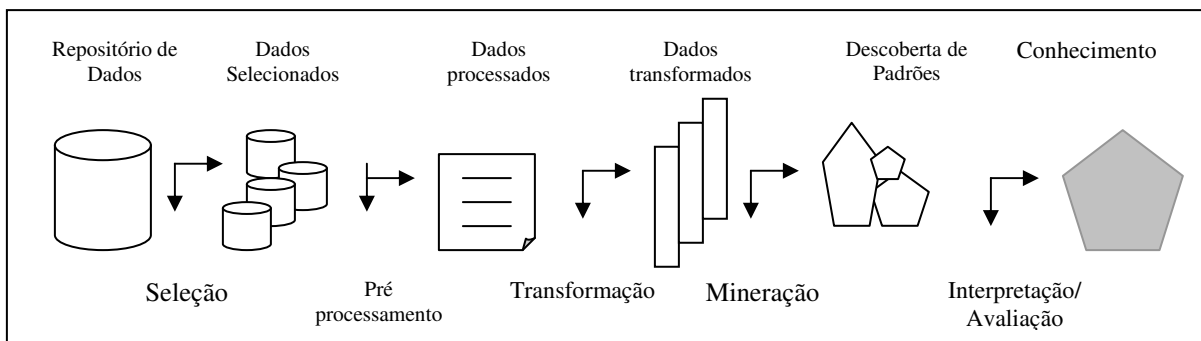
Este fato ilustra a necessidade de tais empresas começarem a desenvolver ações estratégicas para a coleta de dados sobre seus clientes, bem como das transações destes com a empresa, de modo que tal atividade possa ser realizada de forma ordenada e planejada.

Assim, evidencia-se a importância da coleta de dados para a descoberta de novos conhecimentos que possam auxiliar na formulação de estratégias de negócio. Portanto, a existência de dados é fator chave para quem deseja descobrir conhecimento por meio da mineração.

### 3.4.1. Definição

A mineração de dados está inserida no contexto da descoberta de conhecimento em bancos de dados. A descoberta de conhecimento em banco de dados – ou KDD, da abreviação do termo em inglês – *Knowledge Discovery in Databases* – pode ser entendida como o processo de identificar padrões válidos, novos, potencialmente úteis e compreensíveis em banco de dados (FAYYAD, PIATETSKY-SHAPIRO, SMYTH, 1996).

A figura 3 ilustra o processo de mineração de dados para a descoberta de conhecimento



**Figura 3: Processo de Descoberta de Conhecimento em Bancos de Dados (FAYYAD, PIATETSKY-SHAPIRO, SMYTH, 1996)**

Já a mineração de dados, propriamente dita, é uma etapa do processo de descoberta de conhecimento que consiste na aplicação de análises de dados por meio de algoritmos que, sob aceitável limitação da eficiência computacional, poderão produzir uma série de padrões ou modelos sobre determinada base de dados (FAYYAD, PIATETSKY-SHAPIRO, SMYTH, 1996).

Trata-se de uma análise em busca de um novo tipo de conteúdo presente nos bancos de dados, que não pode ser visualizado de maneira imediata pela simples leitura de seus dados, e que pode expressar significados mais profundos sobre os mesmos. Assim, é o processo de transição dos “dados” em “informação”, chegando-se ao “conhecimento”, que é obtido através da exploração das regularidades eventualmente presentes na base de dados (MUNARI, 2001).

No entanto, freqüentemente os termos *descoberta de conhecimento em bancos de dados* (KDD) e *Mineração de Dados* são usados como sinônimos. Algumas definições falam

da mineração de dados como uma das etapas do processo da descoberta de conhecimento, ou seja, a etapa em que é montado um modelo utilizando-se de um algoritmo de mineração de dados. Outras definições mencionam mineração de dados como sendo todo o processo de descoberta de conhecimento (que será detalhado mais adiante), o qual inclui: as etapas de seleção, limpeza e formatação dos dados, além da própria etapa de mineração.

SWIFT (2001) sugere a seguinte definição para contexto empresarial: “A mineração de dados é um processo de análise detalhada de dados, para extrair e apresentar informações recentes, implícitas e que possam ser utilizadas para resolver um problema de negócio”, de modo que tal informação tenha um valor significativo para a empresa.

Neste contexto, a mineração de dados é uma tecnologia utilizada cada vez mais pelas empresas como solução para lidar com o montante crescente de dados. Ela se apresenta como principal aliada para o desenvolvimento do CRM analítico, visando a descoberta de diferentes perfis de clientes e de seus padrões de comportamento de consumo.

### 3.4.2. Tipos de análises

A mineração de dados é a descoberta de padrões presentes em bancos de dados como maneira de se explicar o comportamento desses dados de acordo com determinados eventos.

No entanto, antes de apresentar como tais padrões podem ser descobertos é importante que se ilustre como os dados podem ser apresentados nos bancos de dados.

Tomando-se como exemplo um fictício banco de dados de clientes de uma dada empresa, ilustrado pela tabela 2:

**Tabela 2: Exemplo de banco de dados fictício para ilustração de possíveis atributos**

Identificação do cliente	Tipo de cliente (pessoa física, jurídica)	Data da Compra	Valor Gasto (em R\$)	Lucro/Retorno (R\$)	Produto comprado (Banana, maçã, mamão, abacaxi)
01	Jurídica	01/03/2003	1.000,00	Alto	Banana
02	Física	03/03/2003	20,00	Baixo	Maçã, mamão
03	Física	12/03/2001	100,00	Alto	Maçã, morango, kiwi
01	Jurídica	01/04/2003	500,00	Médio	Banana
03	Física	05/04/2003	120,00	Alto	Banana, maçã, mamão
04	Jurídica	15/04/2003	560,00	Médio	Abacaxi, morango
02	Física	15/04/2003	50,00	Médio	Banana, morango, abacaxi
05	Jurídica	16/04/2003	900,00	Alto	Maçã
06	Física	16/04/2003	99,00	Médio	Laranja, maçã, morango, abacaxi
04	Jurídica	15/05/2003	999,00	Médio	Abacaxi, morango
07	Jurídica	15/05/2003	499,00	Baixo	Maçã, morango

Pode-se notar que, neste exemplo, clientes do tipo pessoa física e do tipo pessoa jurídica recebem classificações diferentes em relação ao item **Lucro/Retorno**. Clientes do tipo pessoa **física** são considerados de **alto retorno** quando o lucro obtido com suas compras é igual ou excede o valor de **R\$ 100**; já clientes do tipo pessoa **jurídica** são considerados com **alto retorno** quando o lucro é igual ou excede o valor de **R\$ 1000**. Este exemplo, portanto, ilustra os tipos de padrões que podem ser encontrados por meio da mineração de dados.

Este banco de dados é composto de vários registros, esses registros (representados pelas linhas) podem significar, por exemplo, uma transação comercial efetuado por determinado cliente em determinada data. Além disso, cada registro possui vários atributos (representados pelas colunas), os quais podem ser do tipo categórico ou do tipo contínuo.

Um atributo categórico é aquele que é representado por símbolos e apresenta um número finito de possibilidades, não sendo possível ou lógico a comparação de valores. Este



tipo de atributo apresenta duas classes: pode ser ordinal, ou seja, apresenta uma possível ordenação em categorias (como, por exemplo, retorno: alto, médio, baixo); já um atributo categórico do tipo nominal não apresenta possibilidade de ordenação (como, por exemplo, produto comprado). Um atributo é dito numérico ou contínuo quando apresenta possibilidade de comparação de seu tamanho, ou ainda, a realização de equações matemáticas com seus valores, tais como soma, média, etc (WITTEN e FRANK, 2000).

Neste dado exemplo, os atributos nominais são representados pelos atributos tipo de cliente, produto comprado e retorno, já o atributo do tipo numérico é representado pelo valor gasto.

Assim, tais registros juntamente com seus atributos são a matéria prima, ou seja, os valores de entrada para a mineração de dados. Esses registros (que em um banco de dados real são apresentados em milhões) podem revelar, por meio da mineração de dados, padrões de comportamento que são importantes para a obtenção de conhecimento e tomada de decisões.

A descoberta desses padrões é realizada por meio de diferentes tipos de análises da mineração de dados. A utilização de determinado tipo de análise depende do tipo de objetivo que se deseja alcançar, além de como se deseja utilizar o conhecimento a ser descoberto.

Primeiramente, são dois os objetivos que podem ser obtidos por meio do processo de descoberta de conhecimento e mineração de dados: verificação e descoberta de hipóteses.

Na verificação, a análise limita-se em verificar uma hipótese levantada pelo usuário. Já na descoberta, o objetivo é descobrir novos padrões, ou seja, descobrir novas hipóteses em determinado assunto (FAYYAD, PIATETSKY-SHAPIRO, SMYTH, 1996).

Segundo FAYYAD, PIATETSKY-SHAPIRO, SMYTH (1996), a descoberta de hipóteses ainda pode ser dividida em predição e descrição. Na predição o objetivo é encontrar padrões para prever comportamentos no futuro e na descrição o objetivo é encontrar padrões que possam descrever o comportamento de determinados dados, visando adquirir conhecimento sobre determinado assunto.

Sendo assim, segundo FAYYAD, PIATETSKY-SHAPIRO, SMYTH (1996), os dois objetivos acima definidos podem ser alcançados por meio de diferentes combinações de análises de mineração de dados, as quais são as seguintes:

**Classificação:** é um método que classifica os dados em categorias pré-definidas (determinadas pelo usuário), baseadas em determinado atributo, de modo que seja descoberto

um padrão para classificação de novos dados. Por exemplo, a classificação de clientes nas categorias: alto-retorno, médio-retorno, baixo-retorno, com base no atributo retorno.

**Regressão:** usa o comportamento de dados existentes para prever como outros dados se comportarão no futuro. Um exemplo para aplicação desta técnica é o monitoramento de estoque de determinado produto para previsão de sua demanda no futuro.

**Descoberta de agrupamentos (*clustering*):** é um método descritivo que tem como objetivo a identificação de grupos de dados, baseados na similaridade entre eles e que sirva para descrever o comportamento dos mesmos. Ao contrário da classificação, neste método não há classes ou grupos pré-definidos.

**Sumarização:** é um método para identificação de descrições compactadas do comportamento dos dados, envolve a utilização de estatística como, por exemplo, média e desvio padrão. É um primeiro contato para a compreensão dos dados e formulação de novos atributos.

**Descoberta de associações:** esta análise consiste em encontrar um modelo que descreva as dependências entre os atributos. Um exemplo é o estudo de análise de cestas de compras, bastante utilizado no varejo, para a descoberta de quais produtos têm sua venda associada a de outros.

Tomando-se como exemplo a tabela 2, essas análises poderiam revelar os seguintes padrões:

As análises de classificação indicariam, que os clientes poderiam ser distribuídos em classes baseadas no atributo Retorno. Assim, as classes estariam divididas em alto, médio e baixo retorno e cada uma delas representaria um determinado perfil de cliente. Portanto, uma possível classificação seria a seguinte:

Se cliente pessoa jurídica e valor gasto por compra:

- >R\$ 999 => Classe: Alto Retorno;
- Entre R\$ 500 e R\$ 999 => Classe Médio Retorno;
- <R\$ 500 => Classe Baixo Retorno

Se cliente pessoa física e valor gasto por compra:

- >R\$ 99 => Classe: Alto Retorno;
- Entre R\$ 50 e R\$ 99 => Classe Médio Retorno;
- <R \$50 => Classe Baixo Retorno

Já uma análise de regressão iria, por exemplo, se basear em registros (dos anos anteriores) dos valores gastos mensalmente por cada uma das classes de clientes para tentar prever quanto cada uma dessas classes iria gastar no próximo ano.

A análise de descoberta de agrupamentos poderia dividir os clientes por: tipo (pessoa física/jurídica), valor gasto, frequência de compra, retorno, etc; pois não há um atributo pré-definido para a formulação de classes e seu objetivo é somente descrever como os registros se organizariam por meio de suas similaridades.

A análise de sumarização tem como objetivo fornecer informações compactadas para dar subsídios à aplicação de outras análises, uma vez que evidenciaria, por exemplo, os produtos que mais vendem ou que tipo de cliente possui maior gasto por compra. Portanto, tornaria claro que clientes do tipo pessoa jurídica possuem alto valor em suas compras para que, assim, pudessem ser analisados separadamente dos clientes do tipo pessoa física.

A descoberta de associação descreve a dependência entre os atributos. No caso do exemplo da tabela 2, esta análise poderia trazer as seguintes informações:

Na maioria dos casos em que se é vendido morango também é vendido abacaxi;

Em 100% dos casos que o cliente é do tipo pessoa jurídica, a compra é realizada somente uma vez ao mês.

### **3.4.3. Representação do conhecimento**

Se os registros juntamente com seus atributos são os valores de entrada para a mineração de dados, a saída é um padrão de comportamento dos dados descoberto por meio de determinada análise escolhida. Assim, para que tal padrão descoberto possa ser representado, há várias técnicas para a representação desse conhecimento adquirido, as quais são: árvores de decisão, redes neurais artificiais, regras de indução, regras de associação, mineração visual de dados e agrupamento (*clusterização*).

A adequação do tipo de análise ao tipo de problema que se deseja solucionar, juntamente com a quantidade e qualidade dos dados, são fatores fundamentais para se definir a técnica mais adequada de execução. Normalmente, os produtos para mineração de dados combinam as diversas técnicas, para se construir um produto mais preciso e mais rápido (MENDONÇA NETO, 2001).

Para o presente trabalho destacam-se as seguintes técnicas de mineração de dados:

#### **3.4.3.a. Regras de Associação**

Descoberta de associação consiste em identificar quais atributos estão associados com outros em um determinado ambiente (TWO CROWS CORPORATION, 1999), também pode ser entendida como uma função que mostra o coeficiente de afinidade entre os dados (TAKAOKA, 1999).

Segundo SWIFT (2001) é uma técnica que descobre regras que correlacionam a presença de um conjunto de itens em outro conjunto de itens, como a presença conjunta de produtos em cestas de compra.

Assim, um exemplo de regra de associação para análise de cestas de compra pode ser o seguinte:

“Quando são comprados iogurtes naturais, em 55% dos casos é comprado um pacote de biscoito integral, a menos que haja uma promoção, caso em que o biscoito de aveia é comprado em 75% dos casos durante a promoção”.

Os exemplos, bem como o texto a seguir, foram adaptados do texto *Introduction to Data Mining and Knowledge Discovery* (TWO CROWS CORPORATION, 1999).

As regras de associação podem ser representadas da seguinte maneira:

$$A \Rightarrow B,$$

onde **A** é chamado de **antecedente** ou lado esquerdo da regra, e **B** é chamado de **consequente** ou lado direito da regra.

Para se determinar a proporção de transações que contenha um determinado item associado a outro é necessário somente que se conte o número de vezes em que tais itens aparecem juntos. Essa frequência com que determinada associação (por exemplo, iogurte natural e biscoito de aveia) ocorre no banco de dados é chamada de **suporte**.

Sendo assim, Suporte é representado pela *(frequência de A e B)/(nº total de ocorrências de um banco de dados)*.

Assim, se 15 transações entre 1000 consistem na venda iogurte natural e biscoito de aveia, o suporte dessa regra de associação será de 1,5%. Portanto, um baixo nível de suporte (por exemplo, uma transação entre um milhão) pode indicar uma associação muito particular que talvez não seja muito representativa, mas tal conclusão depende muito do tipo de problema em que se está trabalhando.

Além disso, para descoberta de regras significativas também é importante que seja considerada a frequência relativa de ocorrência dos itens e de suas combinações. Assim, dada ocorrência de A (o antecedente), com que frequência B (o consequente) ocorre? Dado o exemplo anterior, a pergunta seria a seguinte: “*quando pessoas compram iogurte natural, com que frequência eles também compram biscoito de aveia?*”

Essa frequência relativa é chamada de **confiança** e pode ser calculada da seguinte forma: *(frequência de A e B)/(frequência de A)*.

Para melhor exemplificação de tais conceitos, considera-se o seguinte banco de dados hipotético, com registro de mil transações, ilustrado pela tabela 3.

**Tabela 3: Banco de dados hipotético para exemplificação do cálculo do suporte e da confiança**

<b>Tipo de transação</b>	<b>Nº vezes que ocorre</b>
Iogurte natural	700
Biscoito de aveia	800
Suco de laranja orgânico	400
Iogurte natural e Biscoito de aveia	600
Biscoito de aveia e Suco de laranja orgânico	200
Iogurte natural e Suco de laranja orgânico	300
Iogurte natural, Biscoito de aveia e Suco de laranja orgânico	100

Assim, pode-se calcular suporte exemplificado na tabela 4 e confiança mostrada na tabela 5:

**Tabela 4: Exemplo do cálculo do suporte**

<b>Suporte</b>	<b>Valor %</b>
Iogurte natural e Biscoito de aveia (600/1000)	60
Biscoito de aveia e Suco de laranja orgânico (200/1000)	20
Iogurte natural e Suco de laranja orgânico (300/1000)	30
Iogurte natural, Biscoito de aveia e Suco de laranja orgânico (100/1000)	10

**Tabela 5: Exemplo do cálculo da confiança**

<b>Confiança</b>	<b>Valor %</b>
Iogurte natural => Biscoito de aveia (600/700)	85
Biscoito de aveia => Iogurte natural (600/800)	75
Biscoito de aveia => Suco de laranja orgânico (200/800)	25
Iogurte natural => Suco de laranja orgânico (300/700)	43
Iogurte natural e Biscoito de aveia => Suco de laranja orgânico (100/600)	16
Suco de laranja orgânico => Iogurte natural e Biscoito de aveia (100/400)	25

Portanto, nota-se que a possibilidade de um comprador de iogurte natural também comprar um pacote de biscoito de aveia (85%) é maior do que a possibilidade de um comprador de biscoito de aveia também comprar iogurte natural (75%). Além disso, a ocorrência da regra de associação *iogurte natural => biscoito de aveia* (com suporte de 60%) é consideravelmente suficiente para sugerir que esta regra é significativa e representativa, já que ela ocorre em mais da metade das vezes em que se foi comprado iogurte natural.

Também, pode-se notar que muitas regras podem ser geradas a partir de um pequeno banco dados, portanto métricas como suporte e confiança são fundamentais para avaliação das regras mais interessantes e representativas.

No entanto, é importante frisar que o grau de interesse de uma regra depende totalmente do tipo de conhecimento que se está procurando, já que, por exemplo, no caso de identificação de padrões de comportamento de consumidores fraudulentos, regras com menor suporte são as mais interessantes, já que o que se procura são casos específicos com pouca ocorrência.

Regras de associação diferem da classificação, pois esta técnica gera regras que podem prever qualquer atributo, enquanto que as regras geradas por análises de classificação irão prever somente um atributo, ou seja, o atributo-classe escolhido para a formulação das classes. Além disso, também gera regras que preveem mais de um atributo de uma mesma vez (WITTEN e FRANK, 2000).

Deve-se ressaltar que regras de associação não são realmente regras, mas sim descrições de relações entre dados em um determinado banco de dados, de modo que há uma suposição implícita de que um comportamento do passado possa se repetir no futuro.

#### **3.4.3.b. Árvore de decisão**

É uma das técnicas mais poderosas e populares para análises de classificação, principalmente porque o modo como alcança determinado resultado pode ser expresso em regras, tornando sua compreensão extremamente fácil.

A árvore de decisão pode ser representada por uma série de perguntas com respostas do tipo *sim/não*, onde a resposta da primeira questão determinará o que será perguntado em sequência. Se as perguntas forem bem escolhidas, uma série pequena de perguntas será suficiente para classificar com acurácia determinado registro (BERRY e LINOFF, 1997).

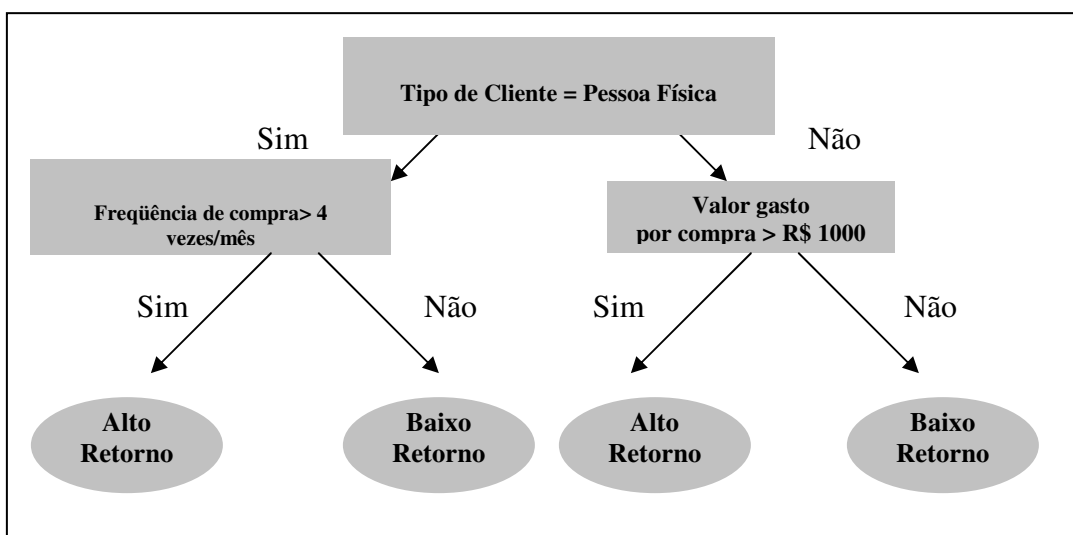
Portanto, árvore de decisão pode ser definida como um modo de representação de uma série de regras que levam até uma determinada classe ou valor a partir da subdivisão sistemática dos dados contidos no repositório de dados analisados.

As árvores de decisão consistem de nodos que representam os atributos, de arcos, provenientes destes nodos e que recebem os valores possíveis para estes atributos, e de nodos

folha, que representam as diferentes classes de um conjunto de treinamento (GARCIA E ALVARES, 2000).

Segundo BRAZDIL (1999), muitos são os algoritmos de classificação que elaboram árvores de decisão. Não há uma forma de determinar qual é o melhor algoritmo, um pode ter melhor desempenho em determinada situação e outro algoritmo pode ser mais eficiente em outros tipos de situações. Os algoritmos que constroem árvores de decisão buscam encontrar aqueles atributos e valores que provêem máxima segregação dos registros de dados, com respeito ao atributo que se quer classificar, a cada nível da árvore.

A figura 4 mostra um exemplo de uma árvore de decisão simples para classificação de clientes em relação ao retorno que geram para empresa.



**Figura 4: Exemplo de árvore de decisão simples**

Pode-se notar que a cada descida dos nodos superiores em direção aos inferiores, os atributos que foram usados são descartados, não sendo utilizados novamente. Esta é uma das críticas referentes a esta técnica, já que uma vez que um atributo é usado, ele será descartado e não será revisto. Assim, a utilização de um atributo para uma nova separação está vinculada ao atributo utilizado anteriormente, sendo conseqüentemente todos dependentes do primeiro atributo utilizado para dar início à árvore. Portanto, se um outro atributo fosse utilizado para dar início à árvore, este traria um resultado diferente para um mesmo conjunto de dados (TWO CROWS CORPORATION, 1999).



Um outro problema referente à técnica da árvore de decisão diz respeito a um fenômeno que recebe o nome de *overfitting*. A árvore cresce até que todos os registros sejam classificados corretamente, porém, há casos em que algumas das regras geradas pela árvore se tornam muito específicas, servindo somente para classificar corretamente aquele conjunto de registros com o qual foi desenvolvida. Assim, ao classificar um novo conjunto de registros, esta árvore trará muitos erros.

Neste caso, existem algoritmos que utilizam técnicas de poda, ou seja, monitoram o crescimento da árvore para que cresça até um tamanho tido como ideal e, assim, possa classificar com acurácia novos conjuntos de registros e não tenha regras específicas para o conjunto de registros do qual se originou.

#### **3.4.3.c. Indução de Regras**

Outra técnica bastante utilizada para análises de classificação é a Indução de Regras.

Como visto anteriormente, na técnica de árvore de decisão os registros de um conjunto de dados são divididos em sub-conjuntos menores, até que as classes sejam formadas e, ao final, dessa divisão é possível que regras sejam formuladas, explicando todo processo de classificação.

Já na técnica de indução de regras, ao invés de se realizar a divisão dos registros em conjuntos menores, o que se procura, por meio de algoritmos, são regras que possam classificar um maior número possível de exemplos de uma só vez, de modo que pertençam a uma mesma classe (MONARD e BARANAUSKAS, 2003).

As regras desta técnica apresentam a seguinte estrutura:

***Se*** <complexo> => ***Então*** <classe>

Onde, complexo representa uma ou mais condições que devem ser atendidas e o conseqüente representa o atributo-classe, com o qual se deseja classificar determinado conjunto de dados.

O processo de indução de regras funciona da seguinte maneira:

De um dado conjunto de registros, é extraída uma lista de regras, de modo que a primeira regra desta lista seja a que cobrir o maior número possível de registros, já a segunda cobrirá um número um pouco menor e assim sucessivamente até que todas as regras possíveis sejam formuladas (WITTEN e FRANK, 2000).

Sendo assim, para a classificação de novos registros estas regras são utilizadas na sequência em que foram geradas. Portanto, se a primeira regra não se aplicar para a classificação de determinado registro, a segunda regra será disparada, caso esta também não se aplique, a terceira será disparada e assim sucessivamente até que seja encontrada a regra que possa classificar este novo registro (WITTEN e FRANK, 2000).

No entanto, pode haver casos em que o registro a ser classificado não se enquadra em nenhuma das regras geradas. Para classificação destes casos, existe uma regra intitulada de regra padrão, esta terá o papel de classificar os casos sem regras, de modo que o registro assuma como classe aquela que tiver o maior número de ocorrências (MONARD e BARANAUSKAS, 2003).

Um outro problema relacionado a esta técnica refere-se ao fato de que há casos em que mais de uma regra se aplicam para a classificação de um mesmo registro. Nestes casos, é necessário que se atribua grau de confiança às regras, devendo-se optar por aquela que apresentar maior valor (TWO CROWS CORPORATION, 1999)

Apesar desse processo ser mais lento do que o processo da técnica de árvore de decisão, seus pontos fortes são a sua compreensibilidade e a necessidade de pouco espaço computacional para seu armazenamento (MONARD e BARANAUSKAS, 2003).

