

Sistema de Apoio à Decisão para o Projeto Segurança nas Praias Baseado em Data Mining

Danilo Domingos dos Santos

ddsantos@zaz.com.br

Anita Maria da Rocha Fernandes

anita@inf.univali.br

Antonio Henrique da Fontoura Klein

klein@cttmar.univali.br

**Centro de Ciências Tecnológicas, da Terra e do Mar – CTTMar
Universidade do Vale do Itajaí – UNIVALI
Itajaí, Santa Catarina – Brasil**

Resumo

O “Projeto Segurança nas Praias” é uma aplicação pioneira no Brasil do “Programa Australiano de Segurança e Gerenciamento das Praias”, iniciado em 1990. O projeto objetiva a criação de uma base de informações sobre as características sócio-ambientais das praias catarinenses e a formulação de um programa de orientação ao público usuário, definindo o grau de periculosidade dos balneários. O problema que surge é como tomar decisões corretas e representativas com base nas informações disponíveis em um banco de dados com inúmeros registros, crescentes a cada ano. Para auxiliar a tomada de decisão, propõe-se: desenvolvimento de um *Data Mining*, que extrairá os conhecimentos “escondidos” no banco de dados, visando definir o grau de periculosidade em relação ao banho de mar dos balneários catarinenses; formulação de programas de orientação ao público usuário das praias; geração de dados para auxiliar a tomada de decisão quanto ao planejamento e administração das praias por parte dos órgãos públicos e instituições privadas; diminuição dos números de acidentes e incidentes relativos ao uso das praias.

Palavras-chave: tomada de decisão, sistemas de informação, *Data Mining*.

1. Introdução

Este trabalho visa apresentar a proposta de desenvolvimento de um sistema de apoio à decisão baseado em *Data Mining* para as instituições envolvidas no “Projeto Segurança nas Praias”, que terá um diferencial perante os projetos de controle e melhor aproveitamento do ambiente.

O Projeto Segurança nas Praias é desenvolvido todos os anos e, conseqüentemente, o número de informações aumenta substancialmente, exigindo a cada ano maior tempo de processamento para consultas, que são altamente dependentes do que o pesquisador quer extrair e do que ele pensa ser relevante. Por isso, torna-se importante o uso de ferramentas e técnicas mais eficientes de pesquisa e auxílio na tomada de decisão, capazes de extrair informações além do imaginado por pessoas, mas que foram recolhidas, estão no banco de dados e parecem escondidas.

A utilização de *Data Mining* é uma alternativa de trabalhar diretamente com os dados de uma base em busca de detalhes, ao invés de seguir caminhos predeterminados, estudando e apresentando resultados de forma automática, não requerendo a administração do usuário. Como *Data Mining* inclui técnicas de Inteligência Artificial (IA), torna-se extremamente adequado para analisar grupos de dados não intuitivos e grandes demais para serem explorados manualmente, trazendo tendências e padrões escondidos num formato mais compreensível, objetivando a obtenção de soluções que produzam previsões úteis e extração de informações críticas capazes de estreitar

ainda mais o relacionamento entre os envolvidos e com os usuários, aumentando agilidade, flexibilidade, qualidade, gerenciamento, melhor investimento, informações padronizadas e maior controle do projeto.

A utilização de *Data Mining* poderá trazer um novo direcionamento ao “Projeto Segurança nas Praias”, eliminando possíveis donos do conhecimento, evitando tornar o projeto altamente dependente de uma pessoa ou grupo delas.

2. Sistemas de Apoio à Decisão

Segundo Oliveira (1993), sistemas de apoio à decisão auxiliam na tomada de decisão através da organização e estatística das informações, respeitando os seus níveis de influência, de decisão e condicionantes. Tais sistemas devem prezar pela qualidade da informação que manipulam, que segundo Oliveira (1993), se resume à precisão (correção da informação), clareza (capacidade de identificar uma situação), relevância (chegar a quem interessa) e rapidez (chegar em tempo hábil).

A tabela 1 exibe o processo de levantamento e análise em sistemas de apoio à decisão.

Tabela 1: Processo de Levantamento e Análise em Sistemas de Apoio à Decisão

ETAPA	OBJETIVOS	COMENTÁRIOS
Identificar focos de decisões	Quem toma decisões? Quais decisões são tomadas? Quais áreas utilizadas para tomada de decisão? Quais os recursos utilizados?	O recurso mais indicado para visualização pelos diretores é a forma gráfica. Relatórios ou tabelas são mais indicados em níveis operacionais.
Análise das Saídas	O que os executivos fazem com as saídas? Qual o volume e frequência das saídas? Qual a influência no processo decisório? Qual o formato mais adequado?	A forma como deverão estar dispostas as informações (ordem alfabética, por valores, por data, etc.) depende da adequação ao modo como aquele que toma decisões mais gosta de visualizá-las, sem abrir mão dos padrões adotados.
Análise das Entradas	Formato Volume Frequência	É um ponto essencial para a credibilidade de qualquer SAD
Análise dos Processos	Sequência das operações Tipos de arquivamento Custos Equipamentos	Quanto mais heterogêneo é o ambiente mais difícil fica de implantar um SAD

Fonte: Oliveira, 1993

Gurovitz (1997) mostra aplicações de sistemas de apoio à decisão baseados em *Data Warehouse*. Tais sistemas guardam e organizam todas as informações que estão espalhadas por vários sistemas dentro de uma organização, ocasionando uma tomada de decisão melhor fundamentada sem depender tanto da intuição.

