

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**SISTEMA ESPECIALISTA PROBABILÍSTICO PARA APOIO
AO DIAGNÓSTICO DE
POTENCIAL ECONÔMICO - SEPE**

Hamilton Pasini

FLORIANÓPOLIS, 2002

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**SISTEMA ESPECIALISTA PROBABILÍSTICO PARA APOIO AO
DIAGNÓSTICO DE
POTENCIAL ECONÔMICO - SEPE**

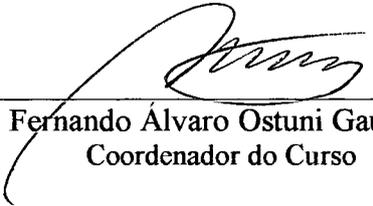
Dissertação submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Mestre em Ciência da Computação sob orientação da Prof^a Silvia Modesto Nassar, Dra.

Florianópolis, 2002

HAMILTON PASINI

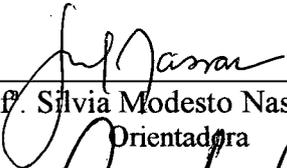
**SISTEMA ESPECIALISTA PROBABILÍSTICO PARA APOIO AO
DIAGNÓSTICO DE POTENCIAL ECONÔMICO - SEPE**

Esta Dissertação foi julgada adequada para a obtenção do título de Mestre em Ciência da Computação, Área de Concentração Sistemas de Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

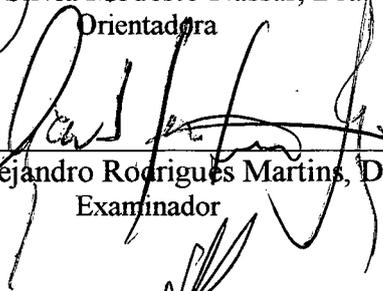


Prof. Fernando Álvaro Ostuni Gauthier, Dr.
Coordenador do Curso

Banca Examinadora:



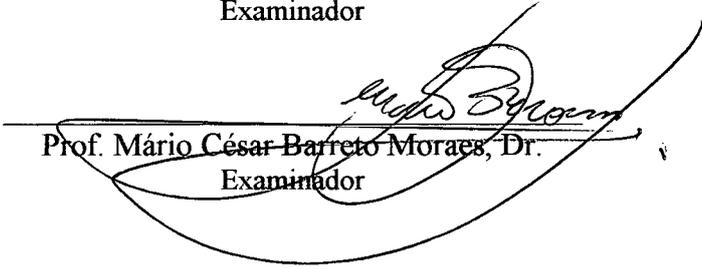
Prof. Sílvia Modesto Nassar, Dra
Orientadora



Prof. Alejandro Rodríguez Martins, Dr.
Examinador



Prof. Bruno H. Kopittke, Dr.
Examinador



Prof. Mário César Barreto Moraes, Dr.
Examinador

“Toda jornada, por mais longa que seja,
sempre começa com um primeiro passo”.
(Sabedoria Oriental)

AGRADECIMENTOS

À minha esposa Sônia, que soube compreender os períodos de ausência.

Aos Correios, por facilitar a realização desta dissertação. Particularmente, agradeço ao gerente da área de *marketing*, Sr. Rosenberg Lucas da Silva, pelo apoio e contribuição.

À professora Silvia Modesto Nassar (Dr^a), orientadora desta dissertação, por seu incentivo, orientações e, principalmente, por sua paciência.

Aos professores Alejandro Rodrigues Martins (Dr.), Bruno H. Kopittke (Dr.) e Mário César Barreto Moraes (Dr.), pelas valiosas contribuições oferecidas até que se atingisse o formato final desta dissertação.

À equipe de trabalho dos Correios, que colaboraram com a realização desta dissertação, incentivando-me.

RESUMO

Realizar o planejamento mercadológico a partir do conhecimento do potencial econômico dos municípios de Santa Catarina é uma das tarefas desempenhadas pela área de *marketing* dos Correios. Na sua elaboração trabalha-se com dados incertos sobre o mercado, ou seja, tomam-se decisões sem ter as informações necessárias. Para que o diagnóstico seja realizado de uma forma mais eficiente, é necessário que os técnicos visualizem os dados relacionados ao potencial de uma forma gráfica e que sejam investigadas relações entre variáveis endógenas e exógenas aos Correios, por meio de evidências observadas. Sendo assim, esta pesquisa volta-se ao desenvolvimento de um sistema especialista probabilístico para o apoio ao diagnóstico de potencial econômico dos municípios de Santa Catarina segundo serviços dos Correios, denominado de SEPE. O sistema explora o raciocínio probabilístico em sistemas especialistas, com uma base de conhecimentos atualizável. Ainda, emprega técnicas de mineração de dados no processo de aquisição e atualização de conhecimento de uma base de dados. O SEPE foi desenvolvido para a plataforma Windows, no ambiente *Visual Basic 5*, utilizando as DLL's da *shell* Netica.

ABSTRACT

One of the tasks for the area of *marketing* of the Post Offices is to carry out the *marketing* planning of the economic potential of the cities of Santa Catarina. Work is done with uncertain data, in other words, decision without having the necessary information are taken. For the diagnosis to be carried out in a more efficient way, it is necessary that the technician visualizes the data related to the potential in a graphical form and that relations between endogenous and exogenous variable of the Post Offices are investigated, by means of evidence. Being thus the case, this research turns to the development of a probabilistic expert system to help the diagnosis of economic potential of the cities of Santa Catarina, called SEPE. As well as it explores the probabilistic reasoning in systems specialists, with a renewable formation of knowledge. Still, it uses techniques of data mining in the process of acquisition of knowledge of a database. The SEPE was developed for the Windows platform, in Visual Basic 5 environment, using the DLL of *shell* Netica.

SUMÁRIO

RESUMO	VI
ABSTRACT	VII
LISTA DE FIGURAS	XI
LISTA DE ABREVIATURAS E SIGLAS	XII
1 INTRODUÇÃO	1
1.1 JUSTIFICATIVAS.....	3
1.2 OBJETIVO GERAL	4
1.3 OBJETIVOS ESPECÍFICOS	4
1.4 ESTRUTURA DA DISSERTAÇÃO	4
2 FUNDAMENTAÇÃO TEÓRICA	5
2.1 SISTEMAS ESPECIALISTAS	5
2.1.1 <i>Os componentes básicos de um Sistema Especialista</i>	5
2.1.2 <i>Sistemas Especialistas Probabilísticos</i>	7
2.2 PROBABILIDADE.....	9
2.2.1 <i>Espaço de Probabilidade</i>	9
2.2.2 <i>Probabilidade Condicional</i>	9
2.3 TEOREMA DE BAYES	10
2.3.1 <i>Independência de Eventos</i>	11
2.3.2 <i>Atualização Bayesiana para uma Nova Evidência</i>	11
2.4 REDES BAYESIANAS	12
2.5 MINERAÇÃO DE DADOS	14
2.5.1 <i>Escolha de Técnica de Mineração de Dados</i>	15
2.5.2 <i>Seleção da Técnica de Agrupamento</i>	15
2.6 CONSIDERAÇÕES FINAIS	17
3 DOMÍNIO DA APLICAÇÃO	18
3.1 A EMPRESA BRASILEIRA DE CORREIOS E TELÉGRAFOS	18
3.1.1 <i>Produtos e Serviços da ECT</i>	20
3.1.2 <i>A Diretoria Regional de Santa Catarina</i>	20
3.1.3 <i>O setor de marketing</i>	21
3.2 A ECONOMIA CATARINENSE	22
3.2.1 <i>Economia</i>	23

<u>3.2.2 Balança Comercial</u>	25
3.3 POTENCIAL ECONÔMICO	26
<u>3.3.1 Marketing</u>	26
<u>3.3.2 Mercado</u>	28
<u>3.3.3 Método para estimar o potencial de mercado</u>	30
3.4 CONSIDERAÇÕES FINAIS	30
4 EXEMPLOS DE SISTEMAS CORRELATOS	32
4.1 SISTEMAS ESPECIALISTAS PROBABILÍSTICOS (SEP).....	32
<u>4.1.1 Baseline Bayesiana para Gerência de Redes de Computadores</u>	32
<u>4.1.2 DIAVAL</u>	32
<u>4.1.3 Sistema de Auxílio ao Diagnóstico Diferencial de Cefaléia</u>	33
<u>4.1.4 SISSPAN</u>	33
<u>4.1.5 SEP de Base de Conhecimento Dinâmica</u>	33
<u>4.1.6 SACI</u>	34
<u>4.1.7 Uma Metodologia de Uso de Técnicas de Indução para Criação de Regras de</u> <u>Sistemas Especialistas</u>	34
4.2 TÉCNICA DE AGRUPAMENTO DE MINERAÇÃO DE DADOS.....	34
<u>4.2.1 WEKA</u>	34
<u>4.2.2 AUTOCLASS</u>	35
4.3 CONSIDERAÇÕES FINAIS	35
5 O SISTEMA	36
5.1 AMBIENTE.....	36
5.2 DADOS	36
5.3 ETAPAS DE DESENVOLVIMENTO DO SEPE	38
<u>5.3.1 Aquisição do conhecimento</u>	38
<u>5.3.2 Representação do conhecimento</u>	40
<u>5.3.3 Implementação do SEPE</u>	40
<u>5.3.3.1 Funções API disponíveis pela shell Netica</u>	40
<u>5.3.4 Desenvolvimento da Interface com o Usuário</u>	41
5.4 MÓDULO MINERAÇÃO DE DADOS	42
5.5 MÓDULO COEFICIENTE DE CORRELAÇÃO	44
5.6 MÓDULO REDE BAYESIANA	44
5.7 ETAPA DE APRENDIZAGEM	45
5.8 AVALIAÇÃO DO SISTEMA	47

<u>5.9 CONSULTA AO SISTEMA</u>	48
6 CONCLUSÃO	50
<u>6.1 TRABALHOS FUTUROS</u>	51
7 REFERÊNCIAS BIBLIOGRÁFICAS	52
8 ANEXO	53

LISTA DE FIGURAS

<u>Figura 1: Módulos usuais de um SE implementado de modo simbólico.</u>	6
<u>Figura 2: Parte qualitativa da rede bayesiana do SEPE.</u>	13
<u>Figura 3: Parte quantitativa da rede bayesiana do SEPE.</u>	13
<u>Figura 4: Diagrama do Fluxo Postal.</u>	18
<u>Figura 5: Rede de Unidades da ECT e de Santa Catarina, 2002.</u>	19
<u>Figura 6: Principais indicadores dos países integrantes do MERCOSUL (1999).</u>	22
<u>Figura 7: População residente em Santa Catarina, por situação de domicílio.</u>	22
<u>Figura 8: Principais produtos exportados por portos catarinenses.</u>	23
<u>Figura 9: Consumo de Energia Elétrica em SC (MWh).</u>	23
<u>Figura 10: Movimento estimado de turistas em Santa Catarina em 2000.</u>	25
<u>Figura 11: Balança Comercial/Santa Catarina (US\$ mil FOB)</u>	26
<u>Figura 12: Mercado potencial.</u>	29
<u>Figura 13: Formação do banco de dados relacional.</u>	37
<u>Figura 14: Variáveis do banco de dados de família de serviços da ECT.</u>	37
<u>Figura 15: Variáveis do banco de dados econômico dos municípios de SC.</u>	38
<u>Figura 16: Aquisição automática do conhecimento.</u>	39
<u>Figura 17: Probabilidades a priori do SEPE.</u>	39
<u>Figura 18: Rede Bayesiana do sistema SEPE.</u>	40
<u>Figura 19: Módulo principal do SEPE.</u>	41
<u>Figura 20: Módulo Coeficiente de Correlação do SEPE.</u>	42
<u>Figura 21: Módulo Consulta de Variáveis Externas do SEPE.</u>	43
<u>Figura 22: Módulo Consulta de Variáveis Internas do SEPE.</u>	44
<u>Figura 23: Representação do protótipo da rede bayesiana na shell NETICA.</u>	45
<u>Figura 24: Rede Bayesiana com máxima entropia.</u>	46
<u>Figura 25: Rede Bayesiana após o aprendizado.</u>	46
<u>Figura 26: Tela de Ajuda do SEPE.</u>	47
<u>Figura 27: Diagnósticos fornecidos pela Rede Bayesiana.</u>	49

LISTA DE ABREVIATURAS E SIGLAS

- BDF** – Banco de Dados Financeiro
- BC** – Base de Conhecimento
- BPI** – *Buying Power Index*
- DCBD** – Descoberta de Conhecimento em Base de Dados
- DR/SC** – Diretoria Regional dos Correios de Santa Catarina
- ECT** – Empresa Brasileira de Correios e Telégrafos
- IA** – Inteligência Artificial
- IAS** – Inteligência Artificial Simbólica
- INE** – Departamento de Informática e de Estatística
- KDD** – *Knowledge Discovery in Databases*
- MD** – Mineração de Dados
- MIB** – *Management Information Base*
- RB** – *Redes Bayesianas*
- SC** – Santa Catarina
- SE** – Sistema Especialista
- SEP** – Sistema Especialista Probabilístico
- SEPE** – Sistema Especialista Probabilístico para Apoio ao Diagnóstico de Potencial Econômico
- UFSC** – Universidade Federal de Santa Catarina

1 INTRODUÇÃO

A inteligência artificial é um ramo da informática que trata de reproduzir a capacidade cognitiva da mente humana mediante sistemas computadorizados. Entre as várias atividades dos pesquisadores desta área pode-se citar a extração de padrões de dados e o desenvolvimento de sistemas especialistas.

Dentre os modelos de sistemas especialistas, os sistemas especialistas probabilísticos trabalham com a incerteza num domínio de conhecimento, por meio da distribuição de probabilidades condicionais em redes *bayesianas*. Segundo Díez (1997), sistemas como esses, têm sido aplicados com sucesso em diversas áreas, inclusive na área de *marketing*.

Os setores de *marketing* das empresas procuram adaptar-se continuamente, buscando oferecer respostas rápidas e corretas às repentinas e bruscas mudanças do mercado. Esta é uma necessidade constante para as empresas, especialmente no ramo da prestação de serviços.

Na busca dessas respostas “a abordagem probabilística proporciona consistência e confiabilidade, além de oferecer uma metodologia bastante adequada à compreensão humana” (CHEESEMAN, 1985).

As questões levantadas nesta pesquisa são as seguintes: como apoiar decisões na área de *marketing* das empresas utilizando sistemas especialistas? Como tratar incertezas por aleatoriedade, presentes na área de *marketing* das empresas? Como estabelecer automaticamente categorias para variáveis contínuas, presentes na base de conhecimento de sistemas especialistas? Como atualizar automaticamente a base de conhecimento de sistemas especialistas probabilísticos (rede *bayesiana*)?

Esta dissertação tem como propósito o desenvolvimento de um sistema especialista probabilístico para a área de *marketing* da Empresa Brasileira de Correios de Santa Catarina. Denominado de SEPE – Sistema Especialista Probabilístico para apoio ao Diagnóstico de Potencial Econômico. O sistema possui também uma etapa de descoberta de agrupamentos naturais das instâncias de cada atributo da base de dados.

A realização desta pesquisa pode proporcionar uma contribuição na aplicabilidade de uma modelagem utilizando inteligência artificial, por meio da abordagem probabilística, na área de *marketing*.

Tendo em vista o ambiente de incerteza em que está inserido o setor de *marketing* e a complexidade que envolve o estabelecimento e acompanhamento de metas de vendas de

produtos/serviços torna-se relevante construir um sistema de apoio ao diagnóstico do potencial econômico.

1.1 Justificativas

O estado de Santa Catarina possui 293 municípios e uma população de 5.028.339 habitantes. Apesar de ocupar apenas 1,12% do território e 3% da população do país, Santa Catarina ocupa uma posição de destaque na formação do PIB Brasileiro.

Inserido neste contexto, a empresa de Correios de Santa Catarina está presente em todos os 293 municípios catarinenses e conta com 2.016 pontos de atendimento. Recebe uma média anual de 135 milhões de objetos, oferecendo uma quantidade de 3.911 produtos e serviços.