#### **3.4.3.d. Agrupamento (Clusterização)**

O texto a seguir foi adaptado do livro *Data Mining Techniques for marketing, sales and customer support* (BERRY e LINOFF, 1997)

Agrupamento é uma das poucas atividades da mineração de dados descritas como descoberta indireta de conhecimento ou aprendizado não supervisionado, já que a maior parte das tarefas de mineração lida com dados pré-classificados de determinado conjunto de treinamento, com o qual desenvolve-se um modelo capaz de prever a classe de um novo registro.

Na atividade de agrupamento não há dados pré-classificados, ao contrário disso, o que se procura é compreender a estrutura de determinado conjunto de dados a partir da formação de grupos – conhecidos como agrupamentos ou clusters – os quais apresentam similaridades entre si. Assim, por exemplo, espera-se que clientes, pertencentes a um mesmo agrupamento, comportem-se de maneira similar.

A atividade de agrupamento é raramente usada isoladamente, pois descoberta de agrupamentos não se caracteriza por ser um fim em si mesmo. Geralmente, está associada a outras atividades de mineração, evidenciando a necessidade de se aplicar outros métodos para a compreensão do que cada grupo significa.

Há um grande número de métodos matemáticos para a descoberta de agrupamentos. Um dos mais conhecidos e utilizados é chamado de método divisivo. Este inicia a atividade considerando todos os registros de determinado conjunto de dados como sendo parte de um único grande grupo, este grupo é dividido em dois ou mais grupos, ou tantos quantos forem necessários, até que cada registro possua seu grupo exclusivo.

A cada etapa do processo, todas as possibilidades de divisão são gravadas para que, ao final, o melhor conjunto (número) de grupos possa ser escolhido, de modo que registros – pertencentes ao mesmo grupo – sejam, entre si, o mais homogêneo possível e o mais heterogêneo possível, quando comparados com registros de outros grupos.

#### 3.4.4. O processo da descoberta de conhecimento

Para que as análises de mineração de dados possam ser efetuadas com sucesso e, assim, possa ser obtido o conhecimento almejado por meio da descoberta de padrões, há uma série de passos que devem ser tomados antes e depois da execução dessas análises. Tais passos fazem parte do processo da descoberta de conhecimento, anteriormente ilustrado pela figura 3.

Segundo FAYYAD, PIATETSKY-SHAPIO, SMYTH (1996) são nove os passos que fazem parte do processo da descoberta de conhecimento, o qual se caracteriza por ser de natureza bastante interativa.

O primeiro deles é compreender do que se trata o tema a ser estudado, buscando adquirir um pré-conhecimento sobre o assunto, de modo que possa ser estabelecido qual será o objetivo a ser alcançado pelo processo de KDD.

O segundo, após conhecimento dos dados disponíveis para o estudo, há necessidade de se escolher quais conjuntos de dados (ou sub-conjuntos) serão trabalhados e com os quais o estudo será desenvolvido.

O terceiro passo diz respeito à limpeza e pré-processamento dos dados. Este passo é crucial para um bom desempenho de todo o processo, pois ele cuidará dos valores de entrada que alimentarão as análises e, mais tarde, gerarão os padrões de comportamento dos dados. Portanto, caso não seja executado de maneira adequada comprometerá todo o resultado do estudo.

Sendo assim, esta terceira etapa inclui tarefas como *Formatação*, ou seja, remoção de dados com “ruídos” (dados duplicados) e a escolha de estratégias para lidar com campos de dados perdidos ou incompletos. Outra tarefa referente a este passo é a de *Adaptação*, que significa tornar os dados e os arquivos utilizáveis pela ferramenta de mineração de dados escolhida.

O quarto passo refere-se à redução e projeção, ou seja, a escolha de quais atributos são os mais interessantes para a representação dos dados. Esta tarefa está intimamente ligada ao objetivo que se pretende alcançar, já que ele indicará que tipo de conhecimento que se busca.

Assim, trabalhando e transformando os atributos é possível não apenas reduzir o número de variáveis a ser considerado (livrando-se daquilo que não é interessante), mas também trazer novos dados ao estudo, o que poderá enriquecer a análise.

O quinto passo é a combinação do objetivo do KDD (passo 1) com uma determinada análise de mineração de dados como, por exemplo, classificação, associação, agrupamento.

O sexto passo refere-se à escolha de qual técnica de mineração será utilizada para a descoberta de padrões, já que algumas técnicas lidam com dados categóricos, enquanto outras lidam com dados numéricos.

O sétimo passo é a aplicação da mineração de dados, ou seja, a procura por padrões por meio de determinada técnica. Neste passo, para melhorar a performance do padrão descoberto, provavelmente o usuário terá que rever todos os passos anteriores, evidenciando a natureza interativa do processo.

No oitavo passo é realizada a interpretação dos padrões minerados por meio da visualização dos mesmos. Também, nesta etapa é possível que novamente haja necessidade de retomar alguns dos passos anteriores.

O último passo, após revisão e melhora de todos os passos anteriores, diz respeito, finalmente, à tomada de decisão por meio do conhecimento descoberto.

É extremamente importante destacar que a qualidade do conhecimento descoberto está completamente vinculada às tarefas de pré-processamento dos dados e à adequação destes e dos atributos às técnicas utilizadas, sendo tais tarefas tão ou até mais importantes que as próprias tarefas de mineração. Assim, se os dados de entrada estiverem errados e desvinculados aos objetivos estabelecidos, o padrão a ser gerado, com certeza, será errôneo e não corresponderá à realidade, comprometendo todo o estudo.

## **4. Material e Métodos**

### **4.1. Material**

#### **4.1.1. Vila Yamaguishi**

##### **4.1.1.a. Yamaguishi no Mundo**

Em 1953 foi criada no Japão, por um agricultor chamado **Miyozo Yamaguishi**, a idéia e o plano de realização de uma sociedade ideal, que tivessem como base a harmonia da Natureza com a ação humana.

Deu-se, então, início ao que se chamou de Movimento *Yamaguishismo* e, por meio das pessoas que se reuniram, concordando e apoiando esse movimento, foi fundada a Associação Yamaguishi Associação Felicidade.

O Yamaguishi dispõe de vilas espalhadas por diversas cidades do mundo todo. Assim, como no Brasil, nas Vilas Yamaguishi dos demais países são produzidas hortaliças e frutas de alta qualidade.

Atualmente, em todo o mundo, atividades de práticas sociais são desenvolvidas pelos seus associados, objetivando um mundo onde todas as pessoas possam viver em alegria, isto é, numa sociedade saudável, com abundância material e espiritual sem disputas.

Este fato pode ser ilustrado pelo caso de Jaguariúna, SP, onde Sítio Vila Yamaguishi, em parceria com a ONG Amigos do Camanducaia, vem realizando anualmente a limpeza dos rios, mobilizando os alunos das escolas da região, numa prática de educação ambiental. Paralelamente à limpeza dos rios, também é realizada a soltura de alevinos nativos da região, com grande participação da comunidade local. As figuras 5 e 6 ilustram este evento.



**Figura 5: Atividade social Yamaguishi Jaguariúna - Limpeza dos rios**



**Figura 6: Atividade social Yamaguishi Jaguariúna - Soltura de alevinos**

Além disso, a Associação Yamaguishi – Associação Felicidade promove o Curso Especial em Reunião de Kensan do *Yamaguishismo* (Curso Tokkou). Este curso, disponível a qualquer pessoa interessada, apresenta-se como uma oportunidade para o auto-conhecimento, bem como para examinar e esclarecer a realidade da sociedade ideal e os métodos concretos para sua concretização.

#### **4.1.1.b. Vilas Yamaguishi**

Existem Vilas Yamaguishi no Brasil, Suíça, Alemanha, EUA, Coreia, Tailândia, Austrália e em mais de 40 localidades no Japão.

Nessas vilas, várias famílias são voluntárias, vivendo sem posses. Moradia, alimentação e roupas são gratuitos. Além disso, não há salários, não existem chefes ou patrões, nem ordens a serem obedecidas.

No Brasil, assim como nos outros países, a Vila pratica a harmonia entre a natureza e a ação humana consciente por meio da produção agrícola, almejando expandir a sociedade baseada no amor fraterno para todo o mundo.

A Vila Yamaguishi também é chamada de “Vila da alegria e amizade onde não é preciso dinheiro”, é o lugar onde a sociedade de amor fraterno, sem fronteiras e posses, se manifesta na prática. As pessoas convivem em um ambiente de amizade e alegria em comunhão com a natureza.

#### **4.1.1.c. Sítio Vila Yamaguishi Jaguariúna**

Para o desenvolvimento do Modelo de Gestão Estratégica para comercialização de produtos orgânicos, esta pesquisa foi realizada por meio de um estudo de caso, tendo como objeto de estudo o Sítio Vila Yamaguishi, produtor de orgânicos, localizado em Jaguariúna, SP, certificado pela ANC (Associação de Agricultura Natural de Campinas e Região).

A ANC é uma associação de produtores orgânicos independente, fundada em 1991 em Campinas.

Ela faz parte do Colegiado Estadual como certificadora credenciada para produtores orgânicos. Por meio de seu corpo técnico, a ANC faz periodicamente a fiscalização dos produtores certificados de acordo com as normas do Colegiado Estadual e do Ministério da Agricultura, garantindo aos consumidores a qualidade orgânica dos produtos certificados em todo o território nacional.



O Sítio Vila Yamaguishi existe desde de 1988 e é responsável pela produção de hortaliças e ovos. Além de comercializar o que produz, também compra e revende frutas e produtos processados (geléias, iogurtes, sucos, etc.) de outros produtores orgânicos.

A Figura 7 ilustra a logomarca da empresa e a 8 mostra algumas de suas hortas.

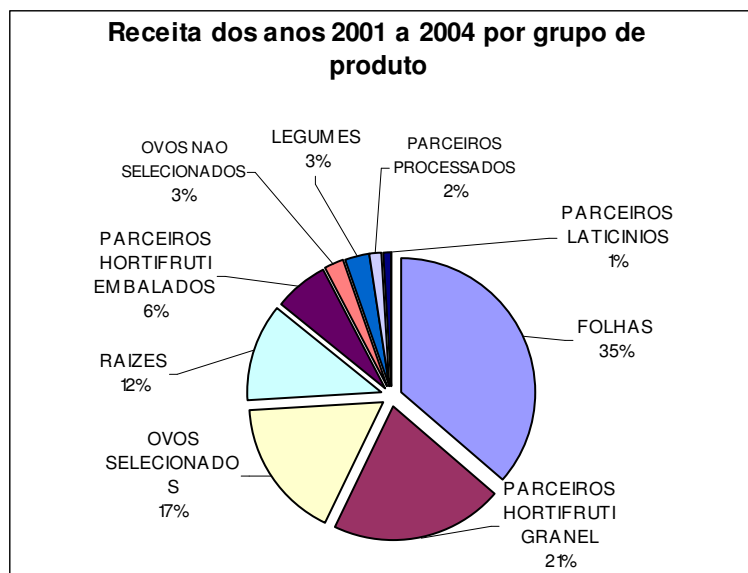


**Figura 7: Logomarca do Sítio Vila Yamaguishi Jaguariúna**



**Figura 8: Hortas do Sítio Vila Yamaguishi Jaguariúna**

O gráfico da figura 9 ilustra a participação de cada um dos grupo de produtos (definido pela empresa) em relação à receita gerada no período de 2001 a 2004.

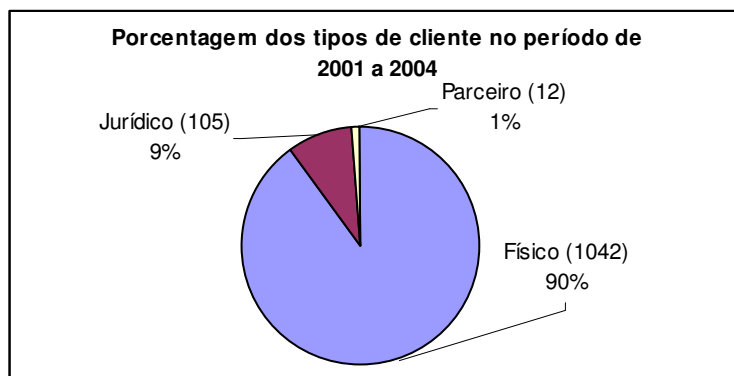


**Figura 9: Gráfico da porcentagem da geração de receita no período de 2001 a 2004 por grupo de produto**

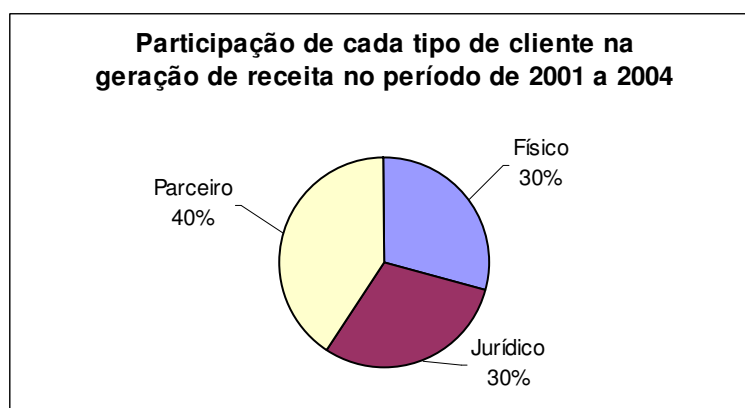
Os produtos do Yamaguishi são vendidos em Campinas e região, além de alguns pontos em São Paulo. Seu principal canal de distribuição é a venda direta, que atende 90% de seus clientes (no caso, clientes físicos); também há venda por meio de feiras de produtos orgânicos, lojas especializadas em produtos naturais e vendas para outros estabelecimentos (supermercados, restaurantes, etc).

O Yamaguishi possui três tipos de clientes: físicos, jurídicos e parceiros. Os clientes parceiros são também clientes jurídicos, no entanto, a diferença entre um e o outro é que clientes parceiros compram os produtos por um preço diferenciado dos clientes jurídicos, pagando menores valores.

O gráfico da figura 10 ilustra o número e a porcentagem de clientes pertencentes a cada um dos tipos e o gráfico da figura 11 mostra a participação de cada um dos tipos na geração da receita bruta no período de 2001 a 2004.



**Figura 10:** Gráfico do número e porcentagem pertencentes a cada tipo de cliente no período de 2001 a 2004



**Figura 11:** Participação de cada um dos tipos de clientes na geração de receita no período de 2001 a 2004

#### 4.1.2. Banco de Dados

Para a construção do modelo, foi utilizado o banco de dados desta empresa, que apresenta os seguintes arquivos e os seguintes dados em cada um deles:

- **Clientes:** dados cadastrais (nome, endereço, telefone), código do cliente, tipo de cliente (pessoa: física, jurídica), código da rota de entrega a que pertence;
- **Produtos:** código do produto, código do grupo a que pertencem, preços (compra e venda), código de seu fornecedor, estoque;
- **Grupos de produtos:** código do grupo, descrição do grupo (hortaliças, frutas, processados, etc.);
- **Fornecedor:** código do fornecedor, dados cadastrais (nome, endereço, telefone, cidade), condições para pagamento;
- **Rotas:** código da rota, descrição da rota (região das entregas);
- **Pedidos:** histórico das vendas;
- **Itens Pedidos:** código do cliente, data da venda, produtos comprados, grupos aos quais esses produtos pertencem, fornecedor, quantidade vendida, preço dos produtos (compra e venda), valor gasto, rota, vendedor, retorno do pedido, perda do pedido.

O arquivo mais importante para a realização deste estudo foi o arquivo: **Pedidos**, uma vez que representa o registro de todas as transações comerciais realizadas entre clientes e a empresa, os registros vão desde o início de janeiro de 2001 até o início de março de 2004 (data da coleta dos dados), contabilizando um total de 27.365 de pedidos.

O arquivo **Itens Pedidos** também foi importante para a realização do estudo, já que contém a descrição dos produtos comprados em cada pedido.

Os demais arquivos são apenas de apoio, já que apresentam a descrição de todos os códigos presentes nos registros.

#### 4.1.3. Recursos Computacionais

As análises de Mineração de Dados, Associação e Classificação, realizadas para a descoberta dos hábitos de consumo e dos perfis de clientes respectivamente, além da fase de análise dos dados, utilizaram os seguintes recursos:

- Microsoft Excel: para mineração visual dos dados
- Gerenciador de banco de dados Microsoft Access: para construção de novas tabelas
- Software de mineração de dados: WEKA (Waikato Environment for Knowledge Analysis) – software gratuito desenvolvido pela Universidade de Waikato da Nova Zelândia, disponível na Internet através do site: <http://www.cs.waikato.ac.nz/~ml/weka/index.html>

É importante destacar que o software WEKA é uma coleção de algoritmos de aprendizagem de máquina e foi desenvolvido para tarefas de mineração de dados. Esses algoritmos podem ser aplicados tanto diretamente na base de dados ou por meio de um código Java próprio.

O WEKA contém ferramentas para o pré-processamento de dados, para tarefas de: classificação, regressão, *clusterização*, regras de associação e visualização. O WEKA é um software “*open source*” e distribuído sob licença GNU (*General Public License*).

## 4.2. Métodos

O Modelo de Gestão Estratégica com foco no Cliente para comercialização de produtos orgânicos teve como base para seu desenvolvimento a teoria da **Inteligência Competitiva**, que propõe quatro fases para o processo de transformação de dados e informações em Inteligência, de modo que se obtenha vantagens frente aos concorrentes.

Das quatro fases que compõem o ciclo de IC, este trabalho focou seus esforços principalmente na terceira fase: **A análise da informação e sua transformação em inteligência.**

Assim, para construção deste Modelo foi utilizado o método de **Descoberta de Conhecimento em Banco de Dados** (KDD) proposto por FAYYAD, PIATETSKY-SHAPIRO, SMYTH (1996).

Este método foi utilizado para a transformação de dados em conhecimento sobre os clientes e para descoberta dos hábitos de consumo.

O modelo de diferenciação de clientes, baseado na estratégia de **Diferenciação** da teoria de **Gestão do Relacionamento com Cliente** (CRM) – que propõe a segmentação dos clientes por valor, ou seja, por meio da medição do retorno que estes geram para empresa – foi desenvolvido a partir da teoria de KDD, por meio da tecnologia da **Mineração de Dados**.

Deste modo, dados e informações puderam ser transformados em Inteligência Competitiva, sendo a base para a concretização do modelo de diferenciação de clientes baseado em CRM e posterior proposição de ações de comercialização de acordo com os perfis descobertos.

#### **4.2.1. Limpeza e Pré-processamento dos dados**

Para a atividade de formatação (remoção de “ruídos” e campos incompletos) foi utilizado o programa Microsoft Access.

Os clientes com uma frequência de compra bastante elevada e visivelmente destoante dos demais foram eliminados para não distorcer o resultado das análises.

Houve um caso de um cliente parceiro (Sabor Natural), que é bastante ativo e possui um padrão de frequência de compra e de receita totalmente atípicos, destoando do restante e, portanto, tal cliente não foi considerado nas análises.

Assim, seus registros foram eliminados do banco de dados, pois caso tivessem sido mantidos, as análises não revelariam um padrão verdadeiro dos outros clientes, já que seu peso é muito maior de que todos os outros juntos.

Foram também excluídos os registros correspondentes às vendas das feiras e vendas internas referentes aos moradores da Fazenda Yamaguishi, já que nestes casos não foi possível obter a identificação individual dos clientes (sendo isto a base para a construção do modelo de diferenciação baseado em CRM).

#### 4.2.2. Escolha dos campos e criação de novos atributos

Dos seguintes campos presentes no conjunto de dados Itens Pedidos, os seguintes foram mantidos para criação de novos atributos de acordo com a tabela 6:

**Tabela 6: Campos mantidos no processo de Mineração de Dados e para a criação de novos atributos**

<b>Campos</b>	<b>Mantidos</b>
Código do item de pedido	
Código do pedido	<b>X</b>
Código do cliente	<b>X</b>
Tipo de cliente	<b>X</b>
Data da entrega	<b>X</b>
Produtos comprados	<b>X</b>
Grupos dos produtos comprados	
Quantidade vendida	
Preço de compra	
Preço de venda	
Valor total gasto	<b>X</b>
Rota	<b>X</b>
Cidade	<b>X</b>
Fornecedor	
Vendedor	
Retorno do pedido	
Perda do pedido.	

Os campos dos códigos foram mantidos apenas para a identificação dos pedidos e dos clientes, não sendo utilizados na atividade de mineração, pois não agregam nenhum tipo de importância para as análises.

A seguir serão apresentados os atributos formulados a partir dos campos existentes no arquivo.



#### 4.2.2.a. Atributo dimensão: Tempo

O campo de Data da entrega não foi utilizado em seu formato original nas análises, mas a partir dele foi possível criar os seguintes atributos da dimensão tempo:

- Dia da semana;
- Mês do ano;
- Período do mês: 1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup>, 4<sup>a</sup> semanas;
- Quinzena do mês: 1<sup>a</sup> ou 2<sup>a</sup>;
- Estação do Ano: neste caso foi estabelecido: **1.Verão** 21/12 a 20/03; **2.Outono** 21/03 a 20/06; **3.Inverno** 21/06 a 20/09; **4.Primavera** 21/09 a 20/12;
- Pré e Pós-feriado: os dias antecedentes e posteriores a feriados. Os feriados, caídos durante o final de semana, não tiveram seus dias de pré e pós-feriado considerados, pois as entregas são realizadas somente durante os dias de semana. Assim, para os feriados de sexta-feira foi considerado pós-feriado a segunda e não o sábado, já para os feriados de segunda-feira foi considerado pré-feriado a sexta-feira e não o Domingo.
- Cliente Ativo e Não-ativo: foi considerado cliente ativo aquele que comprou durante o período que compreende dezembro de 2003 ao início de março de 2004, data que corresponde ao término dos registros no momento da aquisição dos dados.

A motivação por ter utilizado tais atributos da dimensão tempo (com exceção do atributo Cliente ativo) foi verificar sua influência sobre o consumo, como, por exemplo, se a ocorrência de feriados poderia influenciar na venda de determinados produtos, ou ainda, se o tipo da estação do ano influenciaria no consumo de determinadas frutas ou verduras, etc.

Já o motivo da criação do atributo cliente ativo foi verificar quais eram os perfis de clientes ativos e não ativos, para que diferentes estratégias pudessem ser traçadas de acordo com a situação em que o cliente se encontrava.

#### 4.2.2.b. Atributo Recência

A **recência** representa o número de dias que se passaram entre a data da última compra e a anterior de um mesmo cliente. Portanto, para o cálculo deste novo atributo, foi usada a data da compra mais recente em relação a anterior, para a contagem dos dias entre uma e outra.

Depois, a segunda compra mais recente em relação a sua anterior, até que todos os registros de um mesmo cliente recebessem o novo campo da recência.

É importante ressaltar que a primeira compra recebe valor zero para a recência.

A figura 12 ilustra este processo.

Cod. do Pedido,	Cod. Cliente	, Data	, Valor gasto,	Recência
45	, x	, 12/02/04,	10.00	27
38	, x	, 15/12/03,	15.00	45
26	, x	, 01/11/03,	7.00	10
22	, x	, 22/10/03,	20.00	0 → 1ª Compra

Figura 12: Processo para criação do atributo Recência

No entanto, tal atributo não foi utilizado diretamente nas análises, mas sim a discretização dos mesmos em notas de 1 a 5, onde 1 é a pior e 5 a melhor.

Estas notas foram criadas por meio de análises de histograma na qual, após várias tentativas, a melhor curva descendente foi utilizada para estabelecer os valores correspondentes a cada nota.

#### 4.2.2.c. Atributo Frequência Acumulada

Outro atributo criado foi **frequência acumulada** que, neste caso, é o número de vezes que um mesmo cliente realizou compras durante um determinado período de tempo.

No caso de clientes físicos este período foi estabelecido de 90 dias, já para os clientes jurídicos e parceiros foi de 60 dias.

Como os clientes jurídicos e parceiros costumam comprar mais freqüentemente que os clientes físicos, um prazo menor foi estabelecido para o cálculo da sua freqüência acumulada.

Apesar destes períodos parecerem muito longos para o cálculo da freqüência acumulada, por se tratar da venda de produtos perecíveis, deve-se levar em consideração o problema da rotatividade dos clientes. Portanto, ao se considerar três meses ao invés de somente um, é possível destacar quais são os clientes realmente mais freqüentes.

O cálculo da freqüência acumulada obedeceu ao seguinte raciocínio: num prazo de 90 dias (no caso de clientes físicos), considera-se o registro de compra mais recente e conta-se o número de vezes em que houve compras dentro deste período. Deste modo, este novo número é acrescentado ao registro.

Depois, considera-se a segunda compra mais recente e o número de compras realizadas nos 90 dias anteriores e assim por adiante, até que todos os registros de um mesmo cliente recebam o novo campo, sendo que a primeira compra recebe valor zero para este campo.

A figura 13 ilustra este procedimento. A Sigla FA representa a freqüência acumulada.

Cod. do Pedido, Cod. Cliente	Data	Valor gasto	FA
45 , x	, 12/02/04 ,	10.00	2
38 , x	, 15/12/03 ,	15.00	3
26 , x	, 01/11/03 ,	7.00	2
22 , x	, 22/10/03 ,	20.00	0

1ª Compra

**Figura 13: Processo para criação do atributo Freqüência Acumulada**

Da mesma forma que a recência, o novo campo da frequência acumulada não foi utilizado diretamente nas análises. Estes valores foram discretizados em notas de 1 a 5, onde 1 é a pior e 5 a melhor.

O objetivo de ter analisado os clientes em relação aos aspectos da recência e frequência acumulada é que estes atributos permitem uma análise sob diferentes pontos de vista, já que dois clientes podem, por exemplo, comprar 4 vezes em um mês, tendo a mesma frequência. No entanto, um pode efetuar as compras em intervalos regulares de 7 dias e outro pode apresentar uma recência com intervalos totalmente irregulares como, por exemplo, realizar compras com intervalos de 3, 15 e 7 dias.

#### **4.2.2.d. Atributo Valor Acumulado**

Para trabalhar a dimensão Valor dos clientes, o ideal seria ter utilizado o lucro gerado por venda de cada cliente.