O sucesso do *Data Warehouse* está diretamente ligado ao problema. Se este for concreto e sua solução depender da associação lógica de informações que estão espalhadas, o *Data Warehouse* torna-se uma escolha e tanto, pois as suas principais promessas são agilidade e precisão na tomada de decisões, com base na informação *on-line* e no formato desejado pelos executivos, proporcionando ganho de tempo e segurança. (Inmon, 1996)

Para aumentar a eficiência na busca de informações em banco de dados, principalmente nos de grande porte, há uma crescente aceitação da tecnologia de *Data Mining* ou mineração de dados.

3. Data Mining

O nome *Data Mining*, conforme Santos (1999), surgiu da semelhança entre a procura de informação importante para o mundo dos negócios escondida num banco de dados e o minerar de uma montanha para encontrar um veio de ouro. A tabela 2 expressa a evolução do *Data Mining*.

Tabela 2: Evolução do Data Mining

EVOLUÇÃO	PERGUNTAS
1960: Coleção de dados	Qual foi meu rendimento total nos últimos cinco anos?
1980: Acesso aos dados	Qual foi meu rendimento no Brasil no último janeiro?
1990: <i>Data Warehousing</i> & suporte à decisão	Qual foi meu rendimento no Brasil no último janeiro do Sul ao Nordeste?
Atualmente: <i>Data Mining</i>	Por que alguns produtos são mais vendidos na região Sul?

Fonte: Fidalgo (1998)

Ciclo et al (2000) apresenta *Data Mining* como um processo descendente de três vertentes: a estatística, a inteligência artificial e o aprendizado de máquina, conforme ilustra a figura 1.

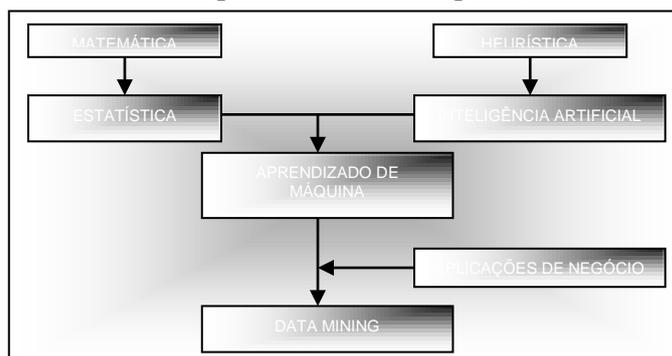


Figura 1: Origem do Data Mining

Fonte: Ciclo (2000)

Ciclo et al (2000) afirma que sem a estatística não seria possível ter-se *Data Mining*, uma vez que no “coração” das suas atuais ferramentas e técnicas a análise estatística clássica desempenha um papel fundamental, pois dela vem conceitos como distribuição normal, variância, análise de regressão, desvio simples, análise de conjuntos e discriminantes e intervalos de confiança, todos usados para estudar os dados e os relacionamentos entre eles.

A inteligência artificial é construída à partir de fundamentos de heurística. A heurística é uma ciência que tenta imitar o comportamento humano durante a resolução de problemas, ou seja, ela trabalha em oposição à estatística (Kremer, 1999).

O aprendizado de máquina é a união entre a estatística e a inteligência artificial, procurando fazer com que os computadores tenham a capacidade de aprender por meio dos dados que manipulam e tomar decisões baseadas nas características destes dados (Kremer, 1999).

Santos (1999) diz que as técnicas de *Data Mining* resultaram de um longo processo de investigação e desenvolvimento que, neste momento, tem o suporte de três tecnologias: massificação do armazenamento de dados, poderosos computadores com múltiplos processadores e algoritmos de *Data Mining*. Os principais componentes desta tecnologia, como a inteligência artificial, estão em pleno desenvolvimento há décadas.

3.1 Conceitos de Data Mining

Data Mining, de acordo com Groth (1998), é um processo de descoberta automática de informações. Figueira (1998) define como um processo automatizado de captura e análise de enormes conjuntos de dados e conseqüente extração do significado. Para Ciclo (2000), *Data Mining* é o processo de extrair informações válidas, desconhecidas e de máxima abrangência a partir de grandes bases de dados, descobrindo relacionamentos escondidos. Mas é Ávila (1998) quem coloca *Data Mining* como uma área de pesquisa da IA responsável por estudar metodologias que gerem um produto capaz de encontrar padrões em bases de dados. É considerado uma forma de descobrimento de conhecimento em banco de dados (KDD – *knowledge discovery in databases*), uma área de pesquisa em evidência nos dias atuais que envolve IA e banco de dados.

Como já foi dito, *Data Mining* é um processo de descoberta, isto é, examina-se um banco de dados para encontrar padrões sem uma idéia ou hipótese predeterminada. O programa toma a iniciativa de encontrar aquilo que interessa aos padrões sem que o usuário verifique se isto realmente interessa.