Para conseguir adequar-se a esse grande mercado e oferecer serviços de qualidade os Correios de Santa Catarina contam com uma área de *marketing* em sua estrutura. Além de administrar toda a rede de pontos de venda, também gerencia o relacionamento com clientes de maior volume de negócios. Neste caso, mantém uma carteira com 2.214 clientes. São grandes clientes situados nos principais municípios do estado e que mantém com a ECT 4.768 contratos.

Todos os anos os Correios de Santa Catarina recebem uma meta de receita a ser alcançada. No ano de 2000 esta meta era de R\$ 87.536.000,00. A meta é distribuída para os pontos de atendimento e para os profissionais da área de vendas em função do histórico alcançado pelos municípios.

Para que as metas sejam coerentes é necessário um sólido planejamento de *marketing* para nortear as ações futuras. A série histórica do montante arrecadado pelas suas agências é uma ferramenta importante, mas que pode gerar equívocos quando mantém situações não desejadas ao longo dos anos, como por exemplo a acomodação das pessoas às mudanças do mercado.

A criação de uma ferramenta que possa apoiar o diagnóstico do potencial econômico dos municípios, segundo os serviços da Empresa Brasileira de Correios e Telégrafos, será de grande utilidade para a área de *marketing*, pois poderá facilitar o trabalho de planejamento e, ao mesmo tempo, apoiar o direcionamento dos esforços mercadológicos.

Esta dissertação propõe o desenvolvimento de um protótipo que procure apoiar uma boa alternativa para decisão de problema de planejamento mercadológico. Algumas áreas de conhecimento são aplicadas para sua solução, tais como: sistemas especialistas, mineração de dados e redes *bayesianas*. Desta forma, este trabalho está inserido no contexto da área de Inteligência Computacional.

1.2 Objetivo Geral

Desenvolver um protótipo de um sistema especialista probabilístico para classificar o potencial econômico dos municípios de Santa Catarina para os Correios – SEPE.

1.3 Objetivos Específicos

São objetivos específicos desta pesquisa:

- Utilizar as técnicas e ferramentas disponíveis na implementação de sistemas computacionais para aplicação em domínios com a presença de incerteza;
- Desenvolver um módulo para aquisição de conhecimento, com atualização automática da rede *bayesiana*;
- Desenvolver os módulos de *interface* e ajuda do sistema SEPE;
- Integrar o protótipo do sistema especialista à *Application Program Interface* (API) do Netica, utilizando funções que facilitam o trabalho com redes *bayesianas*;
- Avaliar qualitativamente a base de conhecimento do protótipo junto a especialistas do setor.

1.4 Estrutura da Dissertação

Esta dissertação está estruturada em seis capítulos. Cada capítulo constitui parte essencial sobre o estudo que foi realizado para a concretização desta pesquisa.

O capítulo 2 traz a fundamentação teórica desta pesquisa. Na ordem são apresentados os conceitos de Inteligência Artificial, de Sistemas Especialistas, de Reconhecimento de Padrões em base de dados e de Redes *Bayesianas*.

O capítulo 3 objetiva mostrar o domínio de aplicação desta pesquisa. Nele, destaca-se a Empresa Brasileira de Correios e Telégrafos e a Diretoria Regional dos Correios de Santa Catarina - foco desta pesquisa. O capítulo apresenta também o estado de Santa Catarina, fornecendo informações sobre a sua geografia e economia. Ainda, tendo em vista que o projeto insere-se no contexto da área de *marketing*, o capítulo apresenta alguns conceitos sobre *marketing* e mercado e apresenta um método para cálculo de potencial econômico.

O capítulo 4 apresenta exemplos de sistemas correlatos. São mostrados alguns sistemas desenvolvidos com abordagens *bayesianas*, particularmente na UFSC, e são apresentados sistemas que realizam agrupamentos utilizando o método *Bayesiano*.

O capítulo 5 apresenta o sistema, as características do desenvolvimento, os recursos necessários e os resultados atingidos.

Finalmente, no capítulo 6 tem-se as conclusões e perspectivas para aprimoramento do presente trabalho e recomendações para futuras pesquisas, no momento vislumbradas, na área de sistemas especialistas probabilísticos aplicados ao *marketing*.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os seguintes tópicos: conceitos referentes a Sistemas Especialistas, tendo em vista sua relevância na investigação de adequação do enfoque *bayesiano*; conceitos relativos a Probabilidade, ao Teorema de Bayes e Redes *Bayesianas*, tópicos esses essenciais para o desenvolvimento de um sistema especialista probabilístico; processo de mineração de dados, procurando determinar as relações de interdependências escondidas nos dados.

2.1 Sistemas Especialistas

Em relação aos principais modelos de IA, esta dissertação utiliza os sistemas baseados em conhecimento, também denominados de sistemas especialistas, e o raciocínio probabilístico de forma a tratar o conhecimento de um especialista humano num domínio de conhecimento cujas informações apresentam incerteza por aleatoriedade.

Na década de setenta, o objetivo dos pesquisadores em inteligência artificial era desenvolver programas computacionais que resolvessem problemas de modo que fossem considerados inteligentes. Surgem então, os primeiros sistemas especialistas, oriundos de vinte anos de pesquisas na área (RUSSEL, 1995).

Os Sistemas Especialistas (SE) são sistemas de informação que contêm dados ou sentenças descritivas sobre determinado ramo da atividade humana (COELLO, 1997).

Segundo Kandel (1992), esses sistemas reproduzem o conhecimento de um especialista, devendo portanto serem desenvolvidos juntamente com o especialista humano no assunto que se pretende resolver, que fornecerá as informações necessárias para a construção da base de conhecimento.

2.1.1 Os componentes básicos de um Sistema Especialista

Um sistema especialista apresenta cinco componentes básicos: base de conhecimento, motor de inferência, subsistema de aquisição do conhecimento, subsistema de explicações e interface com o usuário, como mostrado na figura 1.

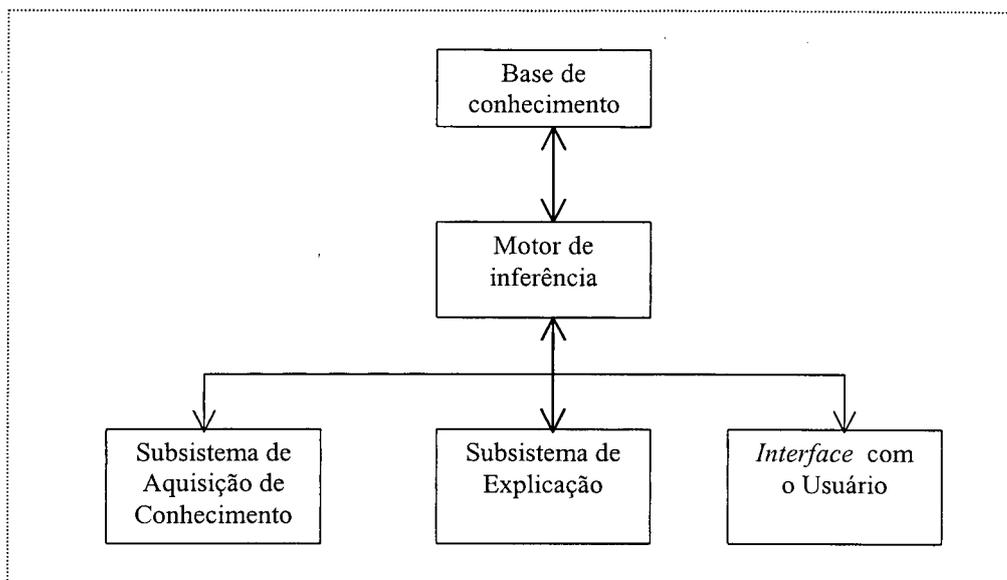


Figura 1: Módulos usuais de um SE implementado de modo simbólico.
 Fonte: Inteligência artificial – no limiar do século XXI

A seguir, apresentam-se os principais componentes de um SE, segundo Barreto (2001):

- **Base de conhecimento:** fornece as características de funcionamento do sistema, armazena os fatos e regras a serem usados. Consiste de conhecimento específico sobre algum domínio;
- **Mecanismo de inferência:** indica a operação a ser realizada sobre o conhecimento contido no SE; busca as regras necessárias a serem avaliadas, ordena-as de uma maneira lógica e direciona o processo de inferência; compara *strings* com padrões, e se a comparação for satisfatória, uma atitude será tomada, buscando determinado objetivo;
- **Subsistema de aquisição de conhecimento:** é usado para alimentar a base de conhecimento. Nele pode-se introduzir, alterar ou excluir regras de acordo com a necessidade. Auxilia os especialistas a expressarem conhecimentos de forma que possa ser incorporada a base de conhecimento;
- **Subsistema de Explicação:** O subsistema de explicações é designado para explicar ao usuário a linha de raciocínio que o sistema especialista utilizou para chegar à conclusão. Isto obriga o sistema a exibir todas as regras usadas para se chegar a uma conclusão particular;

- **Interface com o usuário:** exibe toda a transação de informações que ocorrem durante a consulta. Auxilia os usuários na consulta ao sistema, providenciando-lhes as informações requeridas para resolver seus problemas, mostrando as conclusões do programa e explanando seu raciocínio.

2.1.2 Sistemas Especialistas Probabilísticos

Especialistas humanos são capazes de tomar decisões baseados em informação incerta, incompleta e, até mesmo contraditória. Para que um sistema seja confiável, o mesmo deve lidar com esse tipo de informação com a mesma facilidade que o ser humano.

O mecanismo de raciocínio de Inteligência Artificial aqui utilizado é o raciocínio probabilístico: apoiando-se em informações probabilísticas sobre fatos de um domínio chega-se a uma conclusão a respeito de um novo fato, conclusão esta associada a uma probabilidade.

Nos sistemas especialistas probabilísticos os valores das probabilidades refletem a crença do especialista sobre o que espera que ocorra em situações similares as que tem experienciado e aprendido. A idéia de que as probabilidades se alteram com a mudança de conhecimento é crucial para estes sistemas. Eles tem em sua base de conhecimentos fatos e regras que representam o conhecimento do especialista num domínio, e é explicitada a crença em sua ocorrência através de valores de probabilidade. O raciocínio realizado pelo sistema deve considerar estas probabilidades para associar o vetor de probabilidades ao conjunto de hipóteses diagnósticas. A hipótese com maior probabilidade de ocorrência pode ser considerada a conclusão do sistema, e esta conclusão está associada ao grau de certeza da resposta do sistema (NASSAR, 1998).

Um Sistema Especialista Probabilístico é composto basicamente pelas características já citadas (fig. 11) e tem em sua base de conhecimento uma distribuição de probabilidades. Entre as variáveis serão necessárias relações de dependência estocásticas, que podem ser especificadas por meio de fatos e/ou regras.

O conhecimento do especialista é disposto numa base de conhecimentos, onde as inferências são representadas através de regras do tipo SE-ENTÃO. Neste caso, o conhecimento do especialista é adquirido e representado de tal forma que permite ao sistema emular o comportamento do especialista ao resolver um problema. Nesses sistemas baseados em regras, as declarações SE-ENTÃO são interpretadas de forma unidirecional. As regras são também chamadas de Regras de Produção e elas assumem o seguinte formato:

Se Premissa Então Conclusão.

Para a utilização do Teorema de Bayes as regras de inferência são apresentadas na forma:

Se [evidência e] Então [hipótese H]

A seguir é apresentado um modelo de como a inferência *bayesiana* é realizada. No modelo não são consideradas todas as evidências que influenciam no diagnóstico, devido a grande quantidade de cálculos necessários.

Evidências para uma consulta, onde:

e1: Município possui de 0 a 2 bancos;

e2: Município possui população entre 6.000 e 15.000 habitantes

O formato da regra de produção da probabilidade à posteriori:

$$P(\mathbf{H}_i | \mathbf{e}_1 \wedge \mathbf{e}_2) = [P(\mathbf{e}_1 \wedge \mathbf{e}_2)]^{-1} * P(\mathbf{H}_i) * P(\mathbf{e}_1 | \mathbf{H}_i) * P(\mathbf{e}_2 | \mathbf{H}_i)$$

sendo:

P(e_i|H_i) : Probabilidades condicionais

P(H_i): Probabilidade à priori da Hipótese diagnóstica H_i.

Formato geral das regras de produção no raciocínio *bayesiano*:

$$P(H/e) = \frac{P(e \cap H)}{P(e)} = \frac{P(e/H) \cdot P(H)}{P(e)}$$

Na abordagem *Bayesiana*, a premissa é referida como evidência e a conclusão como hipótese diagnóstica.

Para se utilizar Métodos Probabilísticos no Tratamento de Incerteza em SE, é necessário obter um conjunto de probabilidades a priori, a fim de descrever a população (conjunto de hipóteses e evidências) que formam o conjunto universo. Quando surgem evidências relacionadas a determinadas hipóteses, há uma atualização nas probabilidades a priori das primeiras. O resultado é um conjunto de probabilidades a posteriori (PACHECO, 1991).

2.2 Probabilidade

A seguir, será feita uma apresentação formal dos conceitos de probabilidade (NASSAR, 1998).

2.2.1 Espaço de Probabilidade

Seja ε um conjunto finito e \mathbf{P} uma função de ε para os números reais não negativos, tal que:

$$\sum \mathbf{P}(\mathbf{e}) = 1 \quad \forall \quad \mathbf{e} : \mathbf{e} \in \varepsilon$$

O par $(\varepsilon, \mathbf{P})$ é chamado de espaço de probabilidade. Os elementos de ε são chamados de eventos simples ou elementares. \mathbf{P} é chamado de distribuição de probabilidade ou função de probabilidade. Desta definição resulta que a probabilidade de qualquer evento \mathbf{e} , denotada por $\mathbf{P}(\mathbf{e})$, é medida por um número no intervalo $[0;1]$.

Intuitivamente, ε é a coleção de resultados que se pode esperar em um domínio de aplicação. O valor $\mathbf{P}(\mathbf{e})$ é uma estimativa da crença de que o resultado \mathbf{e} ocorra.

2.2.2 Probabilidade Condicional

O conceito de probabilidade condicional permite considerar as novas informações de forma a obter as novas probabilidades.

Sejam \mathbf{A} e \mathbf{B} eventos compostos de um espaço de probabilidades $(\varepsilon, \mathbf{P})$. Suponha que um evento simples \mathbf{e} ocorra. A probabilidade $\mathbf{P}(\mathbf{B})$ é a probabilidade de que $\mathbf{e} \in \mathbf{B}$ dado nosso conhecimento inicial refletido por \mathbf{P} . Intuitivamente, $\mathbf{P}(\mathbf{B}|\mathbf{A})$ é a probabilidade que $\mathbf{e} \in \mathbf{B}$ quando se tem a informação adicional de que $\mathbf{e} \in \mathbf{A}$.

Seja $(\varepsilon, \mathbf{P})$ um espaço de probabilidade e seja $\mathbf{A} \subseteq \varepsilon$ tal que $\mathbf{P}(\mathbf{A}) \neq 0$. Define-se um novo espaço de probabilidade $(\varepsilon, \mathbf{f})$ da seguinte forma:

$$f(e) = \begin{cases} P(e) / P(A) & \text{se } e \in A \\ 0 & \text{se } e \notin A \end{cases}$$

Para qualquer $B \subseteq \varepsilon$ a probabilidade condicional de B dado a ocorrência de A é igual a $f(B)$. Observe que neste caso A é o novo espaço de probabilidade, onde B deve ser analisado.

Se $A = \varepsilon$ então $P(B|A) = P(B)$.