Este lucro seria calculado subtraindo-se o preço de compra do preço de venda, multiplicando-se pela quantidade vendida de determinado produto. No entanto, tal atividade não foi possível ser desenvolvida, já que o campo preço de compra estava incompleto na maior parte do arquivo, o que impossibilitou a criação do campo lucro.

Assim, o atributo valor acumulado foi criado por meio do campo Valor total gasto que, na verdade, é a receita bruta gerada em cada transação.

Este atributo foi desenvolvido de modo similar à variável frequência acumulada, considerando-se um período de 90 dias para clientes físicos e 60 dias para clientes jurídicos e parceiros.

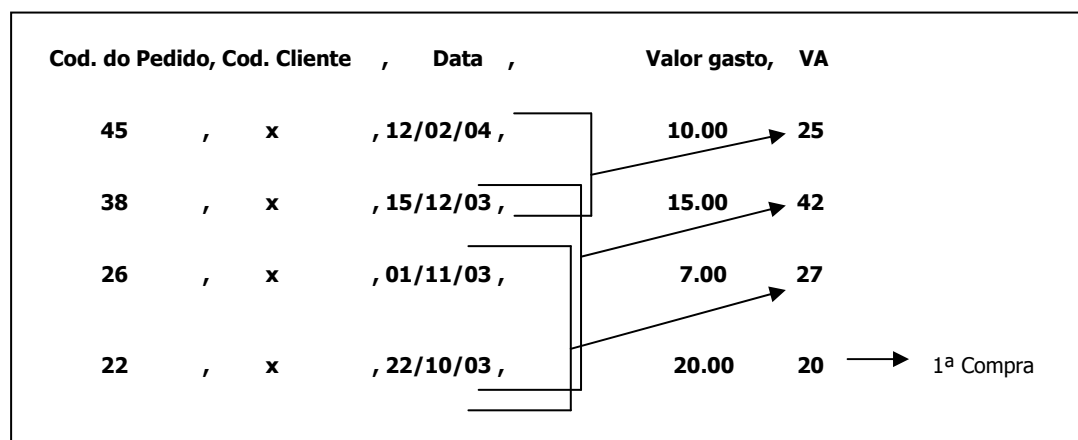
A cada período de 90 dias (clientes físicos) ou 60 dias (clientes jurídicos/parceiros), foi somado o valor gasto do registro mais recente ao valor gasto dos registros com datas de entregas correspondentes a 90 dias anteriores.

Depois, o valor gasto do segundo registro mais recente foi somado ao valor gasto dos registros anteriores, correspondentes aos 90 dias, e sucessivamente até que todos os registros de um mesmo cliente recebessem o novo campo de valor acumulado.

No caso da primeira compra o valor acumulado é igual ao valor gasto.

O mesmo período da frequência acumulada foi utilizado para o cálculo do valor para facilitar a análise dos resultados.

A figura 14 ilustra este procedimento. A sigla VA representa o valor acumulado.



**Figura 14: Processo para criação do atributo Valor Acumulado**

Deste modo, foi possível analisar o cliente não apenas pontualmente, ou seja, somente pelo valor gasto em uma determinada compra, mas sim pelo seu histórico de 3 meses no caso de clientes físicos e 2 meses no caso de clientes jurídicos e parceiros.

Do mesmo modo da frequência, o valor deste atributo não foi diretamente utilizado na fase de mineração, mas sim discretizados em notas de 1 a 5, onde 1 é a pior e 5 a melhor, por meio do mesmo tipo de análise de histograma.

### *Observação*

Para análises que contenham o uso dos atributos valor acumulado e frequência acumulada, deve-se considerar o mesmo período de tempo para o cálculo de ambos, já que este procedimento facilita a compreensão do resultado.

Para a recência não há esta necessidade, pois seu cálculo não envolve o uso de um determinado período, mas somente os números de dias entre um registro e outro.

#### 4.2.3. Tipos de análises realizadas

As seguintes análises de Mineração de dados foram utilizadas para alcançar os objetivos propostos:

- **Classificação:** para identificação dos vários perfis de clientes existentes (criação do modelo de Diferenciação dos clientes, baseado em CRM).  
Dois tipos de arquivos diferentes foram utilizados – do **histórico** e das **transações**.
- **Associação:** para descoberta dos hábitos de compras.  
Três tipos diferentes de arquivos foram utilizados – arquivo geral (com todos os registros), arquivo apenas com os registros onde ocorriam os atributos pré-pós-feriados, arquivos apenas com os registros dos produtos de maior valor agregados, já detalhados anteriormente.

Para todas as análises realizadas, os clientes físicos e clientes jurídicos/parceiros foram analisados separadamente, já que possuem características diferentes quanto aos valores gastos por compra, período entre as mesmas. Portanto, não seria interessante mantê-los juntos.

#### 4.2.3.a. Classificação

Nas análises de classificação, dois formatos de arquivo foram utilizados.

No **primeiro tipo de arquivo, transações**, não houve transformação no formato original do arquivo de Pedidos, pois cada registro analisado (cada linha do arquivo) representa exatamente cada um dos pedidos realizados pelos clientes.

Também, é importante lembrar que o período para o cálculo da frequência e valor acumulado foi de 90 dias para físicos e 60 dias para jurídicos/parceiros.

Além disso, é importante destacar que o objetivo da utilização dos campos cidade e rota foi verificar suas influências no processo de classificação, analisando a possibilidade de determinada cidade ou rota ter algum cliente que se destacasse pelo seu alto valor.

Os atributos Cliente ativo e Nota do valor acumulado foram utilizados como atributos meta de classificação nas análises do arquivo das transações.

As tabelas 7 e 8 ilustram os atributos e seus respectivos valores utilizados neste tipo de arquivo.

**Tabela 7: Atributos utilizados no arquivo das Transações para análise dos clientes físicos**

<b>Atributos</b>	<b>Valores</b>
Nota de recência	0,1,2,3,4,5
Nota da frequência acumulada	0,1,2,3,4,5
Nota do valor acumulado	1,2,3,4,5
Cliente Ativo	S, N
Cidade	Códigos das cidades
Rota	Códigos das rotas

**Tabela 8: Atributos utilizados no arquivo das Transações para análise dos clientes jurídicos e parceiros**

<b>Atributos</b>	<b>Valores</b>
Dia da semana	1,2,3,4,5
Nota de recência	0,1,2,3
Nota da frequência acumulada	0,1,2,3,4,5
Nota do valor acumulado	1,2,3,4,5
Cliente Ativo	S, N
Cidade	Códigos das cidades
Rota	Códigos das rotas

O **segundo formato de arquivo** utilizado para as análises de classificação foi o do **histórico** de clientes.

Neste tipo de arquivo, todos os pedidos de um mesmo cliente foram unidos em somente uma linha, como mostra a figura 15.

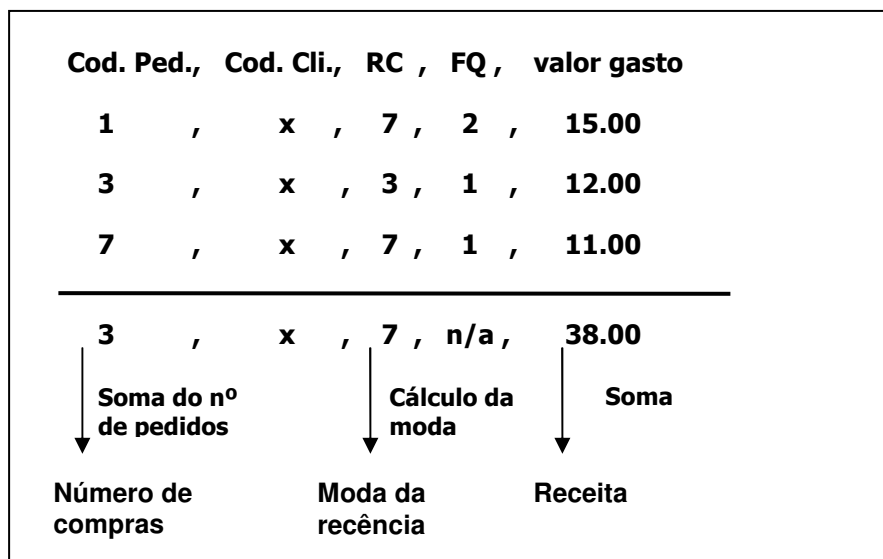


Figura 15: Exemplificação do processo de criação do arquivo de Histórico

A tabela 9 compara os atributos presentes no arquivo de transações com os presentes no arquivo de histórico.

Tabela 9: Comparação entre os atributos presentes no arquivo das Transações e no arquivo de Histórico

Transações	Histórico
Código do pedido	Número total de compras
Código do cliente	Mantido
Recência	Moda da Recência
Frequência	Excluído
Valor Gasto	Receita total
Cliente ativo	Mantido
Cidade	Excluído
Rota	Excluído



No arquivo do histórico, ao unir todos os pedidos de um mesmo cliente em uma única linha, o campo do código do pedido deixou de ter sentido, uma vez que cada linha deste arquivo representa um cliente diferente. Portanto, esta foi a razão deste campo ter sido eliminado e transformado no número total de compras do cliente.

Em relação à recência, houve a necessidade de unir todos os valores deste campo em somente uma linha para adequação ao formato do arquivo.

Para tanto, foi utilizada a função Moda, que representa o número que mais se repete, sendo justamente o tipo de informação mais interessante para o cálculo deste campo, pois caso fosse utilizado a média, esta função poderia ter distorcido o tipo de conhecimento procurado.

A título de exemplificação, sejam consideradas duas pessoas que efetuaram 4 compras, com seguintes intervalos de dias entre elas: 7/7/4 e 10/2/6; calculando-se a recência pela média ambas teriam recência com valor 6, porém pelo moda, os valores seriam 7 para a primeira e sem moda para a segunda, pois neste caso nenhum valor se repete.

O campo do valor gasto, que anteriormente foi utilizado para a criação do atributo valor acumulado do arquivo das transações, teve todos os seus valores somados em relação aos pedidos um mesmo cliente, sendo transformado no atributo Receita.

Neste formato de arquivo, os valores dos campos do número de compras, da moda da recência e da receita também não foram utilizados diretamente nas análises de mineração, mas foram discretizados em notas, por meio de análise de histograma, para que classes pudessem ser criadas.

A tabela 10 destacar os atributos utilizados neste arquivo.

**Tabela 10: Atributos e seus valores, utilizados no arquivo de Histórico**

<b>Atributo</b>	<b>Valores</b>
Nota do Número de Compras	1,2,3,4,5
Nota da Receita	1,2,3,4,5
Nota da moda da Recência	1,2,3,4,5
Cliente Ativo	S, N

É importante destacar que nas análises com o arquivo histórico, o atributo meta de classificação foi Cliente Ativo e 5 representa o valor da melhor nota.

### *Observação*

Os atributos Nota do Número de Compras e Nota da moda da Recência, para simplificação, serão chamados adiante de: Nota de Compras e Nota da Recência.

Assim, Nota de Compras refere-se à nota atribuída ao número de compras efetuadas em todo o período de três anos, pois se trata do arquivo de histórico, no qual todos os registros de um mesmo cliente foram unidos em uma única linha, idem Nota da moda da Recência.

Nas análises do histórico, tanto para físicos como para jurídicos/parceiros, clientes com até duas compras não foram considerados, já que este número de compras não é suficiente para o estabelecimento de um padrão de comportamento.

Além disso, nestas análises o cliente parceiro Sabor Natural foi mantido, pois seu peso não exerce influência na análise, já que os clientes foram agrupados em classes de receita e número de compras.

#### **4.2.3.b. Associação**

Para a descoberta dos hábitos de compra foram realizadas análises de associação. Neste tipo de análise os dois tipos de clientes também foram analisados separadamente.

Os campos e atributos utilizados nestas análises foram os seguintes:

- Código do pedido
- Código do cliente
- Dia da semana
- Mês do ano
- Período do mês
- Quinzena do mês
- Estação do Ano
- Pré e Pós-feriado
- Cidade
- Rota
- Produtos

É importante ressaltar que houve a necessidade da manipulação dos dados de alguns produtos, já que em alguns casos o tipo de produto constituía em si um produto diferente como, por exemplo, o caso da alface: lisa, americana, crespas; ou ainda, o caso de sucos e iogurtes separados por sabor: morango, pêssago, laranja, etc.

Tal nível de detalhamento não seria interessante para análise, pois diminuiria seu peso no processo da construção das regras. Assim, todos os tipos de sucos, por exemplo, foram transformados em um único produto, bem como os demais produtos com mesmo grau de detalhamento.

Além disso, as análises foram realizadas com três tipos de arquivo.

O primeiro deles apenas levou em consideração a divisão entre cliente físico e jurídico/parceiro, sendo nomeado de arquivo geral.

O segundo tipo teve como objetivo medir a influência do fator pré ou pós-feriado no consumo, pois a ocorrência “sim” deste atributo é muito menor que a “não” ocorrência. Portanto, foi necessário construir um novo arquivo onde somente os registros de ocorrência “sim” foram selecionados para que as regras (mostrando este atributo) pudessem ser criadas pelo algoritmo de associação.

Este mesmo processo também foi utilizado para analisar os hábitos de consumo em relação aos produtos de maior valor agregado como, por exemplo, leite, iogurte, sucos, geléias, etc.

Nestes casos, também, houve o mesmo problema de poucas ocorrências. Deste modo, criou-se um arquivo (para clientes físicos e jurídicos/parceiros separadamente) para cada um dos produtos de maior valor agregado, onde somente foram selecionados os registros em que tais produtos ocorreram.

Assim, os produtos de maior valor agregado analisados separadamente foram:

- Amendoim
- Achocolatado
- Açúcar
- Bananada
- Café
- Produtos de confeitaria: bolos, panetones, pães
- Farinha
- Geléia
- Granola
- Iogurte
- Leite
- Mel
- Molho
- Mostarda
- Óleos
- Noz Macadâmia
- Pickles
- Queijo
- Soja
- Sucos
- Tortas

#### **4.2.4. Técnicas utilizadas**

Para a análise de associação foi utilizada a técnica de regras de associação, já para a análise de classificação foram utilizadas duas técnicas: a de árvore de decisão (para o arquivo de histórico) e indução de regras (para o arquivo das transações).

#### **4.2.5. Interpretação dos padrões descobertos**

Para a interpretação dos padrões descobertos, por meio das análises de associação, os resultados foram levados até um dos responsáveis do Yamaguishi para a verificação se tal conhecimento era corriqueiro em relação às suas práticas comerciais.

## **5. Resultados e Discussão**

Primeiramente serão apresentados os dados preparatórios para a construção do Modelo de Gestão Estratégica e posteriormente o resultado deste trabalho, que é justamente a apresentação do modelo.

### **5.1. Resultados preparatórios para a construção do modelo**

O trabalho de manipulação dos dados, já existentes do Yamaguishi, foi a base para a construção do modelo. Portanto, nesta etapa serão apresentados os novos atributos e a discretização dos seus valores em notas, que foram utilizados no processo de Mineração de Dados, bem como o trabalho de pré-processamento dos dados.

### 5.1.1. Pré-processamento dos dados – exclusão do Sabor Natural

Houve a necessidade de exclusão de um cliente (do tipo parceiro: Sabor Natural) nas análises de associação e de classificação do arquivo das transações, devido ao seu peso em relação ao número de pedidos e geração de receita.

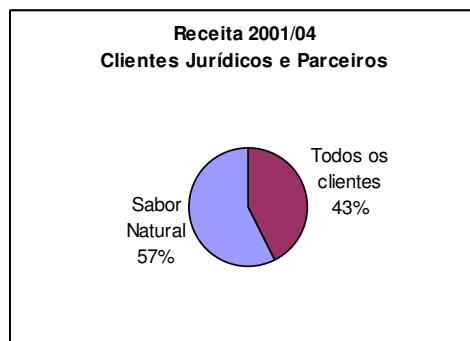
Este fato pode ser evidenciado quando se compara o número de pedidos do cliente em questão (Sabor Natural) com os demais clientes parceiros e, também, a receita gerada por ele em relação à soma das receitas de todos os outros clientes parceiros e jurídicos juntos.

O gráfico da figura 16 ilustra a comparação entre o número de pedidos do Sabor Natural com os demais clientes parceiros no período entre 2001 e 2004.



**Figura 16:** Comparação da porcentagem do número de pedidos do Sabor Natural com todos os demais clientes parceiros, no período de 2001 a 2004

O gráfico da figura 17 ilustra a comparação entre a receita gerada pelo Sabor Natural e os demais clientes jurídicos e parceiros juntos no período entre 2001 e 2004.



**Figura 17:** Comparação da receita gerada pelo Sabor Natural e os demais clientes jurídicos e parceiros juntos, no período de 2001 a 2004

## **5.1.2. Notas dos atributos do arquivo das transações**

### **5.1.2.a. Notas da Recência**

No caso de clientes físicos, foram criadas as seguintes notas de 1 a 5 para a recência, onde 1 é a pior e 5 a melhor:

- Nota 0: identificação da 1ª compra;
- Nota 1: intervalo acima de 28 dias;
- Nota 2: de 22 a 28 dias;
- Nota 3: de 15 a 21 dias;
- Nota 4: de 8 a 14 dias;
- Nota 5: de 1 a 7 dias;

Tais divisões dos dias foram consideradas, pois o ideal é que os clientes comprem ao menos uma vez por semana, já que as rotas de entrega são, na grande maioria das vezes, semanais, podendo ocorrer somente algumas compras em intervalos menores de 7 dias.

No entanto, no caso dos clientes jurídicos e parceiros, a nota da recência recebeu uma outra divisão, já que o comportamento deste tipo de cliente é diferente, pois costumam comprar mais vezes que os clientes do tipo físico. Segue a relação das notas da recência para estes dois tipos de clientes:

- Nota 0: identificação da 1ª compra;
- Nota 1: intervalo acima de 7 dias;
- Nota 2: igual a 7 dias;
- Nota 3: 1 a 6 dias.



### 5.1.2.b. Notas da Frequência Acumulada

Os gráficos das figuras 18 e 19 ilustram os histogramas para a criação das notas da frequência acumulada dos clientes físicos e jurídicos/parceiros respectivamente.

O eixo y do gráfico representa a quantidade de vezes que cada um dos blocos (representados no eixo x) ocorrem.

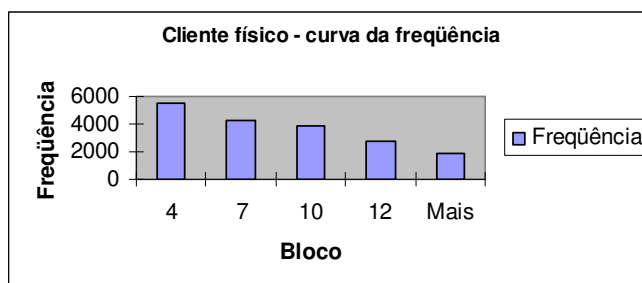


Figura 18: Histograma da curva da frequência acumulada - clientes físicos

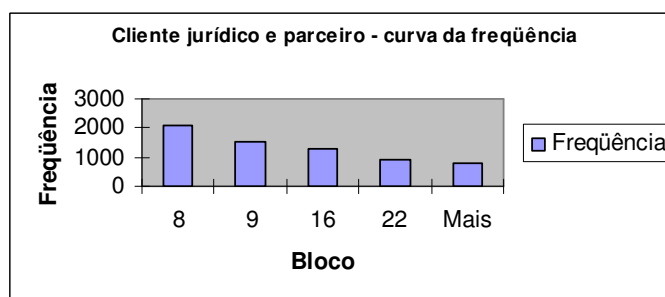


Figura 19: Histograma da curva da frequência acumulada - clientes jurídicos e parceiros

Deste modo, os clientes tiveram suas frequências acumuladas classificadas em notas de 1 a 5, onde 1 é a pior e 5 a melhor.

Assim, clientes físicos receberam nota 1, aqueles que compraram até 4 vezes num período de 90 dias; nota 2 aqueles que compraram de 5 a 7 vezes e, assim sucessivamente como mostra o gráfico da figura 18, até a nota 5 para os que compram mais de 12 vezes no período determinado.

O mesmo aplica-se para cliente jurídicos e parceiros de acordo com os números apresentados no gráfico da figura 19, durante um período de 60 dias.

Sendo assim, as seguintes notas foram estabelecidas para a frequência:

Clientes físicos – período de 90 dias:

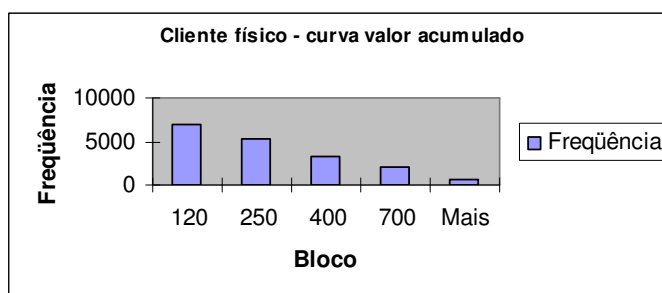
- Nota 1: até 4 compras;
- Nota 2: de 5 a 7 compras;
- Nota 3: de 8 a 10 compras;
- Nota 4: de 11 a 12 compras;
- Nota 5: acima de 12 compras.

Clientes jurídicos e parceiros – período de 60 dias:

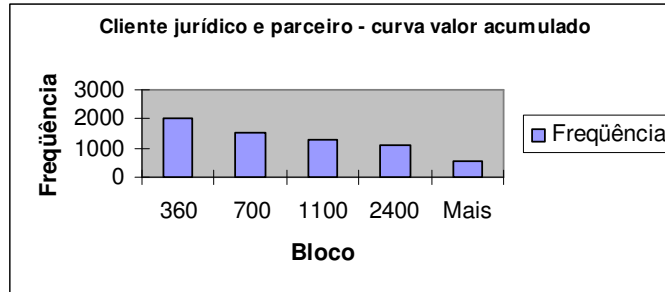
- Nota 1: até 8 compras;
- Nota 2: 9 compras;
- Nota 3: de 10 a 16 compras;
- Nota 4: de 17 a 22 compras;
- Nota 5: acima de 22 compras.

### 5.1.2.c. Notas do Valor Acumulado

Os gráficos das figuras 20 e 21 ilustram a estipulação das notas do valor acumulado para os clientes físicos e jurídicos/parceiros respectivamente, por meio de análise de histograma.



**Figura 20: Histograma da curva do valor acumulado - clientes físicos**



**Figura 21: Histograma da curva do valor acumulado - clientes jurídicos e parceiros**

Assim, as notas para o valor acumulado (Reais) foram as seguintes para clientes físicos, num período de 90 dias:

- Nota 1: até R\$ 120;
- Nota 2: de R\$ 121 a R\$ 250;
- Nota 3: de R\$ 251 a R\$ 400;
- Nota 4: de R\$ 401 a R\$ 700;
- Nota 5: acima de R\$ 700.

As seguintes notas (Reais) foram estabelecidas para clientes jurídicos e parceiros, num período de 60 dias:

- Nota 1: até R\$ 360;
- Nota 2: de R\$ 361 a R\$ 700;
- Nota 3: de R\$ 701 a R\$ 1100;
- Nota 4: de R\$ 1101 a R\$ 2400;
- Nota 5: acima de R\$ 2400.