Santos (1999) afirma que as ferramentas de *Data Mining* servem para prever tendências e comportamentos futuros, permitindo aos gestores tomarem decisões baseadas em fatos, ao invés de suposições. Podem responder a questões de negócio que consomem muito tempo, permitindo prever um futuro cego aos olhos de especialistas humanos devido ao fato desta informação sair dos limites das suas expectativas. A utilidade da tecnologia de *Data Mining* é demonstrada no exemplo dado por Santos (1999) e ilustrada na figura 2:

Imaginemos que você tenha uma rede de lojas do tipo ‘clube de vídeo’ e quer rentabilizar o seu negócio. Para o conseguir seria bom que você soubesse quais são os filmes que os seus clientes vão ver e melhor ainda seria ter um modo de avisar os clientes que chegou um filme que eles querem ver. Para saber quais os filmes favoritos dos seus clientes, poderia armazenar informação sobre todos os filmes que cada cliente alugou. Através de Data Mining – ou de outras técnicas como OLAP ou mesmo simples consultas SQL – seria possível fazer uma previsão sobre quais os melhores filmes a comprar, para que lojas e em que quantidade. Para avisar os clientes, uma simples carta com as últimas novidades seria suficiente, no entanto, uma carta para todos os clientes de todas as lojas pode se tornar uma publicidade bastante cara, levando em conta que muitos dos clientes não estariam interessados em muitos dos filmes anunciados. Poder-se-ia também usar aqui Data Mining para tentar descobrir a quem valha a pena enviar que tipo de publicidade.

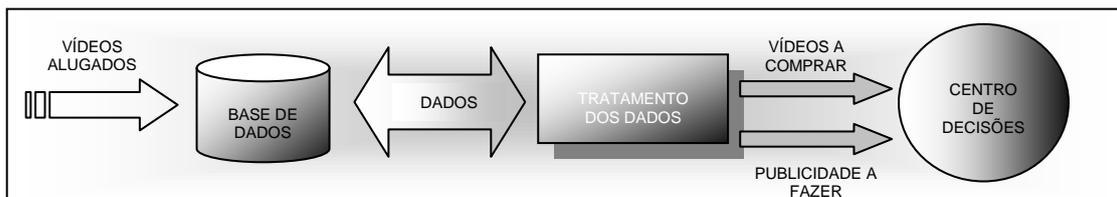


Figura 2: Aquisição da Informação Conforme Exemplo
Fonte: Santos (1999)

As outras técnicas de pesquisa são conduzidas pelo utilizador, como é o caso do OLAP (processamento analítico *on-line*): o analista gera uma hipótese e usa o OLAP para verificar essa hipótese, isto é, quando os utilizadores empregam o OLAP e outras ferramentas de consulta para explorar os dados, eles guiam essa exploração. *Data Mining*, ao contrário, é utilizado nos dados para gerar a hipótese, quando os utilizadores empregam ferramentas de *Data Mining* para explorar os dados são estas ferramentas que exploram os dados. A figura 3 mostra graficamente estas diferenças.

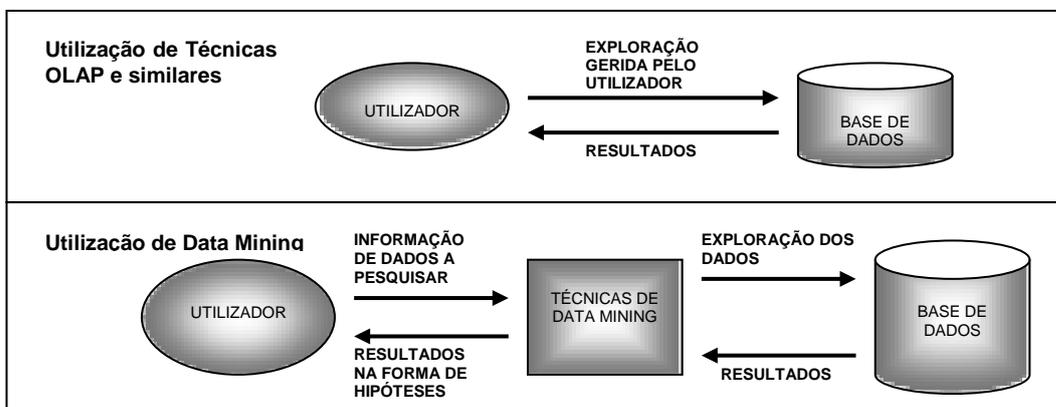


Figura 3: Diferença entre a Utilização de Data Mining e Outras Técnicas
Fonte: Santos (1999)

3.2 Ferramentas de Data Mining

Segundo Santos (1999), as ferramentas de *Data Mining* incluem agentes inteligentes, OLAP e pergunta-resposta. A escolha envolve alguns pontos como a dimensão e o tipo da base de dados.

Os agentes inteligentes podem ser lançados manualmente para executar pesquisas predeterminadas de certos padrões de dados ou executadas automaticamente, obedecendo a períodos predeterminados, para pesquisar ou monitorar tarefas. Desenvolver agentes inteligentes que consigam definir a prioridade e fazer a filtragem da informação ainda é um grande desafio.

OLAP é uma ferramenta que trabalha combinada às ferramentas de análise multidimensional, que representam dados como matrizes de n dimensões chamadas hipercubos. Esta combinação entre OLAP e os hipercubos permitem ao *Data Mining* calcular métricas iterativamente, explorando as várias combinações de uma ou mais dimensões de dados.

Se não for possível usar nenhuma das ferramentas anteriores, pode-se usar o esquema cliente/servidor e consultas SQL a uma base de dados relacional, que podem se tornar muito lentas no caso desta base for de grande dimensão. Algumas SQL incluem módulos inteligentes satisfatórios para a geração de relatórios e análises estatísticas.

3.3 Justificativas para a Utilização do Data Mining

O acúmulo de dados e a crescente valorização da informação é, segundo Figueira (1998), um grande impulsionador de tecnologias capazes de vasculhar bancos de dados com infinitos registros, entre elas a do *Data Mining*. Mesmo ainda possuindo uma análise de dados demorada, dispendiosa e sujeita a erros, Groth (1998) aponta que as quedas no custo de computadores pessoais (PC) e as recentes inovações de metodologias, como os novos algoritmos capazes de aumentar a usabilidade, o poder e a eficiência das ferramentas, tornaram o *Data Mining* cada vez mais acessível ao usuário final.