2.3 Teorema de Bayes

Seja o espaço de probabilidade (ε, P) e os eventos compostos $e, H_1, H_2, \dots, H_k \subseteq \varepsilon$, desde que nenhum desses eventos tenha probabilidade nula, então:

$$P(H_i | e) = \frac{P(e | H_i) \cdot P(H_i)}{P(e)}$$

Se $P(H_i \wedge e) \neq 0$ para todo i então:

$$\frac{P(H_i | e)}{P(H_j | e)} = \frac{P(H_i)}{P(H_j)} \cdot \frac{P(e | H_i)}{P(e | H_j)}$$

Se os eventos $H_1 \cup H_2 \cup \dots \cup H_k = \varepsilon$ e $H_i \cap H_j = \emptyset$ para todo $i \neq j$ então:

$$P(e) = P(H_1) \cdot P(e | H_1) + P(H_2) \cdot P(e | H_2) + \dots + P(H_k) \cdot P(e | H_k)$$

Resultando:

$$P(H_i | e) = \frac{P(e | H_i) \cdot P(H_i)}{\sum_{j=1}^k (P(H_j) \cdot P(e | H_j))}$$

Nas aplicações dos sistemas especialistas probabilísticos os H 'is são as hipóteses concorrentes. O evento e pode ser visto como uma evidência. O conhecimento da ocorrência

desta evidência leva a mudanças na probabilidade *a priori* $P(H_i)$ para a probabilidade condicional $P(H_i|e)$, que por sua vez considera a evidência e .

2.3.1 Independência de Eventos

Seja um espaço de probabilidade (ε, P) . E, sejam os eventos $e_1, e_2 \subseteq \varepsilon$. Segue-se que:

- a) Se $P(e_1 \wedge e_2) = P(e_1) \cdot P(e_2)$ então os eventos e_1 e e_2 são independentes.
- b) Genericamente, para qualquer subconjunto $E = \{ e_{i1}, e_{i2}, \dots, e_{ik} \}$ de $\{ e_1, e_2, \dots, e_n \}$

se $P(e_{i1} \wedge e_{i2} \wedge \dots \wedge e_{ik} | H) = P(e_{i1}|H) \cdot P(e_{i2}|H) \dots P(e_{ik}|H)$ então pode-se dizer que os eventos e_i 's são eventos mutuamente independentes dado a hipótese H .

A idéia básica subjacente ao conceito probabilístico de independência é que o conhecimento de certa informação não traz informação adicional sobre outra coisa. Isto é, se e somente se, saber que o evento e_1 ocorreu não trazer informação sobre o evento e_2 e saber que o evento e_2 ocorreu não trazer informação sobre o evento e_1 , então diz-se que ocorre a independência entre estes eventos.

2.3.2 Atualização Bayesiana para uma Nova Evidência

Seja H_i uma hipótese e $e^n = e_1, e_2, \dots, e_n$ uma seqüência de dados independentes observados no passado e seja e um novo fato. A probabilidade condicional para a nova evidência é dada por:

$$\begin{aligned} P(H_i | e^n \wedge e) &= P(H_i \wedge e^n \wedge e) / P(e^n \wedge e) = \\ &= (P(e^n) \cdot P(H_i | e^n) \cdot P(e | e^n \wedge H_i)) / ((P(e^n) \cdot P(e | e^n))) \end{aligned}$$

Resultando em:

$$P(H_i | e^n \wedge e) = P(H_i | e^n) \cdot (P(e | e^n \wedge H_i) / P(e | e^n))$$

O resultado acima mostra que uma vez calculada a probabilidade condicional da hipótese H dado o conjunto e^n de evidências, isto é o valor $P(H_i|e^n)$, os dados passados e^n podem ser desprezados e assim pode ser obtido o impacto da nova evidência e . A crença velha $(H_i|e^n)$ assume o papel de crença *a priori* no cálculo do impacto da nova informação e ; a probabilidade $P(H_i|e^n)$ sumariza completamente a experiência passada e para sua atualização necessita somente ser multiplicada pela probabilidade $P(e | e^n \wedge H_i)$. Esta probabilidade mede a probabilidade do novo dado e considerando a hipótese H_i e os dados passados e^n .

2.4 Redes Bayesianas

As redes *bayesianas* – conhecidas também como Redes de Crenças, Redes de Bayes e Redes Probabilísticas Causais - são esquemas de representação de conhecimento utilizadas para desenvolver a base de conhecimentos de um Sistema Especialista Probabilístico. Segundo Pearl (1988), as redes *bayesianas* possuem uma parte estrutural refletindo relações causais e valores de probabilidade refletindo a força da relação. A RB é extraída de um especialista que transfere seus conhecimentos do domínio de aplicação para o engenheiro de conhecimentos do sistema.

Segundo GAAG (1996), as redes *bayesianas* são compostas de duas partes complementares:

A parte **qualitativa** (fig. 2) é representada por um grafo, onde as variáveis são os nós e as regras - relações de dependência condicional entre essas variáveis - são os arcos direcionados. Assim um arco ligando as variáveis A e B , na seguinte forma $A \rightarrow B$, indica que a variável B é a consequência e a variável A é a causa, e estas apresentam uma relação de dependência. Por outro lado, se não houver um arco ligando duas variáveis então se supõe que essas variáveis são independentes.

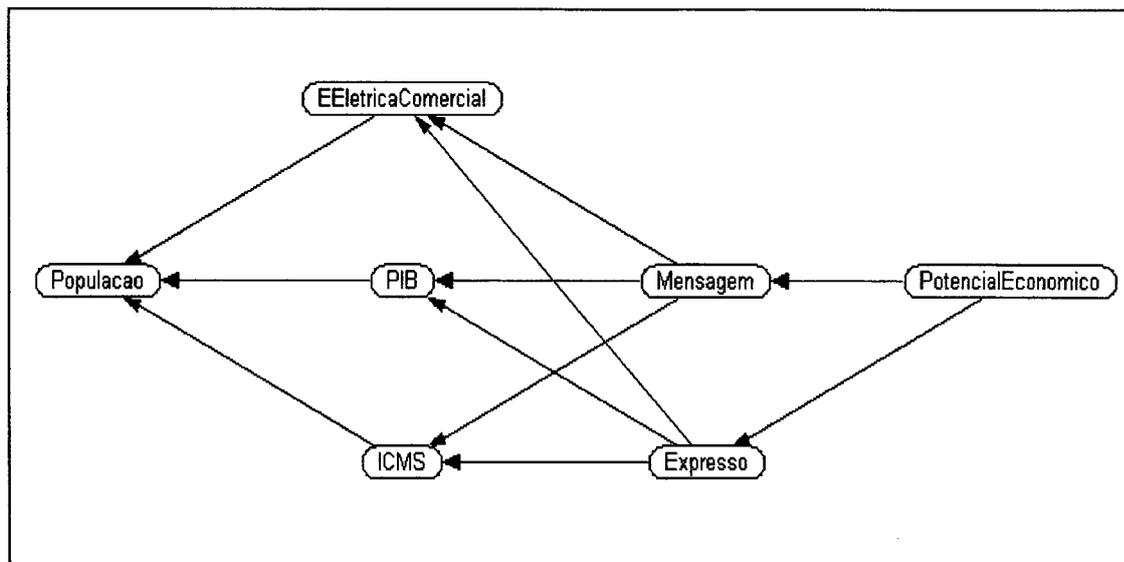


Figura 2: Parte qualitativa da rede bayesiana do SEPE.
Fonte: Shell NETICA

A parte **quantitativa** de uma RB (fig. 3) é o conjunto de probabilidades condicionais associadas aos arcos existentes no modelo gráfico acima descrito e as probabilidades estimadas *a priori* das hipóteses. Associada à parte qualitativa de uma RB está um grupo de funções representando valores numéricos da distribuição em questão. A cada vértice do grafo é associada uma função de atribuição de probabilidades condicionais que descreve a influência dos valores dos vértices predecessores nas probabilidades dos valores deste vértice.

Mensagem	Expresso	G1	G2
G1	G1	82.857	17.143
G1	G2	81.250	18.750
G1	G3	74.074	25.926
G1	G4	100.00	0.000
G1	G5	100.00	0.000
G1	G6		

Figura 3: Parte quantitativa da rede bayesiana do SEPE.
Fonte: Shell NETICA

O SEPE foi desenvolvido com uma base de conhecimento dinâmica. Desta forma o usuário final do sistema pode adicionar e excluir nós, bem como definir novas relações probabilísticas entre esses nós.

2.5 Mineração de Dados

Mineração de Dados, ou *data mining*, consiste na aplicação de técnicas de estatística e inteligência artificial em grandes bases de dados, para encontrar tendências ou padrões entre os dados a fim de apoiar decisões (CABENA, 1998).

Também define-se como um processo de construir modelos baseados em um conjunto de dados, permitindo fazer previsões, controlar ou melhorar algum processo. O objetivo da mineração de dados é descobrir informações e conhecimentos através do reconhecimento de novos fatos e relações existentes em uma base de dados. Ela é apenas uma etapa de um processo maior denominado *Knowledge Data Discovery* (KDD) (MORALES e colaboradores, 2001).

O KDD é um conjunto de tecnologias que envolvem Aprendizagem Automática, tecnologias de bancos de dados, sistemas especialistas e visualização de dados. Ele é empregado para descrever todo o processo de extração de conhecimento dos dados. Já a mineração de dados deve ser empregada somente para o estágio de descoberta de padrões do processo de KDD.

O primeiro passo no projeto de mineração de dados deve ser uma análise superficial do conjunto de dados usando uma ferramenta tradicional de consulta, para se conhecer alguns aspectos básicos e as estruturas do conjunto de dados. Uma boa maneira de iniciar o processo é extraindo algumas informações estatísticas simples do conjunto de dados.

Segundo Skauge (2000), fazer *data mining* significa trabalhar com grandes conjuntos de dados, ou seja, deve-se pegar os dados e transformá-los em informações úteis. E essa transformação deve ser feita segundo uma sequência lógica, isto é, deve ser feita por etapa, e as etapas do *data mining* são: organizar um banco de dados relacional, extrair dados, transformar dados, purificar dados, fazer *data warehouse*, colocar os dados na forma padrão, reduzir as características dos dados, separar os dados em conjunto de treinamento e conjunto de testes, executar os algoritmos, escolher uma técnica de *data mining* e interpretar os resultados.

2.5.1 Escolha de Técnica de Mineração de Dados

O processo de construir um modelo para representar um conjunto de dados é comum para todos os produtos de mineração de dados. Alguns modelos comuns incluem classificação, regressão, séries temporais, agrupamento, associação e regras de seqüência (SKAUGE, 2000).

A seguir, destacam-se três técnicas mais utilizadas na mineração de dados (MORALES e colaboradores, 2001):

- **Classificação:** é uma técnica freqüentemente usada em mineração de dados predizível. É o processo pelo qual examinamos as propriedades (aspectos, estrutura) de um objeto (dados) e o atribuímos a uma das classes predefinidas;
- **Agrupamentos:** a análise de agrupamentos tem como objetivo descobrir agrupamentos naturais de itens. O objetivo nesta técnica é verificar como certos casos estão relacionados e são similares a cada outro (JOHNSON, 1998);
- **Regras de Associação:** é uma técnica que é usada para encontrar relacionamentos entre as variáveis e especificamente eventos que ocorrem ao mesmo tempo. Visa descobrir associações importantes entre os itens, tal que, a presença de um item em uma determinada transação irá implicar na presença de outro item na mesma transação.

Após a escolha da técnica deve-se selecionar o tipo de abordagem a ser utilizada no processo de mineração de dados. Há duas abordagens principais para mineração de dados: supervisionada e não supervisionada. Na abordagem supervisionada, exemplos específicos de um conceito objetivo são dados, e o objetivo é aprender como reconhecer membros de classes usando a descrição de atributos. Na abordagem não supervisionada, um conjunto de exemplos é pego sem nenhuma classificação a priori e mineração de dados é usado para descobrir regularidades e modelos, o mais freqüente é através de identificação de grupos ou subconjuntos de exemplos similares (SKAUGE, 2000).

2.5.2 Seleção da Técnica de Agrupamento

Esta pesquisa utiliza técnicas de agrupamento com abordagem não supervisionada na construção de mineração de dados. Conforme visto anteriormente, técnicas de agrupamento são aplicadas quando não existem classes divididas em grupos naturais.

Dos diversos tipos de agrupamentos podem ser citados três diferentes métodos básicos. O primeiro é o clássico algoritmo *k-means*, que forma grupos em domínios numéricos, particionando instâncias em grupos disjuntos. Ele utiliza o esquema básico do “vizinho mais próximo”. A segunda é uma técnica incremental de agrupamento que foi desenvolvida no final de 1980 e está incorporada em um par de sistemas chamado COBWEB (para atributos nominais) e CLASSIT (para atributos numéricos). O terceiro é um método estatístico de agrupamento baseado no modelo híbrido de diferentes distribuições de probabilidades - uma para cada grupo - e que agrupa instâncias para classes de forma probabilística” (WITTEN, 2000).

O fundamento para o agrupamento estatístico é um modelo estatístico chamado *mixture*. Um *mixture* é um conjunto de K distribuições de probabilidade, onde cada distribuição fornece a probabilidade de que uma particular instância teria um certo conjunto de valores se ele reconhecesse ser um membro deste grupo. Cada grupo tem uma distribuição diferente. Cada instância pertence a um e somente um dos grupos, mas a instância não tem conhecimento de qual. Finalmente, os grupos não são feitos igualmente: existe alguma distribuição de probabilidade que reflete sua população relativa.

Abaixo, apresenta-se o módulo principal do algoritmo EM

```
/**
 * Perform the EM algorithm
 */
private void doEM ()
    throws Exception
{
    if (m_verbose) {
        System.out.println("Seed: " + m_rseed);
    }

    m_rr = new Random(m_rseed);
    m_num_instances = m_theInstances.numInstances();
    m_num_attribs = m_theInstances.numAttributes();

    if (m_verbose) {
        System.out.println("Number of instances: "
            + m_num_instances
            + "\nNumber of atts: "
            + m_num_attribs
            + "\n");
    }

    // setDefaultStdDevs(theInstances);
    // cross validate to determine number of clusters?
    if (m_initialNumClusters == -1) {
        if (m_theInstances.numInstances() > 9) {
            m_num_clusters = CVClusters();
        }
    }
}
```

```

    } else {
        m_num_clusters = 1;
    }
}

// fit full training set
EM_Init(m_theInstances, m_num_clusters);
m_loglikely = iterate(m_theInstances, m_num_clusters, m_verbose);
}

```

Ainda não existem algoritmos que descubram conhecimento. Existem algoritmos que descobrem padrões, percebem coincidências. Apenas os especialistas humanos conseguem interpretar e decidir sobre os resultados obtidos, definindo quando deve ser tomada alguma ação ou seguir uma recomendação dada pelos algoritmos. Portanto, o conhecimento não é descoberto pelos algoritmos, mas sim pelos especialistas humanos, após a utilização das técnicas e algoritmos adequados.

2.6 Considerações Finais

Como é amplamente citado na literatura da área (BARRETO, 2001), (RUSSEL, 1995) o desenvolvimento de um projeto de Inteligência Artificial utilizando um SE não significa a substituição do especialista. O objetivo é oferecer um instrumento de apoio que o ajude a realizar os diagnósticos com maior rapidez e fidelidade.

O emprego de mineração de dados nesta pesquisa tem como base o fato de que é um método útil na análise exploratória em cenários como este da classificação do potencial econômico, onde não existem noções predeterminadas sobre o que irá constituir uma saída interessante.

O uso de redes *Bayesianas* justifica-se pelo seu poderoso formalismo, que representa o conhecimento no domínio e as incertezas, por aleatoriedade, associadas a este domínio. Desta forma, optou-se por utilizá-las para a construção da base de conhecimento do SEPE.

3 DOMÍNIO DA APLICAÇÃO

Neste capítulo apresentam-se: a ECT, empresa na qual está sendo desenvolvido este sistema; a economia catarinense, abrangência de atuação da Diretoria Regional dos Correios de Santa Catarina; o potencial econômico, com conceitos básicos sobre mercado e *marketing*.

3.1 A Empresa Brasileira de Correios e Telégrafos

A Empresa Brasileira de Correios – ECT, é uma empresa prestadora de serviços de correios convencionais e avançados, com atuação em todo o território nacional.