### **5.1.3. Notas dos atributos do arquivo de histórico**

#### **5.1.3.a. Notas da Moda da Recência**

Cliente Físico:

- Nota 0: para os casos nos quais não foi possível estabelecer a moda, já que pode ter havido empate entre dois números ou nenhum número ter ocorrido mais de uma vez
- Nota 1: moda com valor acima de 28
- Nota 2: moda com valor entre 22 e 28
- Nota 3: moda com valor entre 15 e 21
- Nota 4: moda com valor de 8 a 14
- Nota 5: moda com valor até 7

Cliente Jurídico/Parceiro:

- Nota 0: para os casos nos quais não foi possível estabelecer a moda, já que pode ter havido empate entre dois números ou nenhum número ter ocorrido mais de uma vez
- Nota 1: moda com valor maior que 7
- Nota 2: moda com valor igual a 7
- Nota 3: moda com valor menor que 7

### 5.1.3.b. Notas do Número de Compras

Os gráficos das figuras 22 e 23 ilustram a estipulação das notas deste atributo para clientes físicos e clientes jurídicos/parceiros respectivamente:

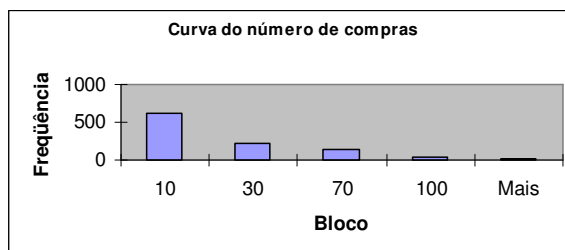


Figura 22: Histograma da curva do número de compras - clientes físicos

As notas para o número de compras são as seguintes para clientes físicos:

- Nota 1: até 10
- Nota 2: de 11 a 30
- Nota 3: de 31 a 70
- Nota 4: de 71 a 100
- Nota 5: acima de 100

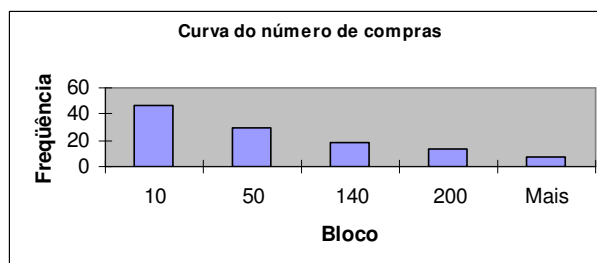


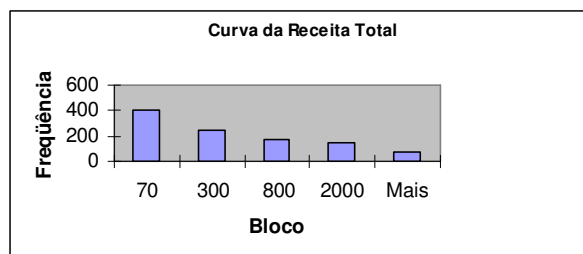
Figura 23: Histograma da curva do número de compras - clientes jurídicos e parceiros

As notas dos clientes jurídicos/parceiros para o número de compras são:

- Nota 1: até 10
- Nota 2: de 11 a 50
- Nota 3: de 51 a 140
- Nota 4: de 141 a 200
- Nota 5: acima de 200

### 5.1.3.c. Notas da Receita

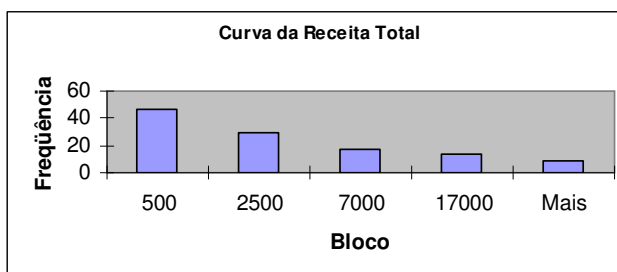
Os gráficos 24 e 25 ilustram a estipulação das notas deste atributo para clientes físicos e clientes jurídicos/parceiros respectivamente:



**Figura 24: Histograma da curva da receita bruta - clientes físicos**

Assim, as notas da receita bruta, em Reais, para os clientes físicos são as seguintes:

- Nota 1: até R\$ 70
- Nota 2: de R\$ 71 a R\$ 300
- Nota 3: R\$ 301 a R\$ 800
- Nota 4: R\$ 801 a R\$ 2000
- Nota 5: acima de R\$ 2000



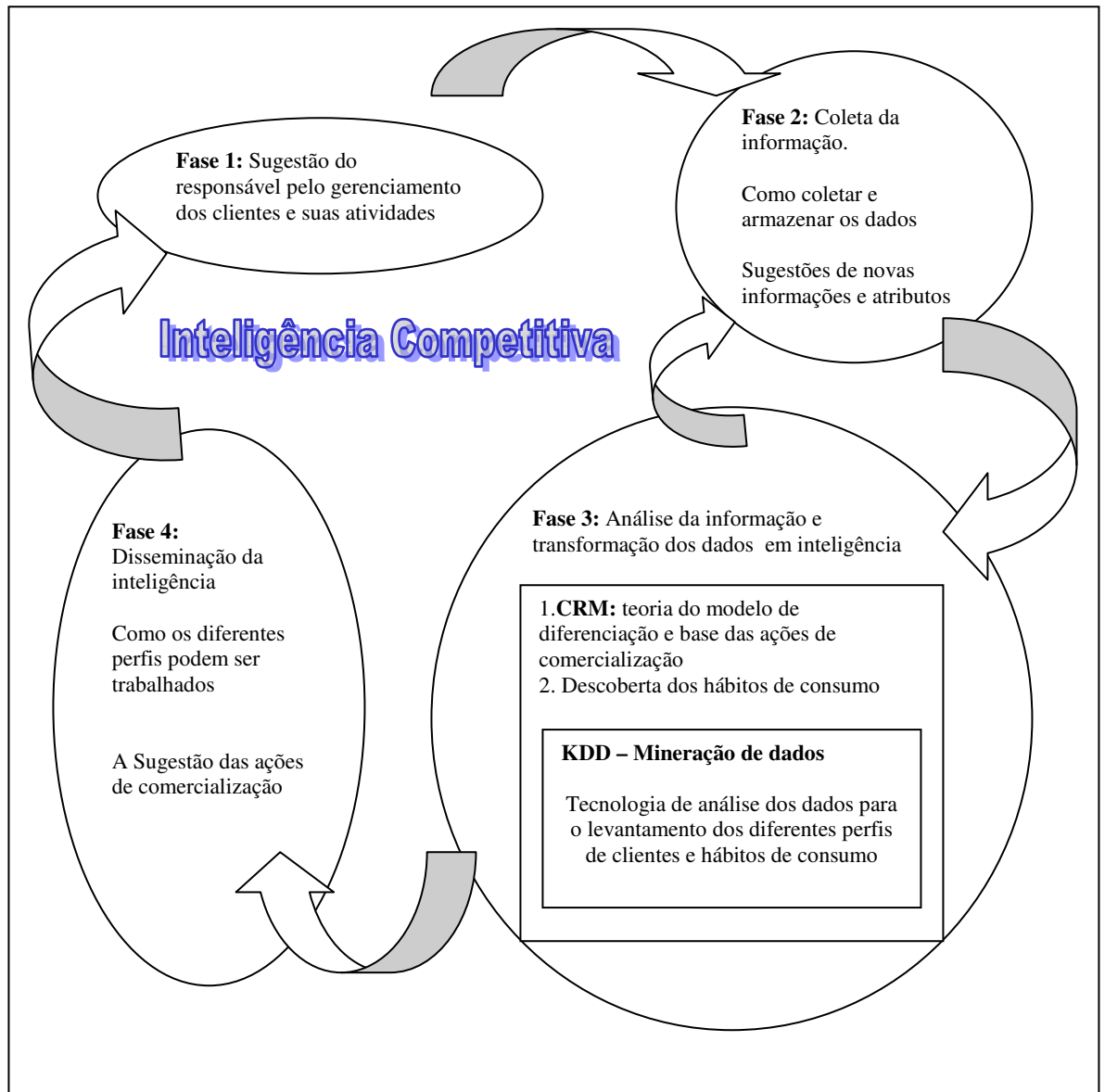
**Figura 25: Histograma da curva da receita bruta - clientes jurídicos e parceiros**

Assim, as notas para a receita bruta, em Reais, são as seguintes para os clientes jurídicos/parceiros:

- Nota 1: até R\$ 500
- Nota 2: de R\$ 501 a R\$ 2500
- Nota 3: de R\$ 2501 a R\$ 7000
- Nota 4: de R\$ 7001 a R\$ 17000
- Nota 5: acima de R\$ 17000

## 5.2. Apresentação do Modelo

A figura 26 ilustra a estrutura do Modelo de Gestão Estratégica com foco no cliente para comercialização de produtos orgânicos.



**Figura 26:** Fases do Modelo de Gestão Estratégica com foco no cliente, baseado nas fases do modelo de IC proposto por MILLER (2002)

Para melhor compreensão do funcionamento deste modelo, deve-se obedecer ao seguinte raciocínio:

Na fase 1 é sugerido um responsável para o gerenciamento dos clientes, bem como as atividades que este deverá desempenhar.

Já na fase 2, são apresentados certos cuidados que devem ser tomados no processo da coleta da informação e no levantamento dos dados para que as atividades, pertinentes à fase 1, possam ser plenamente cumpridas. Além disso, são sugeridas novas informações que podem ser acrescidas aos dados.

Esta fase possui sentido de ida e volta em relação à fase 3, pois foi com base nas dificuldades enfrentadas durante as análises dos dados que se pôde sugerir os cuidados a serem tomados.

A fase 3 envolve a análise dos dados e sua transformação em conhecimento e inteligência. Para realização desta transformação, foram utilizadas a tecnologia de Mineração de Dados e a teoria de CRM, que diz respeito a como este conhecimento pode ser empregado.

Deve-se destacar que esta teoria implica que clientes com perfis diferentes merecem e devem receber tratamentos diferentes, uma vez que também apresentam papéis estratégicos diferentes para a empresa. Este conceito foi a base para a construção do modelo de diferenciação dos clientes.

Para finalizar, a fase 4 corresponde exatamente como esses tratamentos diferenciados podem ser concretizados por meio de ações de comercialização. Esta fase encerra o ciclo, retornado a fase 1, já que com o tempo todo o processo de trabalho e análise dos dados deve ser revisto e retomado.



### **5.2.1. Fase 1: Sugestão do responsável pelas decisões e suas atividades**

Para uma boa gestão do relacionamento com cliente, este modelo sugere que haja um responsável por tal ação, ou seja, um gerente de clientes.

Caberá a este gerente monitorar os clientes por meio dos dados referentes aos registros das vendas. Para tanto, é primordial que ele conheça todos os dados que a empresa dispõe, sabendo como organizá-los.

Deste modo, é importante que seja feito um diagnóstico preliminar da situação atual do negócio em relação aos seus clientes, de maneira que se obtenha um comparativo frente à situação ideal.

Assim, pode ser possível planejar as atitudes a serem tomadas futuramente para que a situação atual possa ser modificada, alcançando a desejada.

Tome-se como exemplo o caso do Sítio Yamaguishi em que poucos trabalhos em relação aos dados de clientes foram desenvolvidos para extração de conhecimento.

O gerente de clientes deve realizar o seguinte questionamento para auxiliar seu trabalho:

#### *1. O que a Empresa possui em termos de dados?*

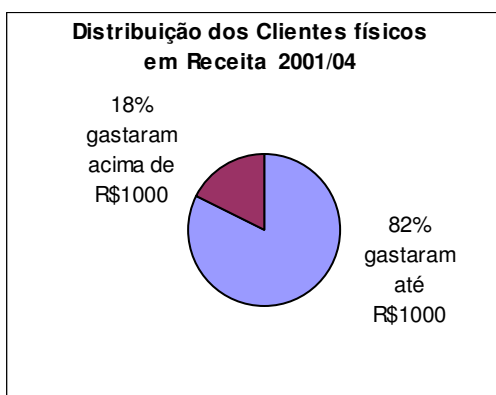
Registros das transações comerciais entre cliente e empresa de um período de três anos, além de informações cadastrais sobre os mesmos. Tais informações foram detalhadas no capítulo anterior.

#### *2. Como organizar os dados?*

Diante de inúmeras linhas de registros acumulados em três anos de armazenamento de dados, é possível extrair algum conhecimento pela simples manipulação dos dados e tabelas, com auxílio do Microsoft Access e Excel, ou qualquer outro programa de gerenciamento de banco de dados e planilhas. Por isso, é importante que o gerente de cliente tenha familiaridade para lidar com estes tipos de programas.

### 3. Qual é a situação atual?

A simples manipulação dos dados, ainda sem o recurso da Mineração de Dados, pôde revelar a seguinte situação, por exemplo, em relação aos clientes físicos, ilustrada no gráfico da figura 27.



**Figura 27: Distribuição do percentual do número de clientes físicos em relação à receita obtida entre 2001 e 2004**

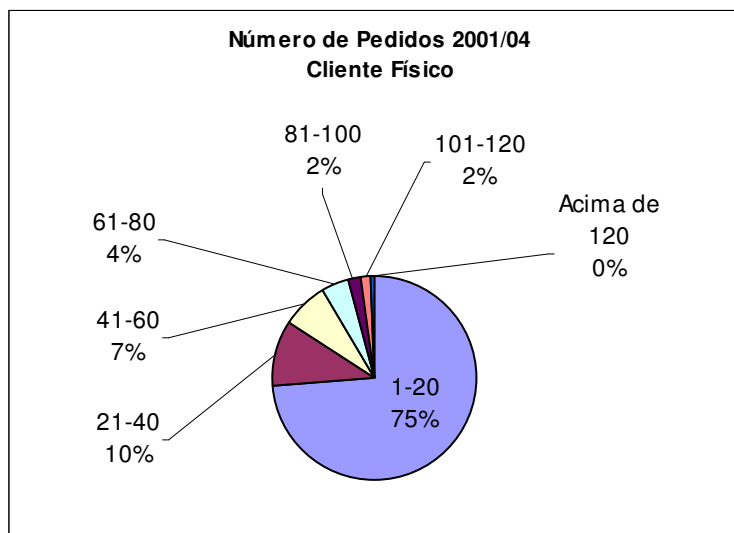
Pode-se verificar que a grande maioria (82%) gastou menos de R\$ 1.000 num período de três anos. Além disso, 54% desta maioria tiveram gastos de até R\$ 100 somente.

Dos 18% que gastaram acima de R\$ 1.000, 56% tiveram gastos de até R\$ 2.000 e somente 2% dos clientes gastaram acima de R\$ 7.000 neste mesmo período.

Soma-se a esta análise, o problema da rotatividade de clientes, ilustrado pelo fato de que grande maioria dos clientes (75%) realizou até 20 pedidos, num prazo de três anos, o que pode ser considerado um número baixo, já que os produtos em questão são perecíveis.

Este número pode ser considerado pequeno, pois é importante destacar que os pedidos podem realizados semanalmente. Ainda, pode-se dizer que somente cinco clientes (de um total de 1042) compraram quase que todas as semanas, já que efetuaram mais de 120 pedidos nos três anos.

O Gráfico da figura 28 ilustra o problema.



**Figura 28: Número e percentual de Pedidos de clientes físicos entre 2001 e 2004**

Estas análises representam o diagnóstico preliminar da situação atual do negócio, ou seja, de modo simplificado pode-se afirmar: são muitos os clientes que compram, porém também são muitos os que gastam pouco e os que deixaram de comprar.

Deste modo, o problema de negócio que precisa ser resolvido refere-se justamente à falta de estratégias mais específicas para conhecer os diferentes perfis dos clientes e seus hábitos de compra, de modo que o ciclo de rotatividade possa ser quebrado.

#### *4. Qual é situação ideal?*

Neste contexto, uma situação ideal seria clientes permanecendo mais tempo com a empresa, de modo que aumentassem o valor gasto em suas compras para, assim, quebrar a rotatividade.

No entanto, não se pode afirmar que tal diagnóstico é suficiente, pois como seria possível distinguir os bons clientes dos demais, já que o termo bom pode ser bastante relativo: “bom” em termos de valor gasto por compra, “bom” porque o cliente é bastante freqüente e compra toda semana, apesar de não gastar muito em cada compra; ou ainda, “bom” porque

apesar de comprar apenas uma vez, gastou uma enorme quantia, ou seja, o termo “bom cliente” pode se apresentar sob diferentes perfis. Assim, a próxima questão é necessária:

5. *O que é preciso para reverter a situação atual?*

Esta pergunta revela a necessidade de angariar novas informações sobre o cliente, que podem ser acrescentadas aos dados já existentes como, por exemplo: os atributos de **recência**, **freqüência acumulada** e **valor acumulado**, no caso do arquivo das transações e os atributos de **nota de número de compras**, **nota de receita** e **nota da moda da recência**, no caso do arquivo de histórico; incorporados por meio do campo data presente em cada um dos registros, além de todos os outros já descritos no capítulo de Materiais e Métodos.

Por fim, a resposta desta questão envolve exatamente a construção do modelo de diferenciação dos clientes baseado em **CRM**.

Deste modo, é possível construir um diagnóstico mais apurado da situação atual desses clientes.

### 5.2.2. Fase 2: Coleta da informação

Apesar destas informações não serem propriamente resultados, mas sim sugestão para o trabalho com os dados, é importante que sejam destacadas e colocadas nesta posição, pois correspondem à segunda fase do Modelo de Gestão Estratégica.

Esta fase propõe alguns cuidados que devem ser tomados em relação à coleta e armazenamento da informação para que o processo de gestão de clientes e o modelo de diferenciação possam ser plenamente executados.

Tais cuidados puderam ser levantados diante de algumas dificuldades enfrentadas em relação aos dados trabalhados.

Algumas informações são pertinentes a qualquer negócio, outras talvez sejam específicas ao caso do Yamaguishi, porém, de qualquer modo é importante ressaltá-las.

As informações expressas, aqui, assumem um papel de alerta em relação aos pontos fundamentais para se obter um bom resultado do trabalho com os dados.

### **5.2.2.a. Cadastro de clientes**

Um dos problemas referentes aos dados do Yamaguishi é em relação ao cadastro de clientes, sendo encontrado casos em que um mesmo cliente possuía mais de um cadastro.

Esta ocorrência se torna um problema para a atividade de Mineração de Dados, pois cada código de identificação representa um cliente diferente.

Deste modo, ao efetuar o cadastro de um mesmo cliente mais de uma vez, este recebe códigos diferentes. Assim, no momento da mineração, ele seria considerado como dois compradores diferentes.

No caso do Yamaguishi, os clientes com este problema foram identificados e passaram a ter somente um código de identificação.

Para evitar tal problema, a primeira informação a ser preenchida para um novo cadastro deveria ser o número do RG, CPF ou CNPJ, no caso de clientes jurídicos.

Este procedimento evitaria que um mesmo cliente recebesse mais de um código de identificação.

Este dado poderia ser levantado por meio de um pequeno formulário a ser preenchido pelo cliente no momento de sua primeira compra, que conteria as seguintes informações: RG, CPF ou CNPJ, nome completo, endereço completo, telefone, já que tais informações também se encontram incompletas em muitos cadastros.

Além disso, uma outra atitude bastante simples que pode ajudar na identificação dos clientes é a divisão em relação ao tipo: clientes físicos, jurídicos ou parceiros.

Por exemplo, os clientes poderiam receber a codificação de identificação de acordo com a seguinte numeração:

- Físicos: de 01 a 1999;
- Jurídicos: de 2000 a 2999;
- Parceiros: de 3000 a 3999.

Este procedimento ajudaria na identificação do tipo de cliente, pois não seria preciso recorrer ao cadastro para a verificação desta informação.

### **5.2.2.b. Novas informações**

Outras informações poderiam ser acrescentadas aos cadastros de clientes físicos, pois ajudariam a trazer novos resultados às análises de associação e classificação. Tais informações seriam o acréscimo de alguns dados demográficos, tais como:

- Cliente casado ou solteiro;
- Filhos ou não;
- Faixa etária dos filhos;
- Número de pessoas na família;
- Preferências por determinados produtos, dia de entrega;
- Vegetariano ou não, etc.

No caso de clientes jurídicos e parceiros, deveria ser acrescentado ao cadastro o tipo de estabelecimento: loja, supermercado, restaurante, etc.

Cabe a cada negócio determinar a melhor forma de angariar tais informações, além do tipo de informação que julgue pertinente para o processo de gestão de seus clientes.

### **5.2.2.c. Campos Importantes**

#### *Preços*

Os campos de preços são de fundamental importância para a determinação do atributo valor acumulado, já que é pela diferença entre o preço de venda e de compra é que se pode calcular o lucro e, assim, determinar o valor acumulado de cada cliente.

No caso de produtos produzidos no próprio sítio seria necessário levantar o custo da produção para o cálculo do lucro.

Tal procedimento não pôde ser realizado neste trabalho, por isso, o atributo valor acumulado foi baseado no campo de valor gasto e não no lucro, já que o campo do preço de compra estava incompleto na maioria dos registros, impossibilitando tal ação.

Portanto, os campos preço de compra, preço de venda, custo, além da quantidade, são campos chaves para o cálculo do lucro e posterior formulação do atributo valor acumulado.

### *Rota*

Um outro problema encontrado nos dados estudados foi em relação ao campo da rota, já que os clientes estavam cadastrados em uma rota, porém nos registros de seus pedidos apresentavam outra e, em alguns casos, mais de uma.

Deste modo, para a atividade de mineração, foi considerada não a rota do cadastro, mas sim a que constava no registro dos pedido.

Para a atividade de mineração este fato dificultou o processo de análise dos resultados.

Além disso, este problema torna inviável a realização da estimativa de quantos são os clientes que compram em cada um dos dias da semana.

Esta informação pode ser importante no processo de monitoramento, ou seja, para que se possa estimar a porcentagem de clientes que estão comprando ou que deixaram de comprar em determinado dia da semana.

Este tipo de análise poderia ajudar no controle dos resultados das ações comerciais.

### 5.2.3. Fase 3: A análise da informação e sua transformação em inteligência

#### 5.2.3.a. Construção do Modelo de diferenciação dos perfis de clientes baseado em CRM

Dos dois arquivos utilizados para as análises de classificação, o arquivo do **histórico** foi aquele que apresentou o melhor resultado para a visualização dos diferentes perfis de clientes existentes, permitindo a visualização da nota da receita, a nota de compras e a nota recência de cada cliente.

Este arquivo alcançou tal resultado, pois ao unir todas as transações de um cliente numa mesma linha foi possível analisá-lo no todo de suas interações com a empresa.

No entanto, o modelo desenvolvido por este arquivo não permitiu visualizar a qualidade destas interações, ou seja, se as notas de receita, compra e recência estavam subindo ou descendo a cada nova interação, justamente porque tais interações foram condensadas em uma única linha.

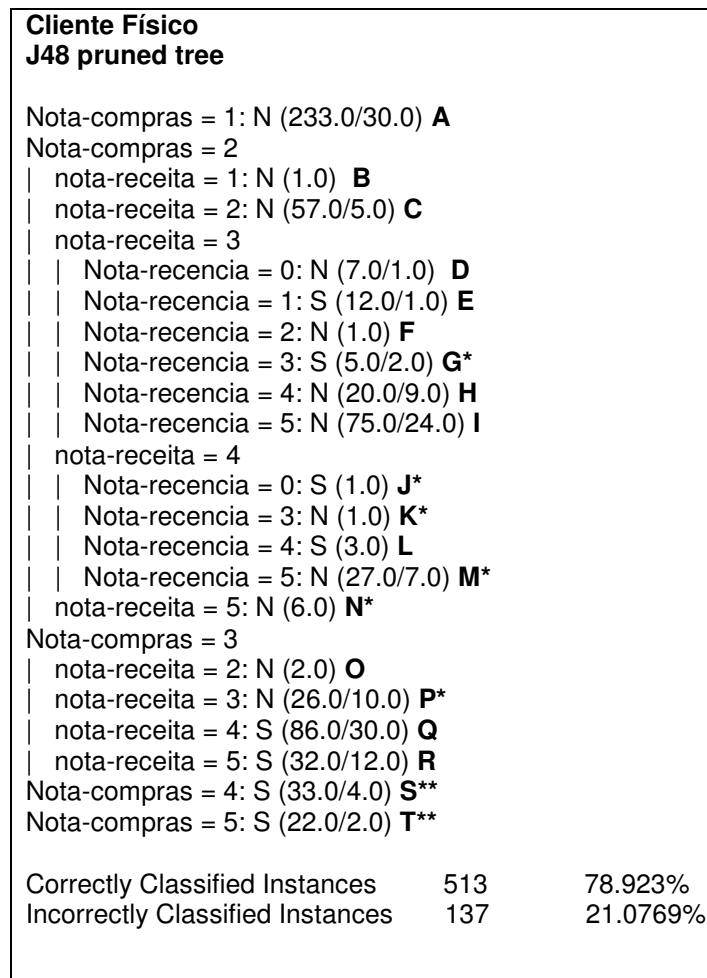
Neste contexto, encontra-se a importância da utilização do outro tipo de arquivo, o das **transações**, onde cada interação de um mesmo cliente foi analisada separadamente, já que neste arquivo cada linha representava uma compra diferente.

Apesar do modelo apresentado pelo arquivo das transações não ter obtido um bom resultado para o levantamento dos diferentes perfis dos clientes, ele apresentou regras que evidenciaram o comportamento de alguns clientes justamente em relação à qualidade de suas interações, mostrando-se um meio eficiente para a **monitoração** dos mesmos por meio de suas interações com a empresa.

Assim, primeiramente será apresentado o modelo gerado pelo software de mineração (WEKA) a partir do arquivo de **histórico**, sendo representado por meio de árvore de decisão.

A seguinte árvore de decisão representa o modelo gerado pelo software para clientes físicos (as folhas marcadas com asterisco serão usadas mais adiante para exemplificação), mostrada pela figura 29:





**Figura 29: Árvore de clientes físicos gerada pelo software de mineração**

O modelo representado acima foi construído tendo como meta de classificação o atributo cliente **Ativo**, deve-se lembrar que para um cliente ser ativo, este deveria ter efetuado compras no período entre dezembro de 2003 até o início de março de 2004 (data do registro mais recente presente no banco de dados).

O atributo nota de **Receita** também foi testado como atributo meta de classificação, porém ele não apresentou bom resultado para este banco de dados, sendo que seu modelo apresentou um grau de acerto menor que o do modelo acima. Este fato não significa que o atributo em questão não possa obter bons resultados em outros bancos de dados.

Além disso, o tipo de árvore gerada não apresentou um ganho de conhecimento tão representativo quanto ao da árvore do atributo cliente **Ativo** como meta. No entanto, o atributo

Receita também se mostra interessante para classificação, pois ele pode trazer melhores resultados para outros negócios e outros bancos de dados, diferentes do caso estudado.

Para melhor compreensão e visualização deste modelo, será apresentado o gráfico desta árvore na Figura 30, onde cada uma de suas folhas foi nomeada com uma letra diferente do alfabeto e os números entre parênteses representam o número de acertos e erros respectivamente, quando há somente um número é porque naquela folha não ocorreram erros.

Erro significa que clientes foram classificados erroneamente em classes que não correspondem a sua situação real.

