

APLICAÇÃO DE MODELO DE MINERAÇÃO DE DADOS EM UM SISTEMA DE  
APOIO A DECISÃO PARA EMPRESAS DE SANEAMENTO

Roberto Queyroi

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM SISTEMAS  
COMPUTACIONAIS EM ENGENHARIA CIVIL

Aprovada por:

---

Prof. Nelson Francisco Favilla Ebecken, D.Sc.

---

Prof. Alexandre Gonçalves Evsukoff, D.Sc.

---

Prof. Elton Fernandes, Ph. D.

RIO DE JANEIRO, RJ – BRASIL

JUNHO DE 2007

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

QUEYROI, ROBERTO

Aplicação de Modelo de Mineração de  
Dados em Um Sistema de Apoio a  
Decisão Para Empresas de Saneamento  
[Rio de Janeiro] 2007.

V, 102 p., 29,7 cm (COPPE/UFRJ,  
M.Sc., Engenharia Civil, 2007)

Dissertação - Universidade Federal do  
Rio de Janeiro, COPPE.

1. Data Mining,
2. Sistema de Apoio a Decisão,
2. Gestão do Conhecimento,
2. Árvores de Decisão,

I. COPPE/UFRJ II. Título (série )

## AGRADECIMENTOS

Agradeço ao meu orientador, Nelson Ebecken, pela confiança em acreditar na possibilidade de desenvolvimento deste trabalho acadêmico e acredito trará complementarmente desdobramentos empresariais.

Gostaria de demonstrar meu reconhecimento especial aos colegas e amigos Cristian Klen, Nicomar Oliveira e Marcos Braga, por me auxiliarem de forma direta e indireta na confecção deste trabalho.

Aos demais colegas com quem passo a maior parte do tempo, agradeço as trocas proporcionadas pelo aprendizado no dia a dia, pelas experiências vividas e principalmente pelo crescimento profissional e pessoal.

A Ivanete, Adriano e Fernando, família que sempre me apóia e incentiva especialmente.

*Caminante no hay camino...*

*se hace camino al andar.*

Antonio Machado

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## APLICAÇÃO DE MODELO DE MINERAÇÃO DE DADOS EM UM SISTEMA DE APOIO À DECISÃO PARA EMPRESAS DE SANEAMENTO

Roberto Queyroi

Junho / 2007

Orientador: Nelson Francisco Favilla Ebecken

Programa: Engenharia Civil

Este trabalho possuiu dois objetivos principais: o primeiro estudar possibilidades de modernização para o segmento de saneamento no Brasil. Este setor que provê serviços essenciais à população que são água potável e tratamento de esgotos e que até então, operava sem necessariamente visar lucratividade e, atualmente se depara com desafios que aumentam esta responsabilidade, exigindo dos gestores e técnicos da área maior profissionalização e desempenho competitivo comparável aos demais serviços prestados à sociedade como transportes, energia elétrica e telecomunicações. Trata ainda de identificar e estabelecer competências essenciais para impulsão de negócios trazendo mais eficiência e menores custos para o consumidor e clientes.

Em segundo lugar, são abordados neste trabalho, técnicas, metodologias e sistemas de apoio à decisão que vem revolucionando os demais segmentos de negocio, suportados principalmente por ferramental de mineração de dados e sistemas inteligentes. Estas ferramentas associadas às praticas de gestão do conhecimento nas empresas de saneamento sem duvida proporcionarão a mudança de paradigmas ultrapassados e até então consolidados, na busca da governança corporativa e da transparência administrativa exigida pela sociedade.

Summary of the dissertation presented to COPPE/UFRJ as part of the necessary requirements for the obtaining of Master's degree in Sciences (M.Sc.)

APPLICATION MODEL OF DATA MINING IN A SUPPORT DECISION SYSTEM  
FOR COMPANIES OF SANITATION

Roberto Queyroi

June / 2007

Advisor: Nelson Francisco Favilla Ebecken

Department: Civil Engeneering

This dissertation has two main objectives: the first to study modernization possibilities for the segment of sanitation in Brazil. This business area that provides essential services to the population in distribution water and treatment of sewers, without necessarily being guided to seek profitability and nowadays challenges extrapolate this mission, demanding from the managers and workers of this especial area, competitiveness and performance comparable to the services as transports, electric power and telecommunications. It still search to identify and to establish essential competences for impulse of businesses bringing more efficiency and smaller costs for the consumer and customers.

In second place, are approached in this work, techniques, methodologies and decision support systems that is revolutionizing the other business segments, supported mainly by data mining tools and business intelligence systems. These tools associated to knowledge Management practices in the sanitation corporations will certainly provide the change from oldest paradigms and until then consolidated, in the search of the corporate governance and management transparency demanded by the society.

## ÍNDICE

<b>1 - INTRODUÇÃO.....</b>	<b>1</b>
1.1 - Objetivos .....	1
1.2 - Desenvolvimento do Trabalho .....	2
1.3 - A Escolha do Tema e Justificativa .....	3
1.4 - Controle de perdas.....	6
1.4.1 Perdas Reais.....	7
1.4.2 Perdas Aparentes .....	7
Ligações Clandestinas .....	8
1.4.3 Os Medidores de Consumo (Hidrômetros) - Características Operacionais .....	9
<b>2 - GESTÃO DO CONHECIMENTO E A ESTRATÉGIA EMPRESARIAL .....</b>	<b>11</b>
2.1 - Competências organizacionais .....	13
2.1.1 Recursos, Competências e Vantagem Competitiva.....	14
2.1.2 Competências essenciais .....	15
2.1.3 Competências organizacionais básicas.....	16
2.2 - Competências, Processos da Organização e Estratégias .....	16
2.3 - A Gestão do Conhecimento e as Estratégias.....	20
2.3.1 A Construção do Conhecimento.....	20
2.3.2 O Desperdício de Talentos .....	23
2.3.3 Quando as Perguntas que nos Ensinam mais do que as Respostas .....	24
2.3.4 O que os funcionários sabem e o que realmente aplicam.....	24
2.4 - Inteligência Aplicada aos Negócios .....	25
<b>3 - SISTEMAS INTELIGENTES DE APOIO A DECISÃO.....</b>	<b>27</b>
3.1 - Dados, Informação e Conhecimento .....	27
3.2 - O Domínio das Regras de Negócios nas Empresas .....	29

3.3 - Data Warehousing .....	30
3.4 - Mineração de Dados.....	33
3.4.1 Conhecimento.....	34
3.5 - Descobrimdo Conhecimento em Banco de Dados .....	36
3.5.1 Fase 1 - Entendimento do Domínio ou Identificação do Problema.....	38
3.5.2 Fase 2 - Pré-Processamento.....	40
3.5.3 Fase 3 - Transformação de Dados .....	42
3.5.4 Fase 4 – Mineração de Dados.....	43
3.5.5 Fase 5 - Pós-Processamento e Interpretação do Conhecimento .....	47
<b>4 - MINERAÇÃO DE DADOS APLICADO À TOMADA DE DECISÃO DE SUBSTITUIÇÃO DE HIDRÔMETROS.....</b>	<b>49</b>
4.1 - Política de Troca de Hidrômetros .....	50
4.2 - Conceituação do Problema.....	51
4.2.1 Modelo do Especialista - Resultados Após a Troca dos Hidrômetros .....	51
4.3 - Aplicação do Modelo de Mineração de Dados ao Problema .....	53
4.3.1 Seleção dos Dados.....	54
4.3.2 Pré-processamento.....	56
4.3.3 Integração dos Dados .....	57
4.3.4 Mineração de dados.....	58
<b>5 - CONCLUSÕES.....</b>	<b>61</b>
5.1 - Interpretação dos Resultados.....	61
5.2 - Trabalhos Futuros.....	64
<b>REFERÊNCIAS.....</b>	<b>65</b>
<b>APÊNDICE 1.....</b>	<b>68</b>
<b>APÊNDICE 2.....</b>	<b>71</b>



CADASTRO DE CONSUMIDORES.....	71
CADASTRO REGISTRO DE CONSUMO.....	72
CADASTRO DE IMÓVEIS.....	73
CADASTRO DE DÉBITOS EM ABERTO.....	74
CADASTRO DE HIDRÔMETRO.....	75
MOVIMENTAÇÃO DE HIDRÔMETROS.....	75
AMOSTRA DOS AGRUPAMENTOS OPERACIONAIS E MÉDIAS.....	76
<b>APÊNDICE 3.....</b>	<b>78</b>
Agrupamento Barra da Tijuca .....	79
Agrupamento Botafogo .....	80
Agrupamento Campo Grande.....	83
Agrupamento Centro .....	85
Agrupamento Ilha do Governador.....	87
Agrupamento Jacarepaguá.....	89
Agrupamento Jardim Primavera.....	91
Agrupamento Meier.....	93
Agrupamento Paquetá .....	95
Agrupamento Ramos .....	97
Agrupamento Rio Bonito .....	99
Agrupamento São Gonçalo.....	101
Agrupamento Teresópolis .....	103

## **1 - Introdução**

O ambiente de negócio que envolve as organizações tem sido alvo de significativas transformações, tanto devido à globalização dos mercados quanto pela transformação da tradicional economia industrial numa economia baseada cada vez mais centrada em conhecimento. Hoje registramos que práticas corporativas até então vencedoras no passado, tem sido insuficiente para responder a tempo e hora as mudanças que estamos assistindo.

Buscando manterem-se eficientes, dentro do mercado e competitivas, as corporações têm sido obrigadas a reorganizar e redesenhar continuamente os seus processos de negócios e a se reinventar adaptando-se às novas estratégias exigidas pelos mercados.

A eletrônica, telecomunicações, informática, telefonia dentre outras importantes evoluções tecnológicas alcançadas nos últimos anos, verifica-se que os sistemas de informação, que suportam o negócio, não respondem necessariamente e eficazmente às constantes mudanças a que a organização está sujeita, causando discrepâncias causado pelo abismo criado pelo conjunto de tecnologias disponíveis e os resultados ainda insuficientes obtidos na gestão dos negócios. Fato este que em última instância, reduz e em alguns casos inibe a capacidade competitiva da organização.

### **1.1 - Objetivos**

Neste trabalho discute-se o papel da *KDD - Knowledge Discovery in Database* e em ultima análise da Mineração de Dados, como ferramenta fundamental no apoio a tomada de decisão no setor de saneamento no Brasil e nesse sentido se busca dispor de um modelo que apóie a decisão de troca de hidrômetros de uma empresa de saneamento visando:

- Diminuir Perdas Aparentes ou não Físicas(volume micro medido)
- Garantindo que a troca permita remunerar o investimento feito (faturamento / arrecadação posterior a troca  $\geq 15\%$  do que o faturamento/arrecadação anterior)
- Explicitar e materializar o conhecimento obtido para que este possa ser entendido, utilizado e multiplicado por diversas pessoas da organização.

## **1.2 - Desenvolvimento do Trabalho**

São abordados nesta primeira parte

específico de domínio do problema é colocado que uma empresa seguindo sua política de diminuir perdas na micro medição e aumentar a arrecadação, estabelecida por especialistas, busca comparar os resultados históricos obtidos com aqueles provenientes de aplicação da modelagem com mineração de dados. Várias simulações são produzidas, apresentadas e discutidas pelo autor no capítulo 5 e as respectivas conclusões obtidas.

Além disso, sugestões para trabalhos futuros e aplicabilidade deste ferramental no segmento de saneamento poderão sem dúvida criar novas oportunidades para a tomada de decisão nestas empresas.

### **1.3 - A Escolha do Tema e Justificativa**

As empresas de saneamento no Brasil perderam em 2002 pelo menos 40,5% da água que corre no seu sistema de abastecimento. Ou seja, dos mais de 12 trilhões de litros produzidos e tratados no ano, cerca de 4,8 trilhões não chegaram a seu destino final ou não foram contabilizados. Numa estimativa aproximada, o volume seria suficiente para atender o Estado de São Paulo por dois anos.

Essa é uma das principais conclusões do relatório do *Sistema Nacional de Informações sobre Saneamento* (SNIS, 2002) o mais completo estudo sobre saneamento já feito no Brasil. A pesquisa abrange mais de 4 mil municípios, que abrigam 94,3% da população urbana nacional.

As perdas podem ser causadas por vazamentos e ausência de medição do consumo. No primeiro caso, os fatores que mais concorrem são; o desgaste ou a falta de manutenção na tubulação e nos demais equipamentos.

Os problemas na medição podem ocorrer por problemas nos medidores - hidrômetros (aparelhos que medem o consumo de água nos domicílios) ou pela falta dele. Há casos em que a empresa concessionária não instala o aparelho e cobra uma taxa estimada, sem medir o consumo com precisão e, nesse caso, pode se configurar a "perda" porque a empresa fornece um determinado volume de água e fatura baseando-se em um volume menor.

Em alguns casos, a leitura do hidrômetro não é feita com regularidade pela empresa concessionária de serviços. Em outros casos, o aparelho fica danificado ou quebra, mas a prestadora não troca nem manda consertarem tempo hábil. Se o medidor não funciona com precisão e registra um consumo menor do que o real, a água também é

contabilizada como "perdida" para a concessionária, porque ela não fatura. Há poucas pesquisas sobre o assunto, mas estima-se por amostragem que metade dos problemas que ocorrem sejam associados a vazamentos e outra metade decorrentes de falhas na medição.

Os níveis de perdas de água no Brasil até 2002 têm se mantido nos níveis de 2001. Os maiores problemas se concentram, segundo o relatório, na região Norte (52,8%), seguida do Nordeste (45,5%), Sudeste (38,2%), Sul (37,9%) e Centro-Oeste (29,2%).

Entre as empresas de abrangência regional, o maior índice de perda é da Companhia de Águas e Esgotos de Rondônia - CAERD — 71,3%. O menor índice é da Companhia de Saneamento do Distrito Federal - CAESB: 21,6%. A Companhia de Saneamento Básico do Estado de São Paulo – SABESP, responsável por 35% do faturamento do serviço de água do Brasil, tem perdas da ordem de 31,7%.

Apesar da perda elevada, o acesso a serviços de água no Brasil é considerado satisfatório pelos técnicos que elaboraram o relatório do SNIS. Em relação ao universo da pesquisa — que corresponde a **94,3%** da população urbana nacional —, 91,7% das pessoas que vivem em cidades têm água encanada em casa.

O volume de investimento ainda é baixo frente às necessidades do setor, mas aumentou de R\$950 milhões em 2001 para R\$1,050 bilhão em 2002. A tendência é de que permaneça crescendo.

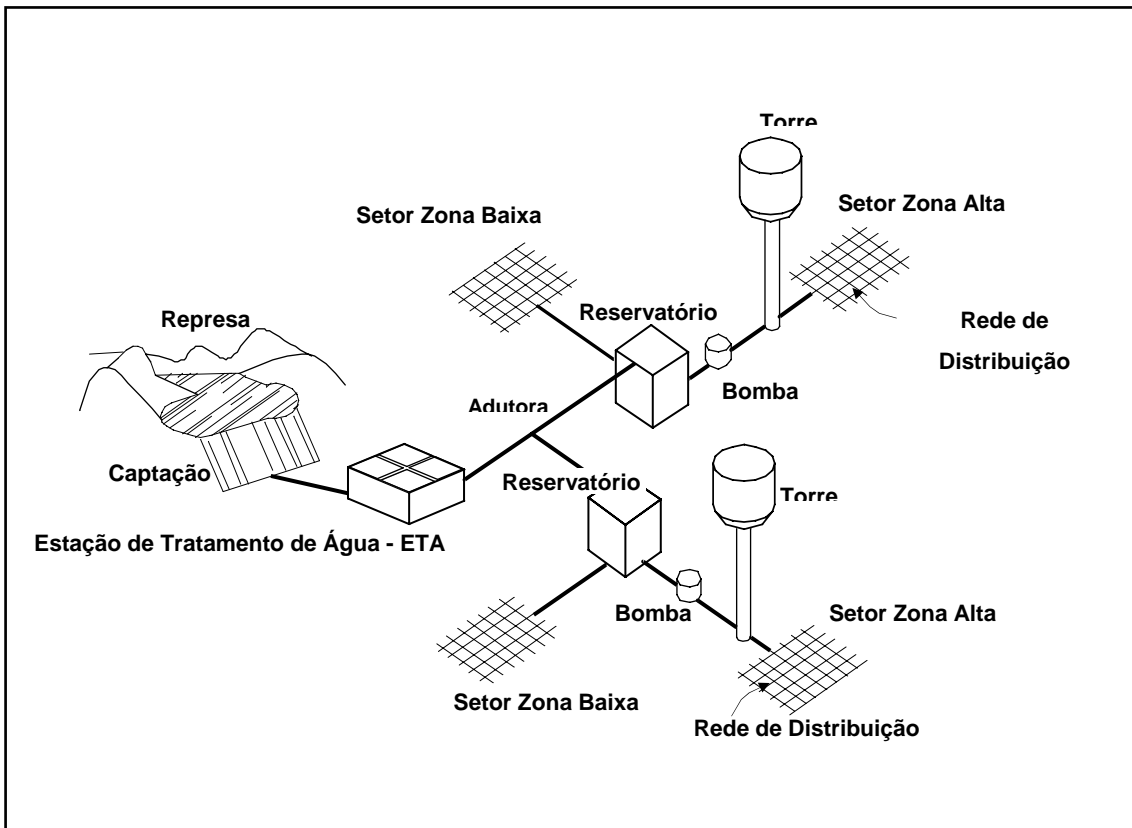


Figura 1 – Sistema de Abastecimento Típico de uma Cidade

Entre as operadoras de abrangência regional, a campeã de abastecimento é a Empresa de Saneamento de Mato Grosso do Sul (SANESUL), que atende a 100% da população urbana. A menos bem colocada é o Departamento Estadual de Água e Saneamento (DEAS), do Acre, que leva água a apenas 51,7% da população urbana.

Na Figura 1, é mostrado através de um desenho esquemático o percurso da água desde a sua captação em um determinado manancial, passando por sua estação de tratamento, sua adução até a sua distribuição ao consumidor.

## O Problema e o Apoio a Tomada de Decisão

Com a sofisticação das operações e controles administrativos proporcionados pela evolução da tecnologia nos últimos anos, muito se tem feito para aumentar a produtividade e racionalização dos custos nas empresas de saneamento. Porém, alguns pontos importantes merecem ser salientados principalmente aqueles que são ligados às perdas físicas provenientes de vazamentos e as perdas de arrecadação que tem como responsável maior com a ineficiência do setor.

Como foi mostrado anteriormente para que as empresas de saneamento melhorem sua eficiência, têm que atacar pontos comuns a estas, qual seja a diminuição de suas perdas que são qualificadas em Perdas Físicas e Perdas Aparentes.

Associadas às perdas aparentes temos questões de decisão a respeito do corte de fornecimento dos consumidores inadimplentes e fraudadores, bem como a **substituição periódica dos hidrômetros**.

### 1.4 - Controle de perdas

De acordo com GONÇALVES e ALVIM (2007) e a literatura a respeito classifica as perdas de fornecimento e abastecimento de água aos consumidores em dois tipos de perdas mostradas na Figura 2: as REAIS ou FÍSICAS e as APARENTES NÃO FÍSICAS:

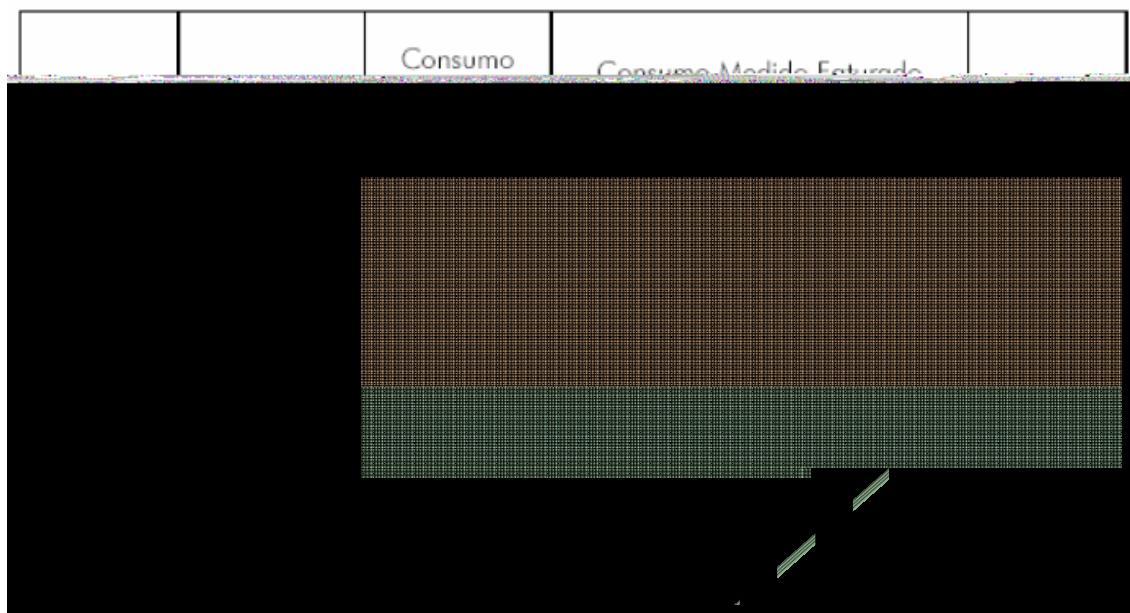


Figura 2 – Subdivisão do Consumo e Perdas

### **1.4.1 Perdas Reais**

Água efetivamente desperdiçada, que não chega ao consumidor, perdida através de vazamentos nas redes de distribuição e seus ramais e pode ser definida como Perdas Reais, também chamadas de Perdas Físicas. Estas perdas podem ser classificadas em:

- Perdas por Vazamentos – São as relacionadas ao rompimento em adutoras, subadutoras, redes, ramais e conexões e às trincas estruturais e fissuras nas impermeabilizações de reservatórios.
- Perdas Operacionais - são as relacionadas à operação do sistema, tais como: lavagem de filtros, lavagem e descargas de redes, descargas em hidrantes, limpezas e extravasamentos em reservatórios. No contexto brasileiro, não está definido o uso do consumo operacional na contabilização das perdas, motivo pelo qual a maioria das companhias considera esse consumo como perda real.

### **1.4.2 Perdas Aparentes**

Onde a água é tratada pela concessionária, fornecida e consumida, porém não é corretamente medida e, portanto não é faturada, nem gera arrecadação correspondente e podem ser decorrentes dos seguintes problemas:

As Perdas Aparentes, também chamadas de Perdas Não Físicas, referem-se a toda água que não é medida ou que não tenha o seu uso definido. Estão relacionadas às ligações clandestinas e/ou irregulares, fraudes nos hidrômetros, erros de micro medição e macro medição, política tarifária, erro cadastral (desatualização do cadastro, inatividade em ligação ativa, ligação não cadastrada por descuido), erro de leitura, etc.

Conforme o trabalho de MIRANDA *et al* (2003), recomenda-se como mais adequado o uso dos termos “perdas reais” e “perdas aparentes” em substituição a “perdas físicas” e “perdas não físicas”, respectivamente. Algumas causas para as Perdas Aparentes são apresentadas a seguir:

#### **Cadastro de Clientes desatualizado ou com erros**

Muita das vezes erros cadastrais proporcionam a devolução da conta por parte do entregador (Correio ou profissional contratado). Erros de endereço, CEP, Nome do responsável pelo consumo, locais de difícil acesso ou inseguros causam prejuízos às empresas e rotineiramente obrigam a concessionária despender recursos para recadastrar consumidores.

#### **Fraudes, Violação ou danificação do medidor**



Infelizmente alguns consumidores e clientes mal-intencionados buscam de diversas formas das mais criativas fraudar os hidrômetros com o objetivo de diminuir ou inibir aquilo que “lido” pelos equipamentos de medição, sendo mais comum:

Furar a cúpula da relojoaria e utilizar uma agulha para trancar os roletes, retirando-a quando na época de leitura.

**Retirar o hidrômetro do cavalete, retirar o filtro e injetar no mesmo impurezas** (como cola, pedrinhas, barbante, etc.) ou furar a câmara de medida com o objetivo de parar a turbina.

**Inverter o hidrômetro no quadro**, deixando-o funcionar apenas metade do tempo no sentido correto.

**Romper o lacre, abrir o hidrômetro e frear os roletes com calços** (palitos, pregos, etc.).

**Golpear ou queimar a cúpula do hidrômetro com a intenção de danificá-lo.**

#### Ligações Clandestinas

Segundo ALVIM e BARBOSA (1997), “As perdas aparentes podem ser estimadas a partir da diferença entre o volume perdido total e a perda obtida através da vazão mínima noturna (perda real). Uma vez que, geralmente, a parcela de contribuição das ligações clandestinas durante o período de vazão mínima noturna é muito pequena comparada ao valor total da vazão mínima noturna, a perda real pode ser considerada como a própria perda obtida pela vazão mínima noturna”.

Especificação técnica, inspeção, dimensionamento e manutenção de hidrômetros –  
Gestão do Parque;

Medidores (Hidrômetros Defeituosos ou Inadequados) - Principais Ações

Algumas ações combinadas com outras se praticadas sistematicamente pelas empresas concessionárias surtem efeito desejado de diminuição das perdas aparentes:

Combate periódico às irregularidades;

Recadastramento periódico de Clientes combinando com cadastros de fontes alternativas (Prefeituras, Receita Federal, SERASA..).

Aferição de hidrômetros a pedido dos clientes e de terceiros;

Monitoramento dos grandes consumidores;

Padronização de ligações de água;

Individualização de ligações de água em condomínios;

Substituição de redes e ramais;

Fiscalização de instalações hidráulicas prediais para liberação de Alvará e Habite-se.

### **1.4.3 Os Medidores de Consumo (Hidrômetros) - Características Operacionais**

A medição, seja ela em qualquer atividade industrial ou comercial é fundamental para controle dos gastos, bem como para prover resultados para a recita das empresas.

Comparando os equipamentos voltados para medição tanto no setor elétrico como no de telecomunicações, percebe-se uma diferença fundamental os medidores destes dois últimos são baseados em tecnologias que lhes oferece vida útil bem maior daqueles utilizados pelas empresas fornecedoras dos serviços de água e esgotos.

O hidrômetro, figura 3, tem sua vida útil esgotada em geral após quatro ou cinco anos de uso, pois seu desgaste é devido principalmente à passagem de água pelo seu hélice ou turbina quer com o tempo de uso cria incrustações e sedimentação de material mineral que dificultam sua movimentação normal.

O Hidrômetro, equipamento fundamental na aferição do consumo de água tratada fornecida aos consumidores, tem sido o vilão quando constatado seu mau funcionamento fato que interfere nas medição daquilo que efetivamente foi entregue pela concessionária.

Problemas de mau funcionamento do hidrômetro devem-se a alguns fatores os quais são apresentados a seguir:



Figura 3 – Hidrômetros de Diversas Bitolas

#### **Desgaste das peças internas**

Ocasionada pelo tempo de uso ou pela passagem da água em uma vazão elevada (acima da capacidade nominal do medidor) durante certo tempo:

Modifica o erro de medição fazendo com que o medidor saia das faixas de medição definidas em norma.

### **Defeito de fabricação**

Os hidrômetros são aferidos em seu processo fabril, mas mesmo assim, por diversos motivos esta aferição pode acarretar folga nos roletes da relojoaria ou permitir que o eixo da mesma caia de seus mancais quando submetido a uma batida forte:

Neste caso, este altera o consumo medido, pois a leitura não é real e sim fruto do posicionamento aleatório dos roletes. Este problema acompanhará sempre o aparelho, sendo detectado quando feito o exame da relojoaria.

Perda da capacidade de magnetização dos ímãs do sistema de transmissão:

O medidor tende a parar, ou apresentar um movimento de "para e anda" em função da vazão de abastecimento. O consumo medido é sempre menor que o real, ou mesmo não é lido. Também, é facilmente confirmado numa inspeção do hidrômetro.

### **Vazamento na cúpula do hidrômetro**

A água sai por cima do medidor, o aperto dado no anel da cúpula quando da montagem não foi suficiente para garantir a vedação da relojoaria. O vazamento de água ocorre somente a partir de determinada pressão e quando isso acontece à mesma não é registrada pelo hidrômetro e nem altera o consumo, pois não ocasiona movimento na turbina da câmara de medida, ou seja, dos totalizadores da relojoaria.

É bastante comum o usuário confundir os vazamentos no quadro ou cavalete, informando que é o hidrômetro que está vazando. Sem dúvida, a troca de um medidor pode acarretar vazamentos junto ao quadro, principalmente nas manutenções preventivas, pois os ramais tendem a esclerosar-se com o tempo e o problema manifestar-se horas após o serviço de substituição ter sido realizado. Somente os vazamentos que ocorrem após o medidor são registrados pelo hidrômetro, podendo alterar o consumo.

## 2 - Gestão do Conhecimento e a Estratégia Empresarial

Em SLACK (1993), o desafio para os gerentes não se limita ao conhecimento e comando da detalhada complexidade das operações das organizações, apesar da importância que isto representa. O grande desafio é se os gerentes podem obter suficiente conhecimento das operações para adequá-las a um contexto estratégico.