A figura 4 apresenta o processo de prestação de serviços pelos Correios desde o recebimento do objeto até a sua entrega.

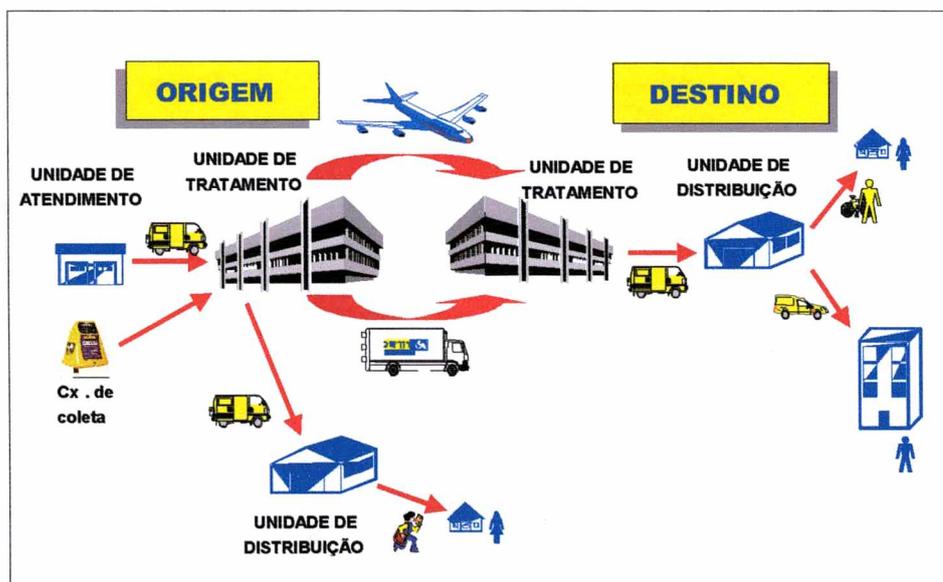


Figura 4: Diagrama do Fluxo Postal.
Fonte: ECT, 2002

A ECT é uma empresa pública, vinculada ao Ministério das Comunicações, com sede em Brasília. Está dividida em 23 Diretorias Regionais, sendo a Diretoria Regional de Santa Catarina o foco desta dissertação.

Segundo pesquisas realizadas, a ECT ocupa um lugar de destaque entre as empresas nacionais, dentre as quais podem ser citadas:

- ◆ 5ª maior empresa de serviço público pelo critério de vendas;
- ◆ 6ª maior empresa de serviço pelo critério de vendas;
- ◆ 7ª maior estatal;
- ◆ 7ª em excelência empresarial.

Em termos de receita operacional bruta, a empresa ocupa o quinto lugar entre as empresas do setor público, com grau zero de endividamento. Apesar de sua estrutura complexa, tem ostentado resultados expressivos nos últimos anos (lucro de duzentos e quarenta e três milhões de reais em 1999). Ainda no ano de 1999, a empresa obteve os seguintes resultados:

- Receita Operacional: R\$ 3.289 bilhões;
- Despesa Operacional: R\$ 2.194 bilhões;
- Quadro de Pessoal: 82,3 mil empregados;
- Quadro de Suplementares: 20 mil colaboradores;
- Objetos distribuídos/ano: 7,4 bilhões;
- Acesso aos serviços: 86% da população;
- Entrega domiciliária: 74% de residências;

Estando presente em todos os municípios do país, possui uma rede de unidades conforme figura 5:

Unidades	Números da ECT	Números de SC
Agências	10.453	288
Caixas de Coleta	25.789	930
Centros de Tratamento	131	4
Centros de Distribuição Domiciliária	651	23
Módulos de Caixa Postal Comunitária	1.973	63
Centros de Operações	827	1

Figura 5: Rede de Unidades da ECT e de Santa Catarina, 2002.
Fonte: ECT

3.1.1 Produtos e Serviços da ECT

Os produtos e serviços universais da ECT (cartas, impressos, telegramas, encomendas) são formatados e normalizados pela Administração Central da empresa em Brasília, de modo a obter uniformidade de procedimentos e padrões em todo o território nacional e ainda atender às orientações governamentais para o setor postal brasileiro.

No caso de necessidade específica dos clientes, a DR/SC estuda e formata adaptações dos produtos e serviços corporativos, ou mesmo novas operações, sob a forma de contratos especiais. Esta forma de atuação é direcionada prioritariamente aos clientes do segmento de atacado, que operam com a ECT através de contrato.

A ECT oferece 3.911 produtos e serviços ao público. Os principais produtos e serviços oferecidos pela ECT são:

- Postagem e entrega de objetos postais: carta simples e impressos;
- Recepção e entrega de objetos telegráficos;
- Recepção e entrega de malotes;
- Produtos Filatélicos;
- Recepção e entrega de Encomendas Expressas: SEDEX, EMS;
- Entrega Especial de Documentos SEED;
- Coleta e entrega de objetos para contratos especiais;

A ECT procurou agrupá-los para facilitar o seu estudo e acompanhamento. Desta forma, foram compostas sete famílias de receita operacional: Mensagem, Encomenda, Expresso, Financeira, Internacional, *Marketing* Direto e Conveniência.

3.1.2 A Diretoria Regional de Santa Catarina

A DR/SC tem por finalidade explorar os serviços postais, correspondência agrupada, telemáticos, encomendas, filatelia e outros.

Os Correios de Santa Catarina atendem uma área territorial de 95.442 km² e uma população residente, estimada em 2001 de 5.028.339 pessoas. Tem sua sede administrativa localizada em Florianópolis. Administrativamente e operacionalmente está dividida em oito regiões de varejo, com sede nos municípios de Florianópolis, Tubarão, Blumenau, Joinville,

Lages, Joaçaba, Chapecó e Itajaí, cujas áreas de atuação são definidas pela zona de influência de suas sedes e pelo sistema de encaminhamento postal.

Na DR/SC, em 1999, foram postados 134 milhões de objetos. A Receita Operacional do ano 2000 alcançou 86,5 milhões de reais, o que representa um crescimento de 13,8% em relação ao ano anterior e uma realização de 98,8% da meta de receita para 2000 (R\$ 87.536.000,00). A DR/SC obteve um lucro de 7,9 milhões de reais no ano 2000.

3.1.3 O setor de marketing

Para viabilizar os negócios postais na DR/SC existe uma estrutura organizacional específica que planeja, orienta e controla as atividades comercial e operacional no âmbito estadual.

O setor de *marketing* de Santa Catarina⁵⁰

está focado em dois processos: administração da rede de pontos de atendimento e vendas no varejo e administração do relacionamento com clientes de maior volume de negócios.

A Gerência de Vendas da Diretoria Regional de Santa Catarina possui em sua estrutura assistentes comerciais e unidades de atendimento espalhadas pelo estado. Esta equipe é responsável pela prospecção e manutenção de clientes em todo o estado. Para esta tarefa, recebem anualmente uma meta de receita, servindo como balizador para a equipe e como uma das formas de acompanhamento periódico pela Gerência de Vendas.

A participação de mercado de cada unidade comercial no estado é definida com base no Índice de Potencial Econômico. A metodologia de cálculo consiste no cruzamento de variáveis econômicas dos municípios e variáveis relacionadas ao faturamento alcançado pelas agências situadas nos respectivos municípios. Periodicamente, a área de *marketing* da DR/SC atualiza as variáveis e, com base no coeficiente de correlação, efetua o planejamento mercadológico. Utiliza-se o Microsoft Excel para realização dos cálculos e geração do índice final.

A partir do Índice de Potencial Econômico são estabelecidas metas de vendas não apenas com base no desempenho histórico de vendas de cada unidade, mas também com base no potencial de mercado definido.

3.2 A Economia Catarinense

O Estado de Santa Catarina ocupa uma área territorial de 95.443 km², correspondente a 16,54% da Região Sul e 1,12% da área do Brasil. Possui uma população de 5.343.060 habitantes (estimativa 2001) e uma densidade demográfica de 55,98 hab/km². As cidades mais populosas são: Joinville, Florianópolis, Blumenau, Criciúma e Chapecó.

Seu quadro Político-Administrativo e Judiciário é constituído por: 293 Municípios, 439 Distritos e 94 Comarcas.

No figura 6 verifica-se que com 3% da população do Brasil, o estado gera 4,2% do Produto Interno Bruto (PIB) do país. Uma comparação com três países vizinhos na América do Sul é bastante ilustrativa: o PIB de Santa Catarina equivale à metade do PIB do Chile, é maior que o do Uruguai e três vezes superior ao do Paraguai.

<i>Países</i>	<i>População (milhões de habitantes)</i>	<i>Área (mil km²)</i>	<i>PIB (US\$ bilhões corrente)</i>	<i>Renda per capita (US\$)</i>
Brasil	168,0	8.512	803,3	5.053
Argentina	36,5	2.780	274,5	7.518
Paraguai	5,4	407	9,6	1.443
Uruguai	3,3	176	29,1	6.291
Bolívia	8,1	1.098	8,2	1.002
Chile	15,0	757	65,6	4.367
Santa Catarina	5,0	95	34,0	6.844

Figura 6: Principais indicadores dos países integrantes do MERCOSUL (1999)
Fonte: IBGE.

Santa Catarina apresenta uma taxa média anual de crescimento populacional de 1,43%, sendo que a taxa de urbanização obteve um crescimento significativo, alcançando 78,70% no ano 2000. A figura 7 compara a população de SC por situação de domicílio.

<i>Ano do Censo</i>	<i>Total</i>	<i>Urbana</i>	<i>(%) s/ total</i>	<i>Rural</i>	<i>(%) s/ total</i>	<i>SC/BR (%)</i>
1980	3.627.933	2.154.238	59,38	1.473.695	40,62	3,05
1991	4.541.994	3.208.537	70,64	1.333.457	29,36	3,09
1996	4.875.244	3.565.130	73,13	1.310.114	26,87	3,10
2000	5.333.284	4.197.287	78,70	1.135.997	21,30	3,14

Figura 7: População residente em Santa Catarina, por situação de domicílio.
Fonte: Fundação IBGE e SDE-SC / DEGE / Gerência de Análise Estatística.

3.2.1 Economia

Três importantes portos marítimos estão instalados no litoral catarinense, permitindo o acesso de navios de até 45 mil toneladas de carga e operando com contêineres. A figura 8 apresenta os produtos exportados pelos principais portos catarinenses.

	São Francisco do Sul	Itajaí	Imbituba
Movimentação: 1998	3.600.000 t	1.700.000 t	600.200 t
Principais produtos	Compressores, carrocerias de ônibus, madeira, papel, móveis, cerâmica, carne congelada, soja (grãos, farelo e óleo), milho, autopeças, têxteis.	Têxteis, açúcar, congelados (frango, carne e pescado), madeira, máquinas, motores, maçãs, veículos, azulejos, derivados de petróleo e papel.	Carnes congeladas, carga geral, açúcar, milho, arroz, fertilizantes, soda cáustica, calçados, fumo, trigo, cerâmica, algodão, coque de carvão, sal.

Figura 8: Principais produtos exportados por portos catarinenses.

Fonte: Secretaria de Estado do Desenvolvimento Econômico e Integração ao Mercosul (1998).

Todos os municípios catarinenses têm energia elétrica. O estado possui o maior índice de eletrificação rural do Brasil: 95% (a média brasileira é de 30%). Seu potencial hidráulico é de 9.5 milhões de KW. A geração de energia em Santa Catarina se dá através de 12 usinas hidrelétricas e uma termelétrica. A figura 9 apresenta o consumo de energia elétrica por classes em SC.

Classes	1999	2000	%
Residencial	2.846.773	2.959.037	3,94
Industrial	4.975.090	5.405.468	8,65
Comercial	1.413.223	1.567.838	10,94
Rural	1.151.481	1.241.001	7,77
Poder Público	216.190	226.569	4,80
Iluminação Pública	396.021	410.686	3,70
Outros	378.789	408.759	7,91
Total	11.377.568	12.219.358	7,40

Figura 9: Consumo de Energia Elétrica em SC (MWh)

Fonte: CELESC e SDE-SC / DEGE / Gerência de Análise Estatística.

Segundo dados da Unidade de Política Econômica e Industrial da FIESC, nos últimas três décadas, a economia catarinense cresceu mais de três vezes, um índice superior ao nacional. A economia se baseia na indústria (agroindústria, têxtil, cerâmica, máquinas e equipamentos), no extrativismo (minérios) e na pecuária. Segundo a (GAZETA MERCANTIL, 2000) a indústria de transformação do estado é a quinta do Brasil em número

de trabalhadores (326 mil) e a sexta em quantidade de empresas. O modelo econômico catarinense se caracteriza pela diversificação de atividades e pela força das empresas familiares de pequeno e médio porte. Há 150 mil empresas familiares, dos mais diferentes portes.

Abaixo, relacionam-se os pólos econômicos mais significativos de Santa Catarina:

- **Pólo agroindustrial:** envolve 1.300 indústrias, que empregam 35 mil pessoas e concentram a maior parte da produção de alimentos do estado. O setor está constantemente investindo milhões para aumentar a produtividade.
- **Pólo florestal:** abrange 1.900 empresas de madeira, 1.180 moveleiras e 117 de papel e papelão. Seu parque moveleiro é o maior da América Latina, com exportações de US\$ 200 milhões em 1998.
- **Pólo eletro-metal-mecânico:** formado por 1.434 indústrias metalúrgicas, 517 mecânicas e 195 de material elétrico e de comunicação, emprega 59 mil trabalhadores. O setor está investindo milhões na expansão de suas atividades.
- **Pólo têxtil/vestuário:** maior concentração de indústrias do setor na América Latina (339, das quais 120 de grande porte) e considerada como a 2ª maior do mundo. O setor fatura US\$ 2 bilhões/ano e responde por 70% das exportações de têxteis do país.
- **Pólo mineral:** os segmentos carboníferos e cerâmico, juntos, têm mais de 1.300 empresas. As indústrias de cerâmica têm faturamento anual de meio bilhão de dólares e respondem por 60% da produção brasileira de pisos e revestimentos. Santa Catarina é o maior produtor de carvão mineral do país.
- **Pólo turístico:** Santa Catarina recebeu 2,2 milhões de turistas no verão 1998/99 e arrecadou 365 milhões de dólares. Há 811 empreendimentos instalados, com capacidade para 69 mil leitos. Os municípios que mais atraem visitantes são Florianópolis, Balneário Camboriú e Blumenau, mas os atrativos turísticos estão distribuídos por todo o estado.
- **Pólo de informática:** O estado tem 600 empresas de *software*, que faturam US\$ 260 milhões/ano e empregam 6 mil trabalhadores. Muitas delas são provenientes de laboratórios universitários e de incubadoras tecnológicas.

O turismo é destaque no setor de serviços. Santa Catarina recebeu 2,7 milhões de visitantes na temporada de verão 2000, dos quais 20,6% estrangeiros e 79,4% nacionais. Este movimento gerou uma receita estimada de US\$ 477 milhões, conforme pode ser visto no figura 10.

<i>Indicadores de Turismo</i>	<i>Receita(*)</i>	<i>Varição anual (%)</i>	<i>Quantidade</i>	<i>Varição anual (%)</i>
Turistas Nacionais	286.663.586,96	13,15	2.194.522	-2,72
Estrangeiros	190.827.808,68	58,75	568.663	22,29
Total de turistas	477.491.395,64	18,99	2.763.185	1,56

Figura 10: Movimento estimado de turistas em Santa Catarina em 2000
Fonte: SDE/SC – SANTUR. (*) Receita estimada em dólar

Conforme Boletim de Arrecadação do Ministério da Fazenda, o setor de comércio e serviços tem grande expressividade no contexto da economia catarinense representando 65,57% da arrecadação do ICMS. Em todo o Estado existem cerca de 92 mil estabelecimentos comerciais que absorvem cerca de 600 mil empregos, representando 23% da população economicamente ativa de Santa Catarina. Do universo de empresas comerciais catarinenses, 90% atuam no setor varejista, compreendendo principalmente estabelecimentos de pequeno e médio porte.