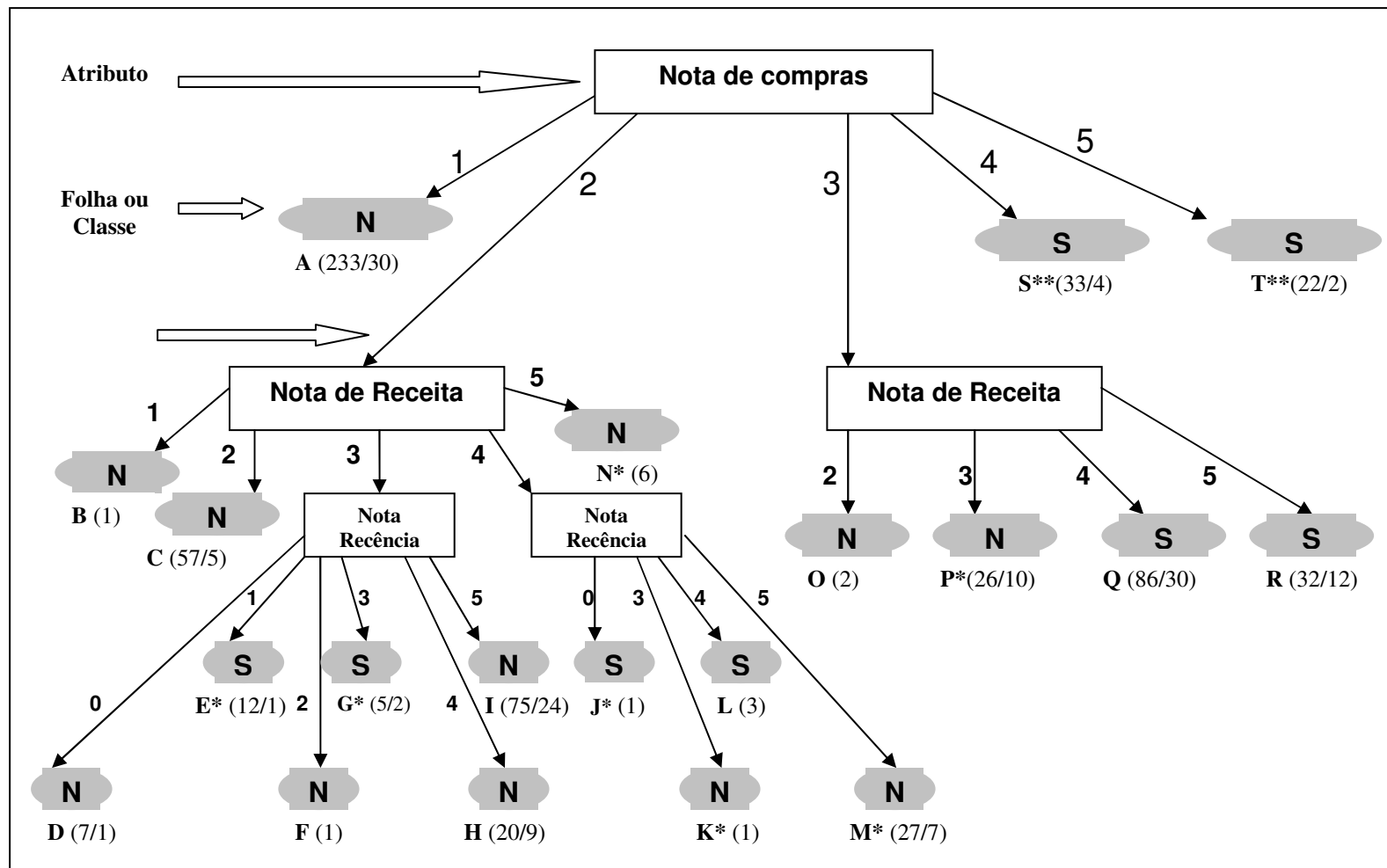


Figura 30: Representação gráfica da árvore de clientes físicos

O modelo acima representado por árvore de decisão deve ser interpretado por meio dos seguintes passos:

1. O primeiro atributo solicitado para a construção do modelo é o de **Nota de Compras** (pois foi aquele que apresentou o melhor poder de segregação dos registros), assim cada valor do mesmo (1,2,3,4,5) gera um galho diferente da árvore. Como se pôde observar, no caso dos valores de nota de compra iguais a 1, 4 e 5; apenas com este atributo, é possível estabelecer três folhas diferentes, ou seja, quando a nota de compra é igual a **1** os clientes **não serão ativos** (Folha: A), já quando a nota de compra é igual a **4** ou **5**, os clientes pertencentes a esta folha **serão ativos** (Folhas: S e T).
2. Quando o valor da nota de compra é igual a **2**, este atributo não foi suficiente para a criação das folhas, portanto um outro atributo é requerido, o de **Nota de Receita**. No caso, dos valores da nota de receita iguais a: **1, 2 e 5**, é possível estabelecer as folhas, ou seja, quando **Notas de compras = 2 e Nota de Receita igual a: 1** (Folha B), **2** (Folha C) ou **5** (Folha N), os clientes **não serão ativos**.
3. Seguindo este mesmo raciocínio, para o galho **Nota de Compra =2 e nota de Receita =3 ou 4**, ainda não é possível atingir as folhas, neste caso mais um atributo foi necessário para o estabelecimento das folhas, a **Nota de Recência**. Assim, tomando-se como exemplo **Nota de compras =2, Nota de Receita =4**, temos as seguintes Folhas para as **Notas de Recência**, quando igual a **0** ou **4**, tem-se **clientes ativos** (Folhas J e L respectivamente ), quando igual a **3** ou **5** tem-se clientes **não ativos** (Folhas K e M respectivamente).
4. O mesmo raciocínio aplica-se para o galho **Nota de compras =2 e Nota de Receita 3**, pois neste caso também houve a necessidade do uso da **Nota de Recência** para o estabelecimento das folhas.
5. No caso do valor **Nota de Compra = 3**, é possível estabelecer as folhas usando-se somente mais um atributo, a **Nota de Receita**.

Deste modo pode-se, por exemplo, levantar as seguintes afirmações:

- Se nota de compras 1, clientes não ativos (Folha A);
- Se nota de compras 2 e nota de receita 2, clientes não ativos (Folha C);
- Se nota de compras 2, nota de receita 3 e nota de recência 3, clientes ativos (Folha G);
- Se nota de compras 3, nota de receita 4, clientes ativos (Folha Q);
- Se nota de compras 4, clientes ativos (Folha S).

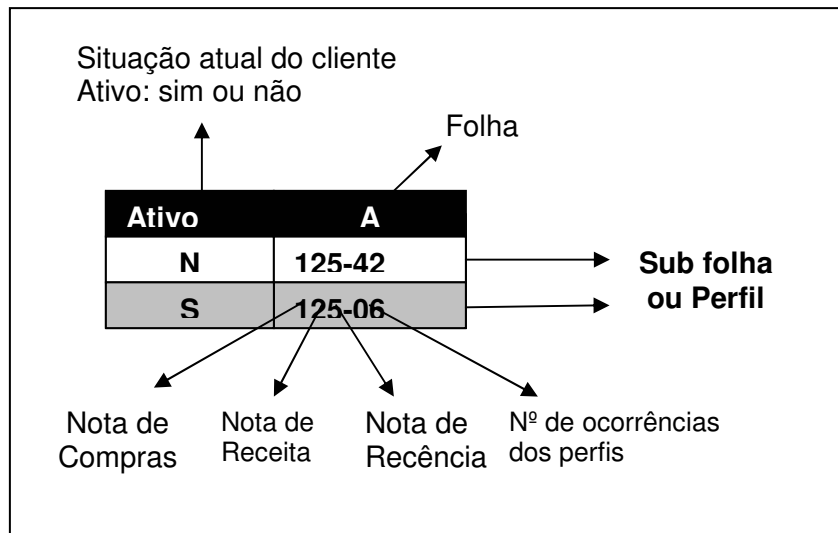
Foi com base neste modelo, representado por árvore de decisão, que se criou o modelo final para diferenciação de clientes baseado em CRM.

Na verdade, este modelo de diferenciação é um detalhamento, por meio de checagem dos dados, das folhas já estabelecidas pela árvore.

Por exemplo, no caso da Folha A, onde os clientes com valor de nota de compra 1 são classificados como não ativos, os valores de nota de receita e de recência, respectivamente, não são mostrados pela árvore, pois somente o valor nota de compras = 1 foi suficientes para classificá-los como não ativos.

No entanto, tais informações (as notas referentes à receita e recência) são informações importantes que devem ser esclarecidas, portanto, como forma de completar este modelo, os dados foram checados para o estabelecimento de sub folhas, onde é possível diferenciar os clientes tendo em consideração todos os outros atributos.

A figura 31 ilustra como este modelo deve ser interpretado.



**Figura 31: Modelo para interpretação da tabela dos perfis de cliente**

As colunas representam as folhas estabelecidas pela árvore de decisão, o primeiro número refere-se à nota de compra, o segundo à nota de receita e o terceiro à nota de recência, já que esta ordem de atributos foi estabelecida pelo modelo gerado através da mineração dos dados, as linhas representam as sub folhas pertencentes a cada uma das folhas.

Este modelo permite que todos os perfis de clientes sejam visualizados em termos de notas compra, receita e recência, inclusive os casos que foram classificados erroneamente pela árvore, já que as sub folhas foram divididas entre não ativo (N) e entre ativo (S).

A tabela 11 ilustra o modelo de diferenciação dos clientes, evidenciando todos os perfis existentes dentro de uma mesma folha.

*Observação:*

Os perfis marcados com asterisco serão usados mais adiante para exemplificação.

**Tabela 11: Modelo de diferenciação dos clientes – Perfis dos clientes físicos**

<b>Ativo</b>	<b>A*</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G*</b>
<b>N</b>	110-49	215-01	220-05	230-06	231-01	232-01	233-02
<b>S</b>	110-01		220-02	230-01	231-11		233-03*
<b>N</b>	113-02		221-03				
<b>S</b>							
<b>N</b>	114-01		222-01				
<b>S</b>			222-01				
<b>N</b>	115-06		223-02				
<b>S</b>			223-02				
<b>N</b>	120-70		224-12				
<b>S</b>	120-13						
<b>N</b>	121-02		225-29				
<b>S</b>	121-01						
<b>N</b>	123-04						
<b>S</b>	123-01						
<b>N</b>	124-10						
<b>S</b>							
<b>N</b>	125-42						
<b>S</b>	125-06*						
<b>N</b>	130-06						
<b>S</b>							
<b>N</b>							
<b>S</b>	131-01						
<b>N</b>	134-02						
<b>S</b>	134-03*						
<b>N</b>	135-08						
<b>S</b>	135-04*						
<b>N</b>							
<b>S</b>	143-01*						

**Continuação da tabela 11**

<b>Ativo</b>	<b>H</b>	<b>I</b>	<b>J*</b>	<b>K*</b>	<b>L</b>	<b>M*</b>	<b>N*</b>
<b>N</b>	234-11	235-51		243-01		245-20*	255-06
<b>S</b>	234-09*	235-24*	240-01		244-03	245-07*	

**Continuação da tabela 11**

<b>Ativo</b>	<b>O</b>	<b>P*</b>	<b>Q</b>	<b>R</b>	<b>S**</b>	<b>T**</b>
<b>N</b>	325-02	334-03	344-03*	354-01*		525-01*
<b>S</b>		334-02*	344-07*	354-03*	435-01	
<b>N</b>		335-13	345-27*	355-11*	445-02*	545-01*
<b>S</b>		335-08*	345-49*	355-17*	445-14*	545-03*
<b>N</b>					455-02*	555-01**
<b>S</b>					455-14*	555-16**

O mesmo procedimento de construção do modelo de diferenciação de clientes aplica-se também para os clientes do tipo jurídico e parceiro, porém, este apresentou uma estrutura mais simplificada que a dos clientes físicos, pois somente com a utilização do atributo nota de compras foi possível estabelecer as folhas.

Além disso, o arquivo do histórico dos clientes jurídicos e parceiros foi formulado com os mesmos atributos presentes no arquivo de clientes jurídicos, lembrando que uma única diferença dá-se em relação à **nota da moda da recência**, que neste caso varia somente de **1 a 3**, como já destacado no capítulo anterior.

A árvore apresentada pelo software de mineração é ilustrada na figura 32 e sua representação gráfica na figura 33.

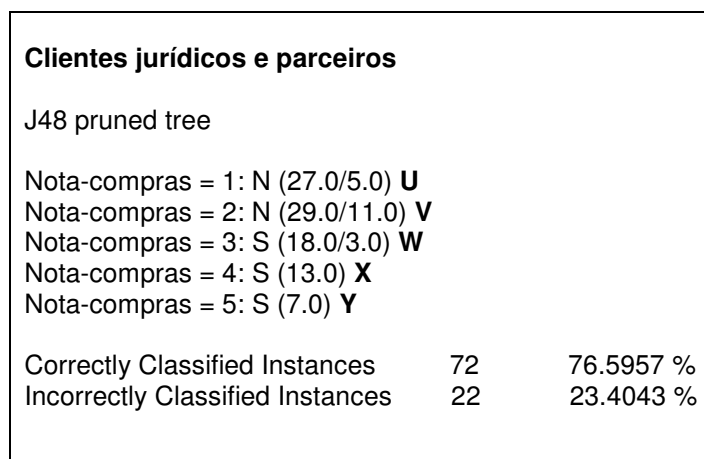


Figura 32: Árvore de clientes jurídicos e parceiros gerada pelo software de mineração

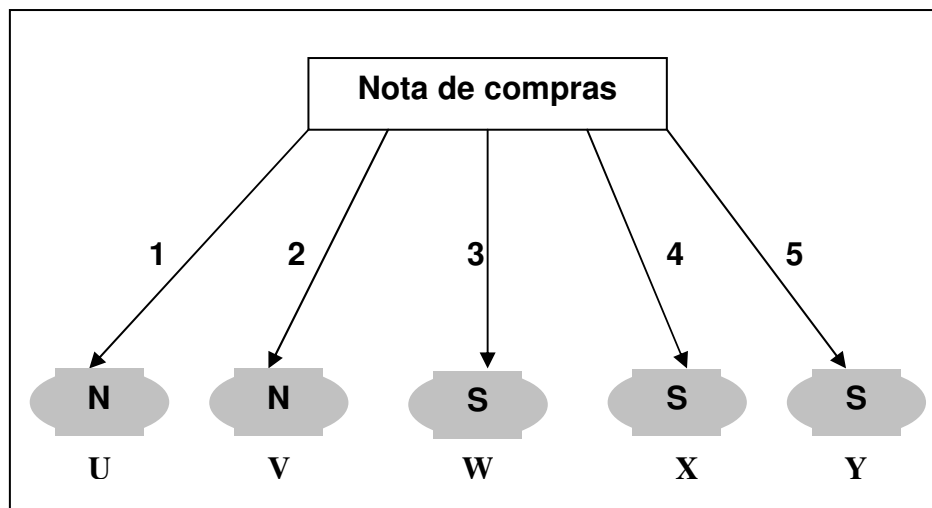


Figura 33: Representação gráfica da árvore dos clientes jurídicos e parceiros



A tabela 12 apresenta o modelo de diferenciação de clientes do tipo jurídico e parceiro, seguindo o mesmo tipo de análise efetuada para os clientes físicos:

**Tabela 12: Modelo de diferenciação dos clientes – Perfis dos clientes jurídicos e parceiros**

Ativo	U	V	W	X	Y
N	110-11	210-01			
S	110-03*		321-01	432-04	543-02**
N	112-03	211-03	322-01		
S	112-01*	211-01*	322-04	433-01*	553-05**
N	120-04	212-02	332-01		
S		212-02*	332-05	442-05*	
N	121-01	221-04			
S		221-01*	333-02	443-01*	
N	122-01	222-06			
S	122-01*	222-05*	342-02*	452-02**	
N	130-01	232-01			
S			343-01*		
N	131-01	233-01	352-01*		
S					
N					
S		242-02*			

### *Interpretação do Modelo (tabelas) de diferenciação e identificação dos perfis*

Antes que qualquer análise seja feita, é importante retomar uma das perguntas levantadas na fase 1 deste modelo de gestão estratégica, sendo a seguinte:

- **Quem são os bons clientes?**

Neste ponto, é importante lembrar a relatividade do termo “bom cliente” e, por isso, para a identificação dos bons clientes, este modelo propõe que eles sejam analisados e caracterizados sob três aspectos diferentes: Nota de **Compras**, Nota de **Receita** e de **Recência**.

Deste modo, é possível que se obtenha uma visão mais detalhada dos clientes, de modo que as estratégias de comercialização sejam adequadas aos perfis encontrados.

Assim, deve-se responder a seguinte questão:

“Cliente *bom* é aquele que compra freqüentemente e gasta pouco ou é aquele que compra poucas vezes, mas gasta muito?”

A resposta desta questão não é simplesmente rotular determinado cliente como bom ou ruim, mas sim compreender que os dois tipos podem ser considerados bons e importantes para o negócio.

Acima de tudo, é importante entender que existem perfis diferentes e, portanto, requerem tratamentos e estratégias diferentes, sendo este o sentido da construção do modelo de diferenciação dos clientes baseado em CRM.

### *Cientes Físicos*

Neste contexto, a tabela 11 evidencia que, mesmo dentro de uma mesma folha, podem existir diferentes perfis como, por exemplo, é o caso da Folha A, que pela árvore de decisão foi classificada como clientes não ativos com valor de nota de compra igual a 1.

O primeiro exemplo desta folha a ser analisado é o perfil **143** (última linha da folha A, marcado com asterisco). Este perfil possui nota de compra 1, porém encontra-se ativo e possui uma nota de receita alta, ou seja, representa um cliente que apesar de ter comprado poucas vezes, o valor gasto em suas compras foi alto. Além disso, sua recência apresenta uma nota intermediária.

Apesar deste perfil apresentar somente a ocorrência de um caso, este merece destaque, pois representa um cliente potencial que pode vir a ser mais freqüente. Assim, ações de **ativação** podem ser tomadas para elevar sua nota de recência.

Há outros perfis desta folha que também chamam a atenção já que estão ativos, como os demais marcados com asterisco (**125, 134, 135**). Apesar de apresentarem notas de compras razoáveis, suas notas de recência são altas, ou seja, podem ser clientes relativamente novos e, por isso, sua nota de compras ainda é baixa. No entanto, estão mantendo uma certa regularidade no período de suas compras (devido à alta nota de recência).

Assim, no caso destes clientes podem ser tomadas ações que os estimulem a elevar sua nota de receita, como, por exemplo, estratégias de venda cruzada.

Este mesmo raciocínio pode ser aplicado para os casos de clientes da Folha H e I, **234** e **235** respectivamente, pois a única diferença dá-se em relação à nota de compra, pois

provavelmente estes clientes são mais antigos, ou então, simplesmente compraram mais que os outros mencionados no mesmo período de tempo.

Há outro caso, como o da Folha G, **233**, que também se encontra ativo e pode ser trabalhado para melhorar sua nota de receita e recência, tendo em vista suas notas razoáveis para estes critérios.

Também, alguns casos da Folha P (**334** e **335**), apesar de possuírem um perfil um pouco mais alto com relação ao número de compras e recência, podem melhorar sua receita.

Num outro extremo da tabela, estão os clientes de perfil **555**, Folha T, ou muito próximos deste perfil: **545**, **455**, sendo este último pertencente à Folha S.

Todos estes clientes estão num patamar elevadíssimo em quase todos os aspectos, ou todos. Portanto, são clientes *ouro*, que devem ser tratados e monitorados pela empresa com extremo cuidado e zelo, devendo direcionar a estes estratégias de **fidelização**.

Existe, também, um outro caso que merece grande destaque, pois se trata de um perfil **555 não ativo**, ou seja, é um cliente *ouro* que a empresa perdeu. Para estes casos, deve haver estratégias de **recuperação**.

Outros casos, que também devem ser mencionados são perfis como: **344**, **354** (Folha Q) e **354**, **355** (Folha R), já que possuem altas notas de receita e recência.

Esta breve explicação visa ilustrar, por meio de alguns exemplos, como as estratégias podem ser tomadas de acordo com os perfis descobertos, sendo que tais ações serão explicadas detalhadamente na fase 4 do modelo. Como nem todos os perfis foram mencionados a título de exemplo, isto não significa que sejam menos importantes.

Deve-se ressaltar, que o planejamento das ações varia de acordo com cada negócio, uma vez que depende do levantamento de suas prioridades e dos perfis de clientes que deseje trabalhar.

Este trabalho assume o papel de somente exemplificar algumas das possíveis estratégias e ações que podem ser desenvolvidas baseadas no conhecimento descoberto com este tipo de análise.

### *Cientes Jurídicos e Parceiros*

Seguindo o mesmo tipo de análise realizada para os clientes físicos, pode-se chegar às seguintes conclusões por meio do modelo de diferenciação:

Na folha U, existem três perfis ativos de cliente: **110, 112, 122**. De modo geral, são clientes com notas baixas em quase todos os aspectos de classificação, porém pelo fato de estarem ativos, ações de ativação e *cross selling* podem ser realizadas como forma de aumentar sua nota de recência e receita respectivamente para que possam ser tornar clientes mais ativos e com maior geração de receita.

Raciocínio semelhante também se aplica aos casos **211, 212, 221, 222** da Folha V, pois se encontram ativos, porém com baixas notas de receita e recência, tendo como única diferença maior nota de compras.

Um outro caso desta folha que merece destaque refere-se ao **242** (Folha V), pois possui uma ótima nota de receita, porém notas baixas de compras e razoável de recência. Este é o tipo de cliente que compra pouco, porém gasta muito, devido à alta receita e a baixa nota de compras. Deste modo, seria interessante torná-lo mais freqüente por meio de ação de ativação.

Em relação à Folha W, no caso **342** sua recência poderia ser melhorada, já que possui uma boa nota de receita. Já para o caso **352**, como não está mais ativo, aplica-se estratégia de recuperação. O caso **343** é um cliente potencial para se tornar fiel, pois apesar da razoável nota de compra e recência, possui boa nota de receita.

Na Folha X, os casos **442, 433, 443** são clientes potenciais a alcançar o patamar máximo (553), portanto os clientes pertencentes a estes perfis devem ser estimulados a comprar mais vezes (apenas no caso do 442) e elevar suas receitas (433,443).

Já **452**, também da Folha X, merece ser fidelizado, além de realizar esforços para melhorar sua nota de recência. Outros casos para fidelização são os clientes da Folha Y, de perfil **543 e 553**.

### 5.2.3.b. Monitoramento dos Clientes

A outra análise de classificação foi realizada com a utilização do arquivo das **transações**.

Diferentemente do arquivo de histórico, o qual apresentou um bom resultado para a diferenciação dos vários perfis de clientes existentes, as análises realizadas com este arquivo não apresentaram um bom resultado para se alcançar este objetivo.

Tal resultado deve-se ao fato de que ao se analisar cada registro de compra de um mesmo cliente separadamente, as regras geradas forneceram um conhecimento fragmentado do comportamento do cliente. Assim, foi bastante comum casos nos quais um mesmo cliente foi classificado por mais de uma regra, o que não é interessante quando se deseja estabelecer diferentes perfis.

No entanto, apesar deste tipo de análise falhar na tarefa de diferenciação, ela se mostrou uma ferramenta interessante para **monitoramento** dos clientes e para a **avaliação da qualidade** de suas interações, uma vez que suas regras funcionaram como uma alerta dos comportamentos que mais se repetiram.

Além disso, é importante ressaltar que as análises dos clientes físicos tiveram como atributo meta de classificação o atributo **cliente ativo**, já que o resultado desta análise apresentou a maior porcentagem de acerto (aproximadamente 66%).

No caso das análises dos clientes jurídicos e parceiros foram utilizados dois atributos meta de classificação diferentes: o valor acumulado e cliente ativo.

Outro fato importante é que as análises foram realizadas utilizando os atributos Cidade e Rota separadamente. Para os casos em estudo, as análises com o atributo Rota não foram bem sucedidas, pois há o problema de que os clientes são atendidos por mais de uma rota, dificultando a análise dos resultados.

Assim, as análises com o atributo Cidade mostraram-se mais adequadas, de modo que o resultado a ser apresentado refere-se somente às análises realizadas com o atributo Cidade.

## *Cientes Físicos*

No caso deste arquivo, o modo para representação do conhecimento foi feito por meio de regras de associação. Na tabela 13 são apresentadas algumas consideradas as mais interessantes, devido ao seu número de casos ou pela não ocorrência de erros, as demais se encontram no capítulo de Anexos.

**Tabela 13: Regras dos clientes físicos geradas por meio de análise de associação**

<p>frequencia-nota = 5 AND cidade = 30: S (148.0/21.0)</p> <p>valor-acumulado-nota = 4 AND cidade = 16 AND frequencia-nota = 5: S (147.0)</p> <p>valor-acumulado-nota = 4 AND cidade = 31: S (86.0/4.0)</p> <p>valor-acumulado-nota = 3 AND cidade = 13 AND frequencia-nota = 4: S (79.0/10.0)</p> <p>valor-acumulado-nota = 4 AND cidade = 14 AND frequencia-nota = 5: S (71.0/6.0)</p> <p>valor-acumulado-nota = 3 AND cidade = 16 AND frequencia-nota = 4: S (79.0/7.0)</p> <p>valor-acumulado-nota = 4 AND cidade = 9: S (1387.0/403.0)</p> <p>valor-acumulado-nota = 3 AND cidade = 18 AND frequencia-nota = 4: N (15.0)</p> <p>valor-acumulado-nota = 3 AND cidade = 9 AND frequencia-nota = 3: S (805.0/269.0)</p>	<p>valor-acumulado-nota = 3 AND cidade = 9 AND frequencia-nota = 4: S (718.0/150.0)</p> <p><b>valor-acumulado-nota = 4 AND cidade = 30: S (180.0/60.0)</b></p> <p>valor-acumulado-nota = 2 AND cidade = 14 AND frequencia-nota = 4: S (35.0)</p> <p>valor-acumulado-nota = 2 AND cidade = 13 AND frequencia-nota = 3: S (104.0/33.0)</p> <p>frequencia-nota = 5 AND cidade = 18 AND valor-acumulado-nota = 2: S (28.0/2.0)</p> <p>cidade = 14 AND valor-acumulado-nota = 3 AND frequencia-nota = 4: S (19.0/3.0)</p> <p>cidade = 16 AND valor-acumulado-nota = 4 AND frequencia-nota = 4: S (49.0/3.0)</p>	<p><b>valor-acumulado-nota = 5 AND cidade = 31 AND frequencia-nota = 3: N (17.0)</b></p> <p>frequencia-nota = 3 AND recencia-nota = 5 AND valor-acumulado-nota = 2 AND cidade = 30: S (76.0/16.0)</p> <p>frequencia-nota = 2 AND valor-acumulado-nota = 4 AND cidade = 13: N (13.0/3.0)</p> <p>frequencia-nota = 2 AND valor-acumulado-nota = 3 AND cidade = 16 AND recencia-nota = 4: N (30.0/13.0)</p> <p>cidade = 31 AND valor-acumulado-nota = 5: S (7.0)</p> <p>cidade = 14 AND valor-acumulado-nota = 4: N (6.0)</p> <p>cidade = 16 AND valor-acumulado-nota = 2 AND frequencia-nota = 3: S (148.0/43.0)</p>
---	--	--

Com base nestas regras, a tabela 14 foi montada e deve ser interpretada de acordo com a seguinte legenda:

- A: Cliente Ativo
- N: Cliente Não Ativo
- F: Frequência acumulada
- R: Recência

Os números após a recência e a frequência correspondem às suas notas e os números entre parênteses correspondem ao número de ocorrência da regra e de erros respectivamente.

**Tabela 14: Radiografia das Cidades - clientes físicos**

Nota Valor/Cidade	Campinas - 9	Itatiba - 13	Jaguariúna - 14
1			
2		AF3(104/33)	AF4(35)
3	AF3(805/269)/AF4(718/150)	AF4(79/10)	AF4(19/3)
4	A(1387/403)	NF2(13/3)	AF5(71/6)/N(6)
5			

**Continuação da tabela 14**

Nota Valor/Cidade	Jundiaí - 16	Mogi-Guaçu - 17	Mogi-Mirim - 18
1			
2	AF3(148/43)	A(20)	AF5(28/2)
3	AF4(79/7)/AF3(74/3)/NF2R4(30/13)	A(17)	NF4(15)
4	AF5 (147)/AF4(49/3)		
5			

**Continuação da tabela 14**

Nota Valor/Cidade	Valinhos - 30	Vinhedo - 31
1		
2	AF3R5(76/16)	
3		
4	A(180/60)	A(86/4)
5		NF3(17)/A(7)

Este formato de expressar os resultados revela uma radiografia das cidades em relação ao comportamento dos clientes, qualificados pelas notas de valor acumulado, frequência acumulada, recência, além da situação ativa ou não.