O maior percentual de ativos e a maior parte de pessoal em geral se concentra na parte operacional da empresa, e que segundo SLACK(1993) na maioria das vezes, se caracteriza como o grande e pesado fardo a ser carregado. O autor argumenta ainda ser um mal-entendido considerar o setor “core” desta forma e descreve: *“Uma função de manufatura saudável dá à empresa a força para suportar o ataque da concorrência, dá o vigor para manter um melhoramento uniforme no desempenho competitivo [...] e proporciona versatilidade operacional... “..*

Todas as decisões e escolhas citadas devem ser baseadas numa estratégia de manufatura, uma vez que, como em outras áreas, no setor de produção também é impossível ter desempenho ótimo em tudo. Assim, deve-se priorizar uma ou poucas dimensões para ser eficiente e competitivo.

Segundo MINTZBERG (2001), boa parte das estratégias se origina na base da corporação, onde se conhece a fundo do negócio, durante a realização de atividades rotineiras. O papel do planejador ou do executivo é apenas traduzir essas estratégias de forma que toda a empresa as veja com a clareza necessária.

SLACK (1993) pondera que vantagem da área de produção se reflete em “fazer melhor”. A expressão que o autor utiliza está associada ao consumidor e ao concorrente.

Nesse caso, os consumidores e os concorrentes são ambos centrais para uma operação competitiva, porque definem suas metas sucintamente: satisfazer o primeiro (fazer melhor para agradá-lo) e ser melhor que o segundo (fazer melhor para manter a eficiência).

Uma operação fabril de sucesso baseia-se em trazer para dentro da área de operação ou produção a mentalidade que considere tanto o consumidor quanto os concorrentes. Os consumidores para atuarem como elementos que constantemente lembram quais os aspectos da competitividade são importantes para eles. Os concorrentes para proporcionar a medida contra o qual o desempenho da operação será julgado. No

entanto, mais do que isso, considerar de forma conjunta o que é importante para o consumidor e a medida de desempenho dos concorrentes, possibilita estabelecer as prioridades para aperfeiçoamento da operação.

Nas operações da empresa uma das metas é atingir uma vantagem preponderante no seu mercado, visando proporcionar o principal diferencial perante a concorrência. Para SLACK (1993), todos os aspectos fundamentais da competitividade estão claramente dentro do campo de atuação da função da produção e ainda, para “fazer melhor” significa cinco coisas:

**Fazer certo ou com qualidade** – não cometer erros, fazer produtos que realmente são o que devem ser, produtos sem erros e sempre de acordo com as especificações de projeto. Através desse conceito o setor que produz bens e serviços dá uma “vantagem de qualidade” à empresa;

**Fazer rápido ou com agilidade** – fazer com que o intervalo de tempo entre o início do processo de manufatura e a entrega do produto ao cliente seja menor do que a concorrência. Setor de produção proporciona uma “vantagem velocidade” à empresa;

**Fazer a tempo ou ser confiável** – manter a promessa de prazos de entrega, ou alternativamente, aceitar as datas de entrega solicitadas pelo cliente e cumpri-las pontualmente. Fazer pontualmente também aborda as entregas internas, ou seja, para clientes internos. Com isso, setor de produção dá à empresa a “vantagem da confiabilidade”;

**Ser flexível ou adaptável** – ser capaz de variar e adaptar a operação, seja porque as necessidades dos clientes são alteradas, seja devido a mudanças no processo de produção causadas, talvez, por mudança no suprimento dos recursos. Significa estar apto a mudar quando e o quanto seja necessário e com rapidez suficiente. Com isso a função manufatura dá à empresa a “vantagem da flexibilidade”;

**Fazer mais barato** – fabricar produtos a custos mais baixos do que os concorrentes conseguem administrar.

A médio e longo prazo, a única forma de se conseguir isto é através da obtenção de insumos mais baratos e/ou dispor de processos de produção mais eficientes e enxutos.

Assim, o setor de produção proporciona à empresa uma “vantagem de custo”.

Portanto, são esses os cinco objetivos de desempenho do setor de produção propostos por: **qualidade, agilidade, confiabilidade, flexibilidade e custo** (SLACK, 1993).

Dentro do setor de produção, o sistema de produção em vigor pode apresentar vantagens e desvantagens para alguns destes objetivos.

Qualquer operação deveria ser capaz de estabelecer a importância relativa dos seus objetivos de desempenho e julgar o seu desempenho alcançado em termos de cada um deles. Dessa forma é importante distinguir entre os aspectos internos e externos de cada objetivo de desempenho.

Se todas as partes da operação (compras, preparação das máquinas, processamento do produto, logística interna) forem ágeis e responderem prontamente nas suas tarefas perante as demais, a habilidade da operação total de responder rapidamente ao consumidor é aumentada. No ambiente interno do setor de produção, alcançar o objetivo de desempenho pode-se traduzir em diversos caminhos.

A redução do tempo da operação como um todo fica impossibilitado se há falhas no fornecimento de matérias-primas, fluxo de produção não balanceado, resultando em pouca credibilidade nas entregas, falhas no fornecimento, dentre outras.

O aprimoramento e a flexibilidade também colaboram com a velocidade. Todavia, os benefícios externos da velocidade são; a possibilidade dos setores oferecerem menores tempos de fornecimento e atendimento. Internamente, porém, esta questão é mais complexa, pois os objetivos de desempenho individuais influenciam e são influenciados pelos demais. “Internamente todos os objetivos de desempenho são importantes. O que varia é o porquê de eles serem importantes” (SLACK, 1993).

## **2.1 - Competências organizacionais**

As práticas internas eficazes, por mais convincentes que sejam os argumentos a seu favor, não estão isentas de dificuldades para a implementação da estratégia. Não importa quanto eficaz seja a estratégia, ela precisa ter sustentabilidade, além de contemplar o crescimento. Depois que uma empresa escolher sua estratégia de crescimento, ela deverá ter condições de se manter crescendo.

Assim, é fundamental que uma organização, seja esta estatal ou privada, para crescer, além da escolha acertada da estratégia, deve empenhar-se profundamente na sua implementação. A eficácia nos processos internos leva as empresas competirem com êxito durante certo período de tempo. Porém, em longo prazo, a eficácia nos processos internos não cria sustentabilidade para a estratégia. Aos poucos, as empresas começam a enfrentar dificuldades cada vez maiores em manter-se à frente dos concorrentes.

### 2.1.1 Recursos, Competências e Vantagem Competitiva

É fundamental que se faça uma avaliação dos recursos humanos, físicos e materiais que deverão suportar uma implementação futura de um projeto determinado. Na tabela 1 adaptada de JOHNSON *et al* (1999), deverão ser medidos e pensados os componentes que alicerçarão uma determinada proposta de mudança:

Tabela 1 – Recursos e Competências adaptado de Johnson et al (1999)

	SEMELHANTE AO DA CONCORRÊNCIA FÁCIL DE SER IMITADO	MELHOR DO QUE O DO CONCORRENTE DIFÍCIL DE SER IMITADO
RECURSOS	RECURSOS BÁSICOS Espaço Físico, maquinário, equipamentos de informática, software e hardware, Pessoal Técnico, Gestores e Usuários	RECURSOS DIFERENCIADOS Equipe para mudança, novos equipamentos, novo hardware e software, ....
COMPETÊNCIAS	COMPETÊNCIAS BÁSICAS Conhecimento da função, das regras de negócio atual, da estrutura de poder, das relações com fornecedores e clientes	COMPETÊNCIAS ESSENCIAIS Domínio das regras de negócio remodelado, pessoal capaz de absorver as novas tecnologias e sistemas, conhecimento profundo das competências dos concorrentes

Isso ocorre devido os concorrentes imitarem com rapidez as suas práticas gerenciais, em operações, inovações ou em gerenciamento com os clientes.

Em vista deste aspecto, o estudo das competências organizacionais pode responder a algumas questões importantes na sustentabilidade da estratégia empresarial.

É consenso no mundo empresarial e acadêmico a importância das competências organizacionais para a construção da vantagem competitiva empresarial. Assim, este assunto tem instigado e desafiado muitos pesquisadores, professores, gerentes e profissionais de diversas áreas do conhecimento.

Compreender e identificar as competências pode trazer inúmeros benefícios à empresa. Por isso, busca-se agora o estudo das competências organizacionais.

Para GREEN (1999) e PARRY (1998), até então os conceitos de competências colocados na literatura têm-se tornado confusos por conta das diferentes maneiras pelas quais o termo competência é utilizado pelos profissionais. Não é raro ouvir-se os termos competências essenciais, capacidades, traços de personalidades, ou outros atributos, utilizados de maneira intercambiáveis.

HAMEL *et al* (1995), definem competências como um conjunto de habilidades e tecnologias. Elas proporcionam as vantagens para a empresa competir e constituem o recurso mais valioso da organização. A perspectiva da abordagem das competências é de movimento empresarial “de dentro para fora” (inside-out).

### **2.1.2 Competências essenciais**

Ao longo do tempo as empresas podem adquirir conhecimentos e habilidades que se transformam em competências essenciais ou recursos internos estratégicos.

HAMEL *et al*(1995) definem competências essenciais (*core competence*), como “um conjunto de habilidades e tecnologias que permite a uma empresa oferecer um determinado benefício aos clientes”. Competências essenciais são habilidades ou capacidades, não são produtos ou serviços.

Esses autores consideram que uma competência essencial da organização deve: ter valor percebido pelo cliente, ou seja, o benefício deve ser visível aos olhos do cliente e não as nuances técnicas, proporcionadas pela competência subjacente ao benefício; mostrar diferenciação entre concorrentes e ser difícil de imitar; possuir capacidade de expansão, ou seja, possibilitar acesso a diferentes mercados. O autor argumenta ainda que competência essencial é um dos conceitos menos compreendidos de todos os tempos. Além de esclarecer o que é uma competência essencial, o autor esclarece também o que não é uma competência essencial, ou seja, competência essencial não é infra-estrutura.

Dispor de um sistema de distribuição nacional não é nenhuma competência essencial. Significa apenas que se tem uma infra-estrutura. Competência essencial também não é tecnologia. O fato de ser muito bom em projetar, não implica uma competência essencial. Geralmente as empresas confundem competência essencial com tecnologia ou infra-estrutura.

Outro ponto comentado por HAMEL *et al*(1995), é que as empresas confundem competências essenciais com capacidade ou aptidão. Por exemplo, a adoção como processo de gerenciamento *just-in-time* de sistemas de suprimentos foi até os anos 80, uma competência essencial, mas hoje, a maioria das empresas já a praticam, o que passou a ser uma é uma capacidade trivial. Com isto ele argumenta que, uma competência pode se tornar uma capacidade tão logo as demais empresas incorporem esta mesma capacidade.



Uma competência essencial é um conjunto peculiar de *know how*, que é o centro do propósito organizacional (GREEN, 1999). Ela está presente nas múltiplas divisões da organização e nos diferentes produtos e serviços, e não será competência essencial se um conjunto de habilidades ou tecnologias estiver restrito a uma única unidade.

A visão expressa unicamente em produtos é limitadora e leva à construção de produtos e serviços semelhantes aos produzidos no passado.

### **2.1.3 Competências organizacionais básicas**

RUAS (2002) argumenta “as empresas que provavelmente não possuam competências essenciais [...] porque não são diferenciadoras, mas possuem competências organizacionais básicas suficientes para permitirem a sua sobrevivência”. Contudo, as competências organizacionais básicas seriam aquelas competências coletivas e comuns que se desdobram em todo o universo organizacional e que contribuem decisivamente para a sobrevivência da organização. Em todo o segmento produtivo se identificam habilidades e capacidades que constituem uma condição *fundamental* para que uma empresa qualquer possa nele sobreviver, mas que não são suficientes para gerar diferenciação (RUAS, 2002).

Na ocasião, DRUCKER (1999) afirmava que a discussão de competências essenciais tem sido, em grande parte, até agora, uma grande piada. Para ele, o conceito não está pronto, porém está evoluindo e mostra como algumas empresas estão desenvolvendo uma metodologia para medir e gerenciar as competências. O primeiro passo é acompanhar com cuidado o desempenho da própria empresa e o dos concorrentes, buscando especialmente sucessos inesperados e mau desempenho em áreas onde elas deveriam ter se saído bem. Eles indicam onde a empresa goza de uma vantagem de liderança. Os insucessos devem ser vistos como a primeira indicação de que o mercado está mudando ou de que as competências da empresa estão enfraquecendo.

## **2.2 - Competências, Processos da Organização e Estratégias**

“A essência da estratégia está nas atividades. A opção de desempenhar atividades de forma diferente ou de desempenhar atividades diferentes em comparação com os rivais” afirma PORTER (1999). Assim, uma estratégia bem-sucedida e sustentável consiste em assegurar o alinhamento entre as atividades internas e a geração de valor para o cliente.

O negócio é rentável quando o valor que a empresa cria é superior ao custo do desempenho das atividades de valor. Para se conquistar a vantagem competitiva, em relação aos concorrentes, a organização precisa desempenhar estas atividades a um custo inferior ou executá-las de modo a conquistar a diferenciação.

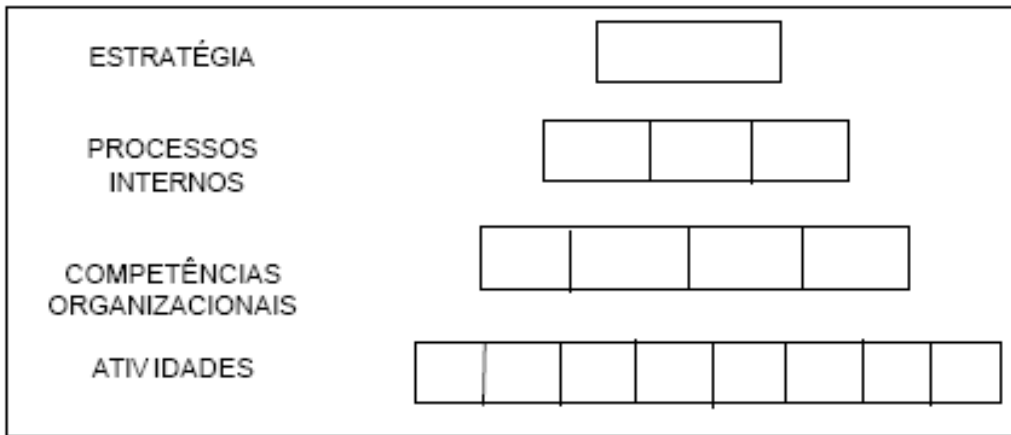


Figura 4 - Estratégia, processos, competências e atividades. Adaptado de Ulrich, Zenger e Smallwood (2000)

A partir da figura 4, é possível inferir que as empresas deveriam investir nas atividades que produzem e desenvolvem as competências e os processos para impulsionar a estratégia. Infelizmente, na busca de sobrevivência ou por desconhecimento, gerentes escolhem uma estratégia empresarial, mas optam por desenvolver processos internos conflitantes. Isso pode ser exemplificado com a opção empresarial pela estratégia em diferenciação, mas escolhem indicadores de desempenho de processos que se concentram na redução de custos, na padronização e eficiência operacional. Estas empresas apresentam um completo desentrosamento entre estratégia e processos internos.

A eficácia operacional quanto a estratégia são essenciais para o desempenho superior de todas as empresas (PORTER, 1999). Contudo, uma e outra atuam de formas diferentes. Uma empresa só é capaz de superar em desempenho os concorrentes se conseguir estabelecer uma diferença que possa preservar. Ela precisa proporcionar *maior valor aos clientes* ou gerar valor comparável, a um custo mais baixo, ou ambos.

O fornecimento de maior valor permite à empresa cobrar preços mais elevados. A maior eficiência resulta em custos mais baixos. A eficácia operacional significa o desempenho de atividades melhor do que os concorrentes. Abrange a eficiência, mas

não se limita apenas a esse aspecto. Diz respeito a qualquer prática pelas quais a empresa utiliza melhor o insumo.

Programas do tipo qualidade total, *empowerment*, *benchmarking*, terceirização, reengenharia, por exemplos, produzem melhorias da eficácia operacional. São feitos para eliminar as ineficiências, aumentar a satisfação dos clientes e melhorar a produtividade. O constante aprimoramento da eficácia operacional é imprescindível para a consecução da rentabilidade superior, contudo, não é o bastante.

Com base apenas na eficácia operacional, poucas empresas competem com êxito durante períodos mais prolongados, enfrentando dificuldades cada vez maiores em manter-se à frente dos rivais. A causa mais óbvia desta situação é a rápida proliferação de melhores práticas. Os concorrentes conseguem imitar com rapidez as técnicas gerenciais, as novas tecnologias, as melhorias nos insumos e as formas superiores de atender às necessidades dos clientes. Eficácia operacional também se refere à utilização de todos os insumos de uma atividade – mão-de-obra, capital, matérias-primas, energia – para gerar produto de maneira mais eficiente. Como mostra a figura 5, a competitividade resulta da interseção de boa estratégia e alta eficácia operacional (FAIRBANKS *et al*, 2000).

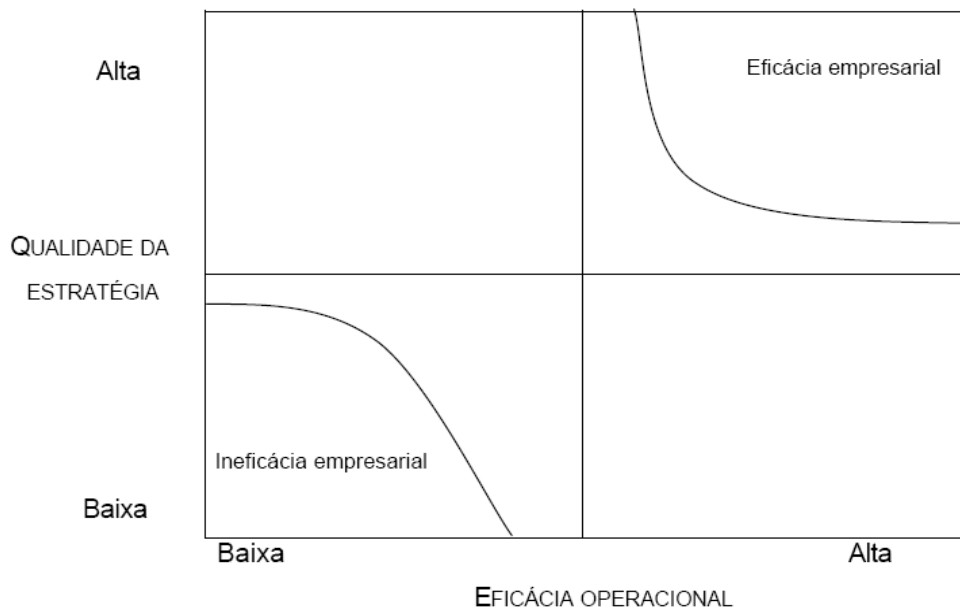


Figura 5 - Estratégia, eficácia operacional. adaptado de (FAIRBANKS *et al*, 2000)

## Vantagem Competitiva em Processos de Negócio Compartilhados

O modelo que trata da vantagem competitiva (PORTER, 1985), nos mostra que era também possível uma organização ou unidade de negócios terem inter-relações ganha-ganha com outra, beneficiando-se ambas da vantagem competitiva compartilhada.

Na Figura 6, encontramos uma visão de PORTER (1985) para a interface da cadeia de valor empresa com as cadeias de valor de seus clientes e fornecedores, distribuidores. Isto mostra que uma empresa pode manter inter-relações compartilhando processo e tecnologias, o que caracterizaria sistemas integrados gerindo a cadeia de suprimentos através de software de SCM (Supply Chain Management), além disso, as inter-relações entre empresa-cliente podem ser fortalecidas através da utilização de ferramentas e processos do tipo CRM (Customer Relationship Management). A próxima geração de ferramentas ERP (Enterprise Resource Planning), estará incorporando módulos complementares que já tratam as atividades internas da organização aperfeiçoando e fortalecendo as relações empresa - fornecedor e empresa - cliente.

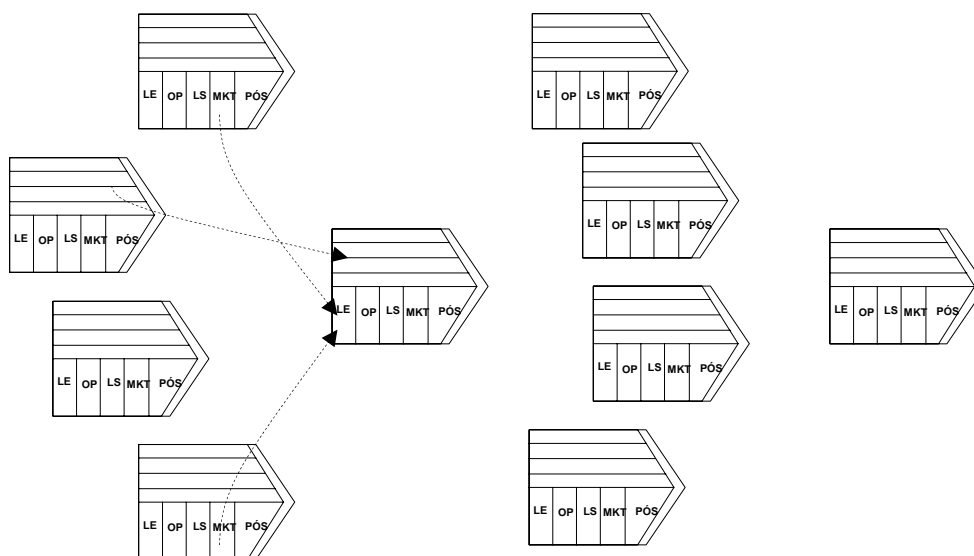


Figura 6 – Sistema de Cadeia de Valor (Porter, 1985)

Nela percebemos os sistemas que permeiam a estrutura da empresa e seus relacionamentos definindo a partir da os denominados sistemas de SCM- Supply Chain

Manufacture, CRM – Customer Relationship Management e o próprio ERP – Enterprise Resource Planning.

## **2.3 - A Gestão do Conhecimento e as Estratégias**

Uma das possíveis definições para esta terminologia Gestão do Conhecimento - GC seria: “juntar, tratar, organizar e sistematizar, em todos os pontos de contato internos e externos, a capacidade da empresa de captar, gerar, criar, analisar, traduzir, transformar, modelar, armazenar, disseminar, implantar e gerenciar a informação, tanto interna como externa. Toda informação útil deveria ser transformada efetivamente em conhecimento e distribuída de forma explícita, tornando-se acessível aos interessados da organização”.

### **2.3.1 A Construção do Conhecimento**

A construção do conhecimento deve ser uma prática coletiva e que permita a integração e socialização do conhecimento.

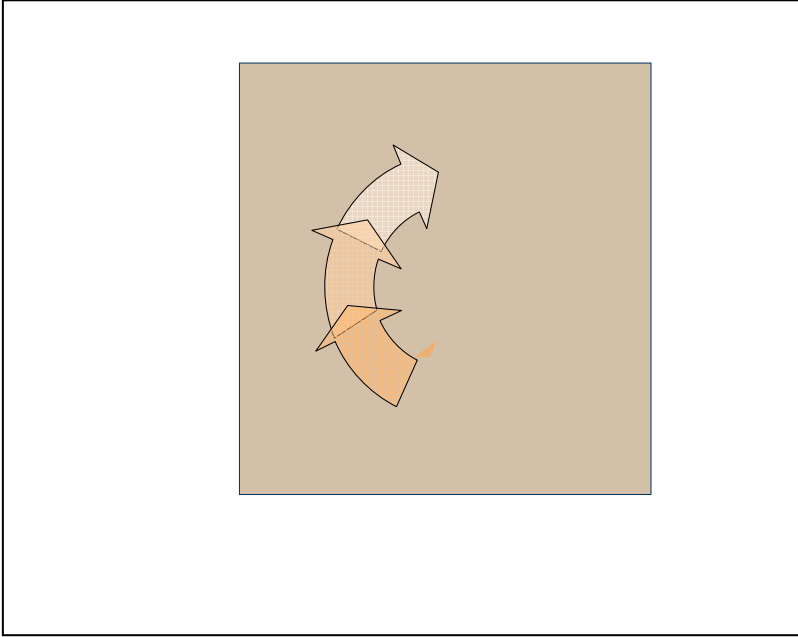
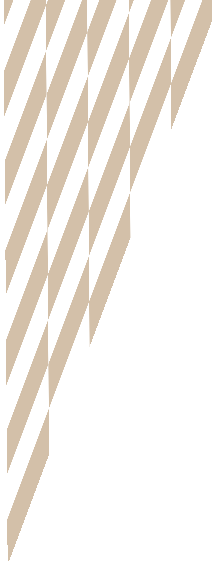
O processo de construção do saber é dividido por NONAKA *et al* (1998) em quatro (4) etapas: socialização, externalização, combinação e internalização e é baseado na transformação do conhecimento explícito em tácito e vice-versa.

A socialização é aquela fase em que acontece a partilha das experiências, criação do conhecimento tácito em modelos mentais ou habilidades técnicas.

Na fase de externalização ocorre a tradução do conhecimento tácito em formas que os outros possam aprender, seja através de metáforas, analogias conceitos, hipóteses e modelos.

A combinação diz respeito à difusão, edição e sistematização do conhecimento fragmentado codificado (análise, categorização e reconfiguração).

A última das fases, internalização é a tradução do novo conhecimento codificado em tácito individual, através de novas experiências e exercícios. É o que se chama de “aprender fazendo”. Esse processo ocorre repetidas vezes, de forma como mostrado na Figura 7, que é denominado como “espiral do conhecimento”, um processo contínuo cíclico. Trata-se de um processo de construção e acumulação. Acontece inicialmente no nível individual, e vai subindo em direção aos níveis superiores, cruzando setores, departamentos, divisões e organizações.



Compartilhar conhecimento, no entanto, ocorre quando as pessoas estão genuinamente interessadas em ajudar umas às outras a desenvolver novas capacidades para a ação e em criar processos de aprendizagem mutua.

Para que uma organização torne-se ou se mantenha competitiva, é necessário que haja uma gestão eficiente do conhecimento, aliado a uma metodologia eficaz de inteligência competitiva.

É importante aqui, destacar a diferença entre esses dois modelos de apoio a gestão, uma vez que, freqüentemente, são usados de maneira inapropriada. Gerir conhecimento significa identificar, analisar e administrar através de um conjunto de técnicas e ferramentas o conhecimento existente dentro da empresa para todos os recursos, e manter este conhecimento na empresa.

Inteligência competitiva é conhecer os concorrentes e o ambiente externo à organização aplicado a processos de tomada de decisão nos níveis estratégico e tático.

Em outras palavras, é um processo de coleta e análise sistemática da informação sobre o ambiente externo, que, por sua vez, é disseminada como inteligência aos usuários em apoio à tomada de decisão, tendo em vista a geração ou sustentação de vantagens competitivas.

O mercado hoje oferece ferramentas para gestão da informação e comunicação que, implementadas dentro de uma lógica de transformação da cultura organizacional, serão bastante úteis e apoiarão seus projetos de gestão de conhecimento.

Entretanto a tecnologia é apenas uma ferramenta, ela não será suficiente por si só. É importante:

- ter uma perspectiva estratégica do seu negócio para compreender o que é conhecimento de valor
- explicitar e coletar as informações necessárias no dia-a-dia da empresa nos níveis operacional, tático e estratégico
- valorizar a inovação, a experimentação e aqueles que compartilham sua informação e conhecimento
- apoiar e estimular seus colaboradores nessa tarefa de organizar as melhores práticas

A organização e sistematização do conhecimento necessário aos níveis operacional, tático e estratégico de sua organização é um processo feito no dia-a-dia, e não um

projeto que tem início, meio e fim. A gestão do conhecimento se inicia, mas não tem prazo certo para terminar. Na realidade não mais termina.

### **2.3.2 O Desperdício de Talentos**

É muito comum percebermos colaboradores atuando em posições que nada tem a ver com suas aspirações, potencial e força de trabalho. A pessoa certa no local certo é, portanto, fruto da mais pura Gestão do Conhecimento. Este tipo de desarranjo pode ser considerado um tipo de desperdício de talentos, que é completamente incoerente com a gestão do conhecimento.

Hoje em dia, muitos dos processos sucessórios são traumáticos não porque as empresas não têm sucessores à altura dos talentos a serem substituídos, mas principalmente por que elas não os conhecem. Os talentos raramente são descobertos e treinados com a devida antecedência.

Além das questões dos processos sucessórios, a atuação coerente em GC exige dos gerentes a descoberta perene de talentos nas empresas também para resolver desafios empresariais mais complexos, resolver problemas dos clientes e uma infinidade de atividades relacionadas com a aplicação do conhecimento.

Muitos gerentes estimulam a atitude (ação) como se bastasse ao sucesso, mas se esquecem que a atitude é sim fundamental quando acompanhada de conhecimentos, habilidades, experiências, métodos e contexto.

Atitude por atitude, sem o conhecimento como suporte, pode ser terrível. Embora importante, a atuação desacompanhada de conhecimentos quase sempre depõem contra a organização.