3.2.2 Balança Comercial

Com relação às exportações, Santa Catarina é responsável por 5,7% das exportações, sendo o quinto maior exportador. Os Estados Unidos, a Alemanha e a Arábia Saudita são os maiores compradores. Produtos têxteis, carnes de aves congelados e derivados, farelo de soja, fumo, motocompressores, calçados, açúcar refinado, óleo de soja, papel, pisos e azulejos, portas, janelas e batentes de madeira, carne suína congelada, algodão, refrigeradores de uso doméstico etc são as mercadorias que mais se destacam nas exportações.

A figura 11 apresenta dados sobre as exportações e importações de Santa Catarina. Nele, verifica-se que em 2000 o saldo da balança comercial foi positivo.

<i>Principais produtos Exportados</i>	<i>Exportações 2000</i>	<i>Importações 2000</i>	<i>Saldo</i>
Máquinas e aparelhos, material elétrico	596.182	299.137	
Animais vivos e produtos do reino animal	542.475	37.945	
Produtos alimentícios, bebidas e fumos	190.986	36.368	
Materiais têxteis e suas obras	301.093	139.208	
Metais comuns e suas obras	63.035	23.306	
Material de transporte	80.781	53.364	
Produtos do reino vegetal	30.822	104.233	
Gorduras, óleos e cêras animais e vegetais	24.640	-	
Madeira, cortiça e suas obras	298.911	-	
Obras de pedra, cerâmica, vidros, etc.	144.444	-	
Pastas de madeira, papel e suas obras	105.739	-	
Produtos da indústria química	-	101.902	
Instrumentos e aparelhos científicos	-	20.083	
Outros produtos	332.595	43.599	
Total	2.711.703	956.716	1.754.969

Figura 11: Balança Comercial/Santa Catarina (US\$ mil FOB)
Fonte: SDE/DIDE/GEDEC.

3.3 Potencial Econômico

As empresas e os mercados cresceram e muitos profissionais perderam o contato com seus consumidores. Assim, as empresas devem procurar conhecer os clientes e suas necessidades através de outros meios, não mais o contato pessoal.

Nesta dissertação, o meio utilizado para conhecer os clientes é o potencial econômico dos municípios. Através de dados econômicos dos municípios e de arrecadação na venda de produtos e serviços da ECT procura-se gerar conhecimento sobre a classificação do potencial econômico dos municípios.

Alguns conceitos sobre marketing e mercado são vistos para auxiliar na compreensão da formação do potencial econômico.

3.3.1 Marketing

O termo *marketing* não deve ser entendido na sua antiga concepção de fazer uma venda – “vender” – mas sim na atual concepção de satisfazer as necessidades do cliente.

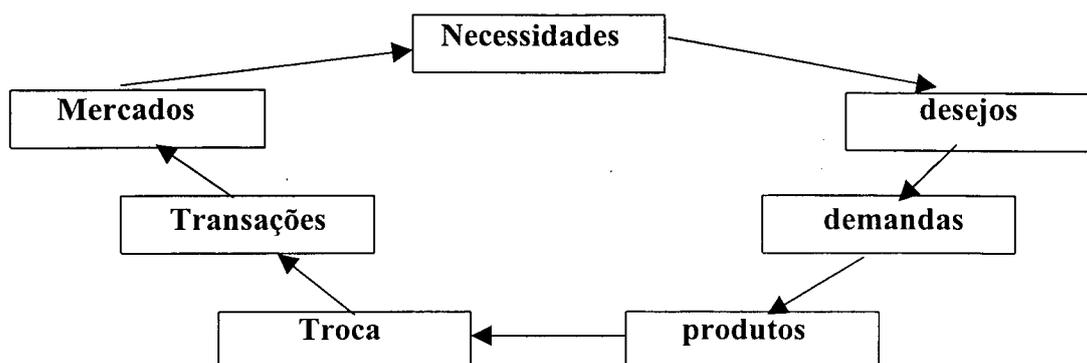


Figura 12 – Conceitos centrais de *marketing*
 Fonte: Kotler (2000)

O conceito de transações leva ao conceito de mercado. Um mercado é o grupo de compradores reais e potenciais de um produto. Desta forma, segundo Kotler (2000), *marketing* significa trabalhar com mercados para conseguir trocas com o propósito de satisfazer necessidades e desejos humanos.

A área de *marketing* da DR/SC procura implementar programas de *marketing* eficientes que proporcionem à empresa uma grande vantagem competitiva em seus mercados-alvo. Na busca de análise e planejamento de *marketing*, a área analisa mercados e o ambiente em que a ECT está inserida com o objetivo de identificar oportunidades atrativas.

Durante todo o processo de *marketing*, comenta Kotler (2000), os profissionais de *marketing* precisam de um vasto suprimento de informações para analisar as oportunidades de mercado. Precisam saber acerca de forças ambientais maiores que afetam a empresa e seus consumidores – fatores demográficos, econômicos, naturais, tecnológicos, políticos e culturais. O SEPE concentra-se unicamente no ambiente dos consumidores, não identificando os demais públicos do ambiente de *marketing* – concorrentes, fornecedores e revendedores.

Segundo Friedman (1999), um grupo de consumidores que respondem de forma similar a um dado estímulo do mercado denomina-se segmento de mercado. Utilizam-se, nesta dissertação, sete segmentos de mercado, conforme os agrupamentos de família de serviços utilizados pela ECT (fig. 14). Após a avaliação dos segmentos, a empresa pode entrar em um ou mais segmentos de um dado mercado.

Um segmento é constituído por fatores que afetam o poder de compra e padrões de gasto do consumidor. Para Kotler (2000), mercados exigem poder de compra tanto quanto exigem pessoas. O poder de compra total depende de renda corrente, preços, poupanças em crédito. Mudanças nas variáveis econômicas mais importantes, tais como rendas, custo de

vida, taxas de juros, poupança e empréstimos, têm grande impacto no mercado. O SEPE procura tirar vantagem do ambiente econômico, buscando variáveis econômicas dos municípios (fig. 15) e oferecendo um método de classificação do potencial econômico para a empresa.

Os administradores da área de *marketing* da DR/SC têm como uma de suas responsabilidades mensurar o tamanho atual e futuro do mercado. Desta forma, a seguir, serão tratados alguns conceitos sobre mercado

3.3.2 Mercado

Cada empresa deve ser capaz de identificar novas oportunidades de mercado. O ambiente, complexo e em constante mudança, sempre oferece novas oportunidades e ameaças. A empresa deve analisar cuidadosamente seus consumidores e o ambiente, de forma a evitar as ameaças e tirar vantagens das oportunidades.

Quando uma empresa encontra um mercado atrativo, ela deve avaliar cuidadosamente o tamanho desse mercado e seu potencial futuro. A empresa pode perder consideravelmente seus lucros se superestimar ou subestimar o mercado.

Segundo Kotler (2000), para um profissional de *marketing*, o mercado é um grupo de compradores reais e potenciais de um produto. Um mercado é o grupo de compradores e uma indústria é o grupo de vendedores. Portanto, o tamanho de um mercado é definido pelo número de compradores que possam existir para uma oferta específica. Aqueles que estão no mercado têm três características interesse, renda e acesso.

A mensuração da demanda do mercado exige uma compreensão clara do mercado envolvido. O mercado potencial é o grupo de consumidores que demonstram algum nível de interesse por um produto ou serviço em particular. O mercado consumidor é constituído por todos os indivíduos e famílias que compram ou adquirem produtos e serviços para consumo pessoal.

Apenas o interesse do consumidor não é suficiente para definir o mercado. Os consumidores em potencial devem ter um renda que lhes permita adquirir o produto. Assim, o tamanho do mercado depende tanto do interesse quanto da renda.

A empresa deve escolher entre atender a todo o mercado qualificado disponível ou concentrar-se em segmentos selecionados, comenta Kotler (2000). O mercado atendido (também chamado mercado-alvo) é a parte do mercado qualificado disponível que a empresa

decide atender. Esta pode, por exemplo, decidir concentrar seus esforços de *marketing* e distribuição em uma certa região, que então se torna o seu mercado-alvo.

A figura 12 reúne os conceitos de mercado com alguns números hipotéticos. A escala à direita mostra a razão entre o mercado potencial – todas as pessoas interessadas – e a população total. Aqui, o mercado potencial é 10%. A escala à esquerda mostra vários desdobramentos do mercado potencial. O mercado disponível – aqueles que têm interesse, renda e acesso – constitui 40% do mercado potencial. O mercado qualificado disponível – aqueles que satisfazem os requisitos legais – constitui 20% do mercado potencial (ou 50% do mercado qualificado disponível). Finalmente, a empresa e seus concorrentes já penetraram em 5% do mercado potencial (ou 50% do mercado-alvo).

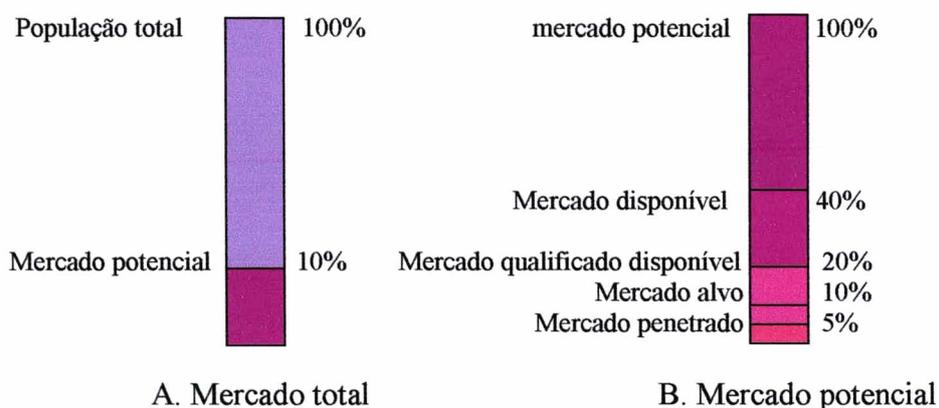


Figura 12: Mercado potencial
Fonte: Kotler (2000)

Estas definições de um mercado são uma ferramenta útil para o planejamento de *marketing* na DR/SC e desenvolvimento do SEPE.

Segundo Kotler (2000), se a empresa não se sente satisfeita com suas vendas atuais, ela pode considerar várias ações. Pode tentar atrair uma porcentagem maior de compradores para seu mercado-alvo. Pode expandir-se para outros mercados disponíveis. Pode reduzir seus preços para expandir o tamanho do mercado disponível. Ou pode tentar expandir o mercado potencial aumentando sua propaganda, para converter consumidores, desinteressados em consumidores interessados.

3.3.3 Método para estimar o potencial de mercado

Existem dois métodos mais conhecidos para estimar o potencial de mercado: o método de desenvolvimento de mercado - utilizado por firmas de produtos industriais; e o método de utilização de múltiplos fatores, empregado principalmente por firmas de bens de consumo. Apresenta-se aqui o segundo, por se adequar melhor aos Correios.

Segundo a revista *Sales and Marketing Management* (2000), um método comum para o cálculo do potencial de mercado de uma área é identificar os fatores de mercado que se correlacionam com o potencial e combiná-los em um índice de pesos. O índice é publicado anualmente e baseia-se em três fatores: a participação da área na renda pessoal disponível, vendas a varejo e população. O índice de poder aquisitivo (BPI) para uma área específica é obtido pela seguinte fórmula:

$$BPI_i = 0,5r_i + 0,3 v_i + 0,2p_i$$

onde

BPI_i = percentagem do poder total de compra na área i.

r_i = percentagem da renda pessoal disponível na área i.

v_i = percentagem das vendas a varejo na área i.

p_i = percentagem da população na área i.

Os três coeficientes na fórmula refletem os pesos relativos dos três fatores. Os pesos utilizados no índice do poder de compra são arbitrários. Pode-se utilizar outros pesos. Pode-se também ajustar o índice de potencial do mercado a outros fatores adicionais, tais como o nível de concorrência no mercado, custos promocionais locais, fatores sazonais na demanda e características próprias do mercado.

Este modelo, encontrado na literatura, permite calcular o índice de potencial econômico para determinada área, comparando-o com outras áreas.

3.4 Considerações finais

Diante de situações como: insuficiente instrumentalização da força de vendas, desconhecimento do posicionamento da empresa no mercado por segmento de negócio e sistema de informações deficiente, os Correios de Santa Catarina procuram se sobressair num mercado cada vez mais exigente e competitivo. Neste sentido, a análise do potencial

econômico é uma forma de procurar conhecer o mercado em que os Correios em Santa Catarina se inserem.

O objetivo do próximo capítulo é apresentar alguns sistemas que possuam situações semelhantes a esta dissertação e que, assim, contribuam com a pesquisa sobre o problema do potencial econômico.

4 EXEMPLOS DE SISTEMAS CORRELATOS

O objetivo deste capítulo é relacionar alguns sistemas e utilitários que apresentem associação com o tema proposto para esta dissertação. Assim, são apresentados alguns exemplos de Sistemas Especialistas Probabilísticos e de técnicas onde são empregadas o agrupamento de mineração de dados.

4.1 Sistemas Especialistas Probabilísticos (SEP)

Nesta primeira parte são mostrados alguns sistemas desenvolvidos nas áreas de aplicação de medicina e engenharia e que utilizam a abordagem *bayesiana*.

4.1.1 *Baseline Bayesiana para Gerência de Redes de Computadores*

Este sistema é um módulo inteligente que utiliza informações de um MIB (*management information base*) para avaliar dinamicamente o tráfego em certo trecho de uma rede de computadores. Oferece como saída as probabilidades das hipóteses diagnósticas do tráfego na rede estar: normal, leve, moderado e intenso a partir dos dados iniciais de hora e dia fornecidos pelo administrador e as informações buscadas na MIB. Foi desenvolvido utilizando a *shell* Netica (VERONEZ, 2000).

4.1.2 *DIIVAL*

É um sistema especialista destinado a ajudar médicos no diagnóstico de enfermidades cardíacas a partir, principalmente, de dados ecocardiográficos. Este sistema faz a introdução dos dados, apresentação dos resultados do diagnóstico e explicações sobre o raciocínio. Ele calcula a probabilidade a posteriori a partir dos sintomas informados e seleciona a doença mais provável, gerando um relatório impresso (DÍEZ, 1997).

4.1.3 Sistema de Auxílio ao Diagnóstico Diferencial de Cefaléia

É um sistema especialista médico que oferece apoio no diagnóstico diferencial de cefaléias (dores de cabeça). Este sistema utiliza o enfoque de rede *bayesiana*. A base de conhecimentos foi construída considerando os Critérios de Classificação da Sociedade Internacional de Cefaléias (IHS), levando em conta os sinais e sintomas dos pacientes e os valores estimados de probabilidades fornecidos pelos especialistas que participaram do projeto. Foi desenvolvido utilizando a *shell* Netica. O sistema desenvolvido no PPGCC da UFPR apresenta um desempenho muito bom ao realizar diagnóstico diferencial de cefaléias (TOLEDO, 2000).

4.1.4 SISPAN

Sistema com domínio na área médica. Foi desenvolvido para auxiliar na avaliação do estado nutricional em crianças de até dois anos de idade. Considera como entradas os dados relativos a: peso, altura, idade, cabelo, pelo, unha, hipotrofia e edema. Classifica o estado nutricional da criança em: sem desnutrição ou com desnutrição, podendo ser leve, moderada ou severa quanto à intensidade, e marasmo, kwashikor ou marasmo-kwashikor quanto ao tipo. Foi desenvolvido no PPGCC da UFSC utilizando a *shell* Netica, com o módulo de interface com o usuário desenvolvido em C++ (KOËHLER, 1998).