Tenham-se como exemplos os casos destacados em negrito das cidades de Valinhos e Vinhedo para compreensão de como esta tabela pode servir para a interpretação do conhecimento descoberto.

No caso **A(180/60)** de Valinhos, com valor acumulado nota 4, ao checar os registros dos dados, vê-se que esta regra corresponde apenas ao cliente de código 986 o qual apresenta um perfil **355**.

Por meio de uma checagem mais detalhada dos dados no arquivo das transações, é revelado que este cliente, em suas últimas interações (ou seja, nas últimas compras), suas notas de valor e frequência acumulada e da recência estão em queda, o que significa que ele está comprando e gastando cada vez menos.

Deste modo, este fato evidencia que é importante haver monitoramento dos clientes, pois se este novo padrão de comportamento tornar a se repetir por mais vezes, a empresa corre o risco de perder mais um cliente.

Outro caso que comprova a utilidade deste tipo de análise é o caso **NF3(17)** de Vinhedo. Neste caso, a regra também representa somente um cliente que possui perfil **355**.

Este cliente apresentava ótimas notas de valor acumulado e recência, porém com o passar do tempo e, principalmente, em suas últimas interações estas notas foram decrescendo até que deixasse de comprar, já que não está mais ativo.

Estes exemplos ilustram como o conhecimento descoberto pode ser trabalhado, de modo que a tabela juntamente com a checagem dos dados sirvam de alerta, já que podem ajudar a visualizar a urgência de ações estratégicas de comercialização em relação a determinados clientes.

Obs.: os registros das últimas compras dos dois clientes mencionados estão presentes no capítulo de Anexos.



### *Cientes Jurídicos e Parceiros*

O mesmo procedimento foi realizado para as análises dos clientes jurídicos e parceiros.

No entanto, neste caso houve necessidade de se realizar duas análises com atributos metas diferentes para a classificação, uma vez que o resultado obtido com o atributo cliente ativo como meta não forneceu, na maioria das regras, detalhes de valor acumulado, frequência ou recência. Alguns exemplos destas regras são apresentados na tabela 15.

**Tabela 15: Regras clientes jurídicos e parceiros - meta: cliente ativo**

cidade = 9: S (4185.0/216.0)
cidade = 14: S (778.0/68.0)
cidade = 30: S (240.0/20.0)
cidade = 31: S (167.0)
cidade = 12: S (147.0)
cidade = 8: N (140.0)
cidade = 10: S (82.0)

Deste modo, com base nos resultados apresentados por esta análise (atributo meta: cliente ativo) a tabela 16 foi construída.

**Tabela 16: Relação das Cidades e seus respectivos clientes em relação à situação Ativo/Não-ativo, tendo como atributo-meta: cliente ativo**

<b>Cidade</b>	<b>Ativo</b>	<b>Não Ativo</b>
Campinas	4185/216	
Jaguariúna	778/68	
Valinhos	240/20	
Vinhedo	167	
Holambra	147	
Cajamar		140
Casa Branca	82	
Americana + Valor acumul.=2	73/2	
Jundiaí + Valor acumul.=1		107
Socorro	69	
Americana + Valor acumul.=1	66/2	
Monte Alegre do Sul		43
Jundiaí + Valor acumul.=2		86/33
Jundiaí	39/1	
São José dos Campos	36/7	
Rio de Janeiro		35
Mogi Mirim		24
Americana		19
Itatiba		15/2
Santo Antônio de Posse	14	
Amparo	13/5	
Piracicaba		9
Petrópolis		8
Guarulhos	7	

Assim, por meio da tabela 16, foi possível verificar as cidades e seus respectivos números de clientes ativos ou não ativos, lembrando que o primeiro número representa o número de casos cobertos por determinada regra e o segundo representa o número de erros.

A outra análise foi realizada com o atributo valor acumulado, tido como meta de classificação. Seu resultado apresentou informações mais detalhadas sobre a frequência e a recência, porém nem todas as regras apresentaram informações sobre o estado ativo ou não ativo dos clientes.

Deste modo, a tabela 16 foi utilizada para complementar as regras da tabela 17 nas quais esta informação não era fornecida. Porém, este procedimento foi somente utilizado para cidades, da tabela 16, em que as regras não apresentaram erros (como por exemplo: Mogi-Mirim, Rio de Janeiro, Piracicaba), pois caso contrário não seria possível ter feito tal dedução.

A tabela 17 apresenta o conhecimento descoberto, tendo o atributo valor acumulado como meta, por meio das regras consideradas mais interessantes, devido ao seu número de casos ou, então, pela pouca ou não ocorrência de erros. As demais se encontram no capítulo de Anexos.

**Tabela 17: Regras clientes jurídicos e parceiros - meta : valor acumulado**

<b>cidade = 8: 5 (140.0/3.0)</b> cidade = 12: 2 (101.0)  cidade = 29 AND recencia-nota = 3: 4 (46.0)  frequencia-nota = 1 AND cidade = 14 AND ativo = N: 1 (43.0/5.0)  frequencia-nota = 1 AND cidade = 13 AND ativo = S: 1 (35.0/2.0)  cidade = 14 AND frequencia-nota = 2 AND dia-semana = 2: 1 (169.0/39.0)  frequencia-nota = 1 AND cidade = 16 AND ativo = N: 1 (72.0/13.0)	frequencia-nota = 1 AND cidade = 9 AND ativo = N: 1 (59.0)  cidade = 2 AND ativo = S AND frequencia-nota = 2: 2 (80.0/36.0)  cidade = 16 AND ativo = S: 4 (19.0/1.0)  cidade = 31 AND frequencia-nota = 2: 4 (103.0/20.0)  dia-semana = 2 AND cidade = 9 AND frequencia-nota = 3: 4 (118.0/33.0)  <b>dia-semana = 3 AND</b> <b>cidade = 26 AND</b> <b>frequencia-nota = 2: 5 (15.0/1.0)</b>
--	---

Com base na tabela 16 e nas regras da tabela 17, a tabela 18 foi construída:

**Tabela 18: Radiografia das Cidades - clientes jurídicos e parceiros**

Nota Valor/Cidade	Americana-2	Cajamar-8	Campinas-9	Holambra-12	Itatiba-13
1			NF1(59)		AF1(35/2)
2	AF2(80/36)			A (101)	
3					
4	F3(9)		F4(118/33)		
5		<b>N (140/3)</b>			

**Continuação da tabela 18**

Nota Valor/Cidade	Jaguariúna-14	Jundiaí-16	Mogi-Mirim-18	M.AlegreS.-19	Petrópolis-22
1		NF1(72/13)	NF1(11/3)	NF1(28/2)	
2	NF1(43/50)				
3					
4		A(19/1)			NF1(8/2)
5					

**Continuação da tabela 18**

Nota Valor/Cidade	R. Janeiro-24	SJ dos Campos -26	Socorro-29	Valinhos-30	Vinhedo-31
1	NF1(33/7)				
2					
3				F2(129/33)	
4			AR4(46)		AF2(103/20)
5		<b>AF2(15/1)</b>			

A mesma legenda deve ser utilizada para interpretação da tabela:

- A: Cliente Ativo
- N: Cliente Não Ativo
- F: Frequência acumulada
- R: Recência

Dois casos serão analisados como forma de exemplificação.

O primeiro, **N (140/3)**, refere-se à cidade de Cajamar, representa somente um cliente **não ativo** e de **valor acumulado 5**, de perfil **352**.

Em relação a este caso, era necessário ter havido ações para que este cliente se mantivesse ativo, devido ao seu ótimo perfil.

Assim, ao checar seus dados nos arquivos de **transações**, é possível notar por meio dos seus registros que, apesar da nota do valor acumulado ter-se mantido estável até o final, nota-se também quedas sucessivas nas notas da frequência acumulada e da recência em suas últimas compras. Este tipo de comportamento, caso tivesse sido monitorado, poderia ter sido um indício da perda deste cliente num futuro próximo.

Deste modo, uma observação mais cuidadosa deste comportamento evitaria que o cliente fosse perdido, sendo que este apresentava um grande potencial para se tornar perfil 553, caso as estratégias necessárias tivessem sido tomadas para que, assim, suas notas de recência e frequência acumulada fossem elevadas.

O segundo caso **AF2(15/1)**, referente à cidade de São José dos Campos, apresenta uma regra cliente ativo, de frequência acumulada 2 e de valor acumulado 5.

Neste caso, esta regra também representa somente um cliente, o qual possui um perfil **242**. Realizando o mesmo procedimento de checagem dos dados, neste exemplo também é possível observar quedas sucessivas em dois dos atributos: valor acumulado e frequência acumulada.

Assim, ao monitorar tal comportamento pode ser possível realizar ações antes que se perca mais um cliente de grande potencial, devido à alta nota de receita presente em seu perfil (242).

Obs.: os registros das últimas compras dos dois clientes mencionados estão presentes no capítulo de Anexos.

### **5.2.3.c. Descoberta dos hábitos de consumo**

As análises de associação foram utilizadas para descobrir os hábitos de consumo existentes, ou seja, para esclarecer quais produtos são comprados juntamente com outros.

Deve-se retomar, que a motivação de ter realizado este tipo de análise era, além de verificar quais produtos tem suas vendas associadas às de outros, foi também revelar a influência dos atributos da dimensão tempo no consumo. Por exemplo, revelar se a ocorrência de feriados traria alguma mudança nos hábitos de consumo ou se algum dia da semana, período do mês, estação do ano possuía alguma característica especial de consumo.

Além disso, esperava-se checar se havia a ocorrência de alguma característica particular de consumo em relação a alguma cidade ou rota em especial.

Porém, no caso estudado foi predominante a venda de seis principais produtos: alface (principalmente), ovos, cenoura, tomate, laranja e banana. Deste modo, estes itens tiveram grande influência no processo de criação das regras de associação, já que estavam presentes em quase todos os registros.

Este fato acabou por revelar-se um problema, pois na maioria das regras somente estes produtos estavam presentes, principalmente do lado conseqüente da regra, ou seja, não constituindo descoberta de conhecimento, além de não possibilitar tomada de decisão no caso do Yamaguishi. No entanto, isto não significa que este tipo de análise não possa apresentar resultados melhores para outros negócios.

Outro ponto que merece destaque é que as regras, que não apresentaram tais itens em seu lado conseqüente, possuíam um grau de confiança bastante baixo o que impossibilitou qualquer sugestão de ação.

Tal fato ocorreu tanto para as análises dos clientes físicos como para dos jurídicos e parceiros.

Para ilustração do problema, seguem na tabela 19 algumas destas regras gerados pelo arquivo geral de clientes físicos. As demais se encontram no capítulo de Anexos.

**Tabela 19: Regras geradas a partir de análise de associação por meio do arquivo geral de clientes físicos**

<b>SE</b> CENOURA CHICORIA OVOS-12-unid 1075 ==> <b>ENTÃO</b> ALFACE 988 conf:(0.92) < lift:(1.84)>
<b>SE</b> ALFACE CENOURA LIMA 1346 ==> <b>ENTÃO</b> LARANJAS 1008 conf:(0.75) < lift:(2.26)>
<b>SE</b> ALFACE LIMA OVOS-12-unid 1308 ==> <b>ENTÃO</b> BANANA 966 conf:(0.74) < lift:(1.96)>
<b>SE</b> ALFACE LARANJAS OVOS-12-unid temperos 1380 ==> <b>ENTÃO</b> CENOURA=37 1055 conf:(0.76) < lift:(2.11)>
<b>SE</b> ALFACE BATATINHA CENOURA 1539 ==> <b>ENTÃO</b> TOMATE 1109 conf:(0.72) < lift:(2.2)>
<b>SE</b> CENOURA LARANJAS temperos 1667 ==> <b>ENTÃO</b> ALFACE OVOS-12-unid 1055 conf:(0.63) < lift:(2.11)>

Estas regras devem ser interpretadas da seguinte maneira, tendo-se como exemplo a primeira regra da tabela, **SE** houver a venda de cenoura, chicória e ovos **Então**, provavelmente ocorrerá a venda de Alface. No caso, os números representam o seguinte: das 1075 vezes que estes três produtos foram vendidos juntos, em 988 das vezes também se vendeu alface, ou seja,

há uma probabilidade de 92% (confiança) de que quando os três produtos forem vendidos juntos também seja vendida alface.

Assim, devido à forte influência daqueles seis produtos no processo de criação das regras, ela também impossibilitou a descoberta de conhecimento em relação à influência dos atributos da dimensão tempo, bem como em relação às cidades e rotas. Deste modo, nas poucas regras em que tais atributos ocorreram do lado antecedente, do lado conseqüente somente ocorreram os seis itens já mencionados, não constituindo descoberta de conhecimento, pois tais produtos são normalmente vendidos independentemente destes atributos.

A tabela 20 mostra algumas regras geradas pelo arquivo de pré-pós feriados de clientes jurídicos e parceiros que ilustram tal fato.

**Tabela 20: Regras geradas a partir de análise de associação por meio do arquivo de pré-pós feriados de clientes jurídicos e parceiros**

<b>SE</b> pre-feriado BANANA 23 ==> <b>ENTÃO</b> LARANJAS 17    conf:(0.74) < lift:(6.53)>
<b>SE</b> pre-feriado ALFACE BANANA 18 ==> <b>ENTÃO</b> CENOURA=37 16    conf:(0.89) < lift:(6.24)>
<b>SE</b> pre-feriado CENOURA 28 ==> <b>ENTÃO</b> BANANA 17    conf:(0.61) < lift:(4.47)>
<b>SE</b> pos-feriado RUCULA 16 ==> <b>ENTÃO</b> ALFACE 15    conf:(0.94) < lift:(3.62)>

Como se pode observar não é possível tomar nenhuma ação com base nestas regras, já que laranja, cenoura, banana e alface são normalmente vendidos.

Seria diferente, caso houvesse uma regra que revelasse, por exemplo, quando ocorre pré-feriado na cidade de Holambra (lado antecedente da regra) há venda de iogurte e geléia (lado conseqüente da regra), pois assim seria possível realizar o planejamento de estoque para este caso e até promover promoções para esta cidade em datas específicas que antecedem os feriados.

No entanto, tais análises não devem ser desconsideradas somente porque não tiveram um resultado positivo para o caso do Yamaguishi, pois para um outro negócio pode ser que estes atributos tragam novos conhecimentos, permitindo a tomada de ações estratégicas.

Assim, a descoberta de conhecimento deu-se em relação aos arquivos de registros dos produtos de maior valor agregado, onde cada um foi analisado separadamente. É nesta direção que os esforços devem ser concentrados quando há predominância da venda de determinados itens.

A tabela 21 apresenta algumas regras dos clientes físicos e a tabela 22 ilustra algumas regras dos clientes jurídicos-parceiros do arquivo dos produtos de maior valor agregado.

**Tabela 21: Regras de associação geradas por meio do arquivo de produtos de maior valor agregado de clientes físicos**

<b>SE</b> ALFACE OVOS-12-unid temperos 110 ==> <b>ENTÃO</b> ACUCAR BANANA 88 conf:(0.8) < lift:(1.31)>
<b>SE</b> BATATINHA TOMATE 365 ==> <b>ENTÃO</b> ALFACE ARROZ-ORG 293 conf:(0.8) < lift:(1.28)>
<b>SE</b> CENOURA QUEIJOS 81 ==> <b>ENTÃO</b> ALFACE BANANADA LARANJAS 71 conf:(0.88) < lift:(1.43)>
<b>SE</b> ALFACE COUVE temperos 124 ==> <b>ENTÃO</b> BANANA confeitaria 96 conf:(0.77) < lift:(1.34)>
<b>SE</b> BANANA temperos 173 ==> <b>ENTÃO</b> ALFACE GELEIAS LARANJAS 110 conf:(0.64) < lift:(1.66)>
<b>SE</b> COUVE QUEIJOS 335 ==> <b>ENTÃO</b> ALFACE IOGURTES 290 conf:(0.87) < lift:(1.36)>
<b>SE</b> CENOURA OVOS-12-unid temperos 303 ==> <b>ENTÃO</b> ALFACE MOLHO-POLPA-TOMATE 284 conf:(0.94) < lift:(1.23)>
<b>SE</b> ABOBRINHA ALFACE LARANJAS 11 ==> <b>ENTÃO</b> NOZ-MACADAMIA temperos 10 conf:(0.91) < lift:(2.36)>
<b>SE</b> ALFACE KOMATSUNA 12 ==> <b>ENTÃO</b> CEBOLINHA CENOURA PICLES 10 conf:(0.83) < lift:(3.25)>
<b>SE</b> BANANA QUEIJOS 37 ==> <b>ENTÃO</b> LARANJAS SOJA 34 conf:(0.92) < lift:(1.72)>
<b>SE</b> LARANJAS MILHOS 221 ==> <b>ENTÃO</b> BANANA SUCOS 179 conf:(0.81) < lift:(1.31)>
<b>SE</b> CENOURA MILHOS 44 ==> <b>ENTÃO</b> AMENDOIM temperos 33 conf:(0.75) < lift:(1.88)>
<b>SE</b> BATATINHA CENOURA 38 ==> <b>ENTÃO</b> ACHOCOLATADO TOMATE 33 conf:(0.87) < lift:(1.53)>
<b>SE</b> COUVE temperos 94 ==> <b>ENTÃO</b> ALFACE CAFE-ORGANICO 91 conf:(0.97) < lift:(1.31)>
<b>SE</b> QUEIJOS temperos 75 ==> <b>ENTÃO</b> FARINHAS LARANJAS 63 conf:(0.84) < lift:(1.49)>
<b>SE</b> ALFACE BANANA CENOURA TOMATE 33 ==> <b>ENTÃO</b> GRANOLA temperos 25 conf:(0.76) < lift:(2.15)>
<b>SE</b> LARANJAS temperos 198 ==> <b>ENTÃO</b> BANANA MEL 158 conf:(0.8) < lift:(1.52)>
<b>SE</b> RUCULA temperos 152 ==> <b>ENTÃO</b> ALFACE MOSTARDA OVOS-12-unid 127 conf:(0.84) < lift:(1.43)>
<b>SE</b> ALFACE MANDIOCA 48 ==> <b>ENTÃO</b> BANANA LARANJAS OLEOS 38 conf:(0.79) < lift:(1.77)>
<b>SE</b> ALFACE BANANA temperos 692 ==> <b>ENTÃO</b> LARANJAS QUEIJOS 539 conf:(0.78) < lift:(1.34)>
<b>SE</b> RUCULA temperos 179 ==> <b>ENTÃO</b> ALFACE TORTAS 166 conf:(0.93) < lift:(1.26)>



**Tabela 22: Regras de associação geradas por meio do arquivo de produtos de maior valor agregado de clientes jurídicos e parceiros**

<b>SE</b> CENOURA TOMATE VAGEM 39 ==> <b>ENTÃO</b> AMENDOIM BETERRABA 36	conf:(0.92) < lift:(1.4)>
<b>SE</b> MANDIOCA MILHOS 9 ==> <b>ENTÃO</b> ALFACE BANANA <b>confeitaria</b> OVOS-12-unid 9	conf:(1) < lift:(3.55)>
<b>SE</b> ALFACE CHICORIA OVOS-12-unid RUCULA 13 ==> <b>ENTÃO</b> AGRIÃO BANANA <b>GELEIAS</b> 12	conf:(0.92) < lift:(5.89)>
<b>SE</b> CHICORIA 53 ==> ALFACE BANANA <b>IOGURTES</b> OVOS-12-unid 46	conf:(0.87) < lift:(2.03)>
<b>SE</b> ALFACE BANANA COUVE LARANJAS 22 ==> <b>ENTÃO</b> CENOURA ESPINAFRE <b>MOLHO-POLPA-TOMATE</b> 20	conf:(0.91) < lift:(5.83)>
<b>SE</b> CEBOLINHA NIRA 17 ==> <b>ENTÃO</b> BARDANA CENOURA <b>OLEOS</b> 16	conf:(0.94) < lift:(2.77)>
<b>SE</b> ALFACE BANANA CHICORIA 49 ==> <b>ENTÃO</b> ESPINAFRE <b>TORTAS</b> 46	conf:(0.94) < lift:(1.54)>
<b>SE</b> LARANJAS RUCULA 31 ==> <b>ENTÃO</b> ALFACE <b>ARROZ-ORG</b> BANANA 28	conf:(0.9) < lift:(5.24)>
<b>SE</b> ALFACE MILHOS OVOS-12-unid temperos 22 ==> <b>ENTÃO</b> ALHO <b>CAFE-ORGANICO</b> MANDIOCA 20	conf:(0.91) < lift:(3.55)>
<b>SE</b> BARDANA CEBOLINHA 21 ==> <b>ENTÃO</b> CENOURA <b>FARINHAS</b> NIRA 20	conf:(0.95) < lift:(4.19)>
<b>SE</b> MANDIOCA 48 ==> <b>ENTÃO</b> ALFACE COUVE <b>MEL</b> temperos 44	conf:(0.92) < lift:(2.66)>
<b>SE</b> LARANJAS 103 ==> <b>ENTÃO</b> BANANA <b>MOSTARDA</b> 98	conf:(0.95) < lift:(1.9)>
<b>SE</b> BANANA MANDIOCA 41 ==> <b>ENTÃO</b> OVOS-12-unid <b>QUEIJOS</b> 37	conf:(0.9) < lift:(1.79)>
<b>SE</b> ALFACE QUEIJOS 23 ==> <b>ENTÃO</b> <b>SUCOS</b> temperos= 20	conf:(0.87) < lift:(6.12)>

Pode-se observar que nem todos os produtos de maior valor agregado, listados no capítulo anterior, estão presentes nas listas acima, pois alguns deles não apresentaram registros de venda seja para clientes físicos ou para clientes jurídicos e parceiros, ou ainda, o número de registro era extremamente pequeno, não sendo interessante analisar.

Outro fato que merece ser mencionado é que, mesmo nos arquivos destes produtos, também é possível observar a forte presença dos seis produtos: alface, ovos, cenoura, tomate, laranja e banana.

De modo geral, a quantidade de regras geradas foi muito grande para todos os tipos de arquivos trabalhados. Sendo assim, algumas regras foram selecionadas para serem analisadas por um dos responsáveis do Yamaguishi.

A seleção se baseou em dois critérios: no grau de confiança, já explicada anteriormente, e o *lift*, que é uma métrica que compara as chances, por exemplo, de ocorrer a venda de *sucos e temperos* (lado conseqüente), dada a venda de *alface e queijos* (lado

anterior) – última regra da tabela 22, com as chances de se vender sucos e temperos aleatoriamente, ou seja, esta métrica representa o ganho em se usar determinada regra para prever a venda de determinado produto versus uma previsão aleatória da venda desse produto.

As regras de grau de confiança inferior a 0.6 foram descartadas, independentemente do *lift*. Com relação ao *lift*, não é possível estabelecer um grau mínimo para seleção, pois o cálculo deste número depende do número de registros totais de um banco de dados sobre o número de vezes que o produto X ocorre, dado produto Y, assim, o *lift* varia de conjunto para conjunto. Portanto, foram selecionadas regras com confiança superior a 0.59 e que tivessem os valores mais altos de *lift* em relação às outras.

Ao serem analisadas pelo responsável do Yamaguishi, as regras dos produtos de maior influência (alfaca, tomate, cenoura, ovos, laranja, banana) foram confirmadas, pois são os produtos mais vendidos. Porém, em relação às regras dos produtos de maior valor agregado, ele não soube avaliar já que são vendidas poucas quantidades e, por isso, não saberia distinguir se há ocorrência mais comum da venda de um destes produtos associada à de outros.

Deste modo, cabe a cada negócio analisar se há produtos que predominam nas vendas, bem como se é interessante analisar algum atributo da dimensão tempo que tenha impactos em seu negócio. Além disso, determinar quais os produtos de maior valor agregado são mais interessantes para serem analisados separadamente.

As regras de tais produtos podem ajudar a planejar as vendas por meio da estipulação de descontos de determinados produtos com o objetivo de impulsionar a venda de outros a que esteja associado, além de fornecer dicas para estratégias de vendas cruzadas.

#### **5.2.4. Fase 4: Disseminação da inteligência**

##### **Sugestão das ações de comercialização**

Finalmente, a fase 4 encerra o ciclo de IC, bem como o processo de KDD, mostrando como utilizar o conhecimento descoberto para a tomada de decisões. Neste caso trata-se de sugerir ações de comercialização baseada no modelo de diferenciação de clientes e nas regras de associação que revelam alguns hábitos de consumo.

É importante ressaltar que devido ao número de perfis de clientes descobertos ser bastante elevado, não se encontra no escopo deste trabalho propor ações para cada um deles, mas sim ilustrar, por meio de alguns exemplos, algumas ações que podem ser tomadas de acordo com determinado perfil, cabendo a cada negócio determinar a prioridade de suas ações e estratégias.

Sendo assim, é interessante retomar a situação atual do caso estudado, onde temos um quadro de grande rotatividade dos clientes, além da grande maioria (82%) ter gastado até R\$ 1000 num período de três anos. Deste montante, mais da metade (54%) gastou até R\$ 100 neste mesmo período, ou seja, como seria possível mudar tal situação de modo que os clientes permanecessem mais tempo comprando e gerando mais receita para empresa.

Neste caso, como forma de quebrar este ciclo de rotatividade, ações como ativação, retenção e vendas cruzadas podem fazer com que o cliente passe mais tempo fazendo negócio com a empresa, além de contribuir para aumentar o valor gasto por compra, desde que tais ações sejam direcionadas aos perfis corretos.

Deste mesmo modo, outras ações também são importantes para condução do negócio e dizem respeito à fidelização e recuperação.

A seguir serão apresentadas como tais ações podem ser implementadas e o perfil de cliente mais adequado a cada uma delas.

#### 5.2.4.a. Ativação

As ações de ativação são destinadas a clientes pouco ativos para que sejam estimulados a comprar mais vezes

Neste contexto, enquadram-se os seguintes perfis:

- Folha A: **143** – cliente físico
- Folha V: **242, 342** – cliente jurídico/parceiro
- Folha X: **442** – cliente jurídico/parceiro

Como se pode observar tais perfis de cliente, possuem uma boa nota de receita, no entanto, suas notas de compras e recência não são tão boas; já no caso 442, apesar de também possuir boa nota de compra seria interessante melhorar sua recência.