Data Mining se apresenta como um método alternativo de trabalhar diretamente com os dados de uma base em busca de detalhes ao invés de seguir caminhos predeterminados. É um processo automático de estudar os dados e apresentar seus achados, isto é, *Data Mining* não requer que o usuário pilote ou administre a ferramenta ao longo do processo de análise dos dados. A única atribuição dada ao usuário é a de fornecer dados úteis.

Data Mining é um componente essencial para a análise e complementa as funções já existentes nos Sistemas de Apoio à Decisão.

3.4 Aplicações de Data Mining

Data Mining tem sido utilizado por analistas de negócios, segundo Groth (1998), para descobrir, por exemplo, por que alguns clientes flutuam entre empresas enquanto outros são extremamente fiéis? *Data Mining*, neste caso, identificará os perfis de clientes fiéis versus os perfis de outros clientes através de um histórico modelado pelo analista de negócios. O papel do analista é supervisionar, ajudar na definição do processo, propor uma meta geral e dados a serem utilizados.

Uma outra aplicação é a procura por similaridades dentro de um conjunto de dados e linhas de grupos similares, podendo ser baseado em dados históricos não supridos com dados de treinamento, ou seja, o aprendizado não é supervisionado.

Groth (1998) afirma que a aplicação do *Data Mining* como benefício estratégico trouxe maior diferenciação, análise e previsão ao mercado. A diferenciação de mercado se aplica à identificação de potenciais clientes com significativos cortes nos custos de *marketing*. A análise de mercado trata-se de descobrir padrões de comportamento nos clientes. A previsão se reflete mais significativamente no mercado financeiro através da combinação com redes neurais.

As áreas de maior aplicação do *Data Mining* são, segundo Santos (1999): vendas (*Marketing*): conhecer e oferecer produtos adequados às necessidades individuais dos clientes, a mais poderosa arma competitiva do mercado; saúde: reduzir os custos e ao mesmo tempo aumentar a qualidade do serviço prestado; ciência e tecnologia: *Data Mining* é um processo de descoberta, a maior motivação da ciência; finanças: principalmente na verificação e prevenção de fraudes.

Na tabela 3 apresenta-se alguns exemplos da aplicação de *Data Mining*.

Tabela 3: Aplicações do Data Mining

ORGANIZAÇÃO	PROBLEMAS ENCONTRADOS	RESULTADOS OBTIDOS
Lojas Brasileiras	Até abril de 1998, a rede mal conseguia pôr em suas gôndolas os 51.000 produtos de seu catálogo. Faltava espaço. Depois de um processo de automação que consumiu 1 milhão de dólares, a cadeia de setenta lojas espalhadas pelo Brasil descobriu que muitas de suas mercadorias só serviam para juntar pó. O mais incrível: expunham guarda-chuvas, sombrinhas e malhas de lã no Nordeste, onde dificilmente faz frio ou chove, e bateadeiras eram vendidas em Santa Catarina unicamente em 110 volts, embora neste Estado a voltagem seja de 220 volts.	Hoje o grupo mantém 14.000 itens nas prateleiras, numa única operação tirou 37.000 produtos de sua lista de compras. Seus executivos utilizaram <i>Data Mining</i> também para, com base no consumo, <i>hobbies</i> e informações sobre transações financeiras dos clientes, traçar associações que revelaram grandes filões do mercado.
Bank of America	Para detectar fraudes, cortar gastos e aumentar sua receita, selecionou entre seus 36 milhões de clientes aqueles com menor risco de dar um calote num empréstimo e, a partir daí, enviou cartas oferecendo linhas de crédito para os correntistas cujos filhos tivessem entre 18 e 21 anos e, portanto, precisassem de dinheiro para auxiliar nos estudos e compra do próprio carro ou casa.	Lucro de 30 milhões de dólares
Wal-Mart	O mais clássico exemplo da aplicação de <i>Data Warehouse</i> e <i>Data Mining</i> . Descobriu em seu gigantesco armazém de dados que a venda de cervejas cresciam proporcionalmente à de fraldas. <i>Data Mining</i> apontou que os maiores compradores de fraldas descartáveis eram homens entre 25 e 30 anos que aproveitavam a volta do trabalho para casa nas sextas-feiras à noite para comprar fraldas e abastecer a geladeira de cervejas para o fim de semana.	Redefinindo as posições dos produtos (foram postos lado a lado) baseada na conexão de hipóteses desenvolvidas pelo <i>Data Mining</i> , as vendas de ambos dispararam 30%.

Fonte: Pilla (1999)

4. Projeto Segurança nas Praias

4.1 Introdução

As praias, conforme Medeiros (1999), podem ou não representar riscos à segurança pública, tudo depende das condições de tempo, da direção e força de ventos, da direção e características das ondas e das marés, da fauna marinha, da época do ano e das características locais, tanto geográficas e geológicas quanto do nível e tipo de ocupação. Estes são fatores que comprovam a importância de se conhecer as praias e seus usuários e, a partir daí, formular um programa de educação e prevenção aos banhistas.

O Projeto Segurança nas Praias é uma aplicação pioneira no Brasil de uma política de prevenção e orientação ao público usuário das praias baseada, segundo Universidade (1999), no Programa Australiano de Segurança e Gerenciamento das Praias iniciado em 1990. Em Santa Catarina, o projeto é executado desde o verão 95/96 através de uma parceria entre a Universidade do Vale do Itajaí (UNIVALI) – destacando-se o trabalho do Centro de Ciências Tecnológicas da Terra e do Mar (CTTMar), principalmente do seu curso de Oceanografia – e empresas públicas e privadas, entre elas o Corpo de Bombeiros da Polícia Militar do Estado de Santa Catarina, a Petrobrás e a Santa Catarina Turismo S.A. (SANTUR).