Localizar colaboradores e recursos “abandonados” ou não considerados pela estratégia tem grande importância corporativa. Se existem pessoas com potencial ilhadas, alocadas em trabalhos paralisantes que não geram crescimento, escondidas em rotinas etc., isto deve ser descoberto e tratado à luz da GC, pois onde se encontram, jamais poderão gerar e ou aplicar o potencial dos conhecimentos reprimidos. O que está abandonado pode estar esquecido e se está esquecido está também à margem da GC, portanto, de total interesse empresarial.



### **2.3.3 Quando as Perguntas que nos Ensinam mais do que as Respostas**

A gestão do conhecimento nos ensina a valorizar as perguntas. Hoje em dia, as boas perguntas feitas no ambiente de negócios são tão importantes quanto as respostas conseguidas. Não importa o quanto não estamos preparados para responder, de pronto, todas as perguntas na empresa, se temos as melhores perguntas e se temos perguntas que valem a pena responder, quase sempre encontramos quem possa respondê-las. As perguntas interessantes questionam paradigmas

organização. E se não houver disposição, motivação, empatia, propósito e atitude da melhor qualidade para aplicação generalizada de conhecimentos, de nada vale o saber.

## **2.4 - Inteligência Aplicada aos Negócios**

Num mercado cada vez mais concorrido e exigente é grande a demanda por soluções capazes de oferecer vantagens competitivas às empresas. Por isso corporações dos mais diversos setores buscam ferramentas estratégicas para:

- entender melhor o nicho de atuação no mercado;
- promover melhoramentos na competência essencial da empresa;
- identificar novas oportunidades;
- responder adequada e eficientemente às mudanças do mercado;
- melhorar o relacionamento com clientes e fornecedores;
- reduzir custos operacionais.

O conceito de Inteligência de Negócios (Business Intelligence (BI)), de forma mais ampla, pode ser entendido como a utilização de variadas fontes de Informação para definir estratégias de competitividade nos negócios.

Existe uma grande problemática ao nível empresarial e de mercado: uma grande quantidade de dados está disponível, provocando muitas dificuldades na extração de informações a partir deles e a enorme quantidade de informações dificulta o processo de tomada de decisão na medida em que a gerência se sente impotente no processo de recuperação e análise.

As informações vitais para tomadas de decisões estratégicas estão escondidas em milhares de tabelas e arquivos, ligadas por relacionamentos de correlações transacionais,

em uma organização inadequada para o estabelecimento de decisões.

O objetivo maior das técnicas de BI, neste contexto, está exatamente na definição de regras e técnicas para a formatação adequada destes volumes de dados, com a finalidade de transformá-los em depósitos estruturados de informações, independentemente da sua origem.

O grande desafio das empresas hoje, não é apenas organizar seus Sistemas de Gestão do Conhecimento, mas estabelecer uma ponte entre ela e a Inteligência de Negócios.

Enquanto o BI transforma dados em Informação, produzindo as visões (views), relatórios, dentre outras, a Gestão do Conhecimento realiza as devidas combinações,

compilações, subscrições e a distribuição, inserindo pontos de discussão, com o objetivo de transformar informações em conhecimento.

Em Inteligência de Negócios existem abismos não superados pelo *Business Intelligence* tradicional. Imagine uma rede de lojas que descobre em seu histórico de vendas, várias ocorrências de compras casadas de TV com DVD players ou pelo menos que ambas ocorram entre intervalos de no máximo 4 meses. Ótimo. A empresa acaba de descobrir um comportamento de consumo (oportunidades) e poderá tirar proveito disso.

O que esta companhia jamais descobriria em seu *Data Mining* é que se oferecesse cabos dos tipos RCA e Componentes, por exemplo, também teria êxito em suas vendas casadas com TVs e DVDs players. Por não oferecer em suas lojas tais acessórios, os sistemas de *Business Intelligence* da empresa não descobririam esta correlação e oportunidade. Portanto, não se deve conformar com os resultados e recursos atuais. A inteligência de negócios esperada pelas empresas quase sempre é bem maior do que a normalmente atingida.

Algumas empresas que por limitações a respeito da gestão do conhecimento, aproveitam e engajam muito pouco do conhecimento corporativo de valor, perdendo negócios, clientes e receitas contínuas.

- O que estamos fazendo na empresa quando modificamos uma rotina de atendimento e implementamos uma solução que viabilize a prontidão de respostas aos clientes?
- O que estamos fazendo quando disponibilizamos um portal e permitimos que conhecimentos tácitos e explícitos sejam trocados entre os funcionários?
- O que estamos fazendo quando emitimos sinais de reconhecimento positivos diante das iniciativas de troca de conhecimento praticadas entre os funcionários dentro da empresa?

Seja o termo adequado ou não, é Gestão do Conhecimento que se praticando.

### **3 - Sistemas Inteligentes de Apoio a Decisão**

Neste capítulo pretende-se abordar as metodologias e sistemas que concorrem para apoiar e melhorar a tomada de decisão nas empresas. Os bancos de dados existentes nas corporações estão tendo utilidade muito além daquela de armazenar as informações processadas pelos seus diversos sistemas administrativos e operacionais.

Sistemas de controle de estoque, cadastros de clientes, ordens de serviço, de manutenção e operações, quando conjugados entre si ou integrado ao outros bancos de dados de terceiros, geram uma fonte infindável de possibilidades que podem enriquecer e até conduzir decisões mais corretas dos dirigentes.

Grandes e médias corporações têm tirados vantagem competitiva do uso de ferramentas de mineração de dados em suas simulações para apoio a decisões corporativas. Alguns conceitos serão aqui tratados visando o nivelamento dos conhecimentos e nomenclatura adotada no trabalho.

#### **3.1 - Dados, Informação e Conhecimento**

Os conceitos de dados, informação e conhecimento estão intimamente relacionados, na Figura 8, é apresentada uma representação do relacionamento entre dados, informação e conhecimento, em função da capacidade de entendimento e da independência de contexto que cada um destes conceitos implica (KOCK et al, 1997).

Antes de estabelecer qualquer ligação desses conceitos com as diferentes tecnologias para seu registro e processamento, faz-se necessária uma breve discussão sobre a distinção entre dado, Informação e conhecimento.

O dado é um elemento puro, quantificável sobre um determinado evento. Dados, são fatos, números, texto ou qualquer mídia que possa ser processada pelo computador. Hoje em dia, as organizações estão acumulando vastas e crescentes quantidades de dados em diferentes formatos e em diferentes tipos de repositórios. Ressalta-se que o dado, por si só, não oferece embasamento para o entendimento da situação. Entre os dados armazenados atualmente estão:

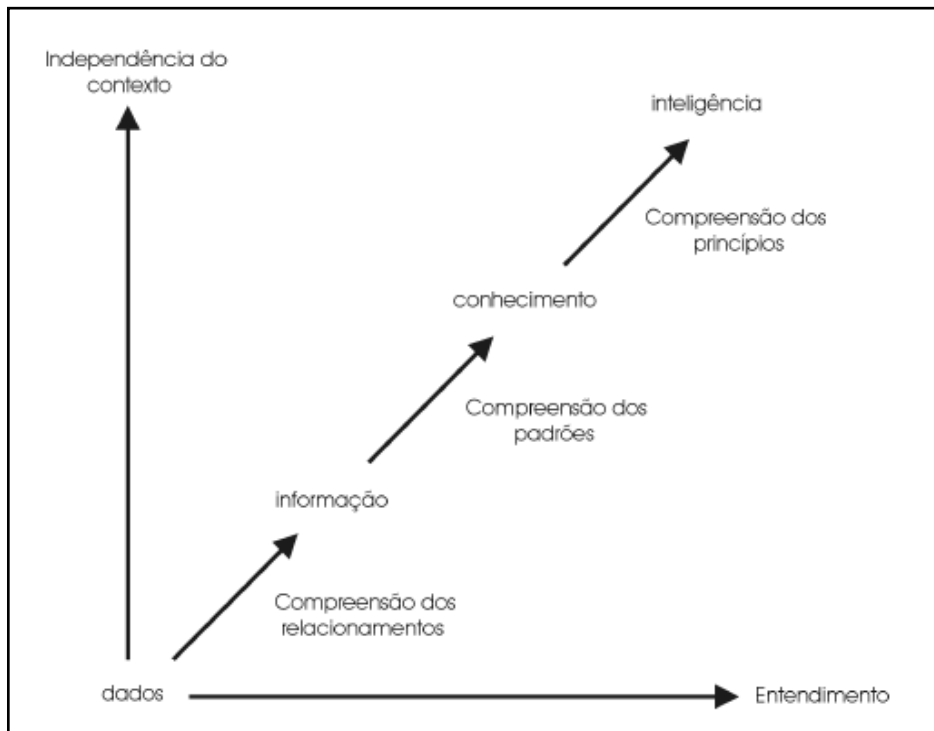


Figura 8 - Gráfico Contexto versus Entendimento adaptado de KOCK et al, 1997

- dados operacionais ou transacionais como vendas, custos, inventários, folhas de pagamento ou contas correntes;
- dados não operacionais como previsões de mercado, vendas ao nível industrial, e dados macro-econômicos;
- metadados, ou dados descrevendo a estrutura dos dados armazenados, tais como projetos lógicos de Bancos de Dados ou dicionários de dados;
- mídia conteúdo imagens, sons ou uma combinação de ambos, que além de ser armazenada, pode ser catalogada eletronicamente.

A Informação é o dado analisado e em determinado contexto. Envolve a interpretação de um conjunto de dados, ou seja, a Informação é constituída por padrões, associações ou relações que todos aqueles dados acumulados podem proporcionar. Por exemplo, a análise de pontos de equilíbrio no mercado pode fornecer Informação acerca de quais produtos estão sendo vendidos e a frequência de tais operações.

A Informação pode gerar conhecimento que ajude na análise de padrões históricos para conseguir uma previsão dos fatos futuros (pelo menos no contexto das variáveis que estão sendo envolvidas na análise). Por exemplo, a Informação dos dados

sumarizados nas vendas de um determinado ambiente comercial pode ser analisada com a finalidade de fornecer informações relacionadas com a natureza dos clientes.

Enquanto que a Informação é descritiva, o conhecimento é utilizado fundamentalmente para fornecer uma base de previsão com um determinado grau de certeza. O conhecimento refere-se à habilidade de criar um modelo mental que descreva o objeto e indique as ações a implementar e as decisões a tomar. Uma decisão é o uso explícito de um conhecimento. O conhecimento pode ser representado como uma combinação de estruturas de dados e procedimentos interpretáveis que levam a um comportamento conhecido.

Este comportamento fornece informações que podem, então, serem utilizadas para planejar e decidir.

A compreensão, análise e síntese, necessárias para a tomada de decisões inteligentes, são realizadas a partir do nível do conhecimento. Assim, é fundamental que se mantenha a coerência dos dados que estão armazenados nos diferentes repositórios e das informações nos diferentes níveis.

Assim, o desafio há duas décadas foi migrar os dados para as informações, por meio do desenvolvimento dos Sistemas de Informação, que tinham por finalidade analisar dados e organizar a Informação para melhorar o processo decisório empresarial. A partir da década de 90, o desafio foi o de criar sistemas que fossem capazes de tratar e processar conhecimento, em resposta às diferentes necessidades da sociedade.

### **3.2 - O Domínio das Regras de Negócios nas Empresas**

A sociedade do conhecimento está impondo uma competitividade cada vez maior entre países e entre empresas, o que leva a uma necessidade de mudança e reflexão contínuas. É preciso inovar e adquirir sucessivamente novos conhecimentos (CAVALCANTI, et al 2001).

A gestão do conhecimento tem o objetivo de estabelecer meios, de maneira integrada e colaborativa, para capturar, criar, organizar e usar todos os ativos de Informação de uma corporação. A gestão do conhecimento é o primeiro passo na criação de uma estrutura lógica para manipular as informações que uma determinada entidade possui e gerenciar tanto as entradas quanto os resultados da mesma. Em outras palavras, a gestão do conhecimento é responsável pela recuperação e organização das práticas,

documentos, políticas, experiências de funcionários, entre outras fontes, de onde é possível obter conhecimento explícito de uma organização.

O foco principal é o conhecimento organizacional que é inerente a todas as empresas e é definido como “a capacidade de executar coletivamente tarefas que as pessoas não conseguem fazer atuando de forma isolada, tarefas essas projetadas para criar valor para as partes interessas na organização” (GARVIN *et al*, 1998).

Há uma diferença entre o conhecimento estar embutido em estruturas, regras e processos de trabalho em grupo – conhecimento explícito – e estar embutido em trabalhos individuais – conhecimento tácito - acreditam que o conhecimento organizacional deve ser explícito e tácito, pois o conhecimento tácito, que inclui o discernimento, o instinto e a compreensão individual, é fundamental para tornar o conhecimento explícito útil.

É importante frisar que antes de investimentos ou decisões em TI, a empresa já deveria saber identificar, desdobrar e atribuir o devido valor aos conhecimentos disponíveis. Além disso, a organização precisa estar ciente de que muitos dos recursos da TI requerem significativos e continuados esforços para se construir, manter e atualizar, tanto em seus aspectos funcionais quanto nos conteúdos envolvidos, bem como é necessário conquistar a aceitação e confiança do usuário corporativo.

### **3.3 - Data Warehousing**

Data Warehousing é um processo, não um produto, para montar e gerenciar repositórios de dados a partir de várias fontes com o propósito de ter uma visão detalhada e singular de parte ou do todo de um negócio. O produto principal obtido de um projeto de Data Warehousing é o seu Data Warehouse (DW).

A concretização de Data Warehousing (GARDNER, 1998) é considerada um dos primeiros passos para tornar factível a análise de grande quantidade de dados no apoio ao processo decisório. O objetivo básico é criar um repositório, conhecido por Data Warehouse, que contenha dados limpos, agregados e consolidados podendo ser analisados por ferramentas *OLAP (On-Line Analytical Processing)*. Tais ferramentas apresentam facilidades para a realização de consultas complexas em Bases de Dados multidimensionais.

Este processo objetiva integrar e gerenciar dados extraídos de diversas fontes, com o propósito de ganhar uma visão detalhada de parte ou do todo de um negócio (GUPTA, 1997).

Um Data Warehouse é a representação de um Banco de Dados cuja função é proporcionar aos seus usuários uma única fonte de Informação a respeito dos seus negócios, servindo também, como ferramenta de apoio ao processo de extração de conhecimento. Compreende uma coleção de dados orientado por assuntos, integrado, variante no tempo e não volátil, que tem como objetivo básico satisfazer as necessidades dos usuários quanto ao armazenamento dos dados que servirão para realizar consultas e análises necessárias para o apoio à tomada de decisão.

O termo orientado por assuntos está relacionado ao conceito de DATA MART, ou seja, um DATA MART implementa um determinado assunto, enquanto um DATA WAREHOUSE - figura 9 - implementa e integra vários assuntos de um organização. A metodologia mais usual é composta por varias fases e devem se repetir para cada nova área de negócio a ser considerada no projeto. A primeira fase corresponde à justificativa de um projeto de DW.

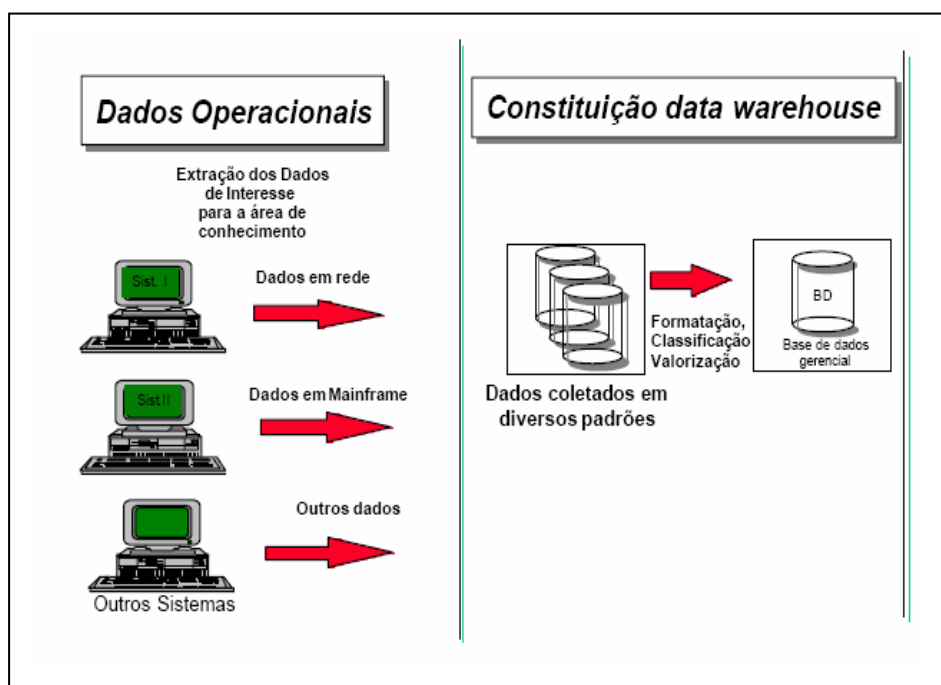


Figura 9 – Modelo de criação de uma base de dados gerencial (DATA WAREHOUSE)

Sugere-se que no DATA MART seja documentado todo o conhecimento obtido durante o projeto da área de negócio selecionada, servindo como modelo para as



próximas iterações do projeto. Um outro tópico que deve ser considerado no projeto de um DATA WAREHOUSE é a sua granularidade, a qual se refere ao nível de detalhe na qual as unidades de dados são mantidas.

As técnicas de análise de dados do Data Warehouse geralmente não extrapolam a realização de consultas SQL (Structured Query Language) simples, a utilização de ferramentas OLAP ou os mecanismos de Visualização de Dados. Por meio desta forma de análise de dados, algumas questões importantes para tomada de decisão não podem ser expressas, como:

- Quais são os potenciais clientes inadimplentes?
- Quais são os usuários potenciais para praticar fraude?
- Quais clientes gostariam de comprar o novo produto?

As ferramentas utilizadas para analisar um DATA WAREHOUSE, normalmente, são orientadas às consultas, ou seja, são dirigidas pelos usuários e especialistas, os quais possuem hipóteses que gostariam de comprovar. Esta abordagem que depende do usuário ou especialista para formular as perguntas ao banco de dados, poderá inibir que “padrões escondidos” nos dados sejam encontrados de forma “inteligente”, uma vez que o usuário não terá condições de imaginar todas as possíveis relações e associações existentes em um grande volume de dados. Por isso, faz-se necessária a utilização de técnicas de análise dirigidas por computador que possibilitem a extração automática (ou semi-automática) de novos conhecimentos a partir de um grande repositório de dados (BRADLEY et al, 1998).

As consultas OLAP geram informações obtidas a partir do DW enquanto determinado conhecimento pode ser extraído utilizando as técnicas de mineração de dados - MD aplicadas nos dados do DATA WAREHOUSE - DW ou diretamente nos dados da Base de Dados, como mostrado na Figura 10. Esta figura retrata a hierarquia de questões que em geral surgem nas corporações. Desde questões levantadas na base da pirâmide por meio de relatórios provenientes de sistemas departamentais e consultas aos bancos de dados fonte, até aquelas que somente podem ser respondidas por meio de ferramentas e técnicas de mineração e de visualização de dados que apóiam a tomada de decisão da organização.

A obtenção do conhecimento é um processo que envolve o uso de múltiplas técnicas e métodos que evoluíram à medida que os requisitos de Informação tornaram-se prioridade dos ambientes de negócios. Enquanto as tecnologias de processamento da

Informação evoluíram separadamente dos sistemas analíticos, o processo de Mineração de Dados pode fornecer um elo entre ambos, apoiando a Gestão do Conhecimento e Inteligência de Negócios.

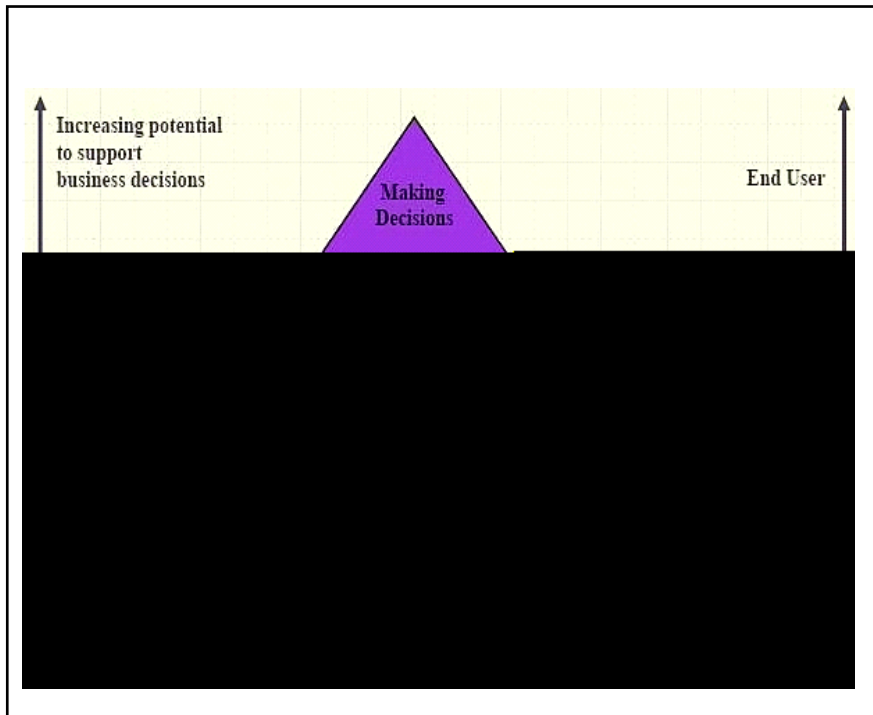


Figura 10 – Data mining e Business Intelligence (extraído de notas de aula Alexandre EvsuKoff)

O objetivo do processo de Mineração de Dados é a extração do conhecimento implícito por meio da descoberta de padrões e da criação de modelos de maneira automática a partir dos dados. A Figura 10, nos mostra ferramentas de *Business Intelligence* e os requisitos de profissionais que atuam em cada etapa do processo.

### 3.4 - Mineração de Dados

A Inteligência Artificial têm propiciado a possibilidade de utilizar diferentes técnicas para o reconhecimento/extração de padrões. Essa extração está acompanhada de técnicas de manipulação de dados e de análises posteriores. Esse conjunto de diferentes técnicas está encaixado em um processo denominado Mineração de Dados.

A área de Visualização de Informação, por sua vez, têm contribuído com técnicas de visualização que permitem reconhecer informações em diferentes níveis e estruturas. FAYYAD et al ( 1996) estabeleceu a primeira definição de Mineração de Dados aceita por diversos pesquisadores como sendo: “Extração de Conhecimento de Base de Dados

é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”.

### **3.4.1 Conhecimento**

Este termo complexo na literatura de Business Intelligence é definido em termos dependentes do domínio e do contexto estudado, relacionados fortemente com medidas de utilidade, originalidade e compreensão.

O processo de Extração de Conhecimento de Bases de Dados tem o objetivo de encontrar conhecimento a partir de um conjunto de dados para ser utilizado em um processo de tomada de decisão. Dessa forma, um requisito importante é que esse conhecimento descoberto seja compreensível aos humanos, além de útil e interessante para os usuários finais e especialistas do processo e que ainda forneça um suporte ao processo de decisório (FAYYAD et al, 1996).

Todo o processo de MD é orientado em função de seu domínio de aplicação e dos repositórios de dados inerentes aos mesmos. Para usar os dados é necessário que estes estejam estruturados de forma a serem consultados e analisados adequadamente.

Os sistemas conhecidos por *OLTP - On-Line Transaction Processing* -, processam dados armazenados em Base de Dados relacionais usadas para armazenar, consultar e alterar informações do negócio.

O processo de KDD envolve áreas relativas a aprendizado de máquina, bases de dados, estatística e matemática, sistemas especialistas e visualização de dados.

Este processo utiliza métodos, algoritmos e técnicas oriundos destas diversas áreas, com o objetivo principal de extrair conhecimento a partir de grandes bases de dados, conforme pode ser observado na figura 11.



### 3.5 - Descobrendo Conhecimento em Banco de Dados

Este capítulo introduz os assuntos na área onde mineração e modelagem de dados se fundem. Modelar é um processo contínuo ao longo do processo de mineração. Primeiro, um modelo estrutural tem que ser descoberto para moldar o problema. Com progresso da mineração, o modelo vai sendo revisado, e é preciso se obter vários tipos diferentes de modelos para entender e utilizar o resultado que é a descoberta nos dados.

Embora modelos informem a direção de progredir na mineração enquanto minerar é a atividade que clarifica e dá substância aos modelos.

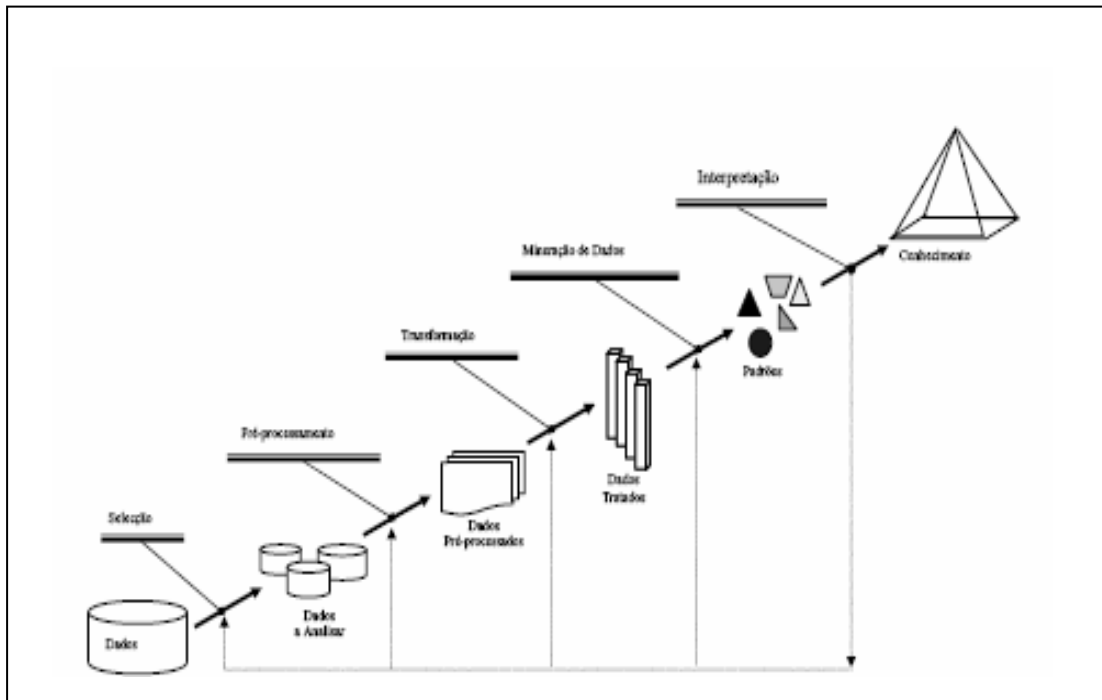


Figura 12 – Processo de Mineração de dados adaptado de FAYYAD et al (1996)

Os modelos vão definindo linhas de investigação; mineração reduz a incerteza aproximadamente, e explica as relações nas que existem, áreas importantes que o modelo identifica. Em outra palavra, mineração esclarece as hipóteses e suposições que inicialmente estabelecidas e estrutura para o problema. Em resumo, mineração reduz incerteza. FAYYAD *et al* (1996), através da figura 12 apresenta seu processo de mineração subdividido nas etapas de: seleção dos dados, pré-processamento, transformação, mineração e pós processamento.

Toda a mineração e modelagem são atividades que buscam uma meta: lidar com alguns questões que identificaram um determinado problema.

A extração de conhecimento a partir de grande quantidade de dados é vista como um processo interativo e iterativo, e não como um sistema de análise automática, sendo centrado na interação entre usuários, especialistas do domínio e responsáveis pela aplicação. Dessa maneira, não se pode esperar que a extração de conhecimento seja útil simplesmente submetendo um conjunto de dados a uma “caixa preta” (MANNILA, 1997).

Existem diversas abordagens para a divisão das etapas do processo de Extração de Conhecimento. Em uma visão particular podemos explicitar as duas mais importantes: a fase anterior ao processo de Mineração de Dados, que se refere ao conhecimento do domínio e identificação do problema, e uma fase posterior ao processo, que se refere à utilização do conhecimento obtido. A Figura 13 - conforme REZENDE *et al* ( 2003), ilustra essas etapas.