Assim, uma forma de estímulo para a efetuação de novas compras seria o oferecimento de descontos nos produtos que tais clientes costumam comprar corriqueiramente.

Portanto, faz-se necessário criar um histórico das preferências de cada cliente, ou pelo menos, dos mais importantes em termos estratégicos. Tal levantamento pode ser realizado por meio do arquivo de Itens de pedidos, onde constam todos os produtos comprados.

Deve-se destacar que tais descontos devem ser oferecidos somente para estes perfis de clientes, assim, a comunicação entre empresa e cliente deve ocorrer de forma individualizada, como prega o CRM.

#### 5.2.4.b. Vendas Cruzadas (*Cross Selling*)

Essas ações têm o objetivo de realizar a venda de produtos relacionados entre si para um mesmo cliente de modo que este proporcione mais receita para a empresa.

Neste caso, as regras descobertas pela análise de associação podem ser bastante úteis para a determinação de qual produto deve ser oferecido juntamente com outro. Além disso, pode também oferecer desconto na compra do produto X, caso também seja comprado o produto Y.

Por exemplo, considerando a seguinte regra:

BATATINHA TOMATE 365 ==> ALFACE **ARROZ-ORG**= 293

Pode ser oferecido desconto de X% na compra de arroz orgânico para os clientes que tiverem comprado batatinha e tomate.

Assim, os perfis aos quais este tipo de ação se destina são os seguintes:

- Folha A: **125,134,135** – cliente físico
- Folha H e I: **234, 235** respectivamente – cliente físico
- Folha P: **334, 335** – cliente físico
- Folha X: **433, 443** – cliente jurídico/parceiro

Em todos os casos, apesar das diferentes notas de compras, eles possuem ótimas notas de recência, sendo que suas notas de receita podem ser melhoradas, portanto, a venda cruzada seria a melhor estratégia para estes casos.

#### **5.2.4.c. Ativação e Vendas Cruzadas**

Casos como os seguintes:

- Folha G: **233** – cliente físico
- Folha U: **110, 112, 122** – cliente jurídico/parceiro
- Folha V: **211, 212, 221** – cliente jurídico/parceiro

Todos apresentam notas de receita e recência que podem ser melhoradas. Portanto, ambas ações devem ser utilizadas, já que são perfis de clientes que, mesmo apesar das baixas notas, estão ativos.

No caso, a data da primeira compra deve ser checada para que os esforços destas ações sejam concentrados nos clientes mais recentes, ou seja, nos clientes mais novos, pois estes estão mais propensos em se tornarem mais ativos, aumentando seu número de compras, do que os mais antigos que compram eventualmente.

Os clientes mais recentes, ao serem estimulados, podem permanecer mais tempo fazendo negócio com a empresa, de modo que alcancem um perfil de notas mais altas.

Além disso, no caso de um cliente antigo, talvez não seja interessante realizar ação alguma, pois pode tratar-se de um cliente que somente compra esporadicamente e, no momento, tornou-se ativo novamente, ou seja, tal situação não reflete a ocorrência de uma mudança de comportamento duradoura.

#### 5.2.4.d. Retenção

São ações destinadas a clientes rentáveis, ainda ativos, que estão prestes a abandonar o negócio com a empresa.

Há dois casos ilustram bem esta situação:

- Folha R: **355** – cliente físico
- Folha V: **242** – cliente jurídico/parceiro

Neste caso, ambos perfis ainda estão ativos. No entanto, como foi mostrado pela análise de monitoração realizada com o arquivo das transações – correspondentes aos casos de Valinhos (físico) e São José dos Campos (jurídico) respectivamente, os clientes com estes perfis estão apresentando quedas sucessivas nas notas de alguns de seus atributos (valor, frequência ou recência), ou seja, tal comportamento pode significar uma futura perda desses clientes.

Assim, é importante que haja uma ação antes que tal predição se concretize, para tanto pode ser oferecido a estes clientes isenção da taxa de entrega por um determinado período, desde que seja um benefício que atenda às suas expectativas, como forma de motivá-los a continuar comprando, não abandonando o negócio com a empresa.

#### 5.2.4.e. Recuperação

São ações destinadas a clientes que deixaram de fazer negócio com a empresa. Nesta situação encontram-se os seguintes perfis:

- Folha T: **555** – cliente físico
- Folha R: **355** – cliente físico
- Folha W: **352** – cliente jurídico/parceiro

Deve-se ressaltar que todos estes perfis não estão ativos. Pode-se observar que são clientes com ótimas notas de receita e recência e deveriam ter sido fidelizados, porém foram perdidos.

Para recuperação destes clientes, é importante investigar o motivo de terem deixado de comprar como, por exemplo, por considerarem alto o preço da taxa de entrega. Além disso, deve ser levantado qual o incentivo mais interessante para que estes clientes voltem a comprar.

Assim, ações que podem ser realizadas como tentativa de resgate destes clientes é o oferecimento de isenção da taxa de entrega por determinado período e descontos nos principais produtos comprados por tais clientes, desde de que tais incentivos correspondam às expectativas destes clientes.

É importante destacar que as análises com o arquivo das transações trouxeram regras que puderam indicar a situação não ativa dos perfis **355** e **352**, correspondendo aos casos de Vinhedo (físico), Cajamar (jurídico), mostrando mais uma vez a importância deste tipo de arquivo para a monitoração dos clientes.



#### **5.2.4.f. Fidelização**

São ações para reforçar o sentimento de conveniência do cliente para com a empresa, de modo que a mudança para uma outra seria inconveniente.

Os perfis que se enquadram em tal ação são os seguintes:

- Folha S: **545, 455** – cliente físico
- Folha T: **555** – cliente físico
- Folha X: **452** – cliente jurídico/parceiro
- Folha Y: **543, 553** – cliente jurídico/parceiro

Estes clientes são especiais e, portanto, merecem ser tratados como tal, já que são muito importantes para a empresa, em termos de geração de receita.

Sendo assim, algumas ações poderiam ser tomadas como: isenção total da taxa de entrega, produtos em lançamento poderiam ser oferecidos gratuitamente para degustação, programa de desconto para determinados produtos de acordo com a capacidade do negócio.

Além disso, o sítio Yamaguishi promove alguns eventos culturais em seu espaço como visitas no sítio e cursos de auto-conhecimento.

Tais cursos poderiam ser oferecidos gratuitamente no mês de aniversário do cliente, ou ainda, oferecer convites a toda sua família para uma visita ao sítio. Tais ações se aplicariam para os clientes do tipo físico e seriam uma forma de aproximá-los, tornando-os não apenas clientes, mas sim amigos do Yamaguishi, fortalecendo, assim, o vínculo com a empresa.

Já para os clientes do tipo jurídico/parceiro poderia ser oferecido sorteio de cursos e de cestas de produtos para seus clientes, pois assim, além de estimular a venda dos produtos Yamaguishi, também seria uma forma de atrair novos clientes físicos para compra de produtos orgânicos diretamente com o Sítio.

É importante observar que:

- Os perfis **545** e **543** (cliente físico) ainda podem ser estimulados a aumentar sua nota de receita através da venda cruzada;
- O **452** (cliente jurídico/parceiro) pode ser estimulado a melhorar sua recência por meio de ações de ativação;

Outros potenciais clientes a atingirem o patamar de fidelização são os clientes com perfil do tipo: **442**, **433**, **443** (todos físicos); já que suas notas de receita e recência são boas ou razoáveis e que por meio de ações de ativação e vendas cruzadas podem ser elevadas.

#### **5.2.4.g. Observação**

Outro ponto importante é em relação ao controle destas ações o que, no caso, corresponderia à avaliação dos processos e produtos de inteligência, sendo a quinta fase do ciclo de IC.

Assim, após a realização das ações propostas é necessário que seja feito um acompanhamento, semanal ou mensal, dependendo do tipo de ação e dos objetivos de cada negócio.

É importante que seja verificado se os clientes tiveram uma resposta positiva ao estímulo e se estão melhorando a qualidade de suas interações, a qual pode ser medida pelo arquivo das transações através do valor e frequência acumulados e da recência.

Deste modo, com o passar o tempo todos os dados dos clientes devem ser revisto por meio de novas análises de Mineração de Dados.

Além disso, deve-se também ser verificado se os clientes trabalhados tiveram aumento em suas notas, de modo que alcançassem um perfil mais elevado, sendo que esta informação pode ser medida por meio do arquivo de histórico.

Deve-se ressaltar que, por meio deste trabalho, não é possível afirmar com que período todo o processo de Mineração deve ser revisto, já que esta informação depende do planejamento e das particularidades de cada empresa e negócio.

## 6. Considerações Finais

### *Diferenciação dos clientes*

O arquivo do Histórico foi o mais importante para a construção do modelo de diferenciação dos clientes.

Os atributos Nota de Compra, Nota de Receita e Recência mostraram-se bastante úteis para a construção dos perfis de clientes.

No entanto, deve-se ressaltar que a determinação do intervalo de tempo em que tais atributos são trabalhados, bem como os valores para determinação das notas dos mesmos, cabe às características de cada negócio.

Além disso, pôde-se observar que foram muitos os perfis definidos pelo modelo de diferenciação proposto e, que em alguns casos, tais perfis se sobrepõem, mudando apenas o número da nota de compras.

Deste modo, é natural que clientes com perfis que tenham mesma nota de receita e recência, porém notas diferentes de compras, tenham ações estratégicas iguais (como, por exemplo, no caso dos perfis: 242, 342).

Assim, um outro ponto que merece destaque é o ideal da filosofia do CRM que implica justamente no conhecimento individualizado dos clientes, o que explica o fato do levantamento desses vários perfis. Portanto, este procedimento foi a base para a proposição das ações comerciais adequadas a cada cliente.

Além disso, o fato de um cliente ter nota baixa de compras pode indicar duas coisas: o cliente é recente e começou a comprar há pouco tempo e, por isso, sua nota de compras ainda é baixa; ou então, o cliente comprou há muito tempo atrás e só, recentemente, voltou a fazer negócios com a empresa.

O mesmo raciocínio pode ser aplicado no caso de uma nota alta de compras, que pode indicar que um cliente é antigo, já que compra há bastante tempo e, por isso, sua nota é alta ou, então, é um cliente recente, mas que em pouco tempo já realizou muitas compras. Para os dois casos, tal questionamento pode ser respondido verificando-se a data de seu primeiro pedido.

Também, é importante que haja atenção especial com os clientes atualmente ativos, mesmo que apresentem notas baixas em todos os atributos, pois são clientes potenciais que podem ter suas notas melhoradas, servindo para combater a indesejada rotatividade.

Outro ponto a ser destacado diz respeito ao perfil atual do cliente, que não deve ser entendido como um comportamento estático, já que é passível de alteração, tanto para melhor como pior.

Caso as devidas ações estratégicas sejam tomadas, com base no modelo de diferenciação, espera-se que o perfil sempre mude para um patamar melhor. Portanto, é importante que haja constante monitoração para verificar mudanças no perfil, evidenciando a necessidade de revisão, com o passar do tempo, dos dados, das análises e das ações tomadas.

Neste contexto, reside a importância do arquivo das Transações que, apesar não ter tido um bom resultado para a diferenciação dos clientes, mostrou-se extremamente útil para auxiliar no processo de monitoração dos clientes, indicando a urgência em se voltar a atenção para determinados clientes.

Pode-se, também, sugerir uma nova possibilidade de análise com a inclusão do perfil, descoberto por meio do arquivo de histórico, nos registros do arquivo das transações. Este tipo de manipulação dos dados poderia ajudar a descobrir novos padrões de comportamento.

### *Descoberta dos hábitos de consumo*

No caso das análises de associação houve a ocorrência de um enorme número de regras o que poderia ter dificultado a compreensão dos resultados. A seleção dessas regras baseou-se na confiança e no *lift*, porém há outros meios que podem ser utilizados para auxiliar no trabalho de seleção.

Além disso, a predominância da venda dos seis principais produtos (alface, ovos, cenoura, tomate, laranja e banana) dificultou a descoberta de conhecimento e, portanto, a confirmação das hipóteses da influência do tempo sobre o consumo (estação do ano, pré ou pós feriado, quinzena do mês, etc.).

No entanto, este resultado não implica que este tipo de análise não seja importante, já que para outros negócios pode apresentar um outro resultado.

No caso do Yamaguishi, a descoberta de conhecimento, por meio da Mineração de Dados, deu-se por meio dos arquivos dos produtos de maior valor agregado.

Além disso, é importante destacar que havendo alguma Cidade, Rota, ou qualquer outro atributo bastante predominante em todo o conjunto de dados, seria interessante realizar análises separadamente. Por exemplo, separar e analisar somente os registros dos clientes de Campinas, ou os clientes da rota 25; sendo que cabe a cada negócio decidir sobre a necessidade de tal análise.

### *Sugestões*

As sugestões deste trabalho já foram descritas anteriormente, pois estas fazem parte da segunda fase do Modelo de Gestão Estratégica.

Um destaque a ser feito, diz respeito aos cuidados com os dados, tais como: a correta armazenagem dos mesmos – no caso, os campos dos preços e rotas, bem como os meios para identificação dos clientes, pois são peças fundamentais para a construção do modelo de diferenciação dos clientes.

### *Observações Finais*

É importante destacar que o modelo de diferenciação de clientes permite que os esforços sejam direcionados de modo adequado a cada perfil. Assim os recursos podem ser investidos naqueles clientes que sejam os mais interessantes estrategicamente para empresa e que tenham mais garantia de proporcionar retorno, evitando que seja gasto tempo e dinheiro indiscriminadamente.

Deve-se ressaltar que a determinação do tipo de cliente mais importante depende dos objetivos de cada negócio. Por exemplo, se o objetivo for aumentar a receita, o negócio deve concentrar seus esforços em ações de vendas cruzadas; já se tiver como objetivo aumentar o número de clientes, o ideal seria investir em ações de ativação.

O Modelo de Gestão desenvolvido teve como objetivo fornecer as bases para o gerenciamento inteligente da informação, ou seja, a coleta, a manutenção e a análise dos dados, antes acumulados e sub-aproveitados.

Desta maneira, ao se extrair o conhecimento proveniente dos dados, este pode indicar o caminho rumo à manutenção dos atuais clientes, proporcionando a criação de outras oportunidades de negócio.

Este trabalho, desenvolvido por meio do Modelo de Gestão Estratégica com foco no cliente para comercialização de produtos orgânicos, atingiu os objetivos propostos, mostrando como informações e dados dispersos e acumulados em bancos de dados podem ser trabalhados, de modo que sejam transformados em inteligência de negócio para obtenção de vantagens comerciais e competitivas.

Por último, deve-se salientar que dependerá de cada negócio a decisão de treinar seu gerente de clientes para assumir todas as tarefas pertinentes à Mineração de Dados ou, então, a contratação de terceiros para o desenvolvimento desta tarefa e para aconselhamento sobre as ações de comercialização.

## **7. Trabalhos Futuros**

São três as sugestões para trabalhos futuros.

A primeira delas diz respeito ao desenvolvimento dos modelos de: Identificação, Interação e Personalização de clientes, que correspondem justamente às outras três estratégias apresentadas pela teoria de CRM e podem servir para complementar o modelo de gestão proposto por este trabalho.

Tal trabalho poderia abordar quais os procedimentos utilizados para Identificação dos clientes, bem como abordar as falhas existentes, além de propor melhoras e cuidados para este procedimento.

Para as estratégias de Interação e Personalização, seria interessante fazer um levantamento por meio, por exemplo, de pesquisa qualitativa e quantitativa junto aos clientes, dos meios e procedimentos mais adequados para realização destas atividades em relação a determinados perfis de clientes. Tal levantamento poderia gerar idéias para criação de novos produtos ou serviços.

A segunda sugestão refere-se ao desenvolvimento de um sistema computacional para auxiliar no processo de gerenciamento de clientes.

Para tanto, este sistema deveria possuir um recurso que mostrasse o perfil em que os clientes se encontram no presente momento e, também, pudesse monitorar semanalmente a variação das notas recência e do valor e frequência acumulados, de modo que fosse possível alertar quando um cliente estivesse apresentando quedas sucessivas nos atributos mencionados.

Este sistema ajudaria a determinar a prioridade e urgência de determinadas ações, gerando relatórios que mostrassem a porcentagem de clientes pertencentes a cada perfil.

Ele também serviria para avaliação e controle das ações tomadas, ajudando a verificar se estão obtendo os resultados esperados por meio do monitoramento das notas de recência, frequência acumulada e valor acumulado.

A terceira sugestão refere-se à utilização de uma metodologia para a avaliação das regras encontradas por meio das análises de associação dos produtos orgânicos.

O número de regras geradas por meio deste tipo de análise é imenso e nem sempre as métricas de Confiança e *Lift* são suficientes para avaliá-las, revelando o que é realmente importante.

Assim, o trabalho poderia aplicar uma abordagem conhecida e difundida no meio acadêmico como Grau de Interesse de Regras (“*Interestingness*”), que diz respeito à avaliação do quanto uma regra pode ser mais interessante que a outra para descoberta de conhecimento. Este procedimento é realizado por meio de modelos matemáticos que ordenam as regras de acordo com o grau de interesse. Maiores detalhes sobre esta metodologia podem ser obtidos no trabalho de HILDERMAN e HAMILTON (2000).



## 8. Referências

BERRY, M. J. A.; LINOFF, G. **Data mining techniques for marketing, sales and customer support**. New York: Wiley Computer Publishing, 1997. 454p.

BESSA, J. **Inteligência competitiva e contra-inteligência**. 2001. Disponível em: <<http://www.abraic.org.br/noticias>. Acesso em 04 de out. de 2004.

BRAZDIL, P. **Construção de modelos de decisão a partir de dados**. 1999. Disponível em: <<http://www.ncc.up.pt/~pbrazdil/Ensino/ML/DecTrees.html>. Acesso em 01 maio de 2003.

BRETZKE, M. **Marketing de relacionamento e competição em tempo real com CRM**. São Paulo: Atlas, 2000. 224p.

CERVEIRA, R.; CASTRO, M.C. Consumidores de produtos orgânicos da cidade de São Paulo: características de um padrão de consumo. **Informações Econômicas**, São Paulo, v.29, n.12, dez. 1999. p. 7-15.

COBB, P. Competitive intelligence through data mining. **Journal of Competitive Intelligence and Management**, New Orleans, v.1, n.3, Winter, 2003. p. 80-88.

DAROLT, M.R. **Agricultura orgânica: inventando o futuro**. Londrina: IAPAR, 2002. 250p.

DINIZ, D.; NAIDITCH, S. Criando raízes. **Revista Exame**, São Paulo, jan. de 2003. p. 74-78.

ESCOBAR, H. Orgânico: saudáveis e cada vez mais acessíveis. **O Estado de São Paulo**, São Paulo, 10 nov. de 2002. Disponível em: <<http://www.estado.com.br/editorias/2002/11/10/ger012.html>. Acesso em 09 abril de 2003.

FARREL, A. **Competitive intelligence basics**. 2003. Disponível em: <<http://www.worksys.com/ci101.htm>. Acesso em 10 ago. de 2004.

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, Menlo Park v. 17, n. 3, Fall 1996, p. 37-54.

GARCIA, S. C.; ALVARES, L. O. **O uso de árvores de decisão na descoberta de conhecimento na área da saúde**. Disponível em: <<http://www.inf.ufrgs.br/pos/SemanaAcademica/Semana2000/SimoneGarcia/>. Acesso em: 23 de maio de 2003.

GUIVANT, J. S. Os Supermercados na oferta de alimentos orgânicos: apelando ao estilo ego-trip. **Ambiente & Sociedade**, São Paulo, v.6, n.2, jul de 2003. p. 64-81.

INFO SOCIETY. **Ações de negócio**. 2004. Disponível em: <<http://www.infosociety.com.br/visita-guiada.php>. Acesso em 14 jan. de 2004.

INSTITUTO BIODINÂMICO. **Certificação**. 2004. Disponível em:  
<<http://www.ibd.com.br/certifica.htm>. Acesso em 25 jan. de 2004.

INSTITUTO BIODINÂMICO. **Lei 10.831 de 23 de Dezembro de 2003**. 2005. Disponível em: <<http://www.ibd.com.br>. Acesso em 03 mar. de 2005.

ISTO É. **Saudáveis**. 07 Jan de 2004. Medicina & Bem Estar. Disponível em:  
<[http://www.terra.com.br/istoe/1787/medicina/1787\\_cardapio\\_do\\_futuro\\_02.htm](http://www.terra.com.br/istoe/1787/medicina/1787_cardapio_do_futuro_02.htm). Acesso em 20 jul. de 2004.

HARTMAN GROUP. **The wellness lifestyle shopper**: mapping the journeys of wellness consumers. 2000. apud GUIVANT, J. S. Os Supermercados na oferta de alimentos Orgânicos: apelando ao estilo ego-trip. **Ambiente & Sociedade**. São Paulo, v.6, n.2, jul de 2003. p. 64-81.

HERRING, J.P. Tópicos fundamentais de inteligência: processo para identificação e definição de necessidades de inteligência. In PRESCOTT, John E.; MILLER, Stephen H. **Inteligência Competitiva na Prática**. São Paulo: Campus, 2002. p. 274-291.

HILDERMAN, R.J.; HAMILTON, H.J. Applying objective interestingness measures in data mining systems. In: [EUROPEAN SYMPOSIUM ON PRINCIPLES OF DATA MINING AND KNOWLEDGE DISCOVERY (PKDD-2000)] 4, Lyon, France, Springer Verlag, September, 2000, pp. 432-439.

LIU, S.; YAP J. Beyond intuition. 2001. **DB2 Magazine**. Disponível em:  
<[http://www.db2mag.com/db\\_area/archives/2001/q4/miner.shtml](http://www.db2mag.com/db_area/archives/2001/q4/miner.shtml). Acesso em 27 mar. de 2004

MENDONÇA NETO, M. G.de. Mineração de dados. In: ESCOLA REGIONAL DE INFORMÁTICA DA SBC REGIONAL DE SÃO PAULO, 6., 2001, São Carlos. **Anais**: Minicursos - coletânea de textos: São Carlos: USP-ICMC, 2001. p. 189-218.

MILLER, J.P. **O milênio da inteligência competitiva**. Porto Alegre: Bookman, 2002. 293p.

MILLER, S. H. **Competitive intelligence**: an overview. Society of Competitive Intelligence Professionals. 2005. Disponível em <<http://www.scip.org/Library/overview.pdf>>. Acesso em 03 mar. de 2005

MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvore de decisão. In: REZENDE, S. O. **Sistemas inteligentes**: fundamentos e aplicações. Barueri: Manole, 2003. p. 114-134.

MUNARI, C. de B. **Uso de técnicas de classificação automática na análise ambiental**: um estudo de caso, 2001. Dissertação (Mestrado). Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas. Campinas. 142p

O ESTADO DE SÃO PAULO. **Mercado de orgânicos cresce mundialmente**. São Paulo, 19 Fev. de 2003. Suplemento Agrícola . Disponível em:  
<<http://www.icepa.com.br/observatorio/noticias0203/no2402.htm> .Acesso em 15 jan. de 2004.

ORMOND et al. **Agricultura Orgânica: quando o passado é futuro**. Rio de Janeiro: BNDS Setorial, 2002.

PLANETA ORGÂNICO.**Quem certifica**. 2004. Disponível em:  
<<http://www.planetaorganico.com.br/qcertif.htm>. Acesso em 28 jan. de 2004.

PEPPERS AND ROGERS, GROUP. **CRM series, marketing one to one**. Peppers and Rogers Group, 2000. 77p. Disponível em: <<http://www.1to1.com.br/>. Acesso em 18 mar. de 2003.

PRESCOTT, J.E.; MILLER, S.H. **Inteligência competitiva na prática: técnicas e práticas bem sucedidas para conquista de mercados**. Rio de Janeiro: Campus, 2002. 371p.

RUCHINSKI, J. **Consumidores de alimentos orgânicos em Curitiba**, 1999. Monografia (Graduação em Ciências Sociais). Setor de Ciências Humanas, Universidade Federal do Paraná. Curitiba. apud: DAROLT, M.R. **Agricultura Orgânica: inventando o futuro**. Londrina: IAPAR, 2002. 280p.

SCIP. **What is competitive intelligence**. Society of Competitive Intelligence Professionals. 2004. Disponível em: <<http://www.scip.org>. Acesso em 21 out. de 2004.

SWIFT, R. **CRM: o revolucionário marketing de relacionamento**. 4ed. Rio de Janeiro: Campos, 2001. 493p.

TAKAOKA, H. Aplicação de data warehouse no varejo. In: ANGELO, C. F. de; GIANGRANDE, V. **Marketing de relacionamento no varejo**. São Paulo: Atlas, 1999. p. 87-103.

TWO CROWS CORPORATION. **Introduction to data mining and knowledge discovery**. 3 ed. Potomac: Two Crows Corporation, 1999.36p.

VALOR ECONÔMICO. **Alimentos orgânicos rendem US\$25 bilhões**. São Paulo, 23 Jul. de 2003. Disponível em: <<http://www.ibd.com.br>. Acesso em 19 ago. de 2004

WITTEN, I. H.; FRANK, E. Data Mining: **Practical machine learning tools and techniques with Java implementations**. San Francisco: Morgan Kaufmann Publ., 2000. 369p.

WILLER, H.; YUSSEFI, M (Ed). **The world of organic agriculture: statistics and emerging trends**, 2004. IFOAM Publication, 6th, revised edition, February 2004. 167p. Disponível em:<<http://www.soel.de/oekolandbau/weltweit.html>. Acesso em 24 ago. de 2004.