Muitas comunidades do litoral catarinense vivem quase que exclusivamente do imenso número de turistas que invadem suas praias especialmente na alta temporada. Estes turistas se defrontam com problemas de infra-estrutura, inclusive na segurança, tornando essencial um programa que treine e oriente as pessoas. O “Projeto Segurança nas Praias” proporciona, ainda,

levantar informações sobre as características de cada localidade, trabalhando em cima das características dos acidentes e da diminuição do número de ocorrências.

Da temporada 95/96 à 98/99, três mil acidentes e aproximadamente seiscentas entrevistas com o público usuário foram registrados, que demonstraram grande desconhecimento dos perigos oferecidos pelas praias, tornando necessário um criterioso estudo de como passar as informações e dicas para os banhistas de uma forma que realmente sejam compreendidas e absorvidas.

A tabela 4 mostra os procedimentos de ação adotados pela equipe do Projeto Segurança nas Praias, em que quatro estratégias são divididas em duas vertentes.

Tabela 4: Procedimentos de Ação no Projeto

PROCEDIMENTOS	ESTRATÉGIAS	EXECUÇÃO
<u>Natureza física e histórica:</u> busca identificar características físicas e ambientais das praias e seus potenciais riscos, estabelecer um perfil de tendências e padrões de acidentes e auxiliar na elaboração de ações preventivas	Identificar e caracterizar os tipos de acidentes de banho verificados, as condições da vítima e as características ambientais	Equipe do CTTMar e salva-vidas do Corpo de Bombeiros da Polícia Militar do Estado de Santa Catarina
	Obter informações gerais sobre as praias e identificar potenciais pontos com riscos ao banho	Equipe do CTTMar e salva-vidas do Corpo de Bombeiros da Polícia Militar do Estado de Santa Catarina
<u>Natureza psicográfica e sócio-econômica:</u> busca identificar o perfil dos usuários das praias, padrões de comportamento e desenhar ações preventivas e corretivas	Caracterizar o público usuário das praias, definindo perfil psicográfico e sócio-econômico dos frequentadores	Equipe do CTTMar
	Capacitar recursos humanos para atuação na prevenção e administração e elaborar meios de esclarecimento dos riscos à população	Treinamento de pessoal para prevenção e atendimento de acidentes e elaboração de campanhas educativas pela equipe do CTTMar

Fonte: Universidade (1999)

Os objetivos, conforme Universidade (1999), são “a criação de uma base de informações sobre as características sócio-ambientais das praias catarinenses e a formulação de um programa de orientação ao público usuário, definindo o grau de periculosidade dos balneários”.

4.1.1 Perigos Associados às Praias

Segundo Medeiros (1999), os perigos associados às praias podem ser de caráter permanente ou não. Os primeiros podem ser facilmente evitados quando conhecidos, enquanto os outros exigem bastante atenção e cuidado.

Os perigos permanentes, por serem mais facilmente evitados, oferecem maior perigo aos usuários mais desatentos e àqueles que superestimam suas habilidades. Eles correm o risco de ficarem expostos às ações de ondas e correntes e serem facilmente dominados: em águas profundas perde-se o apoio dos pés; qualquer obstáculo (molhe, trapiche, destroço, rocha ou recife) pode escavar buracos e formar correntes mais fortes, além de representar riscos de choques e ferimentos; o encontro de rios e lagoas com o mar modifica ondas, correntes e o relevo do fundo da praia.

Os perigos não permanentes oferecem maior perigo aos usuários das praias e exigem um trabalho efetivo dos salva-vidas quanto à prevenção e educação. Eles podem ser ocasionados pela arrebentação das ondas, tipos diferentes de praias, buracos, correntes e organismos marinhos.

A arrebentação é o que acontece quando as ondas se aproximam da costa, quebrando e gerando grande turbulência e correntes. Existem três tipos de arrebentação: tubular ou caixote: a onda forma um tubo de grande turbulência que quebra violentamente; deslizante: a onda começa a quebrar suavemente longe da beira da praia; sem arrebentação: ocorre em praias com inclinações ou proteções naturais.

As praias, a nível do “Projeto Segurança nas Praias”, podem ser classificadas em três tipos: rasa, intermediária e de tombo. Na praia rasa a profundidade aumenta à medida que o banhista se afasta da praia, podendo ou não apresentar bancos e o tipo de arrebentação normalmente é do tipo deslizante. Na praia intermediária o fundo apresenta bancos irregulares cortados por canais em que

se desenvolvem correntes de retorno e a arrebentação, do tipo caixote, ocorre bem próximo à beira da praia. As praias de tombo apresentam um rápido aumento da profundidade logo após a beira da praia devido à forte inclinação, possuindo uma areia grossa e pouca ou nenhuma arrebentação.

Buracos e correntes podem ser percebidos quando as ondas não quebram, dando uma falsa impressão de segurança. As correntes podem ser transversais ou paralelas. A presença de buracos combinada com maior incidência de ventos e ondas fortes aumentam os perigos das correntes paralelas.

As correntes de retorno (ou repuxos) possuem intensidade variada, partindo da beira d'água em direção ao mar, terminando no fim da zona de arrebentação, podendo ser visualizadas por diferenças na coloração da água (mais escura devido à areia levantada do fundo do mar) e pela ausência de arrebentação.