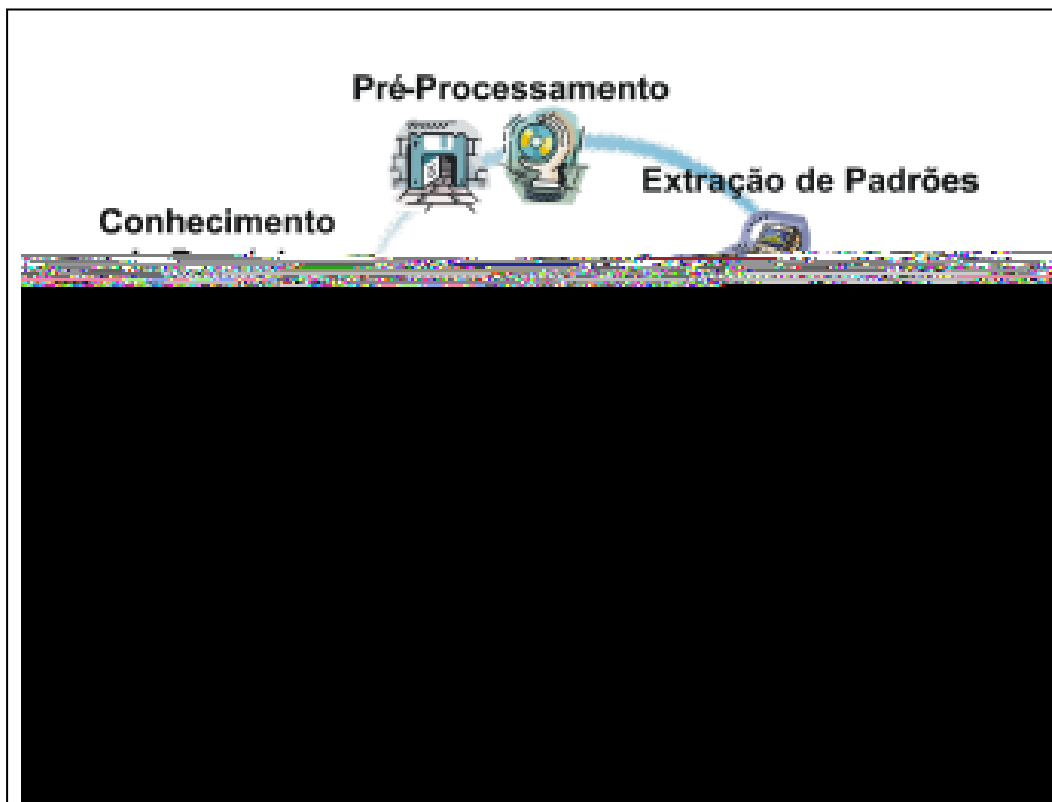


Figura 13 – Etapas do Processo de Mineração de Dados - (REZENDE et al, 2003)

Assim, o processo de Extração de Conhecimento de Bases de Dados, ou Mineração de Dados, inicia-se com o entendimento do domínio ou do contexto da aplicação, considerando aspectos como os objetivos dessa aplicação e as fontes de dados (Base de Dados de onde se pretende extrair conhecimento).

Por ser um processo iterativo, as etapas da Mineração de Dados não são estanques, ou seja, a correlação entre as técnicas e métodos utilizados nas várias etapas é considerável, a ponto da ocorrência de pequenas mudanças em uma delas afetar substancialmente o sucesso de todo o processo. Portanto, os resultados de uma determinada etapa podem acarretar mudanças a quaisquer das etapas posteriores ou, ainda, o retomar o processo inicial (FAYYAD et al, 1996).

O processo de MD é centrado na interação entre as diversas classes de usuários, e o seu sucesso depende, em parte, dessa interação. Os usuários do processo podem ser divididos em três classes:

**Especialista do Domínio ou do Problema** que são aqueles que possuem amplo conhecimento do contexto da aplicação e deve fornecer informações valiosas para a execução do processo;

**Analista** que é um dos técnicos que vai interagir com o especialista no processo de Extração de Conhecimento e responsável por sua execução devendo conhecer profundamente as etapas que compõem o processo de mineração;

**Usuário final** que pode estar representando um conjunto de usuários que utiliza o conhecimento extraído no processo para auxiliá-lo em um processo de tomada de decisão. Este no caso de tomada de decisão poderá ser um gerente nível médio ou mesmo Diretor da corporação.

Em algumas empresas dependendo do assunto objeto, o especialista do domínio também é o próprio usuário final, ou que este auxilie ou execute funções do analista.

### **3.5.1 Fase 1 - Entendimento do Domínio ou Identificação do Problema**

O objetivo é a aprendizagem do domínio de aplicação que tem como atividades:

**A. Selecionar os dados a analisar**

**B. Agrupar os dados ou conjunto de variáveis sobre os quais se pretende trabalhar.**

O estudo do domínio da aplicação e a definição de objetivos e metas a serem alcançadas no processo de Mineração de Dados são identificados nesta fase.

O sucesso do processo de Extração de Conhecimento depende fundamentalmente da participação dos especialistas verdadeiros do negócio que se esteja tratando na sua tarefa de encontrar os padrões e regras para o problema colocado.

Previamente ao início das atividades deste processo é imprescindível a realização de um estudo preliminar com o intuito de se adquirir um conhecimento questão sob investigação (FAYYAD et al, 1996).

Algumas questões importantes devem ser respondidas nessa fase de identificação do problema, como:

- O Especialista indicado conhece o domínio a ser estudado e pesquisado?
- Já se tem idéia do que se deseja resolver?
- Quais são as principais metas do processo?
- Quais indicadores de desempenho buscados?
- O conhecimento obtido é compreensível ou um modelo “caixa-preta” é o mais adequado?
- Qual deve ser o nível de precisão do conhecimento extraído?

Na etapa de pré-processamento, esse conhecimento obtido no entendimento do contexto e do domínio, poderá auxiliar os analistas na seleção do melhor conjunto de dados para se realizar a extração de padrões. Sabendo quais valores são válidos para os atributos listados, os critérios de preferência e prioridade entre os possíveis atributos, as restrições de relacionamento ou informações para geração de novos atributos, dentre outros.

Na etapa de extração de padrões, o conhecimento sobre o domínio pode ajudar os analistas na escolha de um critério de preferência entre os modelos gerados, no ajuste dos parâmetros do processo de indução, ou mesmo na geração de um conhecimento inicial a ser fornecido como entrada do algoritmo de Mineração.

Já na etapa de pós-processamento, o conhecimento extraído pelos algoritmos de extração de padrões deve ser avaliado. Alguns critérios preceptivos de avaliação utilizam o conhecimento do especialista para saber, por exemplo, se o conhecimento extraído é interessante ao usuário.

O entendimento do domínio dos dados é pré-requisito para o sucesso da extração de resultados úteis. A escolha de especialistas com profundo conhecimento sobre domínio do problema é fator crítico para o sucesso da empreitada.

No problema que será colocado no capítulo seguinte, constatamos que na área de saneamento, por ser ainda um setor em sua maioria estatal carece de especialistas acostumados a um ambiente de competição e, portanto acreditam que “aquilo que deu certo até hoje continuará a dar resultados”. Esta não é a realidade que se constata, pois



como verificamos no capítulo 1, o nível de “perdas físicas e aparentes”, não é encontrado em nenhum outro segmento da indústria.

No caso das concessionárias operadoras de energia elétrica, telecomunicações e saneamento, na qual os técnicos têm uma idéia geral das características dos sistemas e das regras do negócio que vem praticando ao longo do tempo, mas alterações significativas em tecnologia, em equipamentos e sistemas de informação implicam que especialistas com conhecimento no tratamento de dados não são necessariamente encontrados nestas empresas.

### **3.5.2 Fase 2 - Pré-Processamento**

Nesta etapa o objetivo é o pré-processamento e limpeza de dados e se caracteriza por:

- operações básicas de remoção de ruído nos dados;
- decisão de estratégias no caso de se obter campos omissos nos dados;

Rotinas de limpeza de dados tentam suprir valores ausentes, reduzir discrepâncias de valores ruidosos e corrigir inconsistências. Para valores ausentes, algumas técnicas aplicáveis são (HAN E KAMBER, 2001):

1 - Ignorar a tupla

2 - Suprir valores ausentes

- a) manualmente;
- b) através de uma constante global;
- c) utilizando a média do atributo;
- d) utilizando a média do atributo para todas as instâncias da mesma classe;
- e) com o valor mais provável (regressão, inferência etc.).

As técnicas 2b, 2c, 2d e 2e podem "viciar" os dados. A técnica 2e é uma estratégia interessante, pois em comparação com outros métodos utiliza um maior número de informações dos dados disponíveis.

Ruídos nos dados são erros aleatórios ou variâncias numa variável mensurada. A eliminação de ruídos pode ser realizada através de: 1 - Interpolação; 2 - Agrupamento; 3 - Inspeção humana e computacional combinadas; 4 – Regressão.

Alguns tipos de inconsistências podem ser corrigidos manualmente através de referências externas. Rotinas de consistência evitam a inserção de dados incorretos através da interface do banco de dados (infelizmente, a maioria dos softwares não são projetados e desenvolvidos levando em conta KDD). Ferramentas de engenharia do

conhecimento podem detectar a violação de restrições de dados. Tanto redundâncias como discrepâncias podem ser combatidas através de dependências funcionais.

Normalmente, os dados disponíveis para análise não estão em um formato adequado para a extração de conhecimento e padrões. Contudo, considerando as limitações físicas de equipamentos de memória ou tempo de processamento, muitas vezes inviabiliza a aplicação direta dos algoritmos de extração de padrões daqueles dados.

Torna-se então necessária a aplicação de métodos para tratamento, limpeza e redução do volume de dados antes de iniciar a etapa de Extração de Padrões. É importante salientar que a execução das transformações deve ser reproduzir os objetivos originais do processo de extração de modo que o subconjunto de dados os quais se vai trabalhar retrate suas características específicas.

A redução do número de exemplos deve ser feita a fim de manter as características do conjunto de dados original, isto é, por meio da geração de amostras representativas dos dados. A abordagem mais utilizada para redução do número de exemplos é a amostragem aleatória, pois este método tende a produzir amostras representativas (WEISS e INDURKHYA, 1998).

A redução do número de atributos pode ser utilizada para reduzir o espaço de busca pela solução. Objetiva-se com isto selecionar um subconjunto dos atributos existentes de forma que isto não tenha grande impacto na qualidade da solução final. Esta redução pode ser realizada com o apoio do especialista do domínio, uma vez que, ao remover um atributo potencialmente útil para o modelo final, a qualidade do conhecimento extraído pode diminuir consideravelmente. Além disso, por não se saber inicialmente quais atributos serão importantes para atingir os objetivos, deve-se remover somente aqueles atributos que, com certeza, não têm nenhuma importância para o modelo final.

Existe outra forma de reduzir o número de atributos que é por meio da indução construtiva, em que um novo atributo é criado a partir do valor de outros. Caso os atributos originais utilizados na construção do novo atributo não estejam presentes em um novo modelo, eles podem ser descartados, reduzindo assim o número de atributos.

A terceira forma de redução dos dados consiste na redução do número de valores de um atributo. Conceito este realizado geralmente, por discretização dos valores de um atributo contínuo.

O processo de discretização de um atributo consiste na substituição de um atributo contínuo (inteiro ou real) por um atributo discreto, valendo de faixas de agrupamento de

seus valores. Essencialmente, um algoritmo de discretização aceita como entrada os valores de um atributo contínuo e gera como saída uma pequena lista de intervalos.

Trabalha-se assim com intervalos de dados originalmente contínuos e que são representados na forma de seus limites superior e inferior. Para que os modelos finais sejam representativos da realidade expressa nos dados brutos as transformações apresentadas devem ser criteriosamente preparadas com o devido cuidado, uma vez que é fundamental garantir que as informações presentes nos dados brutos continuem representados nas amostras.

### **3.5.3 Fase 3 - Transformação de Dados**

Objetiva:

- procurar características úteis nos dados tendo em consideração os objetivos a que se destinam;
- utilizar métodos de transformação com vista à redução do número efetivo de variáveis em consideração;
- procurar representações para os dados

O processo de mineração geralmente demanda a integração de dados (combinação de diferentes bases de dados) e a transformação destes (modificações de formato e enriquecimento semântico).

No caso da integração de dados, várias fontes podem ser utilizadas (diferentes bancos de dados, cubos de dados, arquivos diversos). Segundo HAN E KAMBER (2001), alguns tópicos relevantes neste processo são:

- 1 - Integração de esquemas - casamento de entidades relevantes do mundo real (utilização dos metadados);
- 2 - Redundância de atributos (análise de correlação - medida de quanto um atributo implica em outro);
- 3 - Identificação e resolução de valores de dados conflitantes (especialmente devido a diferenças na representação, escala ou codificação);

Pode-se reduzir e evitar redundâncias e inconsistências no conjunto de dados resultante através de uma integração de dados criteriosa, aumentando a precisão e velocidade do processo de mineração de dados.

No caso da transformação de dados, estes são modificados ou transformados em formatos apropriados à mineração:

- Agregação: geração de totalizadores levando em conta determinados atributos ou instâncias. Por exemplo, vendas mensais, anuais, sazonais etc.;
- Generalização: substituição de dados por conceitos de mais alto nível. Idades, por exemplo, podem ser representadas por faixas etárias, localidades por regiões etc.;
- Normalização: atributos são escalonados para uma faixa específica como -1.0 a 1.0, ou 0.0 a 1.0;
- Construção de atributos: novos atributos são construídos a partir de informações pré-existentes (ex.: classificação de crédito a partir de renda e histórico).
- Redução de dados via Agregações; Redução dimensional: detecção e remoção de atributos irrelevantes; Compressão de dados: utilização de mecanismos de codificação para reduzir o tamanho do conjunto de dados e Redução numérica das instâncias através da amostragem (já explicado da fase de pré-processamento).

#### 3.5.4 Fase 4 – Mineração de Dados

Esta etapa objetiva a escolha do foco da mineração de dados (*previsão, descrição*) tendo em vista o processo global. Como visto anteriormente, as etapas de mineração de dados utilizam técnicas e algoritmos de diferentes áreas do conhecimento, principalmente inteligência artificial (especialmente aprendizagem de máquina), banco de dados (recursos para manipular grandes bases de dados) e estatística (comumente na avaliação e validação de resultados), conforme a figura 14.

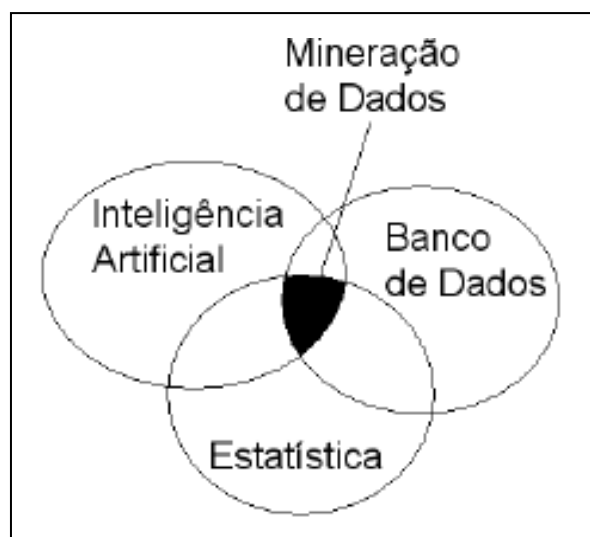


Figura 14 - Mineração de dados utiliza recursos de diferentes áreas

Conforme mencionado, procedimentos estatísticos são utilizados nas etapas de KDD e mais especificamente na mineração de dados. Entretanto, o volume, complexidade e peculiaridades dos eventos e dos dados por eles originados impõem severas limitações a metodologias puramente estatísticas, dentre elas:

- Dados nem sempre possuem independência estatística entre eles, ou seja, muitos domínios possuem inter-relação entre seus objetos e respectivos atributos, comprometendo a aplicação de métodos estatísticos;
- A análise estatística demanda um grau de conhecimento e domínio desta área que apenas estatísticos e profissionais de áreas correlatas possuem, restringindo assim a atuação da grande maioria dos potenciais usuários de procedimentos analíticos;
- Métodos estatísticos manipulam muito bem dados numéricos, mas não manipulam bem valores simbólicos, não conclusivos ou incompletos;
- Estes métodos são computacionalmente caros quando se trata de grandes bases de dados.

Desta forma, percebe-se claramente que a mineração de dados possui grande relevância, contribuição e abrangência no que diz respeito a aplicações e de sistemas. São citados abaixo descrição simplificada dos principais métodos de mineração de dados utilizando aprendizagem de máquina – figura 14.

**Classificação** (aprendizagem de uma função que faça o mapeamento de um elemento dos dados em uma ou várias classes)

**Regressão** (aprendizagem de uma função que faça o mapeamento de um elemento dos dados na valoração de uma variável de previsão)

**Clustering/Segmentação** (identificação de um conjunto finito de categorias ou *clusters* para descrição dos dados)

**Sumarização** (utilização de métodos para procura de uma descrição compacta para um subconjunto de dados)

**Modelagem de Dependências** (associações: procura de um modelo que descreva dependências significativas entre variáveis)

**Deteção de Alterações e Divergências** (descoberta das alterações mais significativas nos dados a partir de valores medidos previamente ou normativos)

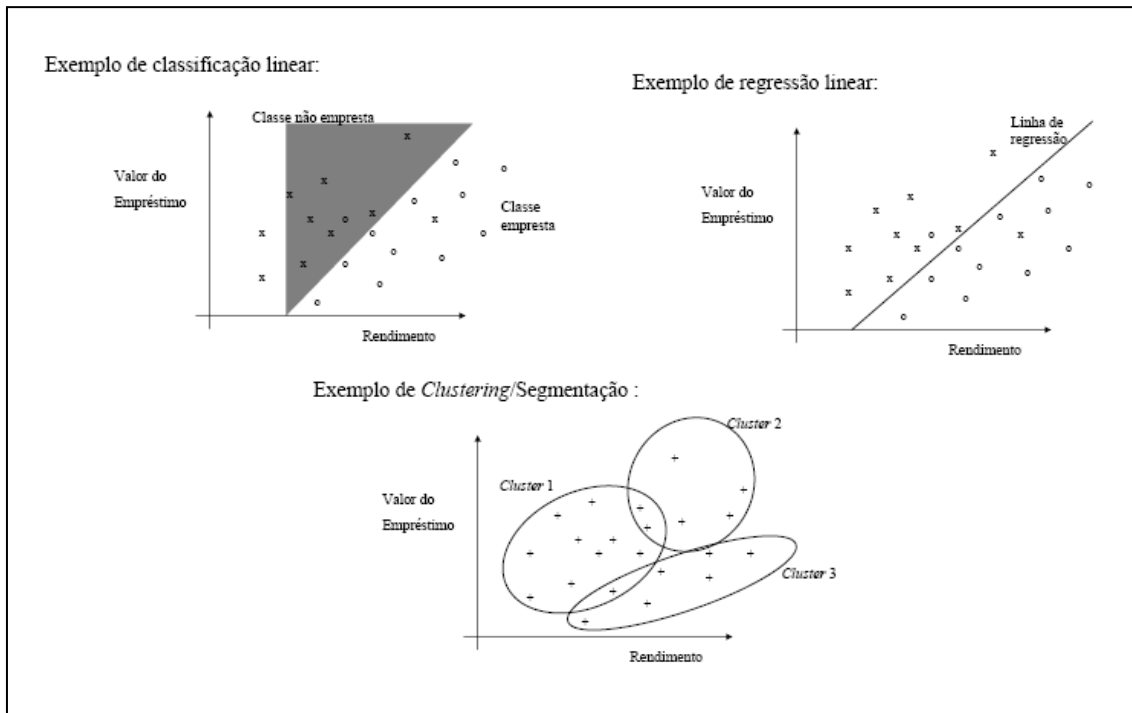


Figura 15 - Exemplos de Métodos de Mineração de Dados

A extração de padrões consiste na aplicação dos algoritmos de Mineração – figura 15 escolhidos para a extração dos padrões embutidos nos dados. É importante ressaltar que dependendo da função escolhida pode ser necessária a execução dos algoritmos de extração de padrões diversas vezes. A disponibilidade do conjunto de padrões extraídos nesta etapa ao usuário ou a sua incorporação a um Sistema Inteligente ocorre após a análise e/ou o processamento dos padrões na etapa de pós-processamento.

#### Escolha da tarefa

A escolha da tarefa é feita de acordo com os objetivos desejáveis para a solução a ser encontrada. As tarefas possíveis de um algoritmo de extração de padrões podem ser agrupadas em atividades **preditivas** e **descritivas**, conforme mostrado na figura 16.

A Mineração de Dados Preditiva consiste na generalização de exemplos ou experiências passadas com respostas conhecidas ou regras de negócio estabelecidas por especialistas em uma linguagem capaz de estabelecer a classe de um exemplo.

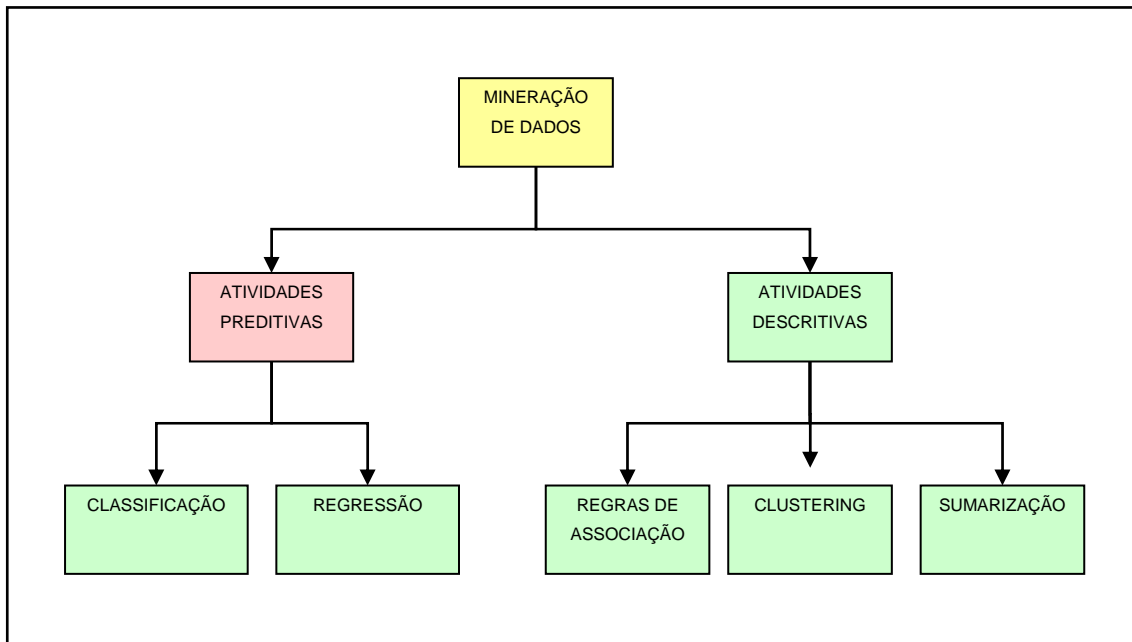


Figura 16 – Taxonomia de Tarefas de Mineração de Dados

Os dois principais tipos de tarefas para predição são classificação e regressão.

- A **classificação** consiste na predição de um valor categórico como, por exemplo, prever se vai chover ou não, se o cliente provavelmente vai pagar em dia ou não.
- A **regressão**, estabelece que o atributo a ser predito consiste em um valores não discretos ou contínuos como, por exemplo, prever o lucro ou a perda em um empréstimo (WEISS e INDURKHYA,1998).

#### A Mineração de Dados Descritiva

Esta atividade consiste na identificação de comportamentos intrínsecos do conjunto de dados, sendo que estes dados não possuem uma classe especificada. Algumas das tarefas de descrição são clustering, regras de associação e sumarização.

Uma vez eleita a tarefa a ser empregada, existe uma variedade de algoritmos para executá-la. A definição do algoritmo de extração e a posterior configuração de seus parâmetros também são realizadas nesta etapa. Por isso, a escolha de vários algoritmos para realizar a tarefa desejada pode ser feita, levando a obtenção de diversos modelos que, na etapa de pós-processamento, são tratados para fornecer o conjunto de padrões mais adequado ao usuário final.

### 3.5.5 Fase 5 - Pós-Processamento e Interpretação do Conhecimento

Objetivo: Consolidação do conhecimento descoberto através da interpretação dos padrões minerados com o possível regresso a uma das fases anteriores para maior validação do conhecimento

- **incorporação deste conhecimento** no sistema ou elaboração de relatórios para as partes interessadas.
- **Verificação e resolução de potenciais divergências** com conhecimento tido como verdadeiro (ou previamente extraído)

A obtenção do conhecimento não é o passo final do processo de Extração de Conhecimento de Bases de Dados. O conhecimento extraído pode ser utilizado na resolução de problemas da vida real, seja por meio de um Sistema Inteligente ou de um ser humano como apoio a algum processo de tomada de decisão. Para isso é importante que algumas questões sejam respondidas aos usuários (LIU e HSU, 1996):

- O conhecimento obtido corrobora ou contesta o conhecimento do especialista?
- Em que partes o conhecimento do especialista está correto e onde não está correto?

Independentemente disto, nem sempre se pode responder essas questões, já que os algoritmos de extração de padrões podem gerar uma quantidade enorme de padrões, muitos dos quais podem não ser importantes, relevantes ou interessantes para o usuário.

Sabe-se também que fornecer ao usuário uma grande quantidade de padrões descobertos não é produtivo, pois, normalmente, ele procura uma pequena lista de padrões interessantes. Portanto, é de vital importância desenvolver algumas técnicas de apoio no sentido de fornecer aos usuários apenas os padrões mais interessantes (SILBERSCHATZ & TUZHILIN 1995).

Diversas medidas para avaliação de conhecimento têm sido pesquisadas com a finalidade de auxiliar o usuário no entendimento e na utilização do conhecimento adquirido. Estas medidas podem ser divididas entre medidas de desempenho e de qualidade. As medidas de desempenho são em geral precisão, erro, confiança negativa, sensibilidade, especificidade, cobertura, suporte, satisfação, velocidade e tempo de aprendizado.

Já as medidas de qualidade são necessárias, pois, um dos objetivos do processo de Extração de Conhecimento é que o usuário possa compreender e utilizar o



conhecimento recém descoberto, contudo, há casos em que os modelos são muito complexos ou não são interpretáveis pelos os especialistas.

De todo modo, a compreensão do conhecimento que foi produzido é fundamental para validação do processo. A compreensão extraída de um conjunto de regras está relacionada com a facilidade de interpretação dessas regras por um ser humano.

Um fato determinante é o de que os usuários especialistas têm a tendência a compreender melhor modelos que não contrariam seu conhecimento prévio.

Medidas objetivas são aquelas que estão relacionadas somente com a estrutura dos padrões e do conjunto de dados de teste. Elas não levam em consideração fatores específicos do usuário nem do conhecimento do domínio para avaliar um padrão.

Diferentes usuários finais podem ter diferentes graus de interesse para um determinado padrão, e assim, medidas subjetivas são necessárias. Estas medidas consideram que fatores específicos do conhecimento do domínio e de interesse do usuário devem ser tratados ao selecionar um conjunto de regras interessantes ao usuário.

Os procedimentos de extração pode ser repetido varias vezes, desde que, ao final da análise, o conhecimento obtido não seja de interesse do usuário final ou não cumpra as premissas propostas. Desse modo permanece-se ajustando os parâmetros até que se obtenha resultados melhores.

#### **4 - Mineração de Dados Aplicado à Tomada de Decisão de Substituição de Hidrômetros**

No capítulo 1, foram citadas questões que têm sido uma constante nas empresas de saneamento, envolvendo decisões para diminuir as perdas e aumentar arrecadação.

No que tange as perdas aparentes, que são aquelas provenientes de problemas na medição podem ocorrer por mau funcionamento do hidrômetro (aparelho que mede o consumo de água nos domicílios) ou pela ausência deste. Neste último caso, a empresa não tem instalado o aparelho de medição e cobra do consumidor uma taxa genérica, seja esta por estimativa ou por média de consumo estimado e, neste caso, sem a medição do consumo preciso pode ser configurada alguma "perda". Isto porque a empresa fornece um determinado volume de água e pode estar faturando com base em um volume normalmente subestimado.

Outros motivos também frequentes são:

- os prepostos da empresa concessionária não registram com regularidade a leitura do hidrômetro ou,
- o hidrômetro fica obsoleto ou danificado, sendo que a concessionária não age em tempo hábil de substituí-lo.

Se o hidrômetro não funciona com precisão e registra um consumo menor do que o real, a água também é considerada como "não contabilizada" para a prestadora de serviço, porque esta não fatura aquilo que coletou, tratou e entregou ao consumidor. No segmento de saneamento estima-se que metade dos problemas estejam ligados a vazamento( perdas físicas) e outra metade decorrentes de falhas na medição( perdas aparentes).

O dimensionamento correto de um medidor consiste na determinação do aparelho a ser instalado em uma ligação específica, de forma que o mesmo opere dentro da faixa de vazão ideal para a qual ele foi projetado. Para isto, é recomendado que o dimensionamento seja fundamentado nas vazões nominal, mínima e máxima, às quais o medidor estará sujeito. O dimensionamento é necessário para ligações novas ou readequações de ligações já existentes.