4.1.5 SEP de Base de Conhecimento Dinâmica

Foi concebido a partir do sistema especialista probabilístico SISPAN, cuja base de conhecimento é estática. Ambos os sistemas oferecem um diagnóstico para avaliar o estado nutricional em crianças com até dois anos de idade. O sistema SEDIN pode ser usado para verificar o grau de importância das variáveis introduzidas no sistema, bem como oportunizar a aquisição continuada do conhecimento. Foi desenvolvido no PPGCC da UFSC utilizando a *shell* Netica, com o módulo de interface com o usuário desenvolvido em C++ (EFRAIN, 2000).

4.1.6 SACI

Sistema de Apoio na Avaliação de Distúrbios do Crescimento Infantil, é voltado ao diagnóstico de alguns distúrbios nutricionais associados à falência do crescimento em crianças com até dois anos, evidenciados pela Deficiência de Ferro, Síndrome de Má Absorção Alimentar e Etiologias a Esclarecer. Explora o raciocínio probabilístico em sistemas especialistas. Foi desenvolvido no PPGCC da UFSC para a plataforma DOS e Windows, no ambiente Delphi 5, utilizando as DLL's da *shell* Netica. A base de conhecimento é composta por 4 nós, sendo um referente às hipóteses diagnósticas e os demais relacionados aos sinais e sintomas apresentados. O SACI permite também acompanhar por uma análise gráfica a evolução do crescimento infantil, auxiliando dessa forma o diagnóstico e acompanhamento do crescimento infantil (SIMÕES, 2001).

4.1.7 Uma Metodologia de Uso de Técnicas de Indução para Criação de Regras de Sistemas Especialistas

Sistema especialista para auxílio dos tomadores de decisão empresarial. Utiliza o algoritmo de mineração de dados ID3 para desenvolver o sistema GARP, que proporciona a geração automática de regras probabilísticas podendo ser usada em qualquer *Shell* de sistemas especialistas baseadas em regras (OLIVEIRA, 2001).

4.2 Técnica de Agrupamento de Mineração de Dados

Tendo em vista a existência de uma etapa de aquisição de conhecimento por meio da técnica de agrupamento de mineração de dados, aqui estão apresentados exemplos de um utilitário e de um sistema para reconhecimento de padrões em base de dados. Como o tipo de agrupamento empregado nesta dissertação é o probabilístico, deu-se preferência na pesquisa à busca de exemplos que utilizassem a mesma técnica.

4.2.1 WEKA

O *Waikato Environment for Knowledge Analysis* (WEKA) foi desenvolvido na Universidade de Waikato na Nova Zelândia. Ele é o resultado de um projeto de pesquisa em Aprendizado de Máquina feito no Departamento de Ciência da Computação. WEKA é um utilitário para análise de conhecimento. Inclui uma variedade de ferramentas para transformação de dados, como algoritmos de discretização. Utiliza algoritmos para as três técnicas mais conhecidas de mineração de dados: classificação, agrupamento e regras de associação (WITTEN, 2000).

4.2.2 AUTOCLASS

O *Automatic Class Discovery from data* (AUTOCLASS) é um sistema de classificação com aprendizagem não supervisionada, suplementado pelo método *bayesiano*, para determinação de classes otimizadas. Este sistema busca uma máxima classificação probabilística a posteriori. Ele trabalha tanto com atributos nominais quanto numéricos e usa o algoritmo EM (*expectation-maximization*) para estimar os parâmetros da distribuição de probabilidade para ajustar melhor os dados. Como não existem garantias que o algoritmo EM convirja para o ótimo global, o algoritmo é repetido para diversos e diferentes conjuntos de valores iniciais (CHEESEMAN, 2001).

4.3 Considerações Finais

Este capítulo apresentou o resultado de uma pesquisa sobre SEP, agrupamentos e métodos de estimativa de demanda de mercado.

Dos SEP vistos pode-se constatar a importância de sua utilização no apoio à resolução de problemas em domínios com a presença de incerteza e com a participação de especialistas. Com relação às técnicas de agrupamento com emprego da probabilidade, destaca-se a sua importância como técnica de agrupamento com aprendizagem não supervisionada. Sobre o método de estimativa de demanda, ele é uma importante ferramenta dos profissionais de *marketing*.

No próximo capítulo será apresentado o desenvolvimento do sistema SEPE.

5 O SISTEMA

O sistema SEPE, objeto desta pesquisa, é uma aplicação voltada para o apoio ao diagnóstico do Potencial Econômico na área de *marketing* dos Correios de Santa Catarina, onde a base de conhecimentos é atualizável a partir de informações armazenadas em bancos de dados.

A seguir são apresentados detalhes sobre o seu desenvolvimento.

5.1 Ambiente

O sistema especialista SEPE pode ser utilizado em microcomputadores sob ambiente Windows, recomendando-se a utilização de um Pentium, com no mínimo 32 MB de RAM para a utilização das bases de dados.

A implementação foi feita em *Visual Basic*, ambiente de desenvolvimento utilizado na Diretoria Regional de Santa Catarina.

5.2 Dados

As fontes dos dados do sistema SEPE são bancos de dados operacionais da ECT e dados econômicos de Santa Catarina. Com relação aos Correios utiliza-se o Banco de Dados Financeiro (BDF) e os dados do Sistema de Gerenciamento de Contratos (SGC).

O BDF é um banco de dados da área financeira que agrupa dados de faturamento de todas as unidades da ECT. Diariamente as Agências dos Correios enviam os dados de sua movimentação financeira para um banco de dados central, onde ficam armazenados. Assim, o BDF contém dados de arrecadação diária de venda de produtos e serviços de todas as unidades dos Correios.

O SGC é um sistema que recebe dados oriundos das áreas de *marketing* relativos ao faturamento de grandes clientes. Os dados são concentrados em um banco de dados em Brasília.

Para que a manipulação dos dados não interfira no desempenho dos bancos de dados operacionais, os dados do BDF e SGC são filtrados e armazenados em um banco de dados relacional (fig. 13)

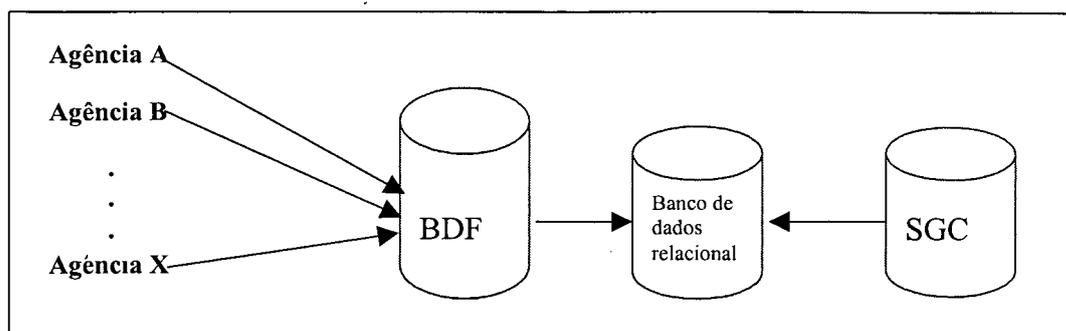


Figura 13: Formação do banco de dados relacional.

Um banco de dados relacional é composto de uma série de tabelas. Cada tabela é armazenada como um arquivo. Uma tabela relacional consiste de linhas e colunas.

No final deste processo têm-se uma tabela com o faturamento do período solicitado de todos os clientes de varejo e atacado da DR/SC por segmento de família (nome dado ao agrupamento de produtos e serviços da ECT), conforme figura 14.

Nome	Descrição
Município	Nome do município
Família Mensagem	Agrupamento de serviços para esta família.
Família Encomenda	Agrupamento de serviços para esta família.
Família Expresso	Agrupamento de serviços para esta família.
Família Internacional	Agrupamento de serviços para esta família.
Família <i>Marketing</i> Direto	Agrupamento de serviços para esta família.
Família Financeiro	Agrupamento de serviços para esta família.

Figura 14: Variáveis do banco de dados de família de serviços da ECT.

Os dados econômicos de Santa Catarina são retirados de diversas fontes. Eles são agrupados em um banco de dados que já contém algumas informações sobre os municípios catarinenses. A figura 15 apresenta algumas variáveis que foram selecionadas para este protótipo.

Nome	Descrição
Município	Nome do município
ICMS	Imposto sobre Circulação de Mercadorias e Serviços
Energia Elétrica Industrial	Consumo industrial de energia elétrica dos municípios

Nome	Descrição
Energia Elétrica Comercial	Consumo comercial de energia elétrica dos municípios
Energia Elétrica Residencial	Consumo residencial de energia elétrica dos municípios
Energia Elétrica do Poder Público	Consumo de energia elétrica dos órgãos públicos dos municípios.
Telefones Comerciais	Número de terminais telefônicos comerciais dos municípios.
Telefones Residenciais	Número de terminais telefônicos residenciais dos municípios.
População	População por município
PIB	Produto Interno Bruto do município.

Figura 15: Variáveis do banco de dados econômico dos municípios de SC.

5.3 Etapas de Desenvolvimento do SEPE

O desenvolvimento do sistema foi feito em cinco etapas: aquisição do conhecimento, representação do conhecimento, implementação do sistema especialista probabilístico, desenvolvimento da interface com o usuário e avaliação do sistema.

5.3.1 Aquisição do conhecimento

O conhecimento a ser adquirido para o desenvolvimento do sistema referiu-se ao diagnóstico do potencial econômico dos municípios de Santa Catarina.

Essa etapa do desenvolvimento foi realizada de duas formas: através de entrevistas com o especialista do domínio da aplicação e de forma automática – sem a presença do especialista. Neste último caso, a partir das bases de dados, foram realizados cruzamentos entre as variáveis econômicas dos municípios e as variáveis de faturamento de produtos e serviços dos Correios (Fig. 14)

Figura 16: Aquisição automática do conhecimento.
Fonte: Shell NETICA

A figura 16 apresenta o resultado do cruzamento entre a variável econômica PIB e as variáveis de família de serviços MENSAGEM e EXPRESSO. As probabilidades apresentadas foram calculadas pelo sistema de forma automática. Assim, por exemplo, a porcentagem 82.857 representa a probabilidade de ocorrência do estado G1 de PIB quando ocorrem os estados G1 de MENSAGEM e G1 de EXPRESSO.

Nessa etapa foram definidas as hipóteses diagnósticas relevantes a esse domínio de estudo, bem como as evidências que são necessárias para se chegar ao diagnóstico do potencial econômico. As hipóteses diagnósticas definidas foram: Alto Potencial Econômico, Médio Potencial Econômico e Baixo Potencial Econômico.

Em seguida, o especialista estimou as probabilidades *a priori* do sistema. Por tratar-se de estudo com pouco material disponível para pesquisa, o especialista preferiu manter-se cauteloso, adotando a máxima entropia para as probabilidades (fig. 17).

Hipóteses Diagnósticas	P(H _i)
Alto Potencial Econômico	0,333
Médio Potencial Econômico	0,333
Baixo Potencial Econômico	0,333

Figura 17: Probabilidades a priori do SEPE

5.3.2 Representação do conhecimento

Uma vez definidas as hipóteses diagnósticas, evidências e probabilidades, foi realizada a representação do conhecimento por meio de uma rede *bayesiana*, construída na *shell Netica*, representada pela fig. 18, composta de um nó para representar as hipóteses diagnósticas (Alto, Médio e Baixo), e outros seis nós para representar as evidências necessárias para chegar às hipóteses diagnósticas (Mensagem, Expresso, Energia Elétrica Comercial, PIB, ICMS e População).

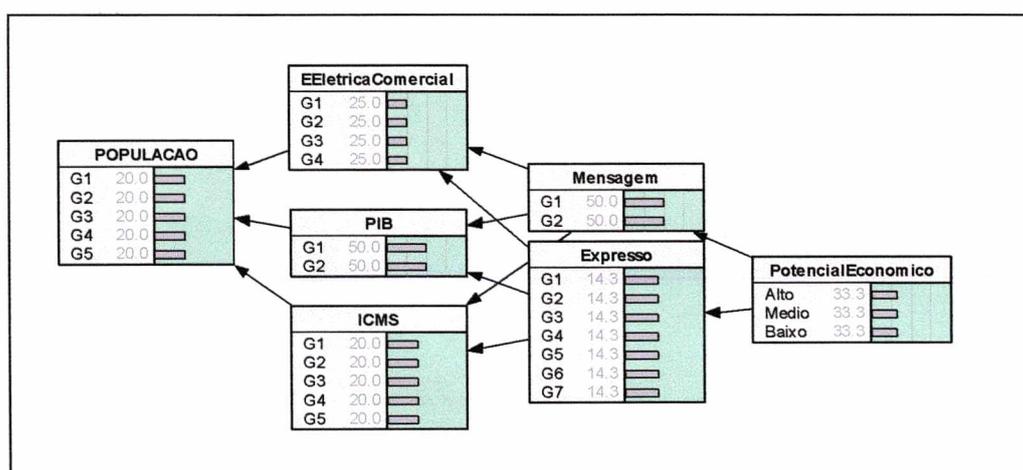


Figura 18: Rede *Bayesiana* do sistema SEPE.

Fonte: *Shell NETICA*

5.3.3 Implementação do SEPE

O primeiro passo para a implementação do sistema SEPE foi a escolha de um ambiente de desenvolvimento adequado e que fosse utilizado pelos Correios. Escolheu-se o ambiente *Visual Basic 5.0* que atende aos requisitos necessários para a construção deste sistema.

A implementação foi iniciada realizando-se a comunicação do ambiente de desenvolvimento *Visual Basic 5.0* com a base de conhecimento representada na *shell Netica*, por meio de funções disponíveis em DLL para o desenvolvimento nesse ambiente.

5.3.3.1 Funções API disponíveis pela *shell Netica*

As funções API disponíveis na *shell Netica* podem ser utilizadas tanto no ambiente de programação C, quanto em *Visual Basic*. O fabricante da *shell Netica*, Norsys Inc.,

disponibiliza em sua *home page* (<http://www.norsys.com>) diversas bibliotecas disponíveis para o desenvolvimento de sistemas integrados ao Netica em diferentes plataformas e ambientes de desenvolvimento.

A *shell* Netica foi desenvolvida em C, o que permite uma maior facilidade de integração da ferramenta com esse ambiente. Escolheu-se utilizar o ambiente *Visual Basic* por atender aos requisitos para a construção do sistema SEPE, além de oferecer uma contribuição a futuras pesquisas de sistemas especialistas probabilísticos que utilizarão a *shell* Netica com o ambiente *Visual Basic*. Isso se deve ao fato de existir poucas aplicações deste tipo, bem como carência de documentação na área. Outro motivo pelo qual optou-se utilizar o *Visual Basic* é o fato de tratar-se de um ambiente de desenvolvimento utilizado pelos Correios.

O anexo 1 apresenta uma parte do algoritmo utilizado para a integração entre o sistema SEPE e a *shell* NETICA. O algoritmo foi adaptado para facilitar a sua compreensão.

5.3.4 Desenvolvimento da Interface com o Usuário

A seguir são ilustradas as principais telas e algumas funções do SEPE. Ao inicializar o SEPE aparece uma estrutura de menus, conforme figura 19

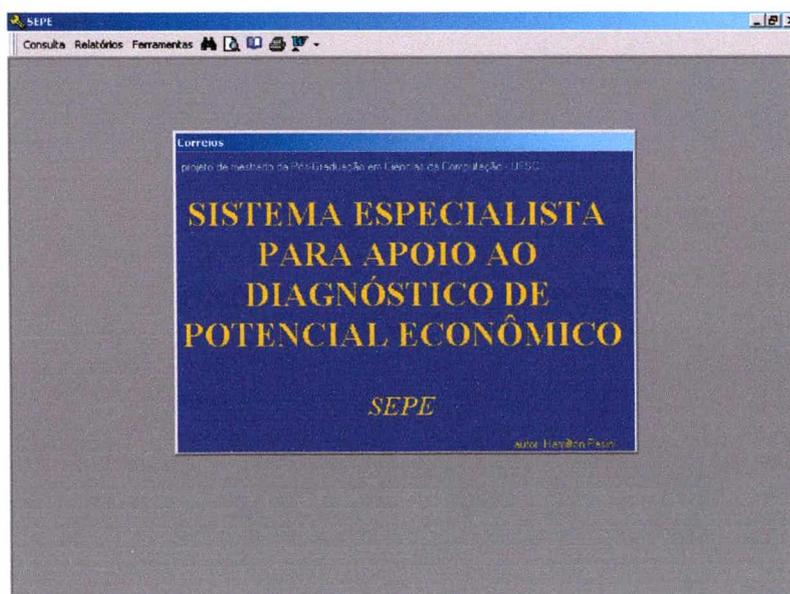


Figura 19: Módulo principal do SEPE.