Os organismos marinhos são em sua maior parte inofensivos, mas alguns podem oferecer riscos aos banhistas, como mexilhões, cracas, ouriços e águas-vivas, podendo causar desde cortes até queimaduras.

4.1.2 Informações Relativas ao Usuário

Quanto às informações relativas ao usuário, o “Projeto Segurança nas Praias” considera informações como familiaridade da vítima com a praia, tipo de usuário (banhista, surfista, nadador, mergulhador), influência de álcool, comportamento da vítima, lesões (como cortes ou parada respiratória, entre outras), sistema de sinalização e tempo.

Quanto à familiaridade com a praia, considera-se dois tipos: o visitante ocasional (aquele que está na praia pela primeira vez ou raramente a frequenta) e o veranista (aquele que mora ou costuma passar determinado período na praia onde foi socorrido). O comportamento da vítima diz respeito a se ela manteve-se calma, ficou descontrolada ou inconsciente.

Para Medeiros (1999) a falta de conhecimento do sistema de sinalização usado nas praias catarinenses e a pouca familiaridade com a praia aumenta os riscos de acidente. Somado a isto, há uma diferença de sinalização com relação a outros estados, como o Rio Grande do Sul.

4.2 Levantamento de Dados

O levantamento de dados/informações é realizado em conjunto pelo CTTMar e pelo Corpo de Bombeiros, preenchendo-se a dois questionários: um referente ao perfil dos usuários e outro aos registros de ocorrência. Para Medeiros (1999) estas informações servem de base para a formulação de campanhas educativas e preventivas, fornecendo subsídios para o planejamento do patrulhamento nas praias.

4.2.1 Questionário de Perfil do Usuário

No questionário de levantamento do perfil dos usuários das praias, os entrevistados fornecem informações sobre: município, praia, data e horário da entrevista; idade e sexo; procedência: estrangeiro, catarinense ou de outros estados e a cidade; escolaridade; familiaridade com a praia: visitante ocasional ou veranista; tempo de permanência na praia; local de permanência ou residência; habilidade de natação; toma banho de mar: sempre, às vezes ou nunca; até que altura do corpo entra na água; o que considera como sinal de perigo: ondas, repuxo, profundidade, buracos, bandeira, coloração do mar ou outros; noção de perigo da praia; percepção de riscos: observa o mar, as bandeiras, conversa com salva-vidas, toma banho em locais patrulhados e entra na água após as refeições; se os filhos sabem nadar e entram na água acompanhados; envolvimento anterior em acidentes nesta ou em outra praia, que tipo de acidente e de socorro; depois do acidente:

nunca mais retornou àquela praia, nunca mais entrou na água ou retornou normalmente à praia e à água; renda mensal líquida.

Através deste questionário, o Projeto Segurança nas Praias busca levantar informações sobre todos os usuários das praias, independente de acidentes, tornando possível conhecer o público das praias em geral e o público de cada praia e, a partir daí, planejar procedimentos e estratégias de ação ao combate de acidentes e fornecer uma prestação de serviço mais adequada, principalmente no sentido de educação deste público.

4.2.2 Questionário de Registro de Ocorrências

O questionário de registro de ocorrências compreende quatro partes: identificação, dados da vítima, dados do resgate e dados da praia. Através deste questionário, conhece-se os fatores do ambiente que oferecem risco e as pessoas envolvidas em cada acidente de forma individual, possibilitando ao Projeto Segurança nas Praias propor soluções a nível ambiental e comportamental.

No item identificação são fornecidas informações relativas ao local de ocorrência do acidente e tipo de ocorrência. O primeiro refere-se a município, praia, número do posto, identificação dos salva-vidas, data e hora. O segundo refere-se às seguintes alternativas: arrastamento, afogamento com recuperação em água doce ou em água salgada, afogamento seguido de morte em água doce ou em água salgada, embarcação à deriva, prevenção ou encontro de criança perdida.

No item dados da vítima são fornecidas informações de: identificação e procedência da vítima: nome, cidade, estado, país, idade e sexo; familiaridade da vítima com a praia: visitante ocasional ou veranista; tipo de usuário: banhista, nadador, surfista ou mergulhador; habilidade de natação; vítima sob influência ou não de álcool e drogas; durante o resgate a vítima manteve-se calma, descontrolou-se ou esteve inconsciente; se houve lesões associadas como parada respiratória, cortes ou câibras, entre outras.

No item dados do resgate são fornecidas informações sobre: resgate realizado dentro ou fora da área patrulhada; número de salva-vidas envolvidos no resgate; equipamentos empregados no resgate: homem com nadadeiras, *life-belt*, prancha ou bóia, *jet ski*, lancha, helicóptero ou outros; o acidente ocorreu antes, depois ou na zona de arrebentação; bandeira sinalizada no posto: verde, amarela ou vermelha; se havia sinalização na praia, como bandeira vermelha; perigos da praia associados ao acidente: correntes de retorno, correntes longitudinais, proximidade de costão rochoso, desembocadura de rio, proximidade de estruturas rígidas.

No item dados da praia, o questionário pede informações sobre: céu limpo, com nuvens, nublado, chuvoso ou tempestade; intensidade do vento: ausente, fraco, moderado, forte ou muito forte; direção de incidência do vento: norte, sul, leste, oeste, nordeste, noroeste, sudeste ou sudoeste; altura da onda de arrebentação; tipo de arrebentação: deslizante, caixote ou sem arrebentação; tipo de corrente presente: de retorno, longitudinal para direita ou longitudinal para esquerda; intensidade da corrente: fraca, moderada ou forte; forma da praia: rasa, intermediária ou de tombo.