## 9. Anexos

Para informações complementares sobre os arquivos e resultados gerados pelo software, entrar em contato através do email: tatidn@gmail.com

### 9.1. Definição dos arquivos originais

**Observação:** Os campos em negrito foram utilizados no estudo para criação dos atributos.

#### 9.1.a. Clientes

**1 COD\_CLI NUMERIC 6 1 1**  
2 NOME\_CLI ASCII 50 4 2  
3 FANTASIA ASCII 40 54  
4 CGC\_CLI ASCII 14 94  
5 IE\_CLI ASCII 14 108  
6 END\_CLI ASCII 40 122  
7 COMPL\_CLI ASCII 25 162  
8 BAIRRO\_CLI ASCII 25 187  
**9 CIDADE\_CLI ASCII 30 212**  
10 UF\_CLI ASCII 2 242  
11 CEP\_CLI ASCII 9 244  
12 TEL\_CLI ASCII 14 253  
13 TEL2\_CLI ASCII 14 267  
14 FAX\_CLI ASCII 14 281  
15 MAIL\_CLI ASCII 30 295  
16 CONTATO\_CLI ASCII 25 325  
17 COD\_ROTA NUMERIC 4 350 4 ROTAS.COD\_ROTA  
18 ORD\_ENTR NUMERIC 4 352  
19 COD\_TXENTR NUMERIC 4 354 TAXA\_ENT.COD\_TXENTR  
**20 TIPO\_CLI ASCII 1 356 3**

#### 9.1.b. Produtos

**1 COD\_PROD NUMERIC 4 1 1**  
2 DESC\_PROD ASCII 30 3 2  
3 DESC\_PROD\_NF ASCII 50 33  
4 COD\_GRUPO NUMERIC 6 83 GRUPOS.COD\_GRUPO  
5 UNID ASCII 2 86  
6 PRC\_COMPRA NUMERIC 8,2 88  
7 PRC\_VENDA\_U NUMERIC 8,2 93  
8 PRC\_VENDA\_L NUMERIC 8,2 98  
9 PRC\_VENDA\_S NUMERIC 8,2 103  
10 PRC\_VENDA\_4 NUMERIC 8,2 108  
11 PRC\_VENDA\_5 NUMERIC 8,2 113  
12 COD\_FOR NUMERIC 6 118  
13 ESTOQUE NUMERIC 6,2 121  
14 ULT\_COMPRA DATE 6 125  
15 ULT\_PRECO NUMERIC 8,2 128

### **9.1.c. Rotas**

**1 COD\_ROTA NUMERIC 4 1 1**  
**2 DESC\_ROTA ASCII 30 3 2**

### **9.1.d. Pedidos**

**1 DATA\_ENTREGA DATE 6 1 1**  
**2 COD\_CLI NUMERIC 6 4 1 CLIENTES.COD\_CLI**  
3 COD\_ROTA NUMERIC 4 7 2  
4 ORD\_ENTR NUMERIC 4 9 9  
5 DATA\_PED DATE 6 1 1  
**6 TOTAL\_PROD NUMERIC 8,2 14**  
7 COD\_TAXA NUMERIC 4 19 TAXA\_ENT.COD\_TXENTR  
8 VAL\_TXENTR NUMERIC 8,2 21  
9 VALOR\_DESC NUMERIC 8,2 26  
10 TOTAL\_PED NUMERIC 8,2 31  
11 DATA\_PAG DATE 6 36 3  
12 COD\_PED OVERLAP 6 1 1  
13 ULT\_ITEM NUMERIC 4 39  
14 ROTA\_PAG NUMERIC 4 41 7

### **9.1.e. Itens do Pedido**

**1 DATA\_ENTR DATE 6 1 1**  
**2 COD\_CLI NUMERIC 6 4 1**  
3 COD\_GRUPO NUMERIC 6 7 2  
**4 COD\_PROD NUMERIC 4 10 1 PRODUTOS.COD\_PROD**  
5 COD\_FOR NUMERIC 6 12 2  
**6 QTDE NUMERIC 6,2 15**  
7 PRC\_UNIT NUMERIC 8,2 19  
**8 VAL\_TOT NUMERIC 8,2 24**  
9 RETORNO\_PED NUMERIC 6,2 29  
10 COD\_PED OVERLAP 6 1 1 PEDIDOS.COD\_PED  
11 ULT\_ITEM NUMERIC 4 33  
12 PERDA\_PED NUMERIC 6,2 35  
13 COMPRA NUMERIC 8,2 39  
**14 COD\_ROTA NUMERIC 4 44 3**  
15 PROCESSADO ASCII 1 46  
16 VENDEDOR NUMERIC 4 47

## 9.2. Exemplos dos arquivos usados para Mineração – formato: Arrf (WEKA)

### 9.2.1. Arquivo Histórico – Cliente Jurídico e Parceiro

@relation weka-juridpar

@attribute cod-ped REAL

@attribute tipo-cli {2,3}

@attribute compras REAL

@attribute Nota-compras {1,2,3,4,5}

@attribute recencia REAL

@attribute Nota-recencia {0,1,2,3,4,5}

@attribute Ativo {N,S}

@attribute receita REAL

@attribute nota-receita {1,2,3,4,5}

@data

1,3,26,2,8,1,N,1254.0,2  
222,3,12,2,0,0,N,155.7,1  
295,3,570,5,1,3,S,780161.47,5  
395,3,19,2,14,1,N,279.12,1  
914,3,4,1,0,0,N,83.49,1  
1117,3,5,1,0,0,S,27.86,1  
1167,3,17,2,5,3,N,2544.98,3  
22,2,147,4,7,2,S,13865.5,4

### 9.2.2. Arquivo das Transações – Cliente Físico

@relation GRUPO-FISICO-90dias-ativo

@attribute cod-ped REAL

@attribute periodo-mes { 1,2,3,4 }

@attribute quinzena-mes { 1,2 }

@attribute estacao-ano { 1,2,3,4 }

@attribute pre-feriado { N,S }

@attribute pos-feriado { N,S }

@attribute mes-ano { 1,2,3,4,5,6,7,8,9,10,11,12 }

@attribute dia-semana { 1,2,3,4,5,6,7 }

@attribute recencia REAL

@attribute recencia-nota { 0,1,2,3,4,5 }

@attribute cod-cliente REAL

@attribute Ativo { S,N }

@attribute tipo-cliente { 1 }

@attribute valor-acumulado REAL

@attribute valor-acumulado-nota { 1,2,3,4,5 }

@attribute frequencia REAL

@attribute frequencia-nota { 1,2,3,4,5 }

@attribute cidade

{ 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31 }

@data

16569,2,1,4,N,N,11,6,477,1,2,N,1,19.2,1,1,1,9,12

3175,3,2,3,N,N,7,5,98,1,2,N,1,17.7,1,1,1,9,24

1041,2,1,2,S,N,4,5,0,0,2,N,1,9.3,1,1,1,9,24

19914,4,2,2,N,N,3,5,175,1,3,N,1,28.3,1,1,1,9,24

15484,1,1,4,N,N,10,5,21,3,3,N,1,50.4,1,3,1,9,24

14752,2,1,3,N,N,9,5,42,1,3,N,1,31.2,1,2,1,9,24

13266,1,1,3,N,N,8,5,49,1,3,N,1,69.2,1,3,1,9,24

## 9.3. Resultados do Software das Análises

### 9.3.1. Resultado Classificação

#### 9.3.1.a. Arquivo Histórico

##### *Cliente Físico*

=== Run information ===

```
Scheme:      weka.classifiers.trees.j48.J48 -C 0.25 -M 2
Relation:     weka-final-fisico-weka.filters.unsupervised.attribute.Remove-
R1,2,4,6,7,9
Instances:    650
Attributes:   4
              Nota-compras
              Ativo
              Nota-recencia
              nota-receita
Test mode:    evaluate on training data
```

=== Classifier model (full training set) ===

J48 pruned tree

```
-----
Nota-compras = 1: N (233.0/30.0)
Nota-compras = 2
|  nota-receita = 1: N (1.0)
|  nota-receita = 2: N (57.0/5.0)
|  nota-receita = 3
|  |  Nota-recencia = 0: N (7.0/1.0)
|  |  Nota-recencia = 1: S (12.0/1.0)
|  |  Nota-recencia = 2: N (1.0)
|  |  Nota-recencia = 3: S (5.0/2.0)
|  |  Nota-recencia = 4: N (20.0/9.0)
|  |  Nota-recencia = 5: N (75.0/24.0)
|  nota-receita = 4
|  |  Nota-recencia = 0: S (1.0)
|  |  Nota-recencia = 3: N (1.0)
|  |  Nota-recencia = 4: S (3.0)
|  |  Nota-recencia = 5: N (27.0/7.0)
|  nota-receita = 5: N (6.0)
Nota-compras = 3
|  nota-receita = 2: N (2.0)
|  nota-receita = 3: N (26.0/10.0)
|  nota-receita = 4: S (86.0/30.0)
|  nota-receita = 5: S (32.0/12.0)
Nota-compras = 4: S (33.0/4.0)
Nota-compras = 5: S (22.0/2.0)
```



Number of Leaves : 23

Size of the tree : 28

Time taken to build model: 0.08 seconds

=== Evaluation on training set ===

=== Summary ===

Correctly Classified Instances	513	78.9231 %
Incorrectly Classified Instances	137	21.0769 %
Kappa statistic	0.5215	
Mean absolute error	0.3035	
Root mean squared error	0.3896	
Relative absolute error	66.4913 %	
Root relative squared error	81.5541 %	
Total Number of Instances	650	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.879	0.376	0.811	0.879	0.844	N
0.624	0.121	0.737	0.624	0.676	S

=== Confusion Matrix ===

a	b	<-- classified as
370	51	a = N
86	143	b = S

## *Cliente Jurídico e Parceiro*

=== Run information ===

Scheme: weka.classifiers.trees.j48.J48 -C 0.25 -M 2  
Relation: weka-juridpar-weka.filters.unsupervised.attribute.Remove-V-R2,4,6,7,9-weka.filters.unsupervised.attribute.Remove-R1  
Instances: 94  
Attributes: 4  
Nota-compras  
Nota-recencia  
Ativo  
nota-receita  
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

-----

Nota-compras = 1: N (27.0/5.0)  
Nota-compras = 2: N (29.0/11.0)  
Nota-compras = 3: S (18.0/3.0)  
Nota-compras = 4: S (13.0)  
Nota-compras = 5: S (7.0)

Number of Leaves : 5

Size of the tree : 6

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	72	76.5957 %
Incorrectly Classified Instances	22	23.4043 %
Kappa statistic	0.5369	
Mean absolute error	0.2992	
Root mean squared error	0.4013	
Relative absolute error	60.2633 %	
Root relative squared error	80.5596 %	
Total Number of Instances	94	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.86	0.314	0.698	0.86	0.771	N
0.686	0.14	0.854	0.686	0.761	S

=== Confusion Matrix ===

```
a b <-- classified as
37 6 | a = N
16 35 | b = S
```

### 9.3.1.b. Arquivo das Transações

#### *Cliente Físico*

Obs.: Somente algumas regras são apresentadas.

=== Run information ===

```
Scheme:      weka.classifiers.rules.part.PART -M 2 -C 0.25 -N 3
Relation:    GRUPO-FISICO-90dias-ativo-
weka.filters.unsupervised.attribute.Remove-V-R8,10,12,15,17,18,19-
weka.filters.unsupervised.attribute.Remove-V-R2-6
Instances:   18205
Attributes:  5
              recencia-nota
              Ativo
              valor-acumulado-nota
              frequencia-nota
              cidade
Test mode:   evaluate on training data
```

=== Classifier model (full training set) ===

PART decision list

-----

```
valor-acumulado-nota = 4 AND
cidade = 16 AND
frequencia-nota = 5: S (147.0)
```

```
valor-acumulado-nota = 4 AND
cidade = 31: S (86.0/4.0)
```

```
valor-acumulado-nota = 3 AND
cidade = 16 AND
frequencia-nota = 4: S (79.0/7.0)
```

```
.
.
.
```

```
recencia-nota = 0: N (1022.0/257.0)
```

```
cidade = 25: N (37.0/2.0)
```

```
cidade = 31 AND
valor-acumulado-nota = 1: N (112.0/21.0)
```

```
cidade = 18 AND
recencia-nota = 5: N (121.0/29.0)
```

```
.
.
.
```

```
Number of Rules :      108
```

Time taken to build model: 3.11 seconds

```
=== Evaluation on training set ===
=== Summary ===
```

Correctly Classified Instances	12037	66.1192 %
Incorrectly Classified Instances	6168	33.8808 %
Kappa statistic	0.2942	
Mean absolute error	0.4206	
Root mean squared error	0.4586	
Relative absolute error	86.3979 %	
Root relative squared error	92.9506 %	
Total Number of Instances	18205	

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.745	0.454	0.694	0.745	0.719	S
0.546	0.255	0.606	0.546	0.574	N

```
=== Confusion Matrix ===
```

```

      a      b  <-- classified as
7877 2702 |      a = S
3466 4160 |      b = N
```

*Cliente Jurídico e Parceiro*  
*Atributo meta: cliente ativo*

Obs.: Somente algumas regras são apresentadas.

=== Run information ===

```
Scheme:      weka.classifiers.rules.part.PART -M 2 -C 0.25 -N 3
Relation:     GRUPO-juridpar-60dias-sem295-ativo-
weka.filters.unsupervised.attribute.Remove-V-R8,10,12,13,15,17,18,19-
weka.filters.unsupervised.attribute.Remove-R1,4,8
Instances:    6535
Attributes:    5
               recencia-nota
               ativo
               valor-acumulado-nota
               frequencia-nota
               cidade
Test mode:     evaluate on training data
=== Classifier model (full training set) ===
PART decision list
```

cidade = 14: S (778.0/68.0)

cidade = 30: S (240.0/20.0)

cidade = 31: S (167.0)

cidade = 12: S (147.0)

.  
.  
.

cidade = 8: N (140.0)

cidade = 16 AND  
valor-acumulado-nota = 1: N (107.0)

cidade = 19: N (43.0)

.  
.  
.

Number of Rules : 29

Time taken to build model: 0.31 seconds

```

=== Evaluation on training set ===
=== Summary ===

```

Correctly Classified Instances	6170	94.4147 %
Incorrectly Classified Instances	365	5.5853 %
Kappa statistic	0.7176	
Mean absolute error	0.1007	
Root mean squared error	0.2244	
Relative absolute error	43.9966 %	
Root relative squared error	66.342 %	
Total Number of Instances	6535	

```

=== Detailed Accuracy By Class ===

```

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.992	0.372	0.946	0.992	0.969	S
0.628	0.008	0.923	0.628	0.748	N

```

=== Confusion Matrix ===

```

```

  a    b  <-- classified as
5629  45 |    a = S
 320 541 |    b = N

```

*Cliente Jurídico e Parceiro*  
*Atributo Meta: Valor acumulado*

Obs.: Somente algumas regras são apresentadas.

=== Run information ===

```
Scheme:          weka.classifiers.rules.part.PART -M 2 -C 0.25 -N 3
Relation:        GRUPO-juridpar-60dias-sem295-ativo-
weka.filters.unsupervised.attribute.Remove-V-R8,10,12,13,15,17,18,19-
weka.filters.unsupervised.attribute.Remove-V-R1,2,3,4,5,6,7
Instances:       6535
Attributes:      7
                  dia-semana
                  recencia-nota
                  ativo
                  tipo-cliente
                  valor-acumulado-nota
                  frequencia-nota
                  cidade
Test mode:       evaluate on training data
```

=== Classifier model (full training set) ===

PART decision list  
-----

```
frequencia-nota = 5 AND
dia-semana = 7: 4 (179.0/13.0)
```

```
frequencia-nota = 1 AND
cidade = 28 AND
tipo-cliente = 3: 1 (46.0/1.0)
```

```
frequencia-nota = 5 AND
dia-semana = 5: 4 (164.0/16.0)
```

```
.
.
.
```

```
frequencia-nota = 1 AND
cidade = 9 AND
dia-semana = 2: 1 (221.0/17.0)
```

```
cidade = 8: 5 (140.0/3.0)
```

```
frequencia-nota = 1 AND
cidade = 9 AND
dia-semana = 6: 1 (210.0/28.0)
```

Number of Rules : 129

Time taken to build model: 1.25 seconds

```
=== Evaluation on training set ===
=== Summary ===
```

Correctly Classified Instances	4656	71.2471 %
Incorrectly Classified Instances	1879	28.7529 %
Kappa statistic	0.629	
Mean absolute error	0.1573	
Root mean squared error	0.2805	
Relative absolute error	50.8539 %	
Root relative squared error	71.3128 %	
Total Number of Instances	6535	

```
=== Detailed Accuracy By Class ===
```

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.831	0.084	0.814	0.831	0.822	1
0.557	0.102	0.63	0.557	0.591	2
0.63	0.098	0.609	0.63	0.619	3
0.779	0.051	0.76	0.779	0.769	4
0.772	0.033	0.692	0.772	0.73	5

```
=== Confusion Matrix ===
```

	a	b	c	d	e	<-- classified as
1670	270	31	31	7		a = 1
260	868	331	85	15		b = 2
89	197	800	79	105		c = 3
31	17	134	880	68		d = 4
2	26	18	83	438		e = 5



### 9.3.2. Resultados Associação

Serão apresentadas algumas regras selecionadas, pois não se faz necessário a apresentação de todas, uma vez que as mais importantes já foram mencionadas. Além disso, somente a métrica do arquivo geral será apresentada, não havendo necessidade das demais, já que estas obedeceram aos mesmos padrões.

#### 9.3.2.a. Métrica utilizada

=== Run information ===

```
Scheme:          weka.associations.Apriori -N 20000 -T 1 -C 1.1 -D 0.05 -U 1.0
-M 0.05 -S -1.0
Relation:         assoc-clientel-v2-weka.filters.unsupervised.attribute.Remove-
R1,9-weka.filters.unsupervised.attribute.Remove-R8-
weka.filters.unsupervised.attribute.Remove-R4,5
Instances:        18274
Attributes:        126
                   [list of attributes omitted]
=== Associator model (full training set) ===
```

Apriori  
=====

Minimum support: 0.05  
Minimum metric <lift>: 1.1  
Number of cycles performed: 19

#### 9.3.2.b. Regras – Arquivo Geral

Obs.: Somente algumas regras são apresentadas.

##### *Cliente Físico*

Laranja

ALFACE=10 CENOURA=37 LIMAO=65 1346 ==> LARANJAS=63 1008 conf:(0.75) < lift:(2.26)>  
dia-semana=5 cidade=9 2157 ==> rota=24 LARANJAS=63 1158 conf:(0.54) < lift:(8.43)>

Banana

periodo-mes=4 TOMATE=110 1730 ==> quinzena-mes=2 BANANA=20 1025 conf:(0.59) < lift:(3.14)>  
ALFACE=10 BATATINHA=24 LARANJAS=63 1420 ==> BANANA=20 1066 conf:(0.75) < lift:(1.99)>

Cenoura

rota=20 ALFACE=10 1723 ==> dia-semana=6 CENOURA=37 943 conf:(0.55) < lift:(8.38)>  
periodo-mes=4 TOMATE=110 1730 ==> quinzena-mes=2 CENOURA=37 1062 conf:(0.61) < lift:(3.33)>

Tomate

BATATINHA=24 LARANJAS=63 1755 ==> ALFACE=10 TOMATE=110 1028 conf:(0.59) < lift:(2.26)>  
ALFACE=10 BATATINHA=24 LARANJAS=63 1420 ==> TOMATE=110 1028 conf:(0.72) < lift:(2.21)> ALFACE=10

## Ovos

dia-semana=6 cidade=9 ALFACE=10 1839 ==> rota=20 OVOS-12-unid=87 1027 conf:(0.56) < lift:(8.08)>  
dia-semana=6 ALFACE=10 1868 ==> rota=20 OVOS-12-unid=87 1028 conf:(0.55) < lift:(7.96)>

## Alface

dia-semana=6 CENOURA=37 1194 ==> rota=20 ALFACE=10 943 conf:(0.79) < lift:(8.38)>  
rota=20 OVOS-12-unid=87 1263 ==> dia-semana=6 cidade=9 ALFACE=10 1027 conf:(0.81) < lift:(8.08)>

## *Cliente Jurídico e Parceiro*

## Ovos

cidade=14 rota=25 397 ==> dia-semana=2 OVOS-CASCA-FINA=85 328 conf:(0.83) < lift:(10.04)>  
dia-semana=2 cidade=14 401 ==> rota=25 OVOS-CASCA-FINA=85 328 conf:(0.82) < lift:(9.99)>

## Cenoura

ALFACE=10 LARANJAS=63 596 ==> BANANA=20 CENOURA=37 373 conf:(0.63) < lift:(7.65)>  
dia-semana=7 ALFACE=10 468 ==> cidade=9 CENOURA=37 344 conf:(0.74) < lift:(4.58)>

## Chicoria

LARANJAS=63 OVOS-12-unid=87 529 ==> ALFACE=10 BANANA=20 387 conf:(0.73) < lift:(5.67)>  
ACELGA=5 ALFACE=10 CENOURA=37 483 ==> CHICORIA=38 347 conf:(0.72) < lift:(5.63)>

## Temperos

MANDIOCA=68 509 ==> BANANA=20 temperos=109 338 conf:(0.66) < lift:(7.2)> lev:(0.04)  
CHICORIA=38 RUCULA=102 508 ==> ALFACE=10 temperos=109 379 conf:(0.75) < lift:(4.83)>

## Laranja

BANANA=20 CENOURA=37 535 ==> ALFACE=10 LARANJAS=63 373 conf:(0.7) < lift:(7.65)>  
MANDIOCA=68 509 ==> ALFACE=10 LARANJAS=63 344 conf:(0.68) < lift:(7.41)>

## Banana

ALFACE=10 LARANJAS=63 596 ==> BANANA=20 CENOURA=37 373 conf:(0.63) < lift:(7.65)>  
MANDIOCA=68 509 ==> BANANA=20 temperos=109 338 conf:(0.66) < lift:(7.2)> lev:(0.04) [291]

## Acelga

ALFACE=10 CENOURA=37 CHICORIA=38 601 ==> ACELGA=5 347 conf:(0.58) < lift:(4.44)>

## Mandioca

ALFACE=10 LARANJAS=63 596 ==> MANDIOCA=68 344 conf:(0.58) < lift:(7.41)> lev:(0.05)  
BANANA=20 temperos=109 603 ==> MANDIOCA=68 338 conf:(0.56) < lift:(7.2)> lev:(0.04)

## Tomate

CENOURA=37 MILHOS=76 571 ==> TOMATE=110 337 conf:(0.59) < lift:(4.04)>

Rúcula

CENOURA=37 ESPINAFRE=44 539 ==> ALFACE=10 RUCULA=102 337 conf:(0.63) < lift:(5.09)>  
ALFACE=10 CENOURA=37 CHICORIA=38 601 ==> RUCULA=102 392 conf:(0.65) < lift:(4.97)>

Alface

BANANA=20 CENOURA=37 535 ==> ALFACE=10 LARANJAS=63 373 conf:(0.7) < lift:(7.65)>  
COUVE=41 LARANJAS=63 443 ==> ALFACE=10 BANANA=20 335 conf:(0.76) < lift:(5.86)>

### 9.3.2.c. Regras – Arquivo Pré-Pós Feriados

#### *Cliente Físico*

pre-feriado=S ALFACE=10 BATATINHA=24 70 ==> TOMATE=110 51 conf:(0.73) < lift:(2.61)>  
feriado=S BATATINHA=24 87 ==> TOMATE=110 60 conf:(0.69) < lift:(2.47)>  
pos-feriado=S LARANJAS=63 OVOS-12-unid=87 58 ==> ALFACE=10 CENOURA=37 40 conf:(0.69) < lift:(2.46)>  
pre-feriado=S ALFACE=10 BANANA=20 CENOURA=37 73 ==> TOMATE=110 50 conf:(0.68) < lift:(2.45)>  
pre-feriado=S BANANA=20 BATATINHA=24 61 ==> TOMATE=110 41 conf:(0.67) < lift:(2.41)>

#### *Cliente Jurídico e Parceiro*

pre-feriado=S BANANA=20 23 ==> LARANJAS=63 17 conf:(0.74) < lift:(6.53)>  
pre-feriado=S LARANJAS=63 20 ==> BANANA=20 17 conf:(0.85) < lift:(6.25)>  
pre-feriado=S ALFACE=10 BANANA=20 18 ==> CENOURA=37 16 conf:(0.89) < lift:(6.24)>  
pre-feriado=S ALFACE=10 COUVE=41 17 ==> CENOURA=37 15 conf:(0.88) < lift:(6.2)>

#### 9.4. Checagem dos Dados – utilizado no processo de monitoração dos clientes

As tabelas 23, 24, 25 e 26 apresentam os registros referentes às últimas compras, utilizados no processo de checagem dos dados do arquivo das transações.

##### *Cliente Físico*

**Tabela 23: Caso Valinhos A(180/60) – perfil 355**

Pedido	Cliente	Ativo	Valor	Freq.	Recência	Cidade	Rota
22903	986	S	4	2	4	30	22
23071	986	S	4	2	5	30	22
23564	986	S	4	2	3	30	22
23908	986	S	4	2	4	30	22
24071	986	S	4	2	5	30	22
24237	986	S	4	3	5	30	22
24407	986	S	4	3	5	30	22
25063	986	S	4	2	2	30	22
25398	986	S	4	2	4	30	22
26916	986	S	2	1	1	30	22

**Tabela 24: Caso Vinhedo NF3(17)/A(7) – perfil 355**

Pedido	Cliente	Ativo	Valor	Freq.	Recência	Cidade	Rota
15241	768	N	2	2	5	31	22
15474	768	N	3	2	5	31	22
15690	768	N	3	2	5	31	22
16521	768	N	3	3	2	31	22
18132	768	N	1	1	1	31	22
18299	768	N	2	1	5	31	22
18464	768	N	2	1	5	31	22
18819	768	N	2	1	4	31	22
19534	768	N	2	2	2	31	22

*Cliente Jurídico*

**Tabela 25: Caso Cajamar N (140/3) – perfil 352**

Pedido	Recência	Cliente	Ativo	Valor	Freq.	Recência	Cidade	Rota
19366	3	758	N	5	4	3	8	1
19438	3	758	N	5	4	3	8	1
19560	3	758	N	5	4	3	8	1
19632	3	758	N	5	4	3	8	1
19814	2	758	N	5	3	2	8	1
19933	3	758	N	5	3	3	8	1
20812	1	758	N	5	1	1	8	1

**Tabela 26: Caso São José dos Campos AF2(15/1) – perfil 242**

Pedido	Recência	Cliente	Ativo	Valor	Freq.	Recência	Cidade	Rota
26165	2	1300	S	5	2	2	26	2
26310	2	1300	S	5	2	2	26	19
26425	1	1300	S	4	1	1	26	2
26579	2	1300	S	4	1	2	26	2
26725	2	1300	S	4	1	2	26	2
26884	2	1300	S	4	1	2	26	2
27056	2	1300	S	4	1	2	26	2
27222	2	1300	S	4	1	2	26	2