O menu Consulta permite a realização de consultas das variáveis internas, variáveis externas e dos agrupamentos. O menu Relatórios oferece opções para visualização de relatórios analíticos e gráficos gerenciais.

No menu ferramentas pode-se acessar os seguintes módulos: Agrupamentos, Coeficiente de Correlação e Rede *Bayesiana*.

5.4 Módulo Mineração de Dados

O primeiro passo consiste na purificação dos dados. Este processo envolve remoção de registros errados, padronização dos campos, racionalização da estrutura de dados e manuseio de dados perdidos.

Os dados são enviados posteriormente para uma *data warehouse*. Uma *data warehouse* é um recurso centralizado usado para suporte a decisão e processo analítico. Os dados são transformados em uma forma padrão para serem acessados pelos algoritmos de mineração de dados.

	Mensagem	MixDireto	Encomenda	Expresso	Financeiro	Internacional	Conveniência
ICMS	0,658	0,374	0,432	0,610	0,665	0,410	0,484
PIS	0,764	0,771	0,612	0,945	0,802	0,801	0,802
EEletricaResidencial	0,717	0,721	0,555	0,538	0,763	0,763	0,794
EEletricaIndustrial	0,562	0,578	0,597	0,777	0,780	0,613	0,686
EEletricaComercial	0,788	0,796	0,586	0,984	0,793	0,834	0,863
EEletricaRural	-0,030	-0,028	0,080	0,062	0,216	-0,009	0,107
EEletricaPoderPublico	0,813	0,815	0,581	0,883	0,800	0,844	0,869
TelefonesComerciais	0,739	0,749	0,565	0,824	0,759	0,780	0,791
TelefonesResidenciais	0,761	0,768	0,530	0,805	0,709	0,789	0,760
Populacao	0,573	0,582	0,544	0,785	0,788	0,635	0,719

Figura 20: Módulo Coeficiente de Correlação do SEPE.

Ainda falta agrupar as informações para poder concebê-los em um sistema especialista. Para esta etapa foi utilizado o método de agrupamento probabilístico.

Trata-se de um sistema de classificação com uma abordagem não supervisionada, suplementado pelo método *bayesiano*, para determinação de classes otimizadas. Este sistema busca uma máxima classificação probabilística a posteriori. Ele Trabalha tanto com atributos nominais quanto numéricos e usa o algoritmo EM (*expectation-maximization*) para estimar os

parâmetros da distribuição de probabilidade para ajustar melhor os dados. Como não existem garantias que o algoritmo EM convirja para o ótimo global, o algoritmo é repetido para diversos e diferentes conjuntos de valores iniciais.

A figura 21 apresenta o resultado dos agrupamentos das variáveis externas do município de Lages. A coluna Agrupamento aponta o grupo em que o Valor foi alocado e a coluna Grupos indica a quantidade de grupos formados para cada variável.

The screenshot shows a software window titled 'SEPE' with a menu bar containing 'Consulta', 'Relatórios', and 'Ferramentas'. The main window is titled 'Consulta de Dados' and displays a table of data for the municipality of Lages. The 'MUNICIPIO' field is set to 'LAGES'. The table lists various variables with their values, assigned groups, and the number of groups formed for each.

	Valor	Agrupamento	Grupos
ICMS	0,16802	4	4
PIB	17	3	3
EEletricaResidencial	39834	3	5
EEletricaIndustrial	832	3	4
EEletricaComercial	3349	3	5
EEletricaRural	1123	3	3
EEletricaPoderPublico	313	4	4
TelefonesComerciais	5248	3	5
TelefonesResidenciais	14235	1	6
Populacao	140004	4	5

At the bottom of the window, it indicates 'Registro: 105 de 221'.

Figura 21: Módulo Consulta de Variáveis Externas do SEPE.

A figura 22 apresenta o resultado dos agrupamentos das variáveis internas do município de Lages.

Consulta de Dados

MUNICIPIO: LACES

	Valor	Assessment	Grupos
Mensagem	104839,23	2	2
MktDireto	85248,68	3	4
Encomenda	9000,76	6	6
Expresso	101795,62	6	7
Financeiro	25895,5	7	7
Internacional	13153,42	2	4
Conveniencia	24852,02	3	4

Registro: 105 de 221

SEPE

Figura 22: Módulo Consulta de Variáveis Internas do SEPE.

5.5 Módulo Coeficiente de Correlação

Neste momento já podem ser realizadas análises estatísticas. Elas servem como ferramenta de apoio desde o início dos trabalhos até a obtenção dos resultados. São de fundamental importância para que uma mineração de dados possa ser bem conduzida.

No sistema SEPE utiliza-se o Coeficiente de Correlação como ferramenta de análise estatística. A correlação mede a intensidade da associação linear entre as variáveis econômicas de cada município com o faturamento de serviços/produtos oferecidos pela ECT. A figura 20 apresenta o módulo Coeficiente de Correlação do SEPE.

5.6 Módulo Rede Bayesiana

Uma vez definidas as hipóteses diagnósticas, evidências e probabilidades, é realizada a representação do conhecimento por meio de uma rede bayesiana. Para esta pesquisa a rede bayesiana é construída na shell Netica. No protótipo atual a rede é composta de um nó para representar as hipóteses diagnósticas (classificação de 1 a 3), e outros cinco nós para representar as evidências necessárias para chegar às hipóteses diagnósticas. Na figura 23 estão representados seis nós da modelagem do protótipo.

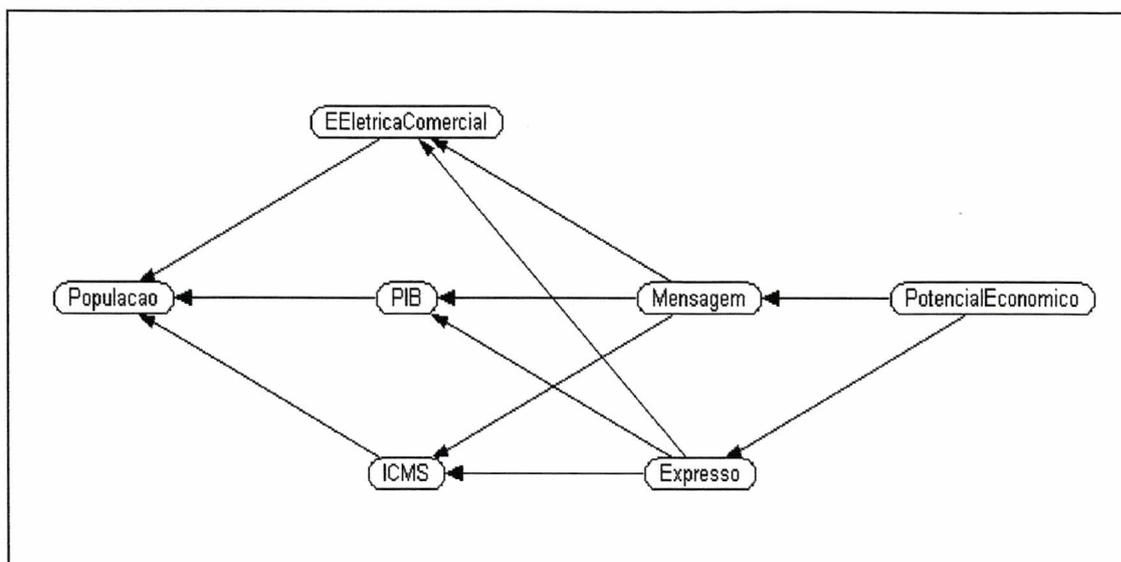


Figura 23: Representação do protótipo da rede *bayesiana* na *shell* NETICA.
Fonte: *Shell* NETICA

5.7 Etapa de aprendizagem

Nesta pesquisa utiliza-se o ambiente de trabalho do Netica API - uma biblioteca que contém funções para trabalhos com redes *bayesianas* - para representar a base de conhecimento do SEPE. Através do Netica API é criada a rede *bayesiana* com seus nós e arcos. Nela são introduzidas as probabilidades a priori das hipóteses $P(H_i)$ e as probabilidades das evidências dado as hipóteses diagnósticas ($P(e/H_i)$).

Depois de sua definição, a rede *bayesiana* é salva em um arquivo *.dne*. Para acessar esta rede pré-gravada utiliza-se a linguagem *Visual Basic 5.0* juntamente com a DLL do Netica.

Assim, a rede *bayesiana* com os seus nós, arcos e probabilidades a priori e condicionais é introduzida na própria aplicação.

No primeiro momento a rede não possui qualquer conhecimento. Assim, ela é apresentada em sua forma de máxima entropia. A figura 24 mostra como a rede se configura nesta situação.

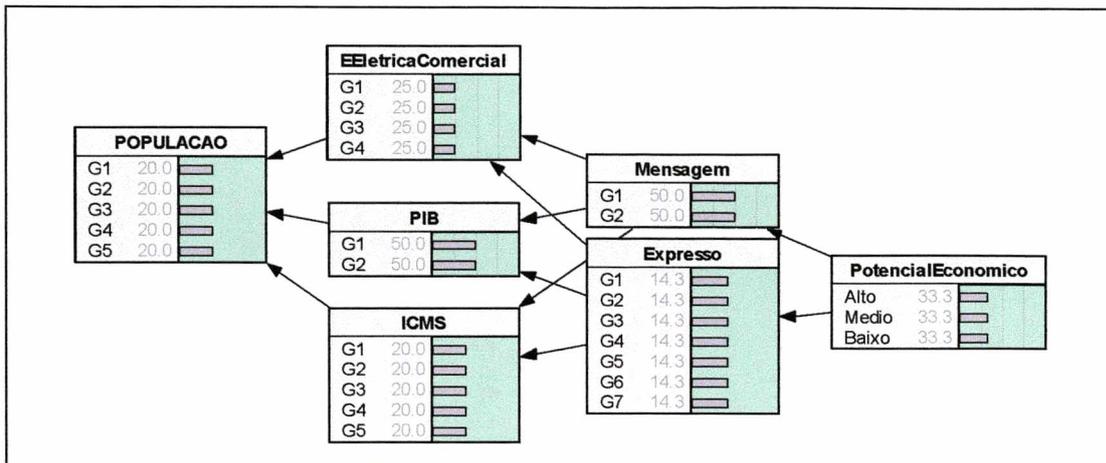


Figura 24: Rede Bayesiana com máxima entropia.
Fonte: Shell NETICA

Em seguida, a rede adquire conhecimento, ou seja, ela aprende. A figura 25 mostra como a rede se configura nesta situação.

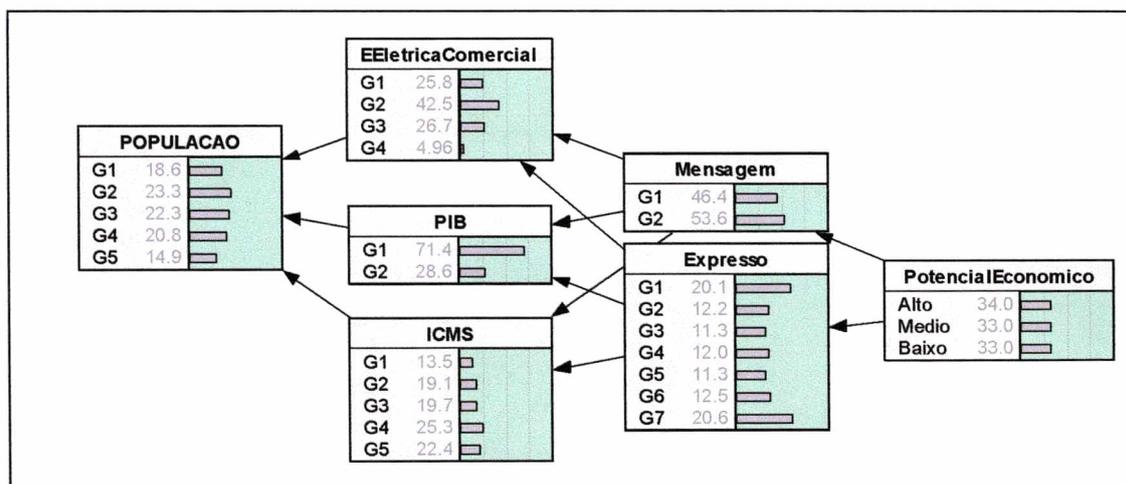


Figura 25: Rede Bayesiana após o aprendizado.
Fonte: Shell NETICA

O SEPE apresenta também o módulo de ajuda que prevê auxiliar na realização das funções computacionais do sistema, como também fornece informações sobre o domínio de aplicação (fig. 26)

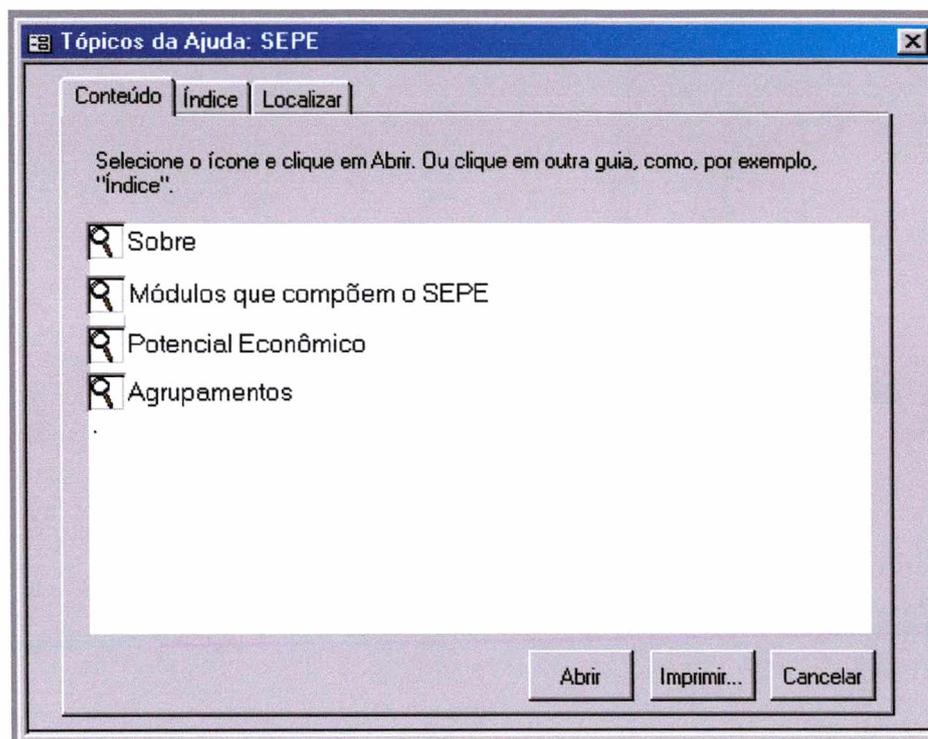


Figura 26: Tela de Ajuda do SEPE.

5.8 Avaliação do Sistema

O sistema SEPE teve os valores de probabilidade estimados diretamente de bases de dados da ECT, no ano de 2001. A avaliação do sistema foi realizada pelo especialista dos Correios, gerente da área de *marketing*, numa abordagem qualitativa.

A seguir estão relacionadas as observações feitas pelo especialista:

- A base de conhecimento do SEPE reflete o domínio de aplicação proposto;
- As categorias encontradas pelos algoritmos do SEPE foram consideradas adequadas;
- A interface é amigável;
- A ajuda do SEPE facilita a utilização do sistema, e fornece informações relevantes ao diagnóstico de distúrbios associados à falência do crescimento.