Conforme citação no Programa Nacional de Combate ao Desperdício de Água PNCD (1999), são consideradas as seguintes metodologias para o dimensionamento dos medidores:

- Dimensionamento por vazão estimada
- Dimensionamento por categoria de consumo
- Levantamento direto do perfil de consumo
- Redimensionamento quando há variações significativas do consumo

Segundo PNCD (1999), o método de dimensionamento mais indicado é o “Dimensionamento por Levantamento Direto do Perfil de Consumo”. Este método, porém é oneroso para ser aplicado a todos os usuários, sendo recomendado apenas para grandes usuários. Contudo, o método pode ser aplicado por amostragem em categorias de consumo homogêneas, por exemplo, residências de uma determinada região, e seus resultados extrapolados para a categoria em função de parâmetros conhecidos, adotando-se então o método de dimensionamento por categoria de consumo, utilizando-se tabelas pré-dimensionadas.

#### **4.1 - Política de Troca de Hidrômetros**

As empresas de saneamento em geral trabalham com as seguintes hipóteses de troca de medidores para diminuir perdas na micro medição e aumentar a arrecadação:

1. Analisar a performance dos macros medidores.
2. Localizar ligações clandestinas e outras irregularidades.
3. Buscar entender, segmentar e tratar clientes em Agrupamentos Operacionais.
4. Utilizar também fatores sócio-econômicos para localizar hidrômetros com problemas
5. Localizar hidrômetros com problemas através de ferramentas que analisem o comportamento e o histórico de consumo, o que ultimamente tem se mostrado mais eficiente do que a troca de um hidrômetro por idade e uso.
6. Analisar os resultados de uma dada política para troca de hidrômetros comparando os respectivos consumos e o faturamento antes e depois das trocas.

## 4.2 - Conceituação do Problema

Em uma determinada empresa de saneamento, levantamentos realizados em anos anteriores a troca dos hidrômetros, comparando-se macro medições com as micro medições apontaram índices de perdas aparentes superiores a 26%.

Em seguida, realizou-se um levantamento do número de medidores vencidos por tempo de uso, em geral com idade superior a cinco anos de uso. A substituição destes hidrômetros foi então conciliada a um estudo para verificação da influência da submedição destes hidrômetros sobre o índice de perdas.

Uma determinada empresa seguindo sua política de diminuir perdas na micro medição e aumentar a arrecadação, estabelecida por especialistas, estabeleceu como meta:

- A substituição e a instalação de 8,4% do parque de hidrômetros de consumidores e clientes da empresa.
- A justificativa principal se deve à não regularidade de troca dos medidores por pelo menos dez anos, considerando-se que o tempo médio de vida útil de medidores é de cinco anos e que estes sofrem bastante com ação do tempo, o efeito de abrasão e corrosão, tornando-se com isto equipamentos pouco confiáveis para o destino que lhes é imputado.

Diante dos fatos apontados e das restrições orçamentárias da empresa, os especialistas propuseram a direção que deveria ser feito investimento na aquisição e instalação de 135 mil hidrômetros dos mais diferentes diâmetros ou bitolas, em um conjunto de cerca de 1,6 milhões de consumidores escolhidos aleatoriamente, segundo as seguintes regras:

**REGRA1** - Substituição de 105 mil hidrômetros de qualquer diâmetro, que estavam “parados” além daqueles com mais de 5 anos de uso;

**REGRA2** - Instalação de 30 mil hidrômetros em clientes ainda não hidrometrados, ou seja, que não dispunham de medidores instalados até a data e que eram cobrados por uma média estimada de consumo.

### 4.2.1 Modelo do Especialista - Resultados Após a Troca dos Hidrômetros

A troca dos hidrômetros ocorreu durante um período que durou cerca de oito meses e proporcionou os seguintes resultados e conclusões:

- A. O volume que continuou não sendo medido deve-se a outros tipos de perdas como vazamentos, ausência de hidrômetros ou fraudes, tornando necessários estudos específicos para sua identificação.

- B. A renovação do parque de hidrômetros proporcionou resultados positivos em relação à redução de perdas não físicas de água de abastecimento, ou seja, o consumo micro medido aumentou em relação ao período anterior à troca. O volume médio medido pelos hidrômetros novos apresentou um acréscimo médio superior a 20% em relação aos que se encontravam instalados.
- C. Constatou-se ser importante a renovação constante do parque de hidrômetros confirmando o estado de deterioração dos hidrômetros vencidos por tempo de uso.
- D. Durante e após a instalação dos novos hidrômetros houve bastante atuação do Serviço de Atendimento ao Consumidor, no sentido de esclarecer dúvidas aos questionamentos dos clientes, quanto à mudança do volume medido e do valor cobrado em geral para maior da conta emitida.
- E. Recomendaram-se para a região avaliada novos levantamentos e acompanhamentos do índice de perdas. A persistência do índice de perdas próximo ao inicialmente encontrado indicará a necessidade da continuidade dos estudos através de georeferenciamento para se detectar vazamentos, irregularidades e fraudes em ligações na região observada.
- F. A arrecadação nos meses subsequentes as trocas dos medidores, frustrou as expectativas de aumento significativo da receita da concessionária. Apesar de ter havido um aumento no volume medido e faturado, obtido comparando-se a média do volume medido nos cinco meses anteriores à troca do hidrômetro de um determinado cliente com a média do período respectivo posterior. É importante registrar que após os primeiros meses após a substituição do hidrômetro, este valor pode se reduzir em decorrência da adaptação dos consumidores ao eventual reposicionamento em outra faixa tarifária resultante do aumento registrado no volume medido.
- G. Na troca dos medidores o tempo de retorno para o investimento na aquisição e instalação foi estimada para ocorrer entre 9 a 16 meses, dependendo da adaptação e aceitação dos consumidores com os valores constantes de suas novas faturas de consumo de água.

Na figura 17, nos dois gráficos apresentados constam; no primeiro a quantidade de clientes atingida mês a mês, nos oito meses de trocas dos hidrômetros e, no segundo a variação percentual da arrecadação considerando-se o mês 1 como base.

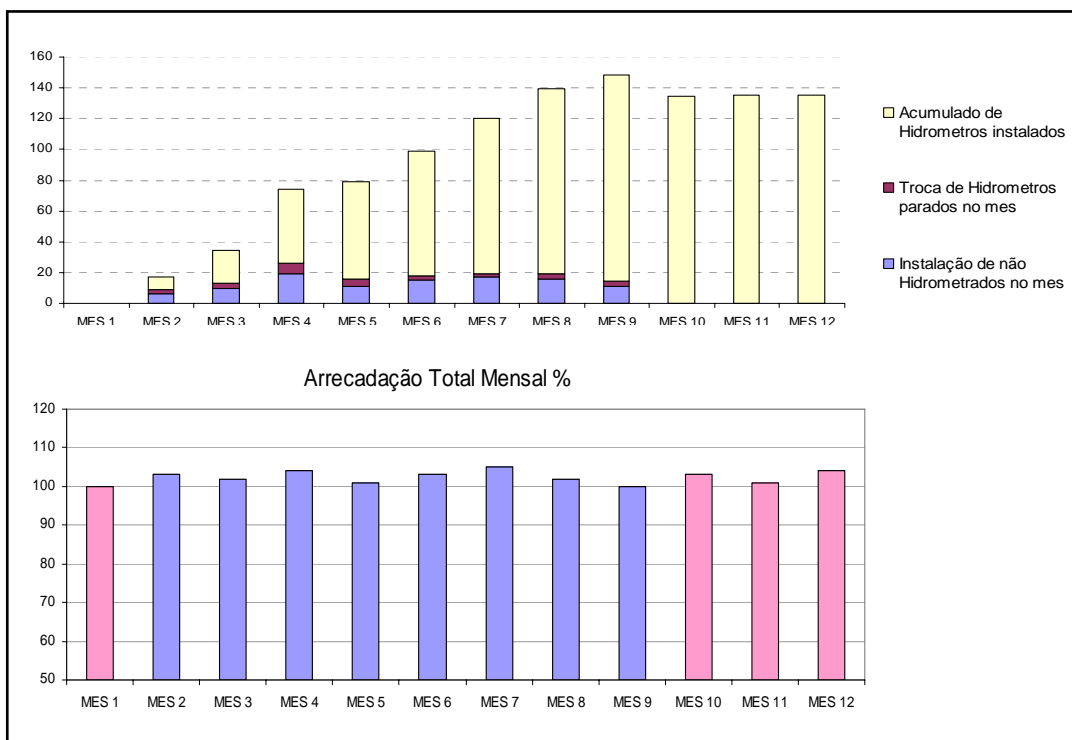


Figura 17 – Hidrômetros substituídos e Evolução da Arrecadação no Período Observado

### 4.3 - Aplicação do Modelo de Mineração de Dados ao Problema

A aplicação de Mineração de Dados apoiando a tomada de decisão de troca de um novo grupo de hidrômetros segue as fases descritas no capítulo 3 considerado neste trabalho. Neste caso, pretende-se obter um modelo, usando técnicas de mineração de dados, que comparado com o que foi implementado pelo Especialista, permita resultados tais que possa propor para a gerencia, no próximo processo de instalação de hidrômetros, que além de diminuir as perdas aparentes (aumentar  $m^3$  medido), alcance primordialmente aumento sustentável da arrecadação deste grupo de clientes. A tabela 2 exprime o que se pretende.

Tabela 2 – Proposta de Domínio do Problema a Resolver

ANTES DA TROCA DE HIDRÔMETRO		APÓS A TROCA DE HIDRÔMETRO	
M³ MEDIDOS	FATURAMENTO 5 MESES ANTERIORES	M³ MEDIDOS	FATURAMENTO 5 MESES POSTERIORES
	ADIMPLENTES		TROCA arrecadação posterior $\geq$ 15% que a anterior a troca
			NAO_TROCA arrecadação posterior $<$ 15% que a anterior a troca
	INADIMPLENTES		TROCA arrecadação posterior $\geq$ 15% que a anterior a troca
			NAO_TROCA arrecadação posterior $<$ 15% que a anterior a troca

Durante o desenvolvimento da solução estará sendo atribuída a classe LUCRO quando a arrecadação posterior à troca for maior ou igual a 15% da arrecadação imediatamente anterior a troca ou instalação do novo hidrômetro. Do mesmo modo, será atribuída a classe PREJUÍZO, quando arrecadação medida posteriormente à troca for menor do que 15% da arrecadação imediatamente anterior à troca ou instalação.

Na evolução da solução foram utilizadas e abandonadas outras tentativas de formulação de classes, como mostrado nos tópicos seguintes, mas que não se mostraram representativas.

Mais adiante chamaremos de TROCA à decisão que envolve LUCRO (arrecadação posterior  $\geq$ 15% que a anterior a troca). E conseqüentemente será utilizado NÃO\_TROCA representando a decisão considerada PREJUÍZO (arrecadação posterior  $<$ 15% que a anterior a troca).

Como o foco central do problema da empresa, consiste em se detectar possíveis fugas no faturamento e na arrecadação, diante da decisão de troca de hidrômetros, não será tratado no domínio o volume (m<sup>3</sup>) medido quer antes ou após a ação de substituição e instalação de novos medidores.

#### 4.3.1 Seleção dos Dados

Para efeito do estudo acadêmico formulado e efeito didático estão sendo utilizados nomes de agrupamentos operacionais, bairros e logradouros característicos do Estado do Rio de Janeiro, mas para preservação da identidade e características das empresas

tratadas, cabe informar que tais dados não têm necessariamente correspondência com as localidades e bairros reais.

Para este trabalho foram segregados dados referentes ao período de 8 meses onde se constatou a troca dos 135mil hidrômetros, onde foram selecionados preliminarmente “atributos sensíveis” à política de troca do especialista, como:

- **DAE** – atributo numérico identifica o Código do Arranjo Operacional de Atendimento físico de responsável pela atuação.
  
- **NOME\_DAE** – atributo alfabético que caracteriza o Nome do Arranjo Operacional de Atendimento físico de responsável pela atuação.

Botafogo  
Centro  
Meier  
São Gonçalo  
São João de Meriti  
Itatiaia  
Rio Bonito  
Ilha de Paqueta  
Passa Três  
Jacarepaguá  
.....  
.....  
Méier  
Tijuca  
Alto da Boa Vista

- **CATEGORIA** – atributo alfabético identifica se o consumidor se caracteriza tipo de economia

Domiciliar  
Comercial  
Industrial

- **SUBCATEGORIA** – atributo alfabético identifica se o consumidor se caracteriza o subconjunto do tipo de economia

Comum  
Baixa Renda  
Conjunto Habitacional  
Entidade sem fim Lucrativo  
Pública Federal  
Pública Estadual  
Pública Municipal  
Decisão Judicial



- **CICLO** - atributo textual identifica os ciclos de medição das leituras nos medidores
- **ECONOMIAS** atributo numérico identifica o numero de domicílios atendido por aquele hidrômetro
- **BAIRRO** atributo textual identifica o Bairro onde se localiza determinado consumidor

São Gonçalo  
 São João de Meriti  
 Itatiaia  
 Rio Bonito  
 Ilha de Paquetá  
 Passa Três  
 Jacarepaguá  
 .....  
 .....  
 Méier  
 Tijuca  
 Alto da Boa Vista

- **HIDROMETRO** atributo textual identifica a bitola para o hidrômetro dimensionado

1/2"  
 3/4"  
 1"  
 1 1/2"  
 2"  
 3"  
 4"  
 6"  
 >6"

- **FATURAMENTO\_5ANTES** atributo numérico identifica o valor em reais (R\$) do somatório de faturamento daquela economia 5 meses antes do mês da efetiva troca do hidrômetro.
- **DECISAO** atributo numérico identifica a CLASSE de decisão do problema de trocar ou não o hidrômetro considerando-se que foi estabelecido para o modelo que:

TROCA se o faturamento dos 5 meses subseqüentes superar ou for ou igual 15% ao faturamento dos 5 meses anteriores a troca  
 NÃO\_TROCA se o faturamento dos 5 meses subseqüentes for menor a 15% do que o faturamento dos 5 meses anteriores a troca

#### 4.3.2 Pré-processamento

Nesta fase preliminarmente utilizou-se a CLASSE = COMPORTAMENTO sendo:

LUCRO	se o faturamento dos 5 meses subsequentes a TROCA superar 10% do faturamento dos 5 meses anteriores a TROCA
ESTÁVEL	se o faturamento dos 5 meses subsequentes a TROCA ficar na faixa entre -10% e +10% do faturamento dos 5 meses anteriores a TROCA
PREJUÍZO	se o faturamento dos 5 meses subsequentes a TROCA for menor a 10% do faturamento dos 5 meses anteriores a TROCA

A escolha desta classe não se mostrou a mais adequada para análise do modelo, pois, os resultados se mostraram indiferentes ou pouco representativos.

Desse modo buscou-se uma CLASSE que representasse efetivamente uma DECISÃO de TROCAR ou NÃO\_TROCAR os hidrômetros, e que apresentou resultados mais representativos, abandonando-se assim a CLASSE anterior.

Foi descartado o atributo ATIVIDADE que descreve o tipo de atividade comercial desenvolvida (hotel, restaurante, clinica, clube, confecção etc..) em uma dada economia ou consumidor, por esta estar com cadastramento comprometido ou “sujo”, ou seja, dados incorretos e repetitivos carecendo de um processo recadastramento para depurar este atributo...

#### 4.3.3 Integração dos Dados

Neste modelo não foram integrados outros dados que caracterizassem a condição socioeconômica do consumidor que pudesse identificar seu perfil de adimplência ou inadimplência comercial perante outras concessionárias ou mesmo fornecedores de bens e serviços. Este dado sem dúvida seria muito útil no sentido de se perceber seu perfil diante da pré-disposição deste consumidor em honrar ou não suas dívidas.

Diferentemente de outras concessionárias de energia ou mesmo telefonia, na área de saneamento como ainda não se consegue a individualização do consumidor, pois o medidor no caso de unidades prediais chega apenas na entrada do edifício e não na unidade habitacional do mesmo o que seria ideal.

Projetos tramitam no sentido de buscar a individualização destes serviços, mas, na pratica isto mesmo para projetos de prédios novos o custo de estabelecimento de ramais específicos para cada unidade condominial se torna muito caro e algumas vezes inexecutável.

O processo de mineração, como neste caso demandaria a integração de dados com a combinação de diferentes bases de dados e a sua transformação partir de modificações de formato e outras.

Algumas parcerias têm sido firmadas entre as empresa de saneamento e as instituições de administração de créditos como o SERASA, buscando-se assim dispor-se de um banco de dados integrado com a condição socioeconômica do consumidor, porem a adaptação e compatibilização dos cadastros não tem se mostrado praticas o bastante. Durante o processo iterativo de mineração, alguns atributos foram descartados logo depois de incluídos por não se mostrarem sensíveis ao modelo.

Outras fontes de dados foram obtidas a partir de bases OLAP (cubos de dados) do Sistema de Informações Gerenciais da empresa.

#### **4.3.4 Mineração de dados**

Foi utilizado na fase de mineração dos dados o software **WEKA** e para geração dos dados em formato CSV e arff o software Ultra Edit 3.2, provenientes dos arquivos amostras em **EXCEL** extraídos do banco de dados comercial corporativo **Oracle**.

Foi utilizado na consecução do problema, algoritmo de classificação utilizando Arvores de Decisão. Em decorrência das características dos dados do problema, contendo atributos contínuos e discretos, foram simulados vários estudos e validações através do algoritmos de classificação o **J48**, pois trata atributos discretos como DAE, BAIRRO, como também os atributos contínuos como FATURAMENTO5\_ANTES.

#### **Deteção de Alterações e Divergências**

Alguns dos atributos não tiveram representação significativa para o modelo e em alguns casos foram excluídos como o CICLO (de leitura) que não agregou significado quando considerado no modelo. No desenvolvimento do trabalho, pela experiência ficou clara a necessidade de se segmentar este universo a ser estudado em agrupamentos operacionais, dada as características geográficas e sócio econômicas das regiões.

Foram escolhidos dos cento e dez (110) agrupamentos mostrados no Apendice2, treze (13) considerados pelo autor e pelos especialistas como os mais significativos para serem elaboradas comparações e conclusões. Na Tabela 3 é apresentado um resumo e evidenciando treze (13) dos agrupamentos escolhidos para estudo e suas principais características.

Tabela 3 – AMOSTRA DOS AGRUPAMENTOS OPERACIONAIS ESTUDADOS

ID	DESC_DAE	Hydrômetros substituídos	Faturamento 5 meses (R\$)	Fatur / Subs (R\$)	Média Mensal (R\$)
1	BOTAFOGO	653	992.830,27	1.520,41	304,08
2	JARDIM PRIMAVERA	33	38.793,40	1.175,56	235,11
3	ILHA GOVERNADOR	1603	818.305,84	510,48	102,10
5	CENTRO	3768	1.551.094,08	411,65	82,33
8	BARRA DA TIJUCA	2709	816.076,35	301,25	60,25
9	PAQUETA	40	9.599,53	239,99	48,00
15	JACAREPAGUA	6981	962.943,94	137,94	27,59
16	RAMOS	16900	2.277.045,92	134,74	26,95
17	MEIER	9247	1.225.129,60	132,49	26,50
23	TERESOPOLIS	292	26.155,37	89,57	17,91
26	CAMPO GRANDE	8510	613.910,75	72,14	14,43
29	RIO BONITO	95	6.369,71	67,05	13,41
47	SAO GONCALO	15752	778.876,32	49,45	9,89

Os agrupamentos estudados têm seus resultados apresentados no **Quadro 1 – Estudos em Agrupamentos Operacionais**, onde foram selecionadas para análise algumas métricas do software **WEKA**:

O detalhamento de cada um dos estudos consta do Apêndice 3, para cada agrupamento analisado, utilizando o algoritmo de classificação **J48**.

ESTUDOS EFETUADOS	Correctly Clas. Instances		TP Rate	FP Rate	Precision	Recall	F-Measure	Class	a b <- classified as		
	Incorrectly Clas. Instances										
Barra da Tijuca	2182	80.5463 %	0.911	0.256	0.673	0.911	0.774	NAO-TROCA	903	88	a = NAO-TROCA
	527	19.4537 %	0.744	0.089	0.936	0.744	0.829	TROCA	439	1279	b = TROCA
Botafogo	585	89.5865 %	0.945	0.215	0.909	0.945	0.926	NAO-TROCA	428	25	a = NAO-TROCA
	68	10.4135 %	0.785	0.055	0.863	0.785	0.822	TROCA	43	157	b = TROCA
Campo Grande	7820	91.8919 %	0.342	0.005	0.902	0.342	0.496	TROCA	339	653	a = TROCA
	690	8.1081 %	0.995	0.658	0.92	0.995	0.956	NAO-TROCA	37	7481	b = NAO-TROCA
Centro	2875	76.3004 %	0.056	0.002	0.914	0.056	0.106	TROCA	53	888	a = TROCA
	893	23.6996 %	0.998	0.944	0.761	0.998	0.863	NAO-TROCA	5	2822	b = NAO-TROCA
Ilha do Governador	1194	74.4853 %	0.383	0.039	0.855	0.383	0.529	TROCA	230	370	a = TROCA
	409	25.5147 %	0.961	0.617	0.723	0.961	0.825	NAO-TROCA	39	964	b = NAO-TROCA
Jacarepaguá	650	92.068 %	0.992	0.823	0.926	0.992	0.958	TROCA	639	5	a = TROCA
	56	7.932 %	0.177	0.008	0.688	0.177	0.282	NAO-TROCA	51	11	b = NAO-TROCA
Jardim Primavera	93	88.5714 %	0.955	0.471	0.913	0.955	0.933	NAO-TROCA	84	4	a = NAO-TROCA
	12	11.4286 %	0.529	0.045	0.692	0.529	0.6	TROCA	8	9	b = TROCA
Méier	7186	77.7117 %	0.995	0.949	0.777	0.995	0.873	NAO-TROCA	7076	35	a = NAO-TROCA
	2061	22.2883 %	0.051	0.005	0.759	0.051	0.096	TROCA	2026	110	b = TROCA
Paqueta	35	87.5 %	0.935	0.333	0.906	0.935	0.921	NAO-TROCA	29	2	a = NAO-TROCA
	5	12.5 %	0.667	0.065	0.750	0.667	0.706	TROCA	3	6	b = TROCA
Ramos	12585	74.4675 %	0.996	0.981	0.745	0.996	0.853	NAO-TROCA	12502	46	a = NAO-TROCA
	4315	25.5325 %	0.019	0.004	0.643	0.019	0.037	TROCA	4269	83	b = TROCA
Rio Bonito	83	87.3684 %	0.973	0.5	0.880	0.973	0.924	NAO-TROCA	73	2	a = NAO-TROCA
	12	12.6316 %	0.5	0.027	0.833	0.5	0.625	TROCA	10	10	b = TROCA
São Gonçalo	13408	85.1193 %	0.998	0.967	0.852	0.998	0.919	NAO-TROCA	13329	31	a = NAO-TROCA
	2344	14.8807 %	0.033	0.002	0.718	0.033	0.063	TROCA	2313	79	b = TROCA
Teresópolis	249	85.274 %	0.618	0.005	0.986	0.618	0.76	TROCA	68	42	a = TROCA
	43	14.726 %	0.995	0.382	0.812	0.995	0.894	NAO-TROCA	1	181	b = NAO-TROCA

Quadro 1 – Estudos em Agrupamentos Operacionais

No capítulo 5, são comentadas as conclusões a que chegaram especialistas, usuários e analistas sobre os estudos efetuados e os benefícios encontrados com este trabalho.

## 5 - Conclusões

Deve ser salientada neste ponto, a razão da escolha do processo de mineração que foi conduzido utilizando-se Árvores de Decisão. Foi considerado o aspecto de facilidade de visualização por parte dos especialistas e a aceitação desta técnica que proporcionando o aprofundamento das análises dos resultados através da interação analista-especialista.

Foram formulados vários estudos concentrando análises nos chamado de “agrupamentos operacionais” e, onde na verdade se tem o maior conhecimento ou sensibilidade do comportamento histórico de um determinado grupo de clientes.

### 5.1 - Interpretação dos Resultados

Observam-se no quadro 1, os vários estudos realizados para os estudos de alguns dos “agrupamentos” neste trabalho:

Houve uma prevalência nos Agrupamentos Operacionais – DAE observados, da classe **NÃO\_TROCA**, ou seja, o faturamento dos 5 meses subsequentes à substituição do hidrômetro foi menor do que 15% do faturamento dos 5 meses anteriores à troca, a menos do agrupamento **Barra da Tijuca e Jacarepaguá** que apresentaram índices de Instâncias Classificadas Corretas acima de 80% para a classe **TROCA**.

Os Agrupamentos **Centro, Méier, Ramos e São Gonçalo** apesar de apresentaram precisão entre 74% e 85% de instâncias classificadas como corretas, para a classe **NÃO\_TROCA**, analisando-se os índices de *Verdadeiro Positivo*, percebe-se taxas de acerto abaixo de 0.05 para a classe **TROCA**, comparando-se os resultados na *matriz de confusão*.

Por outro lado, os agrupamentos **Botafogo, Campo Grande, Jardim Primavera, Paquetá Rio Bonito e Teresópolis**, apresentaram precisão entre 76% e 92% favorecendo também a classe **NÃO\_TROCA**. Analisando-se os índices de *Verdadeiro Positivo* obtivemos taxas na ordem de 0.9, mas, os indicadores de *Verdadeiro Positivo* para a decisão **TROCA** são um pouco melhores do que aqueles obtidos no comentário do item imediatamente anterior ao se observar as respectivas matrizes de confusão.

**Ilha do Governador** foi o agrupamento onde se obteve resultados intermediários, ou seja, indicadores de Instâncias Corretas de 74% , boas taxas de Verdadeiro Positivo oscilando entre 0,961 para NÃO\_TROCA e índices de Verdadeiro Positivo para TROCA de apenas 0,38.

Foram confirmadas algumas das conclusões a que chegaram os especialistas, as de que com a aplicação das regras 1 e 2, obteve-se sim o aumento do volume medido já explicado anteriormente, porém isto não se refletiu na arrecadação, talvez explicado pelo longo tempo em que a empresa não trocou os hidrômetros com regularidade. Informações obtidas no *Call Center* durante no período da troca, apontaram como o maior percentual de reclamações a não divulgação prévia na mídia, dos reflexos que esta troca poderia causar nas contas futuras, tais como o aumento do volume medido bem como o aumento do valor faturado.

O modelo de Descoberta de Conhecimento aplicado ao problema acrescentou ao conhecimento do especialista, padrões de atuação que podem ser levados em conta no futuro. Por exemplo, a partir de então, em trocas futuras se pode privilegiar agrupamentos e subconjuntos de agrupamentos onde os modelos apresentaram maiores índices de precisão e não trocar necessariamente um determinado conjunto de equipamentos (hidrômetros) onde se tem maior imprecisão de resultados.

A integração de outros dados provenientes de fontes tais como características sócio-econômicas da localidade, densidade populacional, taxa de expansão do logradouro, fluxo e índice de mobilidade de habitantes, entre outras poderão enriquecer o modelo trazendo novas conclusões.

A incorporação e integração dos dados como vazão e pressão das redes de abastecimento, permitirão a integração dos dados comerciais e operacionais aprimorando o dimensionamento mais apurado dos hidrômetros a serem instalados.

Ao analisar as Árvores de Decisões obtidas, o especialista verificou e confirmou que os atributos HIDRÔMETRO (diâmetro), FATUTAMENTO5\_ANTES e CATEGORIA,

são os *nós* que mais se repetem nas árvores podadas e, portanto aqueles de maior significado.

Com o aprimoramento deste modelo, certamente os dados históricos obtidos poderão minimizar potenciais divergências e aproximar equipes de KDD ligados ao estudo do modelo e especialistas consagrados da empresa.

SILBERSCHATZ & TUZHILIN (1995), afirmam que é importante desenvolver algumas técnicas no sentido de apoiar os usuários e especialistas munindo-os dos padrões mais interessantes. Pode-se concluir que nas medidas de qualidade resultantes deste processo de Extração de Conhecimento, a *Compreensibilidade* do modelo resolvido com Árvores de Decisão apresentadas no Apêndice 3 foi confirmado através das regras mostradas. A *Interessabilidade* e a *Inesperabilidade* também se comprovaram observando-se principalmente o que foi discutido acima, ao se confrontar as expectativas do especialista frustrando de certo modo a sua decisão original de TROCA.

Para a empresa ficou evidenciada a necessidade de se tratar este tipo de problema com uma equipe interdisciplinar, agregando conhecimentos do Especialista (domínio de informações e resultados históricos, mas com comportamento introspectivos e conhecimento tácitos) com os conhecimentos de Analistas que dominem técnicas de Mineração de Dados provocando assim a explicitação do conhecimento documentando-o na empresa.



## **5.2 - Trabalhos Futuros**

Considerando-se as conclusões anteriores, focando-se na área comercial das empresas, aplicações que exigem a tomada de decisão de corte dos ramais de consumidores inadimplentes são algumas que podem ser exploradas.

Outra aplicação seria uso de modelos na detecção de fraudes ou desvios de consumo que apresentassem variações e comportamentos atípicos.