4.3 Armazenamento das Informações

Todas as informações recolhidas através dos questionários de perfil do usuário e registro de ocorrências estão sendo armazenadas num banco de dados Access[®], disponibilizado no Laboratório de Oceanografia Geológica (bloco 19) da Universidade do Vale do Itajaí. Inicialmente, os próprios

relatórios gerados pelo banco de dados através de consultas estruturadas são suficientes para fornecer informações importantes para a determinação de estratégias de ação.

4.3.1 Resultados Alcançados

Em Universidade (1999) alguns resultados são mostrados, entre eles que a presença de correntes de retorno na praia é o principal causador de acidentes, conformando com dados de outros países como a Austrália. A maior parte dos acidentes ocorre sob bandeira amarela no posto e vermelha na praia, refletindo a falta de conhecimento e respeito à sinalização. Os banhistas superestimam sua habilidade, sua juventude (adolescentes) e o fato de serem do sexo masculino.

A precisão e a riqueza das informações quanto aos procedimentos de assistência, quanto às vítimas e quanto às condições ambientais têm aprimorado as técnicas de salvamento e orientação ao público. A coleta e o registro de dados têm sido aprimorados a fim de aumentar a eficácia dos resultados e proporcionar a maximização do seu uso, além de facilitar o trabalho dos envolvidos.

Com o levantamento dos procedimentos dos banhistas ao chegar na praia e questões sócio-econômicas, cada praia passou a ser caracterizada por suas características ambientais e pelo perfil de seu público usuário. As informações acusam que as pessoas devem ser orientadas antes de chegar na praia, para serem capazes de, por exemplo, identificar uma corrente de retorno e o sistema de sinalização de bandeiras.

Informações importantes relativas aos acidentes levantadas estão em Universidade (1999) e Medeiros (1999):

- a) A maior parte dos acidentes ocorre nos horários de pico de movimento;
- b) 56% dos banhistas são mulheres, mas 63% das vítimas são do sexo masculino;
- c) 81% dos acidentes referem-se a arrastamentos e afogamentos;
- d) Sinalização do posto: 56% dos acidentes foram registrados sob bandeira amarela, 15% sob bandeira vermelha e 13% sob bandeira verde. Sinalização do local: 73% dos acidentes ocorrem dentro da área patrulhada, sendo que 97.5% deles ocorrem com a bandeira de local perigoso;
- e) 6% das vítimas estavam alcoolizadas;
- f) Das vítimas 60% vieram de cidades do interior e 32% de cidades litorâneas; 79% eram brasileiros e 18% argentinos; 50% catarinenses, 11% gaúchos, 11% paranaenses, 10% paulistas;
- g) 32% das vítimas têm entre 16 e 25 anos, 22% entre 11 e 15 anos, 20% entre 26 e 40 anos, 16% são crianças de até dez anos.

Sobre as informações relativas ao público usuário, Universidade (1999) destaca:

- a) 46% dos entrevistados são de outros estados (maioria gaúchos e paranaenses), 40% são catarinenses e 14% são estrangeiros (maioria argentinos e paraguaios);
- b) 34.5% dos turistas estavam pela primeira vez na praia onde foram entrevistados, 54.5% eram veranistas e 11% moradores;
- c) 5% dos entrevistados entram na água até a altura do joelho, 36% até a da cintura, 32% até a do peito, 15% até a do pescoço, 11% ultrapassam a do corpo e 1% não entra na água;
- d) Para saber se o mar está perigoso, 44% dos entrevistados observam as ondas e 31% as bandeiras; em 71% dos resgates realizados as ondas mediam meio metro, no máximo;
- e) 51.5% dos entrevistados eram responsáveis por crianças: destes 50.1% acompanham as crianças no banho e 49.9% dizem observar à distância (em sua maioria estavam lendo ou conversando).

Além disso, Universidade (1999) considera relevante divulgar outros resultados diretos e indiretos gerados, como integração da Universidade com o Corpo de Bombeiros e a comunidade,

estruturação de um banco de dados, levantamento do perfil do turista, formação de pessoal técnico qualificado, melhor avaliação e distribuição do efetivo do Corpo de Bombeiros, melhor atendimento aos usuários e publicação de artigos e de um livro sobre as praias catarinenses.

5. O Sistema Proposto

O primeiro passo para o desenvolvimento deste *Data Mining* foi a modelagem. Esta fase dividiu-se em:

- Manipulação de dados: aqui os dados foram preparados, selecionados, limpos e pré-processados sob o olhar do especialista no assunto. Os dados digitados incorretamente e os valores perdidos (linhas de dados incompletas) foram retirados.
- Definição do estudo: definiu-se aqui o tipo de aprendizado a ser executado. Optou-se pelo aprendizado supervisionado, no qual escolhe-se as variáveis e saídas que satisfarão a meta predeterminada pelo especialista.
- Leitura dos dados e construção do modelo: nesta etapa apresentou-se os dados que compõem a base através de indicadores, como frequência (o número de ocorrência de um certo valor), peso (relevância de um dado, ou seja, o número de saídas proporcionadas por determinado dado), conjunção (quais são as entradas mais relevantes juntas do que separadas), diferenciação (para qual tipo de problema determinada entrada é melhor do que outra) e ruído (erro ou anormalidade gerado pelo *Data Mining*).
- Compreensão do modelo: criado a partir do conjunto de dados, ele envolve o modelo sumário (mostrando pesos, frequências e conjunções importantes para a descrição da meta), distribuição de dados (verificou-se que a quantidade de um dado não interfere diretamente no estudo de um outro dado), diferenciação (se uma entrada prevê algo melhor do que outra é porque há um grau de diferenciação), perfil de validação de sucesso ou falha (o porquê da saída gerada) e perfil dos casos (quais as linhas de dados estão mais propensas a identificar determinada saída).
- Previsão: escolha da melhor situação possível baseada no histórico dos dados.