5.9 Consulta ao Sistema

A seguir é demonstrada a aplicação do teorema de Bayes para propagar as evidências apresentadas, a fim de fornecer o vetor de probabilidades das três hipóteses diagnósticas, ou seja, alto potencial econômico, médio potencial econômico e baixo potencial econômico.

Os valores das probabilidades a priori utilizados para a aplicação do Teorema de Bayes são descritos na figura 15 os cálculos das probabilidades das evidências são demonstrados a seguir:

A probabilidade do estado da família mensagem ser G1 é:

$$\begin{aligned}
 P(\text{Mensagem}_{G1}) &= P(\text{Alto} \cap \text{Mensagem}_{G1}) \\
 &\quad + P(\text{Médico} \cap \text{Mensagem}_{G1}) \\
 &\quad + P(\text{Baixo} \cap \text{Mensagem}_{G1}) \\
 &= P(\text{Alto}) \times P(\text{Mensagem}_{G1} / \text{Alto}) \\
 &\quad + P(\text{Médico}) \times P(\text{Mensagem}_{G1} / \text{Médico}) \\
 &\quad + P(\text{Baixo}) \times P(\text{Mensagem}_{G1} / \text{Baixo}) \\
 &= 0,33 \times 0,20 \\
 &\quad + 0,33 \times 0,40 \\
 &\quad + 0,33 \times 0,80 \\
 P(\text{Mensagem}_{G1}) &= 0,462 \text{ (46,2\%)}
 \end{aligned}$$

As demais probabilidades para cada uma das evidências são obtidas da mesma maneira. A probabilidade do estado da família Mensagem ser G2 é a que segue.

$$P(\text{Mensagem}_{G2}) = 0,538 \text{ (53,8\%)}$$

Uma vez calculadas as probabilidades para as evidências, é calculada a propagação da evidência Mensagem para cada uma das hipóteses diagnósticas.

$$P(\text{Alto} / \text{Mensagem}_{G1}) = P(\text{Alto}) \times P(\text{Mensagem}_{G1} / \text{Alto}) / P(\text{Mensagem}_{G1})$$

$$P(\text{Alto} / \text{Mensagem}_{G1}) = 0,33 \times 0,20 / 0,462$$

$$P(\text{Alto} / \text{Mensagem}_{G1}) = 0,147 \text{ (14,7 \%)}$$

As probabilidades obtidas pela propagação da evidência Mensagem para cada uma das demais hipóteses diagnósticas são obtidas da mesma maneira, resultando nas probabilidades descritas a seguir:

$$P(\text{Média/Mensagem}_{G1}) = 0,284 \text{ (28,4 \%)}$$

$$P(\text{Baixa/Mensagem}_{G1}) = 0,569 \text{ (56,9 \%)}$$

Os cálculos do teorema de Bayes confirmam o diagnóstico fornecido pela inferência à base de conhecimento da *shell* Netica.

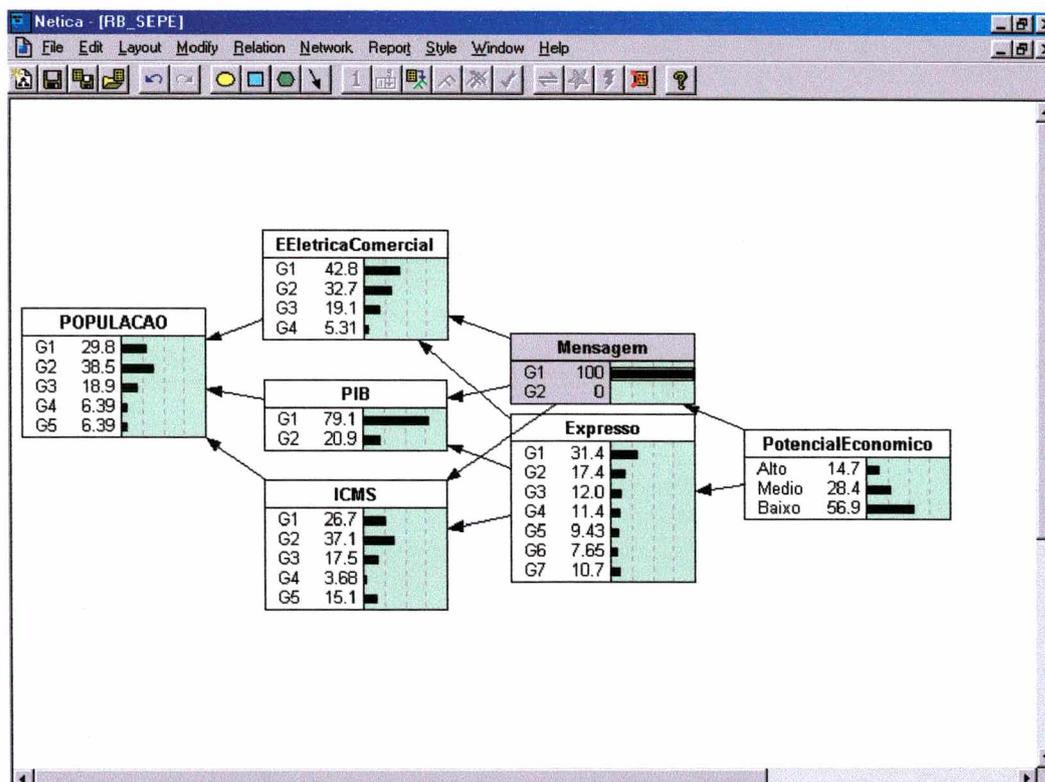


Figura 27: Diagnósticos fornecidos pela Rede Bayesiana.

Fonte: Shell NETICA

6 CONCLUSÃO

Nesta dissertação desenvolveu-se um sistema para apoio ao diagnóstico de potencial econômico, utilizando-se de técnicas de Inteligência Artificial, para auxílio às atividades de *marketing*, apoiando o processo de análise de mercado.

Esse sistema auxilia particularmente no planejamento e análise do mercado para a tomada de decisão dos profissionais de *marketing*. Os resultados mostram que essa tarefa complexa de extrair conhecimento de bases de dados pode ser feita automaticamente, de forma a obter bons resultados.

Em relação às técnicas de Inteligência Artificial utilizou-se Sistema Especialista, Tratamento de Incerteza por meio de modelo probabilístico, usando-se o Teorema de Bayes e as Redes *Bayesianas*.

A abordagem probabilística utilizada na representação do raciocínio do especialista de *marketing* e para compor a base de conhecimento, mostrou-se uma técnica bastante útil por reconhecer e tratar a aleatoriedade das evidências informadas, necessárias para o diagnóstico de potencial econômico.

O processo de raciocínio para o diagnóstico do potencial econômico foi obtido por meio da correlação das variáveis econômicas dos municípios de Santa Catarina com as variáveis de faturamento de produtos e serviços da DR/SC e de entrevistas com o especialista da área.

Levando-se em consideração as ferramentas disponíveis para implementação de sistemas computacionais para aplicações em domínios com a presença de incerteza, utilizou-se a *shell* Netica, já que esta ferramenta possui um interface amigável e de fácil utilização. Isto facilitou o processo de aquisição do conhecimento junto ao especialista da área. Desta forma, pode-se ajustar a base de conhecimento para atender às necessidades de avaliação do potencial econômico.

Nas etapas de desenvolvimento do SEPE, conseguiu-se atingir os objetivos previamente estabelecidos, implementando-se: a rede *bayesiana* para representação do conhecimento; o módulo de interface que se mostrou adequado. Após o desenvolvimento dos diferentes módulos, realizou-se a integração da rede *bayesiana* e da interface.

Para a integração da rede *bayesiana* com o módulo de interface desenvolvido na linguagem de programação *Visual Basic* foi necessário um estudo aprofundado para esse tipo de comunicação, devido ao fato de existir pouca documentação disponível. Concluiu-se com a aplicação das técnicas estudadas que este tipo de integração é viável atendendo em um tempo

satisfatório as consultas realizadas na base de conhecimento por meio da interface desenvolvida.

Foi feito um desenvolvimento de um sistema especialista com Base de Conhecimento Dinâmica, de forma que o usuário final do sistema possa adicionar e excluir nós, bem como definir as novas relações probabilísticas entre esses nós;

A avaliação do sistema junto ao especialista da área foi positiva. O especialista avaliou como satisfatórias tanto a base de conhecimentos quanto as categorias encontradas pelo algoritmo de agrupamento. Quanto à utilização do sistema, o especialista julgou que o SEPE está adequado à realização do diagnóstico de potencial econômico de municípios.

6.1 Trabalhos futuros

Com a conclusão desta dissertação, abrem-se novas possibilidades de pesquisas que podem ser feitas com a concepção de novas versões do SEPE. Algumas das possibilidades sugeridas seguem abaixo:

- Incluir nas bases de dados dos Correios a variável Potencial Econômico.
- Integrar novas bases de dados para aumentar o conhecimento do sistema SEPE;
- Continuar pesquisas em algoritmos de atualização das probabilidades da rede *bayesiana*;
- Incluir na base de conhecimento indicadores qualitativos, tais como: percepção, satisfação e comportamento de clientes;
- Utilizar conjuntos difusos para melhor avaliação dos resultados;
- Implementar o algoritmo próprio de cálculo de probabilidades, ficando o SEPE independente da *shell* Netica.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- BARRETO, J.M. **Inteligência artificial** – no limiar do século XXI. Florianópolis, [s.n.] 2001.
- CABENA, P.; P. Hadjinian; R. Stadler, J. Verhees, et al. **Discovering data mining: from concept to implementation**. Upper Saddle River, NJ: Prentice Hall, 1998.
- CHEESEMAN, P. **In defense of probability**. Proceedings of 9th International Joint Conference on Artificial Intelligence. Los Angeles:[s.n], 1985.
- CHEESEMAN, P.; STUTZ, J. **Bayesian classification (autoclass): teory and results**. Technical Report FIA-90-12-7-01, [S.I]:[s.n], 2001.
- COELLO, J.M.A. Tecnologias e Aplicações Inovadoras de Inteligência Artificial. In: **II Escola Regional de Informática** – SBC Regional de São Paulo. Piracicaba:[s.n], 1997.
- DIEZ, F. J. J. DIAVAL. **Sistema experto bayesiano para ecocardiografia**. Madrid: [s.n], 1997.
- EFRAIN, Carlos. **Sistema especialista probabilístico de base de conhecimento dinâmica**. 2000. Dissertação. (Curso de Pós Graduação em Ciência da Computação). Universidade Federal de Santa Catarina. Florianópolis, 2000.
- FRIEDMAN, Lawrence G.; FUREY, Timothy R. **The channel advantage: going to market with multiple sales channels**. Massachusetts: [s.n], 1999
- GAAG, Linda C. **Bayesian belief networks: Odds and Ends**. The Computer Journal:[s.n], vol 39, nº 2, 1996 , p 97-113.
- GAZETA MERCANTIL**. *Balanço anual 2000 - Suplemento Santa Catarina*. n. 07. São Paulo, 2000.
- JAMES, Barry R. **Probabilidade: um curso em nível intermediário**. Instituto de Matemática Pura e Aplicada. Rio de Janeiro: Hamburg Ltda, 1981.
- JOHNSON, Richard A., WICHERN, Dean W. **Applied multivariate statistical analysis**. New Jersey: Prentice-Hall., 1998.
- KANDEL, A. **Fuzzy expert system**. CRC PRESS. Flórida:[s.n], 1992.
- KOËHLER, C. **Uma abordagem probabilística para sistemas especialistas**. 1998. Dissertação. (Curso de Pós Graduação em Ciência da Computação). Universidade Federal de Santa Catarina. Florianópolis, 1998.
- KOTLER, Philip; ARMSTRONG, Gary. **Princípios de marketing**. New Jersey: Prentice-Hall., 2000.

- MORALES, Aran B. T.; GONÇALVES, A.; MARTINS, A. R. **Mineração de dados**. Disponível em: <<http://stelanel.eps.ufsc.br/aran>>. Acesso em: 15 outubro 2000.
- NASSAR, S.M. **Estatística e informática: Um processo Interativo Entre Duas Ciências**. Trabalho apresentado no concurso para professor titular. Departamento de Informática e Estatística. Centro Tecnológico. Universidade Federal de Santa Catarina. Florianópolis, 1998.
- NETICA**. Disponível em: <<http://www.norsys.com>>. Acesso em: 05 fevereiro 2000.
- OLIVEIRA, Alexandra Faisca Nunes. **Uma metodologia de uso de técnicas de indução para criação de regras de sistemas especialistas**. 2001. Dissertação. (Curso de Pós Graduação em Engenharia de Produção). Universidade Federal de Santa Catarina.. Florianópolis, 2001.
- PACHECO, Roberto C. S.. **Tratamento de imprecisão em sistemas especialistas**.1991. Dissertação. (Programa de Pós-Graduação em Engenharia de Produção) da Universidade Federal de Santa Catarina. Florianópolis, 1991.
- PEARL, Judea. **Probabilistic reasining in intelligent systems: Networks of Plausible Inference**. Califórnia:[s.n], 1988.
- RUSSEL, S.; NORVIG, P. **Artificial intelligence – A Modern Approach**. New Jersey: Prentice-Hall., 1995.
- SALES AND *MARKETING* MANAGEMENT. In: **A User's Guide to the Survey of Buying Power**. Nova York:[s.n], 2000.
- SIMÕES, Priscyla Waleska Targino de Azevedo. **SACI - Sistema de Apoio na Avaliação de Distúrbios do Crescimento Infantil**. 2001. Dissertação. (Curso de Pós-Graduação em Ciência da Computação). Universidade Federal de Santa Catarina. Florianópolis, 2001.
- SÍNTESE ESTATÍSTICA DE SANTA CATARINA**. Florianópolis: DEGE/GERES. Agosto de 2001.
- SKAUGE, Jason. **Data mining**. Disponível em <<http://www.isl.net/~jskauge>>. Acesso em: 20 fevereiro 2000.
- TOLEDO, Renata V. A. **Sistema de auxílio ao diagnóstico diferencial de cefaléia**. 2000. Dissertação. (Curso de Pós Graduação em Ciência da Computação). Pontifícia Universidade Católica do Paraná. Curitiba, 2000.
- VERONEZ, Cleverson. **Baseline bayesiana para gerência de redes de computadores**. 2000. Dissertação. (Curso de Pós Graduação em Ciência da Computação). Universidade Federal de Santa Catarina. Florianópolis, 2000.

WITTEN, I.H. **Data mining**: practical machine learning tools and techniques with Java implementations. New Zealand: [s.n], 2000.

8 ANEXO

Option Compare Database

Option Explicit

Private Sub Interface()

On Error GoTo err

Dim app As Netica.Application

Dim net As BNet

Dim Nodo As BNode

Dim crença As Double

Set app = New Netica.Application

app.Visible = True

Set net = FileNamed("c:\SEPE\SEPE.dne").ReadBNet()

net.Compile

Set Nodo = net.nodes("PIB")

crença = Nodo.GetBelief("G1")

MsgBox "A probabilidade estado PIB=1 é " & crença

net.nodes("População").EnterFinding ("G4")

crença = Nodo.GetBelief("G2")

MsgBox "Dado estado População=4, a probabilidade de estado PIB=2 é " & crença

net.nodes("Potencial Econômico").EnterFinding ("Alto")

crença = Nodo.GetBelief("G2")

MsgBox "Dado estados População=4 e Potencial Econômico=Alto, a probabilidade " _
& "de estado PIB=2 é " & crença

net.nodes("Expresso").EnterFinding ("G6")

crença = Nodo.GetBelief("G1")

MsgBox "Dados estados População=4, Potencial Econômico=Alto e Expresso=6, a probabilidade de estado PIB=2 é " & crença

If Not app.UserControl Then

 app.Quit

End If

Exit Sub

err:

MsgBox "API: Erro " & ((err.Number And &H7FFF) - 10000) & ": " & err.Description

End Sub