Na área operacional ao se integrar os dados das redes de abastecimento de água e esgoto, poderiam trazer novidades em informações para administração e gestão do negócio, bem como a possibilidade no aumento de produtividade.

Na gestão inteligente dos estoques de produtos químicos, insumos essenciais no processo de tratamento da água, validando e comparando desvios associados à qualidade da água que é entregue ao consumidor final.

## Referências

- ADRIAANS, P., ZANTINGE, D., 1996. *Data Mining*. England, Addison Wesley Longman.
- BERTOLDI, J., 2002, *O Painel Estratégico como Ferramenta de Avaliação de Desempenho: Uma abordagem conceitual em uma empresa do ramo metalúrgico. Jaraguá do Sul, SC..89f.* Dissertação de Mestrado – Programa de Pós-graduação em administração – PPGA
- BRADLEY, P., FAYYAD, U. , MANGASARIAN, O.,1998, *Data mining: Overview and optimization opportunities*. Technical Report MSR-TR-98-04, Microsoft Research Report, Redmond, WA.
- CAVALCANTI, M., GOMES, E., PEREIRA, A., 2001, *Gestão de empresas na sociedade do conhecimento: um roteiro para a ação*. Rio de Janeiro, Campus.
- DRUCKER, P. , 1999, *Desafios gerenciais para o século XXI*. São Paulo, Futura.
- FAIRBANKS, M., LINDSAY, S. , 2000, *Arando o mar. fortalecendo as forças ocultas do crescimento em países em desenvolvimento*. Rio de Janeiro: Qualitymark, 2000.
- FAYYAD, U. M., PIATESKY-SHAPIRO, G.; SMYTH, P., 1996 “*From Data Mining to Knowledge Discovery: An Overview*”. In: *Advances in Knowledge Discovery and Data Mining*, AAAI Press.
- GARDNER, S. R.,1998, *Building the data warehouse*. Communications of the ACM 41(9), 52–60.
- GARVIN, D. A., NAYAK, P. R. MAIRA, A. N. BRAGAR, J. L., 1998. *Aprender a Aprender*. HSM Management
- GONÇALVES, E, ALVIM, P.R.A., 2007 ” *G3 - Guia Prático Para Pesquisa e Combate A Vazamentos Não Visíveis* ”, Programa de Modernização do Setor Saneamento–PMSS, URL: <http://www.cidades.gov.br/pncda;>
- GREEN, P., 1999, *Desenvolvendo competências consistentes*. Rio de Janeiro: Qualitymark.
- GUPTA, V. R.,1997, *An introduction to data warehousing. Technical report, System Services Corporation*, Chicago, Illinois.
- HAN, J.; KAMBER, M., 2001, *Data Mining: Concepts and Techniques*. San. Francisco, Morgan Kaufmann

- HAMEL, G., PRAHALAD, C. K., 1995, *Competindo pelo futuro*. Rio de Janeiro: Campus.
- JOHNSON, G.; SHOLES, K., 1999, *Exploring Corporate Strategy*. Prentice Hall Europe
- KOCK, N. F., MCQUEEN, R. J., CORNER J. L., (1997), “*The nature of data, information and knowledge exchanges in business processes: Implications for process improvement and organizational learning*”. *The Learning Organization* 4(2), 70–80.
- MANNILA, H., 1997, *Data mining: Machine learning, statistic and databases*. In *Proceedings of the 8th International Conference on Scientific and Statistical Database Management*, pp. 1–8.
- MINTZBERG, H., AHLSTRAND, B., LAMPEL, J. , 2000, *Safári de estratégia: um roteiro pela selva do planejamento estratégico*. Porto alegre, Bookman.
- MINTZBERG, H, QUINN, J. B. , 2001, *O Processo da Estratégia*. Bookman.
- MIRANDA, E. C., KOIDE, S., 2003, “*Indicadores de perdas de água: o que de fato, eles indicam?*” In: Congresso Brasileiro de Engenharia Sanitária e Ambiental, 22., 2003, Foz do Iguaçu. Anais... Joinville, SC: ABES. p.1-21.
- NEELY, A. et. al. , 1997, “*Designing performance measures: a structured approach*. *International Journal of Operations & Production Management*”. v. 17, n.11, pp. 1131-1152.
- NEELY, A. -1999, “*The performance measurement revolution: why now and next?*” *Int. Journal of Operations and Production Management*, vol. 19, no. 2, pp. 205-228.
- NONAKA, I. TAKEUCHI, H., 1997, *Criação do conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação*. 10 ed., Rio de Janeiro, Campus.
- PARRY, S., 1988, “*Just What is a Competency?*” *Training* , june , p. 58-64.
- PORTER, Michael E., 1985, *Vantagem Competitiva*. Rio de Janeiro: Campus
- PORTER, M. E., 1986, *Estratégia Competitiva: técnicas para análise de indústrias e da Concorrência*, Rio de Janeiro, Campus.
- PORTER, M., 1999, *Competição*. Rio de Janeiro: Campus.
- PORTER, M.E. , 1999, *Vantagem Competitiva*. Rio de Janeiro: Campus
- PRAHALAD, C. K , HAMEL, G. , 1990, “*The Core Competence of the Corporation*”. *Harvard Business Review*, p. 79-91.
- PRAHALAD, C.K.; RAMANSKY, V., 2004, *O Futuro da Competição*, São Paulo: Elsevier

- REZENDE, S. O., PUGLIESI, J. B., MELANDA, E. A., PAULA M. F., 2003, Mineração de Dados., *Sistemas Inteligentes – Fundamentos e Aplicações*. Manole.
- RUAS, R. , 2001, *Gestão estratégica do conhecimento*. São Paulo: Atlas.
- SILBERSCHATZ, A., TUZHILIN, A., 1995, *On subjective measures of interestingness in knowledge discovery*. Proceedings of the First International Conference on Knowledge Discovery and Data Mining 1, 275–281.
- SLACK, N. , 1993, *Vantagem Competitiva em Manufatura: atingindo competitividade nas operações industriais*. São Paulo: Atlas.
- SLACK, N. et al. , 1996, *Administração da Produção*. São Paulo: Atlas.
- SNIS, 2002, *Diagnóstico dos serviços de água e esgotos - Sistema Nacional De Informações Sobre Saneamento*. Brasília.
- STEWART, T.A., 1998, *Capital Intelectual* , 11ed, Rio de Janeiro, Campus
- ULRICK, D. ZENGER, J., SMALLWOOD, N. , 2000, *Liderança orientada para resultados; como os líderes constroem empresas e aumentam a lucratividade*. Rio de Janeiro: Campus.
- WEISS, S. M., INDURKHIA, N., 1998, *Predictive Data Mining: A Practical Guide*. San Francisco, Morgan Kaufmann Publishers.

## APÊNDICE 1

### VERDADES E EQUÍVOCOS SOBRE MEDIÇÃO DE ÁGUA

Neste apêndice as empresas de saneamento buscam responder as perguntas mais freqüentes dos consumidores a respeito das medições realizadas pelos hidrômetros:

#### ***1. Após a troca do hidrômetro, ele começou a andar mais rápido e o consumo aumentou.***

Como qualquer equipamento mecânico, o hidrômetro à medida que o tempo passa vai desgastando-se e diminuindo a sua sensibilidade (capacidade de registrar vazões muito baixas). Por isso, é possível que ocorra uma elevação no consumo após a troca do medidor, principalmente, se o antigo estava instalado inclinado.

Como todo o equipamento moderno, o hidrômetro sofre constantes evoluções tecnológicas, sendo que o emprego da transmissão magnética em substituição a transmissão mecânica proporcionou hidrômetros mais sensíveis, isto é, que registram vazões cada vez menores, com isso consumos que não eram registrados passam a ser medidos.

Nos hidrômetros magnéticos o dispositivo que indica a passagem de água deve por norma funcionar antes de qualquer totalizador, o que não ocorria com os mecânicos, daí a impressão de que o novo medidor está andando mais rápido, não significando um maior ou menor consumo que o hidrômetro anterior.

Caso persista a dúvida quanto ao volume de água medido, o usuário pode fazer a sua própria aferição, para isso basta encher um recipiente graduado em qualquer ponto de consumo com uma quantidade de água (por exemplo, 20 litros), fazendo as leituras do hidrômetro antes de iniciar o enchimento do recipiente e ao final quando fechar a torneira e parar o movimento do orientador.

#### ***2. O hidrômetro está funcionando quando não há consumo de água e mesmo assim ele está girando e marcando.***

O hidrômetro não inventa consumo. Se ele está girando e não existe nenhum ponto de consumo em uso no momento, significa que está ocorrendo um vazamento ou fuga não aparente.

Não havia ninguém na residência naquele período e foi “lido” um consumo de água muito além da média de consumo e deve ser por erro do hidrômetro, pois agora que há gente em casa o aparelho está marcando bem menos.

O hidrômetro não anda sozinho. Significa que naquele período em que foi acusado consumo, algum item da instalação operou de forma irregular. Podendo ser devido à presença de alguma impureza no seu mecanismo que, expulsa após certo tempo pela própria pressão da água, normalizou o funcionamento. Ou mesmo, o esquecimento de alguma torneira aberta, de uma válvula de descarga que não vedou totalmente, da torneira-bóia da caixa de descarga que não assentou corretamente ou estava dando passagem para a água acima de determinada pressão da rede.

***3. O registro antes do hidrômetro está fechado e ele continua girando, portanto isto prova que o medidor está com defeito.***

Este fato pode significar que o registro não está vedando totalmente. Há casos em que o registro está fechado, não sai água em nenhuma torneira e o medidor continua girando. Existe uma fuga não aparente entre o hidrômetro e o primeiro ponto de consumo da instalação, sendo que a quantidade de água que está tendo passagem pelo registro de entrada (aparentemente fechado) e saindo pela fuga tem vazão suficiente para ser detectada pelo medidor, mas insuficiente para atender mais de um ponto além daquele do vazamento.

***4. O hidrômetro está embaçado, a leitura está ilegível, portanto o medidor não está lendo corretamente o consumo medido.***

O embaçamento é causado pela umidade presente no interior do hidrômetro, que condensa na parede interna da cúpula da relojoaria, sendo bastante comum principalmente naquelas situações em que a relojoaria sofre variações brusca de temperatura. A partir de 1995, o DMAE vem adquirindo e utilizando hidrômetros com a relojoaria selada, onde o mecanismo de redução e a própria relojoaria são montados numa cúpula hermeticamente fechada à vácuo e sem contato com a água, eliminando quase que totalmente a possibilidade do embaçamento.

Quanto ao funcionamento do medidor e, principalmente, quanto ao consumo medido não há nenhuma alteração ou prejuízo ao usuário, salvo a possibilidade de uma leitura errada, que na confirmação seria corrigida ou na próxima leitura seria compensada, portanto facilmente verificada..

***5. O hidrômetro gira para trás e para frente, mesmo quando não há consumo de água, portanto ele está com defeito.***

O medidor quando indica este tipo de movimento está acusando o balanceamento de carga da rede naquele momento, principalmente quando há um desnível acentuado e o

hidrômetro está na parte inferior. O ramal tem o comportamento de um pequeno pulmão, quando atua sobre ele uma sucção ou um vácuo originado na rede, devido a sua incapacidade de acompanhar imediatamente a alteração de pressão ocasionada pela variação instantânea de consumo.

Em condições normais é insignificante a diferença acumulada entre a medição da água que retornou e a água que já havia passado, podendo o próprio usuário através de leituras periódicas certificar-se da existência ou não desta diferença e da grandeza da mesma e, se for o caso, recorrer ao Serviço de Saneamento para uma avaliação mais técnica.

***6. O consumo elevado é devido ao ar que passou pelo hidrômetro.***

Sem dúvida, o hidrômetro é movimentado pelo ar e registra a passagem do mesmo em seu totalizador. Isto ocorre quando do retorno do abastecimento após uma falta de água, e se deve a expulsão do ar das tubulações, empurrado pela água que está chegando.

Quando a falha no abastecimento é esporádica (falta de água em função de alguma ruptura na tubulação da rede), o volume de ar é insuficiente para ocasionar alteração significativa no consumo medido, pois, no momento em que faltou o abastecimento, a água e o ar que existiam no ramal foram sugados pela rede, fazendo com que o hidrômetro desmarcasse a leitura acumulada. Este movimento contrário permanece até ser restabelecido o equilíbrio entre as pressões da rede, dos ramais e do ponto da fuga.

De qualquer forma, uma maneira de evitar esta situação é assegurar-se de que as torneiras da residência permanecem fechadas durante a falta de água, não utilizando (ou esquecendo) a torneira de uma pia ou do tanque aberta para saber se o abastecimento normalizou.

Existem exceções em que a constante falta de água, por alguma obra na rede ou deficiência no abastecimento, acarreta alterações significativas no consumo medido.

## APÊNDICE 2

### TABELAS E DADOS UTILIZADOS

#### CADASTRO DE CONSUMIDORES

N	7	Matricula
	N	Digito Verificador
A	30	Nome do consumidor
		Sufixo Pré-Nome
		Sufixo Último Nome
N	1	Categoria (TA08)
N	1	Sub-Categoria (TA09)
N	1	Consumidor Especial
N	3	Código de Atividade (TA06)
	N	Cep
N	3	Distrito Postal
N	1	Corte - Indicador (TA12)
N	1	Entrega de Conta (TA38)
N	2	Ministério / Secretaria (TA02)
N	3	Órgão Interno (TA02)
N	4	Quartos / Salas
N	1	Piscinas % 1 - SIM
		% 2 - NAO
N	3	Banco C/C (TA03)
N	4	Agência C/C (TA03)
A	11	Número C/C
	N	Des (TA32)
N	2	Ciclo de Faturamento
N	4	Quantidade Economias Residencial
N	4	Quantidade Economias Comercial
N	4	Quantidade Economias Industrial
N	4	Quantidade Economias Pública
N	5	Consumo Estimado Mensal
N	5	Percentual de Faturamento Esgoto
N	3	Percentual de Faturamento Residencial
N	3	Percentual de Faturamento Comercial
N	3	Percentual de Faturamento Industrial
N	3	Percentual de Faturamento Pública
N	8	Data de Ligação da Água
N	1	Tipo de Ligação (TA39)
N	1	Diametro (TA21)
N	3	Elevatória (TA29)
N	8	Data da Ligação
N	1	Material (TA14)
N	1	Diametro (TA22)
N	7	Trecho da Rede
N	2	Ciclo de Corte



## CADASTRO REGISTRO DE CONSUMO

N	7	Matricula
<b>Data de Referência da Leitura</b>		
N	4	Ano de Referência da Leitura
N	2	Mês de Referência da Leitura
<b>Data da Leitura</b>		
N	4	Ano da Leitura
N	2	Mês da Leitura
N	2	Dia da Leitura
N	6	Matricula do Leiturista
N	6	Leitura Atual
N	2	Ocorrência (TA24)
N	2	Ciclo de Faturamento
N	1	Realização de Leitura % 1 - SIM
		% 2 - NÃO
N	3	Localidade
N	2	Caso de Consumo Medido (CR10)
N	6	Consumo Medido / Mês
N	3	Dias de Consumo
N	3	Dias de Venda
N	6	Média de Consumo/Dia
N	8	Consumo Mínimo/Dia
N	6	Limite Inferior/Dia
N	6	Limite Superior/Dia
N	1	Flag para Utilização Consumo Medido % 1- CONS. UTILIZ. P/CALC. DA MEDIA
		% 2-CONS. NAO UTILIZ. P/CALC. MEDIA
N(5)V9	6	Consumo Informado/Dia
N	2	Motivo Consumo Informado(TA49)
N	8	Consumo Projetado/Dia
N	2	Caso de Consumo a Faturar (CR16)
N	2	Código de Anormalidade
N	1	Tipo de Faturamento Adotado(TA25)

## CADASTRO DE IMÓVEIS

N	7	MATRICULA
N	1	Dígito DA MATRICULA
		<b>Código do Logradouro</b>
N	3	LOCALIDADE
N	5	NUMERO
N	1	Dígito do COD Logradouro
N	5	NUMERO DA PORTA
A	15	COMPLEMENTO do ENDERECO
N	4	BAIRRO
N	2	R A
		<b>ROTEIRO DE LEITURA</b>
N	3	LOCALIDADE
N	2	CICLO
N	2	CADERNETA
N	4	SEQUENCIAL
N	5	AREA CONSTRUIDA/TERRENO
N	1	AGUA DISPONIVEL % 1 - SIM
		% 2 - NÃO

## CADASTRO DE DÉBITOS EM ABERTO

N	7	MATRICULA
N	1	ORIGEM DO DÉBITO (TA20)
N	1	DÍGITO1 DO DÉBITO
N	1	DÍGITO2 DO DÉBITO
N	3	DAE/LOC
N	2	CICLO DE FATURAMENTO
N	1	CATEGORIA (TA08)
N	1	SUB-CATEGORIA (TA09)
N	5	NÚMERO DE ECONOMIAS
N	3	DIAS DE CONSUMO
N	1	TIPO DE CONSUMO FATURADO (TA25)
N	7	CONSUMO FATURADO
N	14	VALOR TOTAL DO DÉBITO
N	1	QUANTIDADE DE OCORRÊNCIAS
N	1	FLAG
<b>REFERÊNCIA DO DÉBITO</b>		
N	4	ANO DO DÉBITO
N	2	MÊS DO DÉBITO
<b>DATA DO VENCIMENTO</b>		
N	4	ANO DO VENCIMENTO
N	2	MÊS DO VENCIMENTO
N	2	DIA DO VENCIMENTO
<b>OCORRÊNCIAS</b>		
N	3	CÓDIGO DO LANÇAMENTO (TA05)
N	13	VALOR DE LANÇAMENTO
<b>DATA DE COMPETÊNCIA</b>		
N	4	ANO DE COMPETÊNCIA
N	2	MÊS DE COMPETÊNCIA
N	2	DIA DE COMPETÊNCIA OU CÓDIGO DE ORIGEM

## CADASTRO DE HIDRÔMETRO

A	10	NUMERO DO HIDROMETRO
		<b>DATA DE AQUISICAO</b>
N	4	ANO COM 4 POSICOES
N	2	MES
N	2	DIA
N	1	MARCA DO HIDROMETRO (TA15)
N	2	DIAMETRO (TA23)
N	2	CAPACIDADE M3 HH/DIA (TA26)
N	1	TIPO DE LEITURA (TA16)
N	1	TIPO DE FUNCIONAMENTO (TA40)
N	1	CAPACID MAXIMA MEDICAO (TA41)
N	3	DAE/LOC DE LOCALIZACAO(T04/31)
		<b>DATA DESUSO</b>
N	4	ANO COM 4 POSICOES
N	2	MES
N	2	DIA

## MOVIMENTAÇÃO DE HIDRÔMETROS

N	7	MATRICULA
A	10	NÚMERO DO HIDRÔMETRO
		<b>DATA DA INSTALAÇÃO</b>
N	4	ANO DA INSTALAÇÃO
N	2	MÊS DA INSTALAÇÃO
N	2	DIA DA INSTALAÇÃO
N	6	LEITURA NA INSTALAÇÃO
A	20	LOCALIZAÇÃO DO HIDRÔMETRO
A	10	DOCUMENTO
N	1	INDICA HIDRÔMETRO EM MANUTENÇÃO
		<b>DATA DA RETIRADA</b>
N	4	ANO DA RETIRADA
N	2	MÊS DA RETIRADA
N	2	DIA DA RETIRADA
N	6	LEITURA NA RETIRADA
N	1	MOTIVO DA RETIRADA (TA17)
N	1	INDICA HIDRÔMETRO RETIRADO COM DESVIO DE CONSUMO

## AMOSTRA DOS AGRUPAMENTOS OPERACIONAIS E MÉDIAS

ID	DESC_DAE	Hidrometros substituidos	Faturamento 5 meses	Fatur / Subs	Média Mensal
1	BOTAFOGO	653	992.830,27	1.520,41	304,08
2	JARDIM PRIMAVERA	33	38.793,40	1.175,56	235,11
3	ILHA GOVERNADOR	1603	818.305,84	510,48	102,10
4	TIJUCA	2624	1.129.062,01	430,28	86,06
5	CENTRO	3768	1.551.094,08	411,65	82,33
6	SEROPEDICA	33	12.839,39	389,07	77,81
7	ANGRA DOS REIS	66	19.994,98	302,95	60,59
8	BARRA DA TIJUCA	2709	816.076,35	301,25	60,25
9	ILHA DE PAQUETA	40	9.599,53	239,99	48,00
10	CARAPÉBUS	1	195,44	195,44	39,09
11	DUQUE DE CAXIAS	291	47.430,86	162,99	32,60
12	VASSOURAS	39	5.860,70	150,27	30,05
13	S.J. DO PARAISO	1	145,27	145,27	29,05
14	NOVA IGUAÇU	294	41.837,31	142,30	28,46
15	JACAREPAGUA	6981	962.943,94	137,94	27,59
16	RAMOS	16900	2.277.045,92	134,74	26,95
17	MEIER	9247	1.225.129,60	132,49	26,50
18	MAGE	16	1.899,20	118,70	23,74
19	VARRE-SAI	1	107,60	107,60	21,52
20	ITAPERUNA	160	15.932,50	99,58	19,92
21	BONSUCESSO	1	97,73	97,73	19,55
22	CASCADURA	16367	1.580.666,61	96,58	19,32
23	TERESOPOLIS	292	26.155,37	89,57	17,91
24	BOA SORTE	3	256,58	85,53	17,11
25	MARICA	162	12.697,90	78,38	15,68
26	CAMPO GRANDE	8510	613.910,75	72,14	14,43
27	BELFORD ROXO	191	13.730,79	71,89	14,38
28	MACAÉ	2209	157.484,91	71,29	14,26
29	RIO BONITO	95	6.369,71	67,05	13,41
30	SANTANESIA	3	188,66	62,89	12,58
31	BOM J. ITABAPOANA	23	1.433,91	62,34	12,47
32	N.S. APARECIDA	3	183,89	61,30	12,26
33	MIGUEL PEREIRA	34	2.067,37	60,81	12,16
34	SANTA CRUZ	1032	62.367,95	60,43	12,09
35	DEODORO	22981	1.379.210,46	60,02	12,00
36	NILOPOLIS	84	5.025,09	59,82	11,96
37	PORCIUNCULA	8	454,54	56,82	11,36
38	ITABORAI	27	1.533,50	56,80	11,36
39	QUEIMADOS	35	1.939,22	55,41	11,08
40	BOM JARDIM	27	1.478,40	54,76	10,95
41	PINHEIRAL	29	1.521,81	52,48	10,50
42	SÃO JOÃO DE MERITI	219	11.372,29	51,93	10,39
43	PIABETA	2	103,82	51,91	10,38
44	VILA IBITINEMA	2	102,51	51,26	10,25
45	QUISSAMA	15	753,28	50,22	10,04
46	VARGEM GRANDE	1	50,18	50,18	10,04
47	SÃO GONÇALO	15752	778.876,32	49,45	9,89
48	MANGARATIBA	41	1.868,52	45,57	9,11
49	VARJÃO	2	90,55	45,28	9,06
50	MIRACEMA	66	2.968,33	44,97	8,99
51	CAMPELO	1	44,43	44,43	8,89
52	IPITUNA	2	84,17	42,09	8,42
53	ITAGUAI	143	5.917,69	41,38	8,28
54	STO ANTONIO PADUA	53	2.151,04	40,59	8,12
55	ALCANTARA	434	17.238,13	39,72	7,94
56	ITAOCARA	24	945,40	39,39	7,88

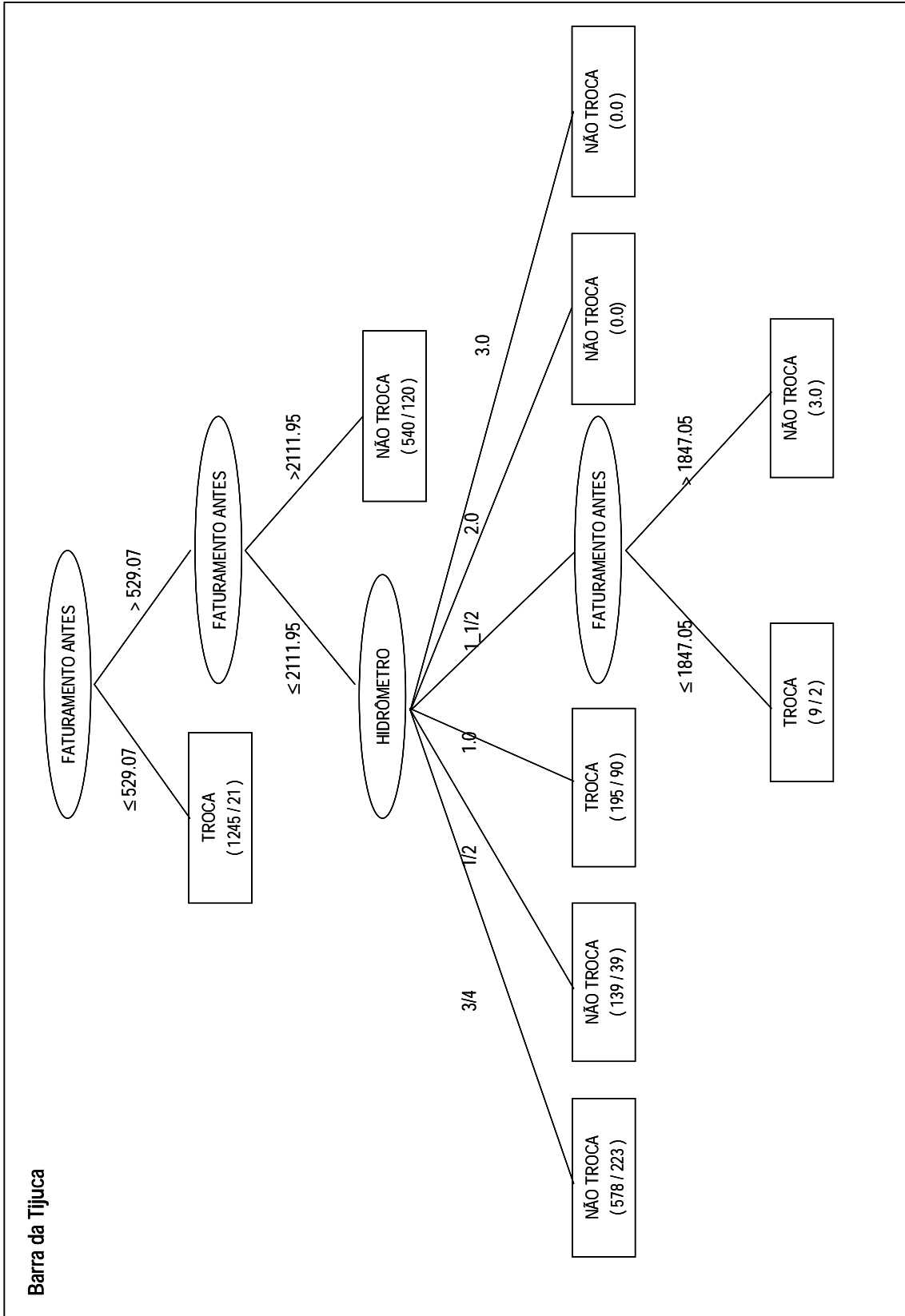
57	CACH. DE MACACU	10	390,57	39,06	7,81
58	NATIVIDADE	12	444,81	37,07	7,41
59	PIRAI	48	1.746,29	36,38	7,28
60	LAJE DO MURIAE'	10	343,00	34,30	6,86
61	PATI DE ALFERES	48	1.569,27	32,69	6,54
62	PARAIBA DO SUL	43	1.369,18	31,84	6,37
63	DEMETRIO RIBEIRO	6	187,45	31,24	6,25
64	SAO FIDELIS	33	1.013,40	30,71	6,14
65	S.SEBASTIAO DO ALTO	11	330,32	30,03	6,01
66	ANTA	16	476,45	29,78	5,96
67	PUREZA	2	57,86	28,93	5,79
68	MACUCO	8	228,17	28,52	5,70
69	CORDEIRO	50	1.422,69	28,45	5,69
70	PAULO DE FRONTIM-ENG	8	216,74	27,09	5,42
71	PORTELA	1	25,91	25,91	5,18
72	SAPUCAIA	27	687,74	25,47	5,09
73	APERIBE	6	151,87	25,31	5,06
74	ITALVA	15	365,81	24,39	4,88
75	PASSA TRES	3	72,17	24,06	4,81
76	CANTAGALO	12	284,40	23,70	4,74
77	DUAS BARRAS	24	568,60	23,69	4,74
78	RIO DAS OSTRAS	50	1.178,62	23,57	4,71
79	GRUSSAI	10	232,96	23,30	4,66
80	S.JOAO DA BARRA	33	762,71	23,11	4,62
81	RIO CLARO	9	206,57	22,95	4,59
82	LIDICE	2	44,96	22,48	4,50
83	CAMBUCI	9	202,10	22,46	4,49
84	TANGUA'	15	332,16	22,14	4,43
85	WERNECK	3	63,18	21,06	4,21
86	BARRA DE S.JOAO	54	1.122,21	20,78	4,16
87	ATAFONA	48	994,32	20,72	4,14
88	PARACAMBI	17	332,75	19,57	3,91
89	MASSAMBARA	2	38,93	19,47	3,89
90	AVELAR	2	35,63	17,82	3,56
91	S.JOSE RIBEIRAO	1	17,76	17,76	3,55
92	BARCELOS	1	17,61	17,61	3,52
93	BARRA ITABAPOANA	1	17,58	17,58	3,52
94	GARGAU	1	17,55	17,55	3,51
95	BARRA DO ACU	1	17,46	17,46	3,49
96	EUCLIDELANDIA	2	34,49	17,25	3,45
97	CONRADO	3	51,72	17,24	3,45
98	VALAO DO BARRO	2	33,71	16,86	3,37
99	S.SEBAST.PARAIBA	2	33,18	16,59	3,32
100	SAO JOSE DE UBA	6	99,33	16,56	3,31
101	BATATAL	1	16,53	16,53	3,31
102	STA.M. MADALENA	2	32,96	16,48	3,30
103	MORRO AZUL	1	16,43	16,43	3,29
104	CARDOSO MOREIRA	7	110,68	15,81	3,16
105	ARROZAL	20	315,77	15,79	3,16
106	JAMAPARA	17	246,57	14,50	2,90
107	SUMIDOURO	126	1.793,12	14,23	2,85
108	VARGEM ALEGRE	196	2.194,74	11,20	2,24
109	PRACA JOAO PESSOA	3	16,15	5,38	1,08
110	ITAKAMOZY	6	28,75	4,79	0,96
		<b>115.379</b>	<b>14.681.078</b>	<b>10.133</b>	