A figura 4 ilustra o processo de modelagem utilizado.

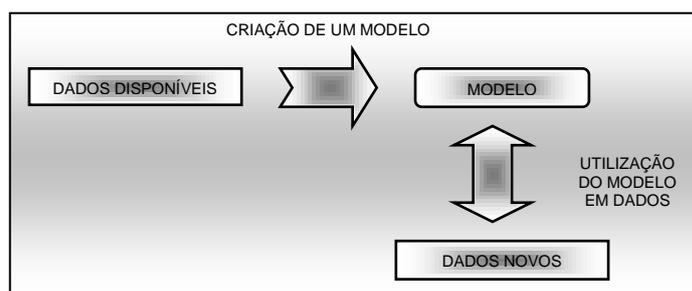


Figura 4: Processo de Modelagem

A segunda etapa do desenvolvimento foi a escolha do método de implementação do *Data Mining*. Aqui, a preocupação recaiu sobre o tamanho da base de dados (base esta que foi criada sem a preocupação futura de usar um *Data Mining*), da qualidade dos dados e do uso de ferramentas de limpeza dos dados e verificação dos erros.

A seguir iniciou-se o estudo das técnicas e algoritmos mais utilizados para implementação de *Data Mining*: *market basket analysis*, *memory based reasoning*, análise de partições, *link analysis*, indução de regras, árvores de decisão, redes neurais artificiais, algoritmos genéticos, OLAP, análise estatística de séries temporais e visualização.

Atualmente, este projeto encontra-se na fase de avaliação das técnicas e algoritmos e, em paralelo, está sendo feita uma análise de desempenho das ferramentas de desenvolvimento de *Data Mining* disponíveis no mercado.

6. Conclusões

A implementação de um sistema de *Data Mining* para o Projeto Segurança nas Praias não significa que os especialistas devam ser descartados, a máquina jamais substituirá o homem. Este sistema simplesmente servirá para beneficiar e auxiliar estas pessoas, proporcionando maior fundamentação e embasamento prático para uma posterior tomada de decisão, que será executada por especialistas.

Com maior velocidade de processamento em uma base de dados crescente, combinando técnicas de inteligência artificial e conhecimento adquirido de situações reais, *Data Mining* proporcionará maior eficiência aos procedimentos adotados para corrigir comportamentos e orientar os usuários das praias catarinenses, buscando reduzir o número de acidentes e tornar o período de verão mais seguro e tranquilo.

Os resultados do sistema de *Data Mining* do Projeto Segurança nas Praias objetivam conhecer o ambiente e mudar o comportamento das pessoas. Comportamentos podem ser facilmente alterados, a natureza não.

7. Bibliografia

- ÁVILA, Bráulio. **Data mining**. Curitiba, 1998. Dissertação (Mestrado em Informática Aplicada) – Pontifícia Universidade Católica do Paraná.
- CICLO, Ivan; PAZ, Luiz C.; MACHADO, Carolina et al. “Data Warehouse”. 18 abr. 2000. [http://www.datawarehouse.inf.br/mining/body_mining.html]. (21 abr. 2000 01:25)
- FIDALGO, Robson N.; JÚNIOR, Fernando F. G. **Data Mining**. Pernambuco, 1998. Trabalho acadêmico (disciplina de Banco de Dados), Universidade Federal de Pernambuco.
- FIGUEIRA, Rafael M. A. **Miner: um software de inferência de dependências funcionais**. Rio de Janeiro, 1998. Trabalho de conclusão de curso (Ciência da Computação) – Instituto de matemática, Universidade Federal do Rio de Janeiro.
- GROTH, Robert. **Data Mining a hands on for business professional**. New Jersey: Editora Prentice Hall PTR, 1998.
- GUROVITZ, Hélio. Competitividade: o que cerveja tem a ver com fraldas?. **Revista Exame**, 9 abr. 1997.
- KREMER, Ricardo. **Sistema de apoio à decisão para previsões genéricas utilizando técnicas de data mining**. Blumenau, 1999. Trabalho de conclusão de curso (Ciência da Computação) – Centro de ciências exatas, Universidade Regional de Blumenau.
- MEDEIROS, Rodrigo P.; MENEZES, João T. **Formação e atualização de salva-vidas: dinâmica costeira**. Itajaí, 1999.
- OLIVEIRA, Djalma P.R. **Sistemas de informações gerenciais: estratégias, táticas e operacionais**. São Paulo: Campus, 1993.
- PILLA, Adriano D.; CAPRETZ, Alexandre L. A.; ALBERTO, Gustavo G. “Data Mining”. *Modelos de Data Mining*. [<http://www.igce.unesp.br/igce/grad/computacao/cintiab/datamine/home.html>] (20 abr. 2000 01:27).
- SANTOS, José; HENRIQUES, Nuno A.C.; REIS, Vanda. **Data mining e data warehousing**. Lisboa, 1999. Dissertação (Licenciatura em Engenharia Informática) – Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
- UNIVERSIDADE DO VALE DO ITAJAÍ. Segurança nas praias: projeto verão 2000. Itajaí, 1999.