**APÊNDICE 3**  
**RESULTADOS DAS SIMULAÇÕES**

## Agrupamento Barra da Tijuca

```
=== Run information ===
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_BarraTijucaCSV-weka.filters.unsupervised.attribute.Remove-R1-6
Instances: 2709
Attributes: 3
    HIDROMETRO
    FATURAMENTO_5ANTES
    DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
FATURAMENTO_5ANTES <= 529.07: TROCA (1245.0/21.0)
FATURAMENTO_5ANTES > 529.07
| FATURAMENTO_5ANTES <= 2111.95
| | HIDROMETRO = 3/4: NAO-TROCA (578.0/223.0)
| | HIDROMETRO = 1/2: NAO-TROCA (139.0/39.0)
| | HIDROMETRO = 1.0: TROCA (195.0/90.0)
| | HIDROMETRO = 1_1/2
| | | FATURAMENTO_5ANTES <= 1847.05: TROCA (9.0/2.0)
| | | FATURAMENTO_5ANTES > 1847.05: NAO-TROCA (3.0)
| | HIDROMETRO = 2.0: NAO-TROCA (0.0)
| | HIDROMETRO = 3.0: NAO-TROCA (0.0)
| FATURAMENTO_5ANTES > 2111.95: NAO-TROCA (540.0/120.0)
Number of Leaves :    9
Size of the tree :    13
Time taken to build model: 0.03 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   2182      80.5463 %
Incorrectly Classified Instances   527      19.4537 %
Kappa statistic                  0.61
Mean absolute error               0.2467
Root mean squared error           0.3555
Relative absolute error           53.1685 %
Root relative squared error       73.8058 %
Total Number of Instances        2709
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.911  0.256   0.673   0.911   0.774   NAO-TROCA
0.744   0.089   0.936  0.744   0.829   TROCA
=== Confusion Matrix ===
  a    b  <-- classified as
903  88 | a = NAO-TROCA
439 1279 | b = TROCA
```





## Agrupamento Botafogo

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: 15%\_BotafogoCSV

Instances: 653 Attributes: 10

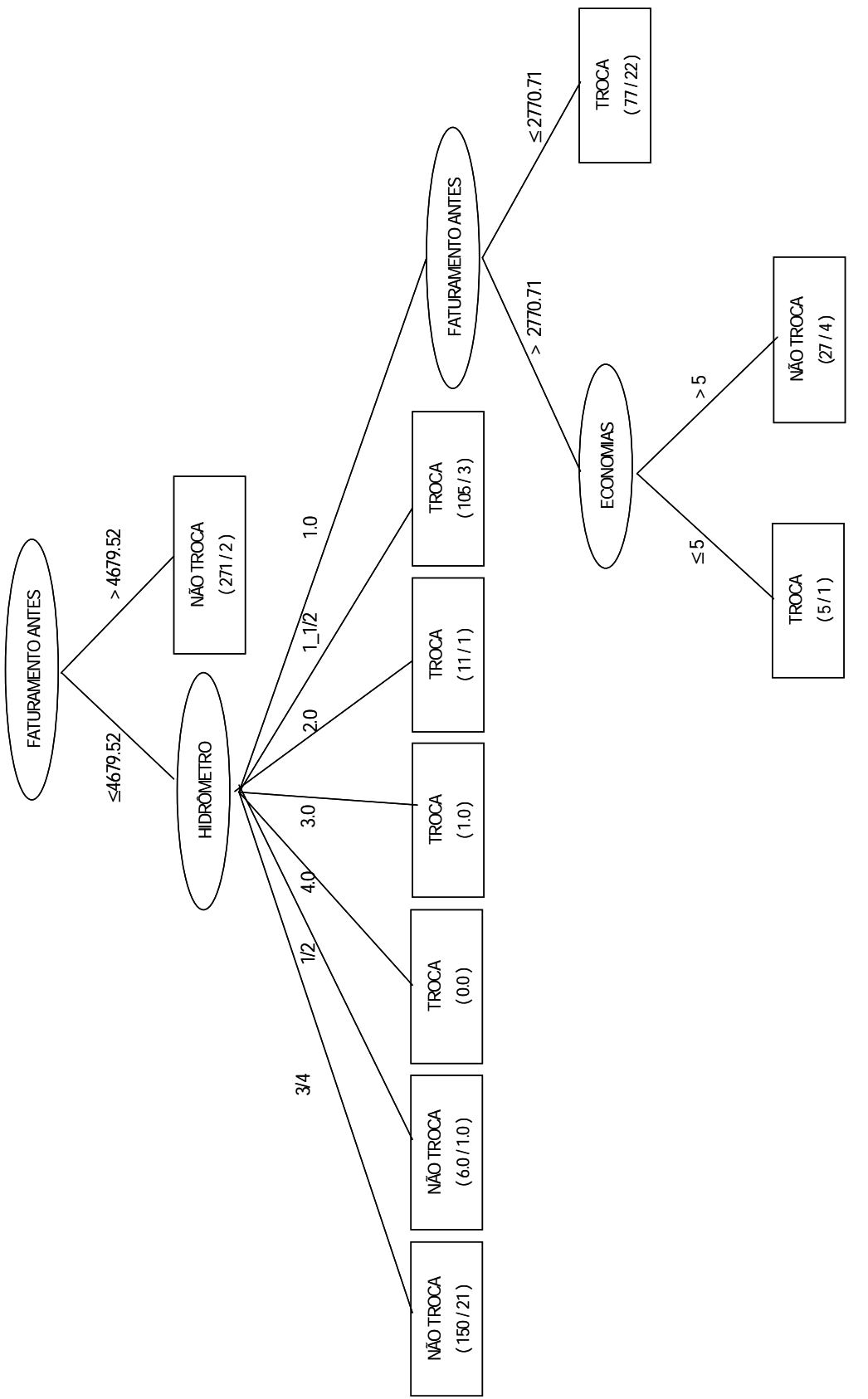
```

DAE
NOME_DAE
CATEGORIA
SUBCATEGORIA
CICLO
ECONOMIAS
BAIRRO
HIDROMETRO
FATURAMENTO_5ANTES
DECISAO

Test mode: 10-fold cross-validation
=== Classifier model (full training set) === J48 pruned tree
FATURAMENTO_5ANTES <= 4679.52
| HIDROMETRO = 3/4: NAO-TROCA (150.0/21.0)
| HIDROMETRO = 1/2: NAO-TROCA (6.0/1.0)
| HIDROMETRO = 4.0: TROCA (0.0)
| HIDROMETRO = 3.0: TROCA (1.0)
| HIDROMETRO = 2.0: TROCA (11.0/1.0)
| HIDROMETRO = 1_1/2: TROCA (105.0/3.0)
| HIDROMETRO = 1.0
| | FATURAMENTO_5ANTES <= 2770.71: TROCA (77.0/22.0)
| | FATURAMENTO_5ANTES > 2770.71
| | | ECONOMIAS <= 5: TROCA (5.0/1.0)
| | | ECONOMIAS > 5: NAO-TROCA (27.0/4.0)
FATURAMENTO_5ANTES > 4679.52: NAO-TROCA (271.0/2.0)
Number of Leaves : 10
Size of the tree : 14
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
Correctly Classified Instances 585 89.5865 %
Incorrectly Classified Instances 68 10.4135 %
Kappa statistic 0.7486
Mean absolute error 0.1467
Root mean squared error 0.2847
Relative absolute error 34.4948 %
Root relative squared error 61.7608 %
Total Number of Instances 653
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure Class
0.945 0.215 0.909 0.945 0.926 NAO-TROCA
0.785 0.055 0.863 0.785 0.822 TROCA
=== Confusion Matrix ===
a b <-- classified as
428 25 | a = NAO-TROCA
43 157 | b = TROCA

```

**Botafogo**



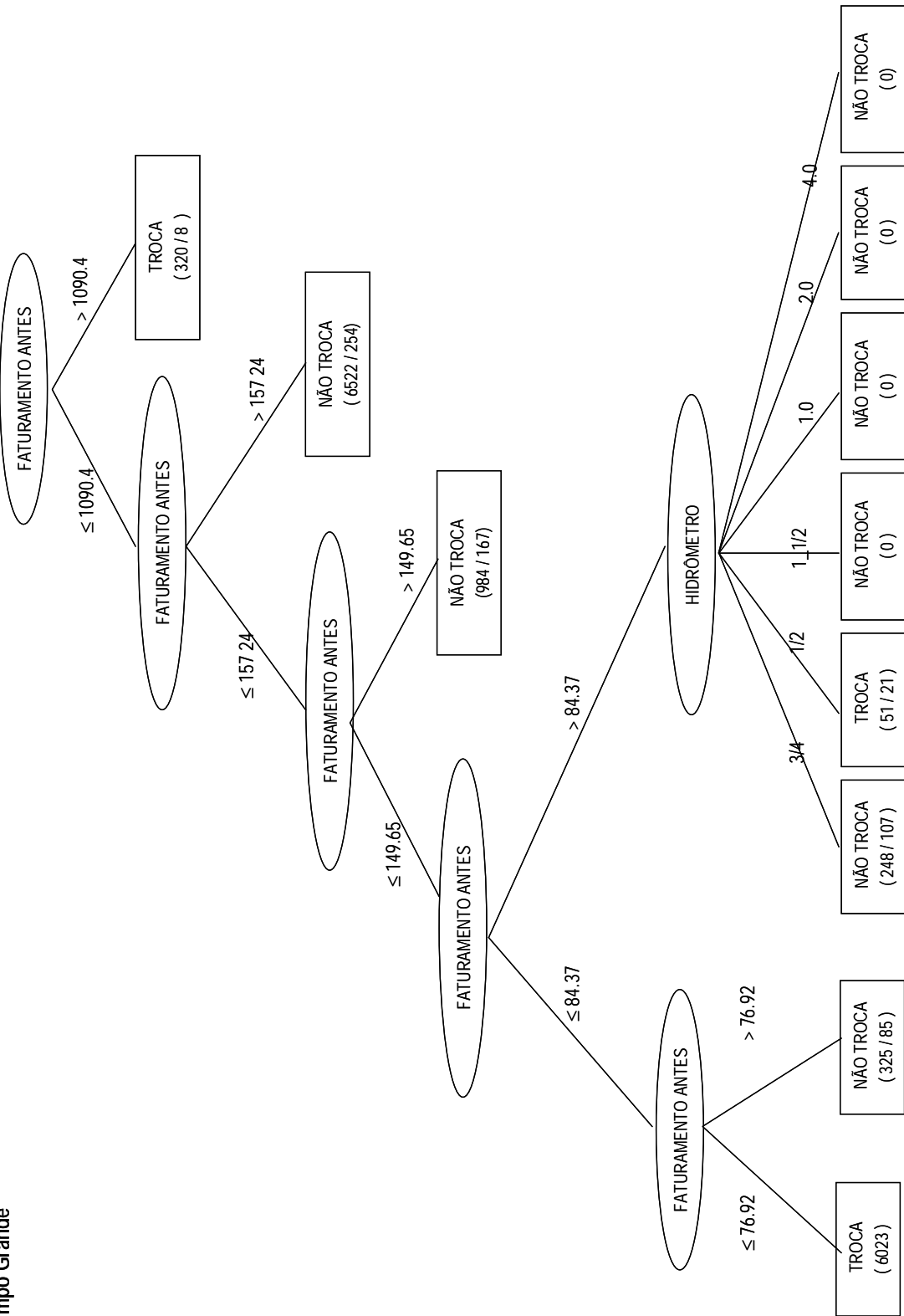
## Agrupamento Campo Grande

```

Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2      Relation: 15%_Campo_GrandeCSV  Instances: 8510  Attributes: 5
  DAE
  SUBCATEGORIA
  HIDROMETRO
  FATURAMENTO_5ANTES
  DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
FATURAMENTO_5ANTES <= 1090.4
| FATURAMENTO_5ANTES <= 157.24
| | FATURAMENTO_5ANTES <= 149.65
| | | FATURAMENTO_5ANTES <= 84.37
| | | | FATURAMENTO_5ANTES <= 76.92: TROCA (60.0/23.0)
| | | | FATURAMENTO_5ANTES > 76.92: NAO-TROCA (325.0/85.0)
| | | | FATURAMENTO_5ANTES > 84.37
| | | | HIDROMETRO = 3/4: NAO-TROCA (248.0/107.0)
| | | | HIDROMETRO = 1/2: TROCA (51.0/21.0)
| | | | HIDROMETRO = 1_1/2: NAO-TROCA (0.0)
| | | | HIDROMETRO = 1.0: NAO-TROCA (0.0)
| | | | HIDROMETRO = 2.0: NAO-TROCA (0.0)
| | | | HIDROMETRO = 4.0: NAO-TROCA (0.0)
| | | FATURAMENTO_5ANTES > 149.65: NAO-TROCA (984.0/167.0)
| | FATURAMENTO_5ANTES > 157.24: NAO-TROCA (6522.0/254.0)
FATURAMENTO_5ANTES > 1090.4: TROCA (320.0/8.0)
Number of Leaves :    11      Size of the tree :    17
Time taken to build model: 0.13 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   7820      91.8919 %
Incorrectly Classified Instances   690      8.1081 %
Kappa statistic                   0.4611
Mean absolute error                0.1369
Root mean squared error            0.263
Relative absolute error            66.4331 %
Root relative squared error        81.9616 %
Total Number of Instances        8510
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.342   0.005   0.902   0.342   0.496   TROCA
0.995   0.658   0.92    0.995   0.956   NAO-TROCA
=== Confusion Matrix ===
  a  b  <-- classified as
339 653 | a = TROCA
37 7481 | b = NAO-TROCA

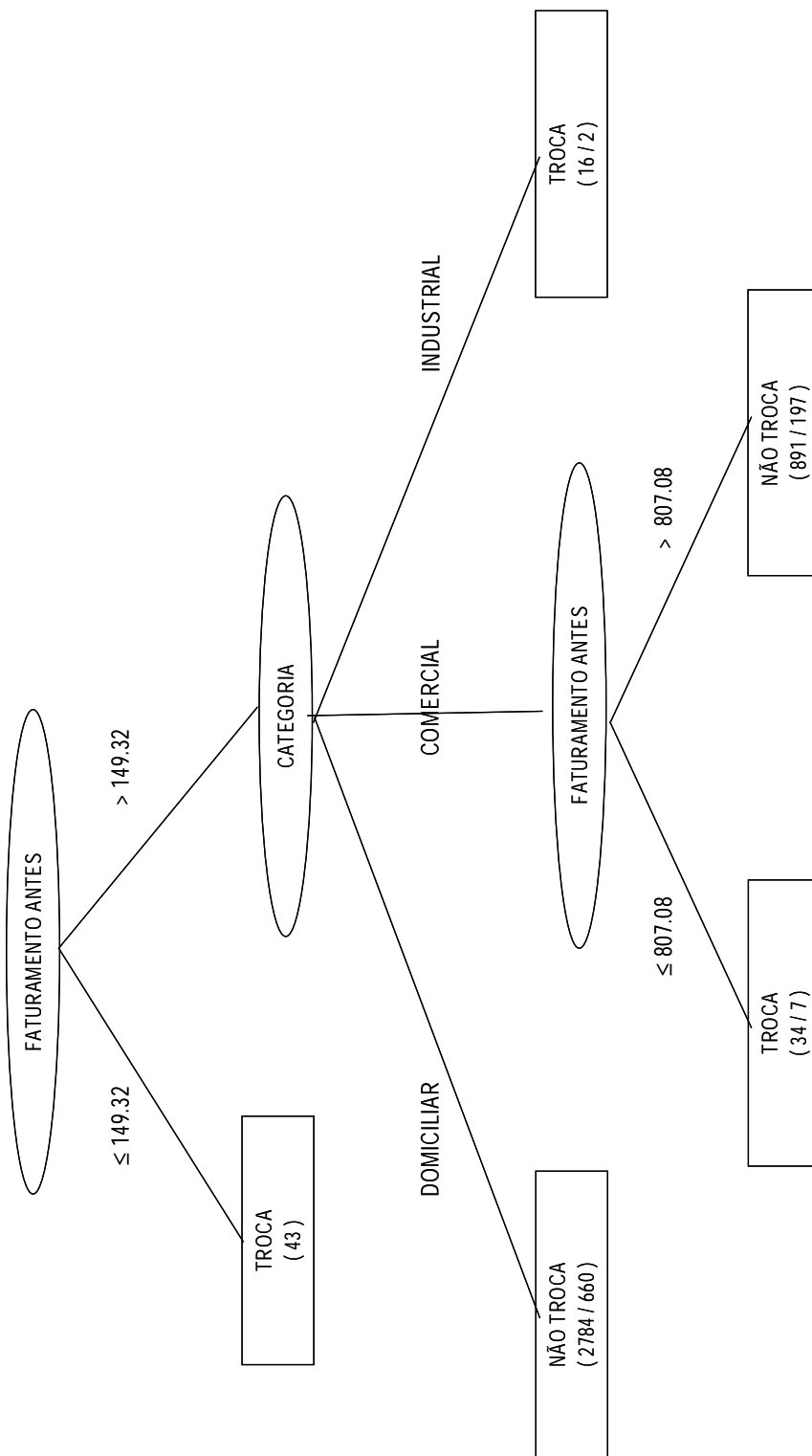
```

Campo Grande



## Agrupamento Centro

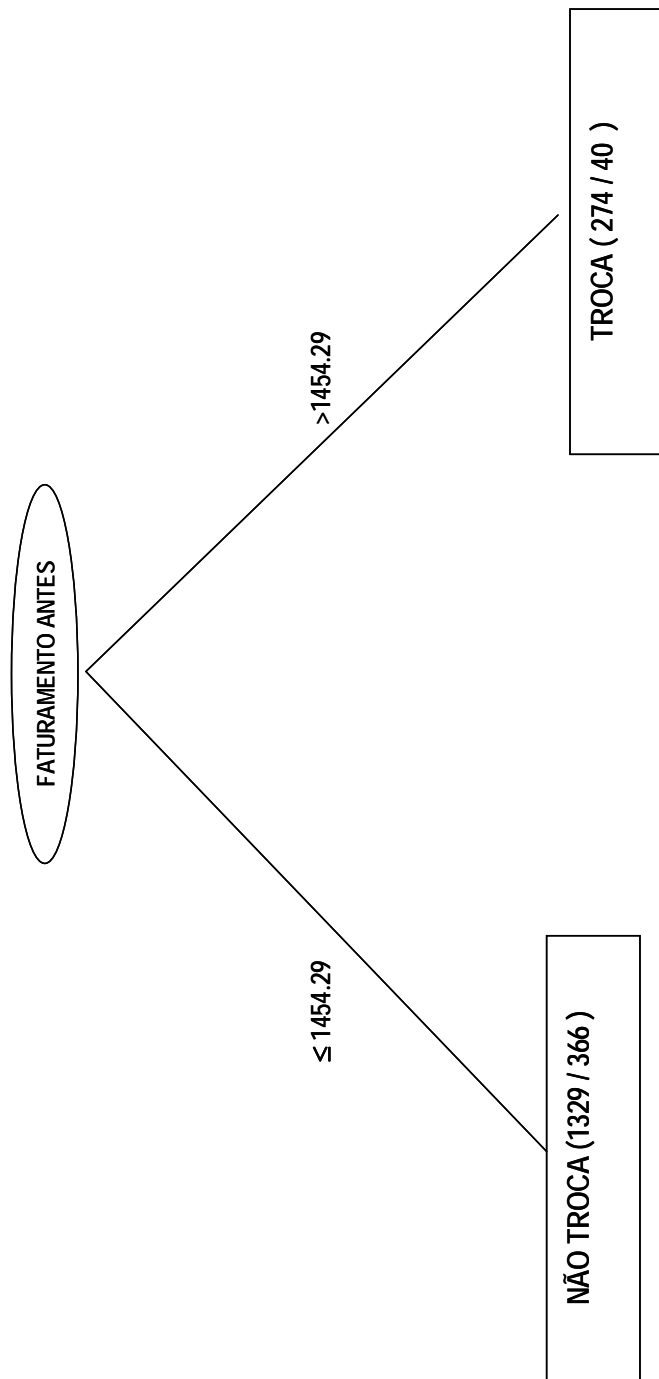
```
=== Run information ===
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_CentroCSV-weka.filters.unsupervised.attribute.Remove-R1-2,5-6
Instances: 3768
Attributes: 6
    CATEGORIA
    SUBCATEGORIA
    BAIRRO
    HIDROMETRO
    FATURAMENTO_5ANTES
    DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
FATURAMENTO_5ANTES <= 149.32: TROCA (43.0)
FATURAMENTO_5ANTES > 149.32
| CATEGORIA = DOMICILIAR: NAO-TROCA (2784.0/660.0)
| CATEGORIA = COMERCIAL
| | FATURAMENTO_5ANTES <= 807.08: TROCA (34.0/7.0)
| | FATURAMENTO_5ANTES > 807.08: NAO-TROCA (891.0/197.0)
| CATEGORIA = INDUSTRIAL: TROCA (16.0/2.0)
Number of Leaves :    5
Size of the tree :    8
Time taken to build model: 0.08 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   2875      76.3004 %
Incorrectly Classified Instances   893      23.6996 %
Kappa statistic                   0.0794
Mean absolute error                0.3594
Root mean squared error            0.4249
Relative absolute error            95.8835 %
Root relative squared error        98.1653 %
Total Number of Instances         3768
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.056   0.002   0.914   0.056   0.106   TROCA
0.998   0.944   0.761   0.998   0.863   NAO-TROCA
=== Confusion Matrix ===
 a  b  <-- classified as
53 888 | a = TROCA
5 2822 | b = NAO-TROCA
```



## Agrupamento Ilha do Governador

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_Ilha_GovernadorCSV
Instances: 1603
Attributes: 10
    DAE
    NOME_DAE
    CATEGORIA
    SUBCATEGORIA
    CICLO
    ECONOMIAS
    BAIRRO
    HIDROMETRO
    FATURAMENTO_5ANTES
    DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
FATURAMENTO_5ANTES <= 1454.29: NAO-TROCA (1329.0/366.0)
FATURAMENTO_5ANTES > 1454.29: TROCA (274.0/40.0)
Number of Leaves : 2
Size of the tree : 3
Time taken to build model: 0.14 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 1194 74.4853 %
Incorrectly Classified Instances 409 25.5147 %
Kappa statistic 0.3874
Mean absolute error 0.3713
Root mean squared error 0.4334
Relative absolute error 79.2544 %
Root relative squared error 89.5509 %
Total Number of Instances 1603
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure Class
0.383 0.039 0.855 0.383 0.529 TROCA
0.961 0.617 0.723 0.961 0.825 NAO-TROCA
=== Confusion Matrix ===
a b <-- classified as
230 370 | a = TROCA
39 964 | b = NAO-TROCA
```





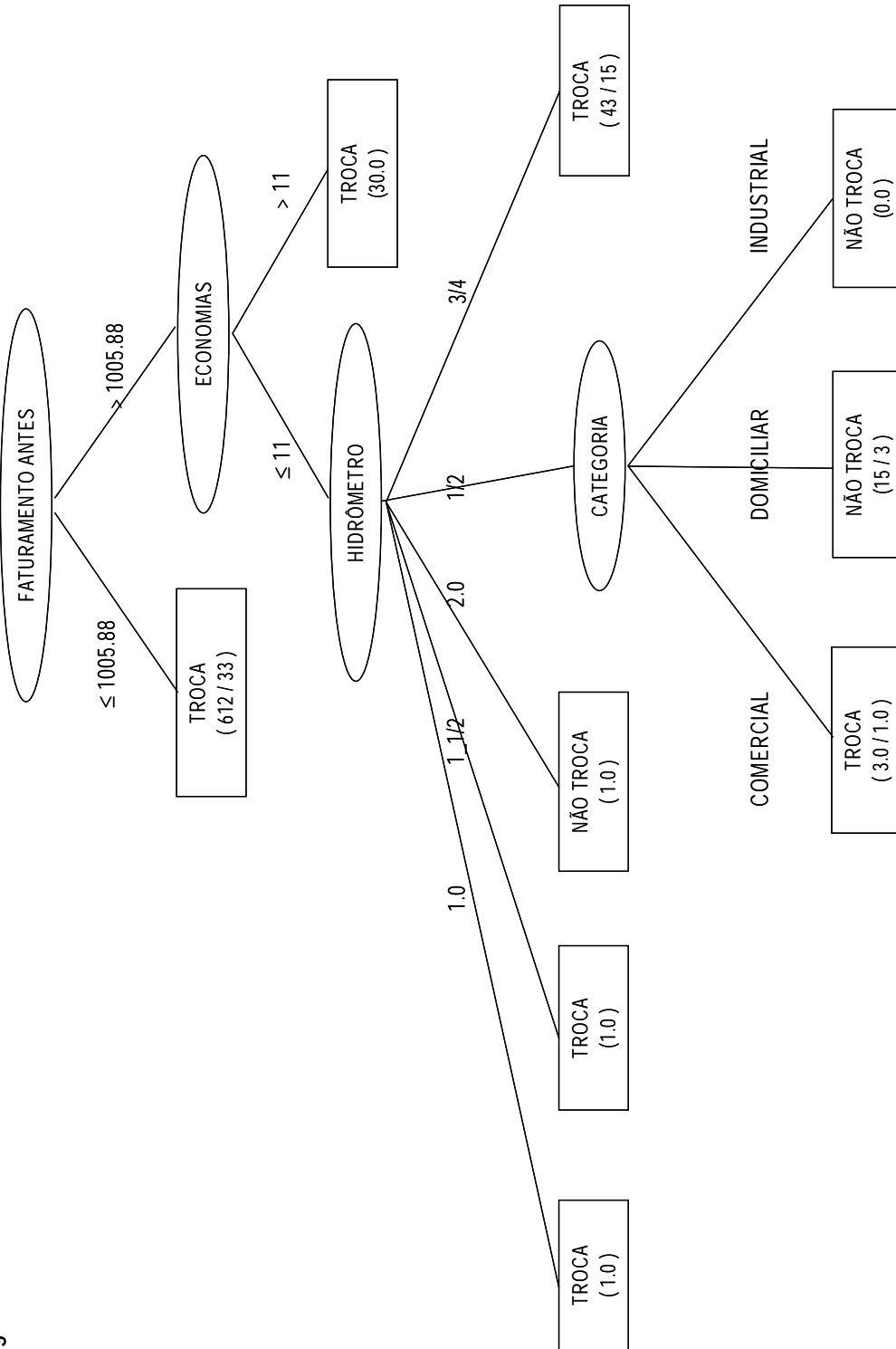
## Agrupamento Jacarepaguá

```

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_JacarepaguaCSV-weka.filters.unsupervised.attribute.Remove-R5 Instances: 807 Attributes: 9
  DAE
  NOME_DAE
  CATEGORIA
  SUBCATEGORIA
  ECONOMIAS
  BAIRRO
  HIDROMETRO
  FATURAMENTO_5ANTES
  DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) === J48 pruned tree
FATURAMENTO_5ANTES <= 1005.88: TROCA (612.0/33.0)
FATURAMENTO_5ANTES > 1005.88
| ECONOMIAS <= 11
| | HIDROMETRO = 1.0: TROCA (1.0)
| | HIDROMETRO = 1_1/2: TROCA (1.0)
| | HIDROMETRO = 2.0: NAO-TROCA (1.0)
| | HIDROMETRO = 1/2
| | | CATEGORIA = COMERCIAL: TROCA (3.0/1.0)
| | | CATEGORIA = DOMICILIAR: NAO-TROCA (15.0/3.0)
| | | CATEGORIA = INDUSTRIAL: NAO-TROCA (0.0)
| | HIDROMETRO = 3/4: TROCA (43.0/15.0)
| ECONOMIAS > 11: TROCA (30.0)
Number of Leaves : 9
Size of the tree : 13
Time taken to build model: 0.06 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 650 92.068 %
Incorrectly Classified Instances 56 7.932 %
Kappa statistic 0.2552
Mean absolute error 0.131
Root mean squared error 0.2655
Relative absolute error 81.2416 %
Root relative squared error 93.7974 %
Total Number of Instances 706
Ignored Class Unknown Instances 101
TP Rate FP Rate Precision Recall F-Measure Class
0.992 0.823 0.926 0.992 0.958 TROCA
0.177 0.008 0.688 0.177 0.282 NAO-TROCA
=== Confusion Matrix ===
 a b <- classified as
639 5 | a = TROCA
51 11 | b = NAO-TROCA

```

Jacarepaguá



## Agrupamento Jardim Primavera

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2 Relation: 15%\_Jardim\_PrimaveraCSV-weka.filters.unsupervised.attribute.Remove-R5-6-weka.filters.unsupervised.attribute.Remove-R2-weka.filters.unsupervised.attribute.Remove-R4

Instances: 105 Attributes: 6

DAE  
 CATEGORIA  
 SUBCATEGORIA  
 HIDROMETRO  
 FATURAMENTO\_5ANTES  
 DECISAO

=== Classifier model (full training set) === J48 pruned tree

```
FATURAMENTO_5ANTES <= 782.81
| FATURAMENTO_5ANTES <= 283.25
| | FATURAMENTO_5ANTES <= 280.83
| | | FATURAMENTO_5ANTES <= 92.09
| | | | FATURAMENTO_5ANTES <= 91.37: NAO-TROCA (3.0)
| | | | FATURAMENTO_5ANTES > 91.37: TROCA (3.0)
| | | | FATURAMENTO_5ANTES > 92.09: NAO-TROCA (32.0)
| | | | FATURAMENTO_5ANTES > 280.83: TROCA (4.0)
| | | FATURAMENTO_5ANTES > 283.25: NAO-TROCA (35.0/3.0)
| FATURAMENTO_5ANTES > 782.81
| | FATURAMENTO_5ANTES <= 1843.42
| | | HIDROMETRO = 6.0: TROCA (0.0)
| | | | HIDROMETRO = 1/2
| | | | | FATURAMENTO_5ANTES <= 1067.62: TROCA (3.0)
| | | | | FATURAMENTO_5ANTES > 1067.62: NAO-TROCA (3.0)
| | | | | HIDROMETRO = 3/4: TROCA (4.0)
| | | | | HIDROMETRO = 1_1/2: TROCA (0.0)
| | | | FATURAMENTO_5ANTES > 1843.42: NAO-TROCA (18.0)
```

Number of Leaves : 11 Size of the tree : 19 === Stratified cross-validation ===

Correctly Classified Instances	93	88.5714 %
Incorrectly Classified Instances	12	11.4286 %

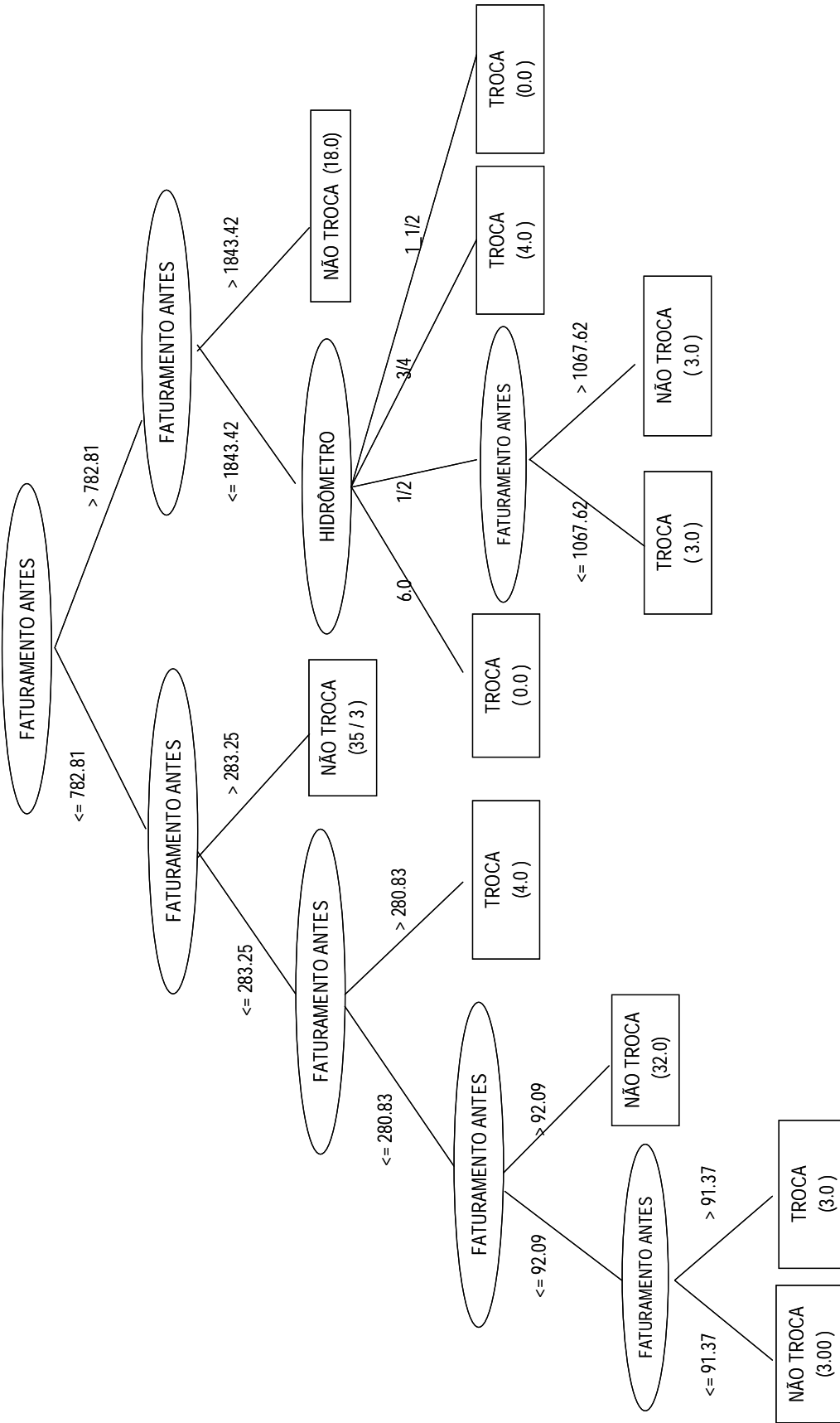
Kappa statistic	0.5347
Mean absolute error	0.1365
Root mean squared error	0.3046
Relative absolute error	49.3985 %
Root relative squared error	82.5783 %
Total Number of Instances	105

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.955	0.471	0.913	0.955	0.933	NAO-TROCA
0.529	0.045	0.692	0.529	0.6	TROCA

=== Confusion Matrix ===

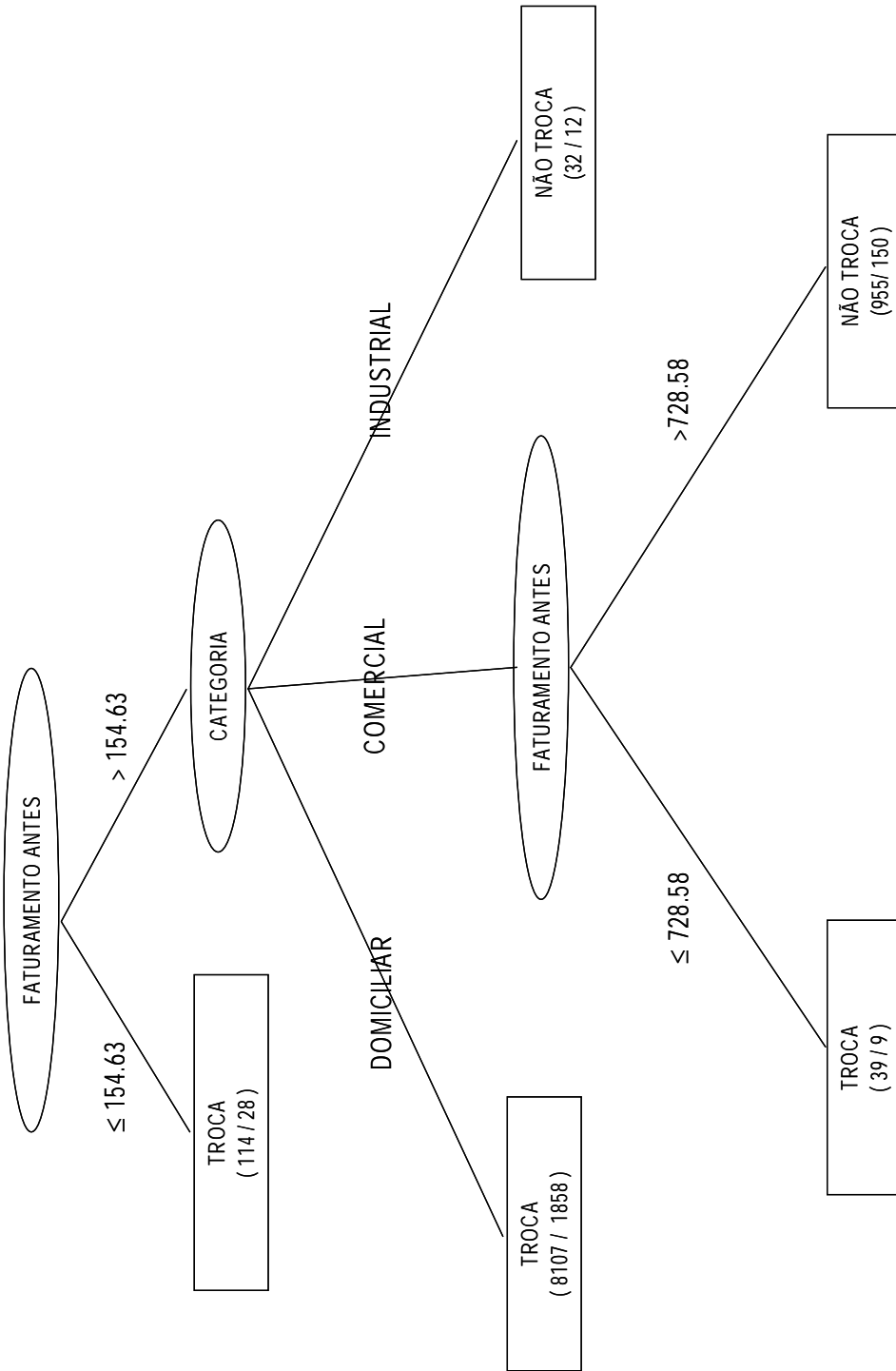
a	b	<- classified as
84	4	a = NAO-TROCA
8	9	b = TROCA



## Agrupamento Meier

```
=== Run information ===
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: Meier
Instances: 9247
Attributes: 10
    NOME_DAE
    CATEGORIA
    SUBCATEGORIA
    CICLO
    ECONOMIAS
    BAIRRO
    DAE
    HIDROMETRO
    FATURAMENTO_5ANTES
    DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
FATURAMENTO_5ANTES <= 154.63: TROCA (114.0/28.0)
FATURAMENTO_5ANTES > 154.63
| CATEGORIA = DOMICILIAR: NAO-TROCA (8107.0/1858.0)
| CATEGORIA = COMERCIAL
| | FATURAMENTO_5ANTES <= 728.58: TROCA (39.0/9.0)
| | FATURAMENTO_5ANTES > 728.58: NAO-TROCA (955.0/150.0)
| CATEGORIA = INDUSTRIAL: NAO-TROCA (32.0/12.0)
Number of Leaves :    5
Size of the tree :    8
Time taken to build model: 0.44 seconds
=== Stratified cross-validation ===
Correctly Classified Instances   7186      77.7117 %
Incorrectly Classified Instances  2061      22.2883 %
Kappa statistic                  0.0691
Mean absolute error              0.346
Root mean squared error          0.4162
Relative absolute error          97.3916 %
Root relative squared error      98.7513 %
Total Number of Instances       9247
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.995   0.949   0.777   0.995   0.873   NAO-TROCA
0.051   0.005   0.759   0.051   0.096   TROCA
=== Confusion Matrix ===
  a  b  <-- classified as
7076  35 | a = NAO-TROCA
2026 110 | b = TROCA
```

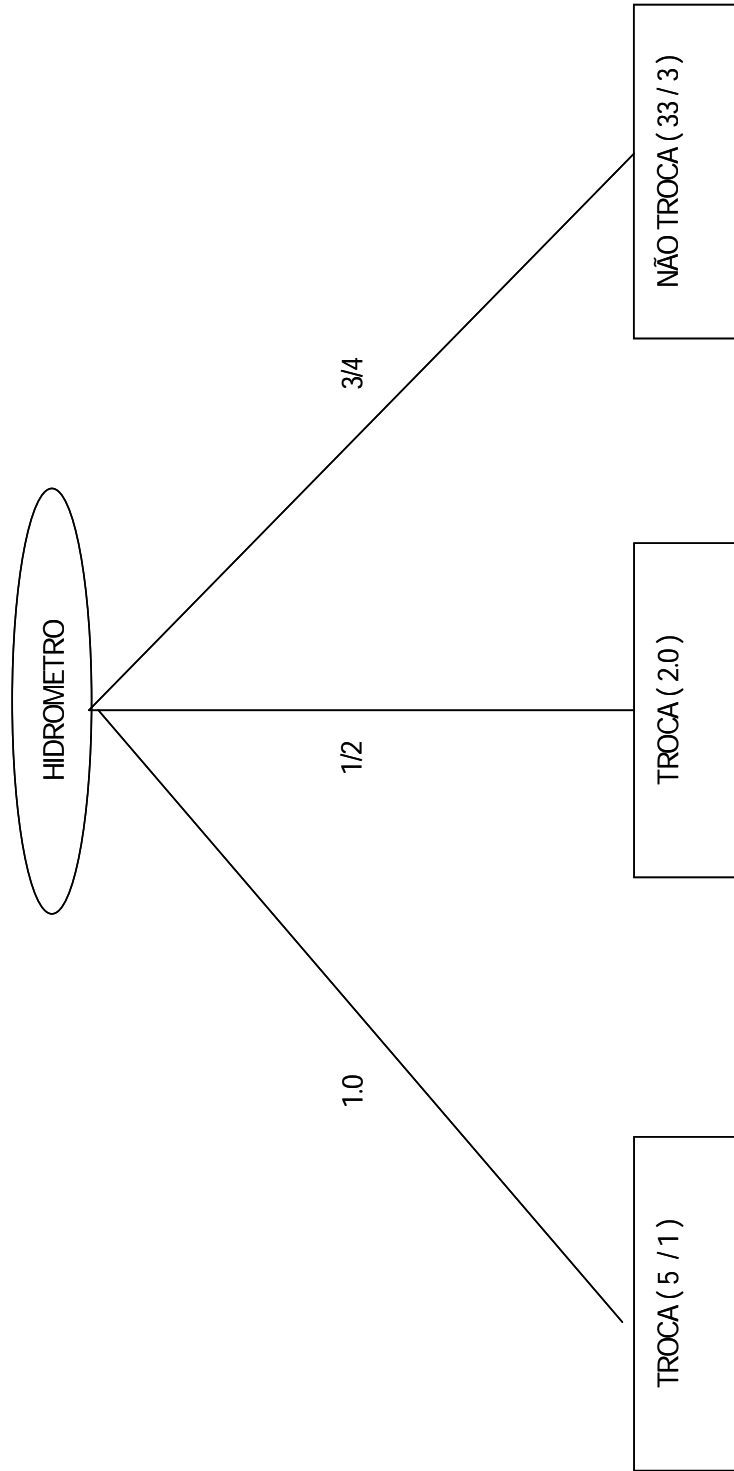
Meier







Paquetá



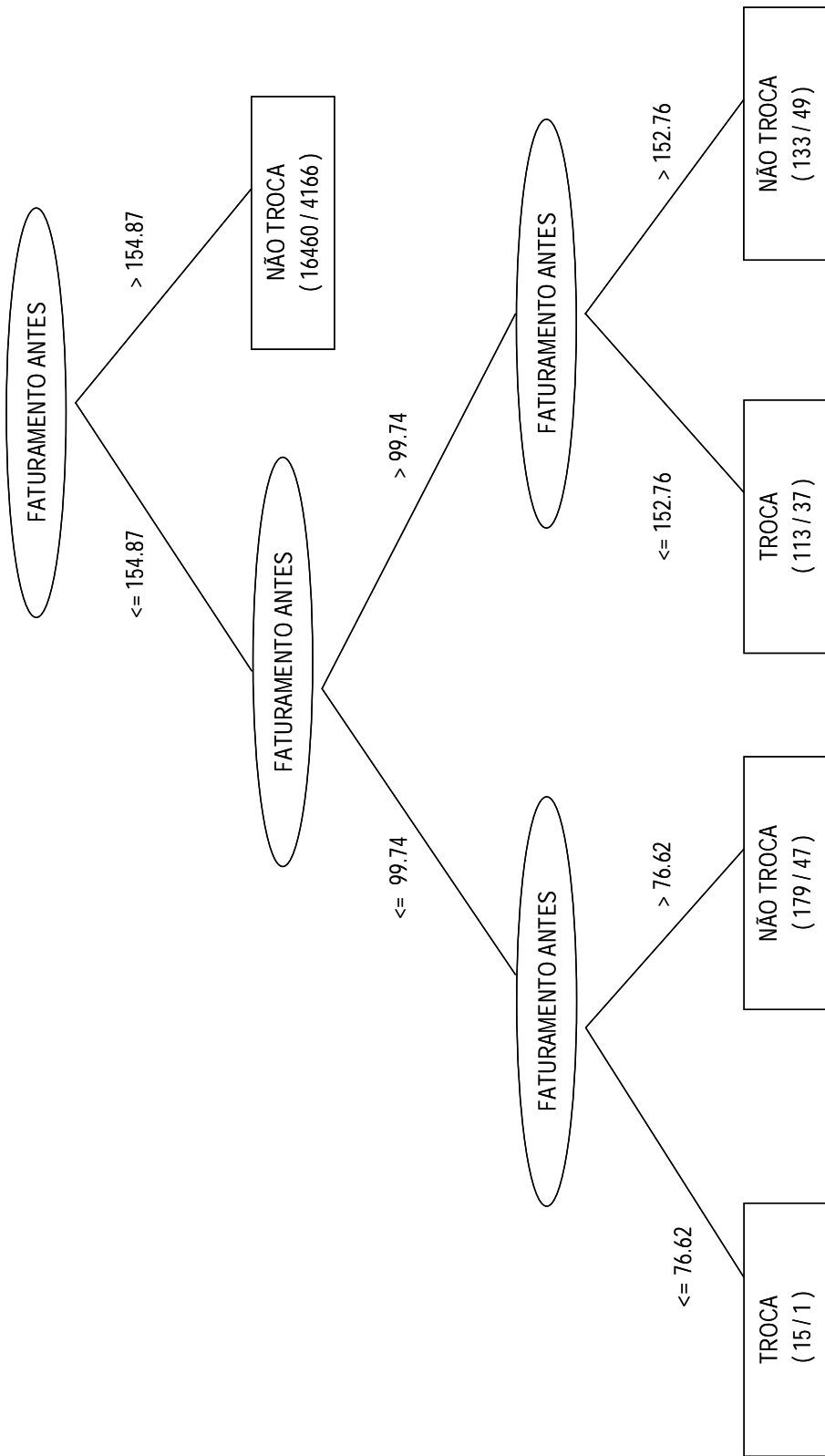
## Agrupamento Ramos

```
=== Run information ===
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_RamosCSV-weka.filters.unsupervised.attribute.Remove-R2-7-weka.filters.unsupervised.attribute.Remove-R1
Instances: 16900
Attributes: 3
    HIDROMETRO
    FATURAMENTO_5ANTES
    DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
FATURAMENTO_5ANTES <= 154.87
| FATURAMENTO_5ANTES <= 99.74
| | FATURAMENTO_5ANTES <= 76.62: TROCA (15.0/1.0)
| | FATURAMENTO_5ANTES > 76.62: NAO-TROCA (179.0/47.0)
| | FATURAMENTO_5ANTES > 99.74
| | | FATURAMENTO_5ANTES <= 152.76: TROCA (113.0/37.0)
| | | FATURAMENTO_5ANTES > 152.76: NAO-TROCA (133.0/49.0)
FATURAMENTO_5ANTES > 154.87: NAO-TROCA (16460.0/4166.0)
Number of Leaves :    5
Size of the tree :    9
Time taken to build model: 0.2 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   12585      74.4675 %
Incorrectly Classified Instances  4315      25.5325 %
Kappa statistic                  0.0226
Mean absolute error               0.3793
Root mean squared error          0.4359
Relative absolute error          99.1967 %
Root relative squared error      99.6859 %
Total Number of Instances       16900

=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.996   0.981   0.745   0.996   0.853   NAO-TROCA
0.019   0.004   0.643   0.019   0.037   TROCA

=== Confusion Matrix ===
a  b  <- classified as
12502  46 | a = NAO-TROCA
4269  83 | b = TROCA
```

Ramos



## Agrupamento Rio Bonito

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_Rio_BonitoCSV
Instances: 95
Attributes: 10
  NOME_DAE
  CATEGORIA
  SUBCATEGORIA
  CICLO
  ECONOMIAS
  BAIRRO
  DAE
  HIDROMETRO
  FATURAMENTO_5ANTES
  DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
FATURAMENTO_5ANTES <= 800.57: NAO-TROCA (85.0/10.0)
FATURAMENTO_5ANTES > 800.57: TROCA (10.0)
Number of Leaves :    2
Size of the tree :    3
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   83      87.3684 %
Incorrectly Classified Instances  12      12.6316 %
Kappa statistic                  0.5547
Mean absolute error              0.2053
Root mean squared error         0.3393
Relative absolute error         61.0515 %
Root relative squared error     83.2069 %
Total Number of Instances       95
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.973   0.5      0.88   0.973   0.924   NAO-TROCA
0.5     0.027   0.833   0.5     0.625   TROCA
=== Confusion Matrix ===
 a  b  <-- classified as
73  2 | a = NAO-TROCA
10 10 | b = TROCA
```

FATURAMENTO ANTES

$< = 800.57$

$> 800.57$

NÃO TROCA  
(85.0 / 10.0)

TROCA  
(10.0)

## Agrupamento São Gonçalo

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_Sao_GoncaloCSV
Instances: 15752
Attributes: 10
  NOME_DAE
  CATEGORIA
  SUBCATEGORIA
  CICLO
  ECONOMIAS
  BAIRRO
  DAE
  HIDROMETRO
  FATURAMENTO_5ANTES
  DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
FATURAMENTO_5ANTES <= 72.15: TROCA (74.0/7.0)
FATURAMENTO_5ANTES > 72.15: NAO-TROCA (15678.0/2325.0)
Number of Leaves :    2
Size of the tree :    3
Time taken to build model: 0.69 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   13408      85.1193 %
Incorrectly Classified Instances  2344      14.8807 %
Kappa statistic                  0.0505
Mean absolute error              0.2523
Root mean squared error         0.3555
Relative absolute error        97.9469 %
Root relative squared error     99.0694 %
Total Number of Instances      15752
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.998   0.967   0.852   0.998   0.919   NAO-TROCA
0.033   0.002   0.718   0.033   0.063   TROCA
=== Confusion Matrix ===
  a    b  <-- classified as
13329  31 | a = NAO-TROCA
 2313  79 | b = TROCA
```

São Gonçalo

FATURAMENTO ANTES

$\leq 72.15$

$> 72.15$

TROCA  
( 74 / 7 )

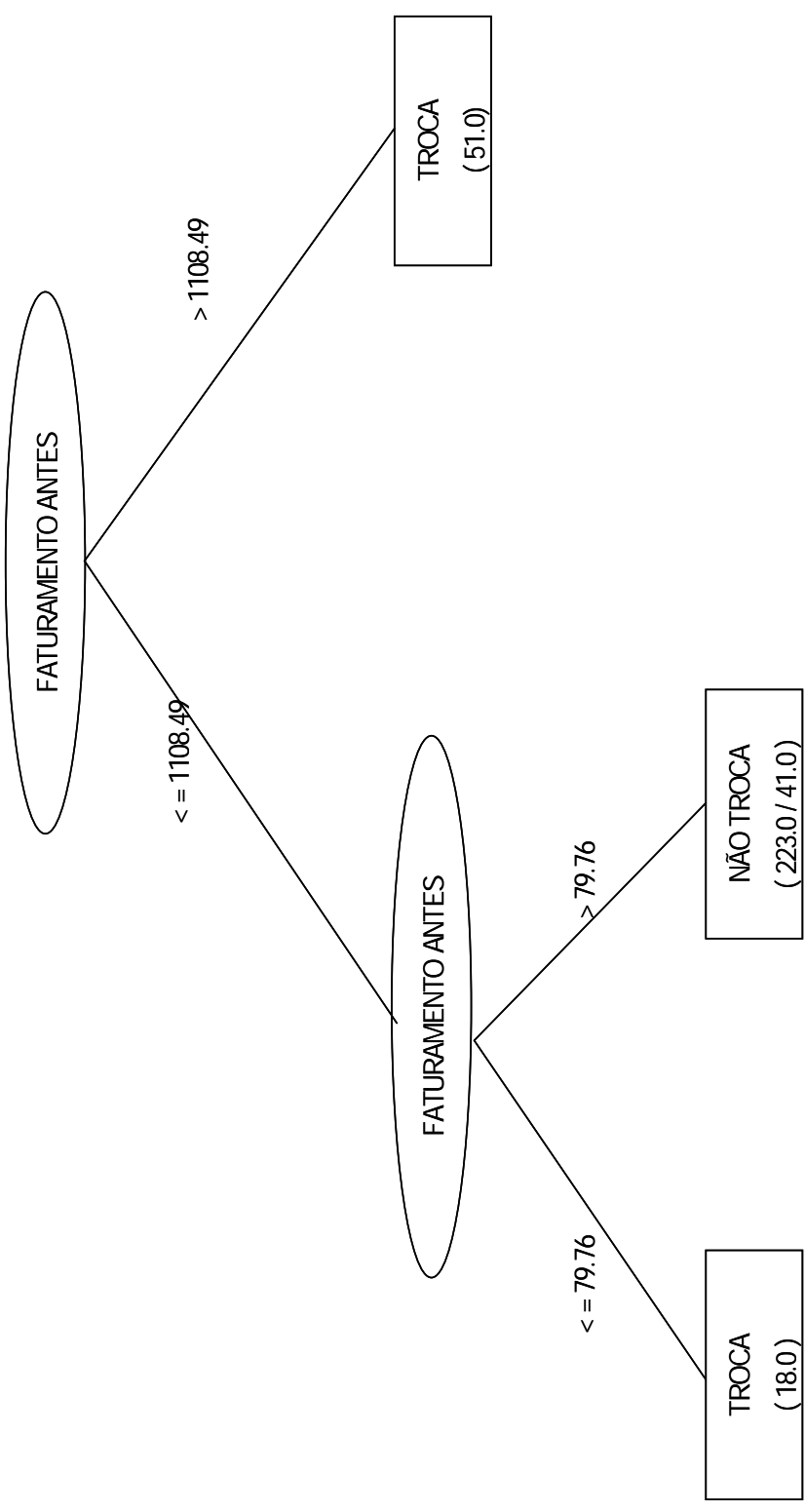
NÃO TROCA  
( 15678 / 2325 )

## Agrupamento Teresópolis

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: 15%_TeresopolisCSV
Instances: 292
Attributes: 10
    NOME_DAE
    CATEGORIA
    SUBCATEGORIA
    CICLO
    ECONOMIAS
    BAIRRO
    DAE
    HIDROMETRO
    FATURAMENTO_5ANTES
    DECISAO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
FATURAMENTO_5ANTES <= 1108.49
| FATURAMENTO_5ANTES <= 79.76: TROCA (18.0)
| FATURAMENTO_5ANTES > 79.76: NAO-TROCA (223.0/41.0)
FATURAMENTO_5ANTES > 1108.49: TROCA (51.0)
Number of Leaves :    3
Size of the tree :    5
Time taken to build model: 0 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances   249      85.274 %
Incorrectly Classified Instances  43      14.726 %
Kappa statistic                  0.6615
Mean absolute error              0.235
Root mean squared error          0.3472
Relative absolute error          50.014 %
Root relative squared error      71.654 %
Total Number of Instances       292
=== Detailed Accuracy By Class ===
TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.618   0.005   0.986   0.618   0.76   TROCA
0.995   0.382   0.812   0.995   0.894   NAO-TROCA
=== Confusion Matrix ===
  a    b  <-- classified as
68  42 | a = TROCA
 1 181 | b = NAO-TROCA
```



Teresópolis



# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)