

NOEMI DREYER GALVÃO

APLICAÇÃO DA MINERAÇÃO DE DADOS EM BANCOS DA
SEGURANÇA E SAÚDE PÚBLICA EM ACIDENTES DE
TRANSPORTE

Tese apresentada à Universidade Federal
de São Paulo, para obtenção do título de
Doutor em Ciências.

São Paulo
2009

Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos desde que na reprodução figure a identificação do autor, título, instituição e ano da tese.

Galvão, Noemi Dreyer

Aplicação da mineração de dados em bancos da segurança e saúde pública em acidentes de transporte. / Noemi Dreyer

Galvão --São Paulo, 2009.

xvi, 120f.

Tese (Doutorado) Universidade Federal de São Paulo. Escola Paulista de Medicina. Programa de Pós-graduação em Enfermagem.

Título em inglês: Use of data mining in the health banks and public security in traffic accidents.

1. Mineração. 2. Segurança "computação". 3. Informática em Saúde Pública. 4. Saúde Pública. 5. Acidentes de trabalho.

NOEMI DREYER GALVÃO

APLICAÇÃO DA MINERAÇÃO DE DADOS EM BANCOS DA
SEGURANÇA E SAÚDE PÚBLICA EM ACIDENTES DE
TRANSPORTE

Tese apresentada à Universidade Federal
de São Paulo para obtenção do título de
Doutor em Ciências, pelo Programa de Pós-
Graduação em Enfermagem.

Orientadora: Prof(a). Dr(a). Heimar de
Fátima Marin

São Paulo
2009

UNIVERSIDADE FEDERAL DE SÃO PAULO
ESCOLA PAULISTA DE ENFERMAGEM
DEPARTAMENTO DE ENFERMAGEM

Chefe do Departamento: Prof(a). Dr(a). Alba Lucia Bottura Leite de Barros
Coordenador do Curso de Pós-graduação: Prof(a). Dr(a). Janine Schirmer

NOEMI DREYER GALVÃO

APLICAÇÃO DA MINERAÇÃO DE DADOS EM BANCOS DA
SEGURANÇA E SAÚDE PÚBLICA EM ACIDENTES DE
TRANSPORTE

Presidenta da banca:

Prof(a). Dr(a). Heimar de Fátima Marin

BANCA EXAMINADORA

Titulares:

Prof(a). Dr(a). Sonia Maria Oliveira Barros

Prof(a). Dr(a). Heloisa Helena Ciqueto Peres

Prof. Dr. Ricardo da Silva Santos

Prof(a). Dr(a). Eneida Abrantes Mendonça

Suplentes:

Prof(a). Dr(a). Iveth Yamaguchi Whitaker

Prof. Dr. Ivan Torres Pisa

DEDICO

A Deus donde vem minha força e socorro bem presente, ao meu esposo, meu filhote e aos meus pais por todo amor, presença e apoio.

AGRADECIMENTOS

Primeiramente, agradeço a Deus que me presenteia diariamente com o dom da vida, alegria e paz, que de forma muito próxima se deixa encontrar em cada detalhe do dia a dia, mostrando sua doce face de Pai celeste.

Ao meu querido esposo Silvano que compartilhou o dia a dia deste estudo, aceitando as viagens, aulas e contratempos, respeitando os espaços, e sobretudo me amando, cujo amor é recíproco. Tenho convicção de que os sonhos que compartilhamos fazem parte de nossas vidas e não são mais meu ou apenas dele, mas fazem parte de nós. Obrigada por ser tão presente nos árduos momentos, compartilhando a construção deste trabalho.

À minha família, principalmente minha mãe, que compartilhou o cuidado do Samuel (filho) nas minhas viagens e também souberam entender a minha ausência em momentos que já não podem mais ser revividos, para me dedicar à elaboração desta tese.

À minha orientadora, Prof(a). Dr(a). Heimar de Fátima Marin, pela compreensão, competência e seriedade com que me transmitiu seus conhecimentos. Pela dedicação, zelo e paciência com que conduziu o trabalho de forma objetiva e clara. Obrigada por atender sempre as minhas incessantes solicitações e pelos longos momentos de discussão que, para mim, sempre foram muito proveitosos e fundamentais para o aprendizado.

Aos gestores da SES/MT que acreditam e autorizaram o meu afastamento para realizar o estudo, o meu muito obrigado, em especialmente destaco os meus amigos e colegas do nível central da SES/MT permitiram que eu realizasse o Doutorado, obrigado pelo apoio e amizade.

Aos amigos, companheiros e professores da FEN/UFMT, cúmplices do DINTER, pelas calorosas discussões que tanto me fizeram crescer. Vocês, sem dúvida, fizeram com que estes três anos de minha vida fossem mais cheios de questionamentos, mas sem alegria e a esperança de acreditar que mudanças são necessárias e que dependem de cada um de nós.

As professoras Dr(a). Sonia Maria Oliveira Barros e Dr(a). Heloisa Helena Ciqueto Peres, pelas observações e ensinamentos na qualificação e defesa, que possibilitaram melhorar o estudo.

Aos Professores Dr(a). Eneida Mendonça e Dr. Ricardo S Santos pelas valiosas contribuições na defesa, que contribuíram aperfeiçoar o trabalho.

Na pessoa da Prof(a). Dr(a). Janine Schirmer e da Prof(a). Dr(a). Áurea Christina de Paula Corrêa, agradeço a todos os professores da DENF/UNIFESP e FEN/UFMT que compartilharei comigo aquilo que acreditam, e por ajudarem a compreender um pouco das bases da ciência da saúde.

Aos gestores da SEJUSP, da Secretaria Municipal de Saúde de Cuiabá autorizaram a utilização dos dados para realizar este estudo. Das equipes destas instituições destaco em especial as seguintes pessoas: a Tenente Tarciana, Sargento Leandro, Soldado Clewton Quintilhano, Prof(a). Dr(a). Ligia de Oliveira, Erlon e Alexandre Perón que contribuíram na coleta dos dados sobre acidente de transporte nos sistemas de informação.

A secretária Rita da pós- graduação, pela dedicação, atenção e carinho nas minhas incessantes solicitações.

A todos os meus amigos que durante estes três anos souberam me acolher, me incentivar e que torceram por mim. É gratificante saber que vocês existem e fazem parte da minha vida.

Aos professores, colegas e amigos do ISC/UFMT que me impulsionaram e sempre me acolheram as minhas indagações.

E a CAPES e a FAPEMAT, instituições que financiaram doze bolsas de estudo ao longo dos três anos.

SUMÁRIO

	Lista de figuras	x
	Lista de tabelas e quadros	xi
	Lista de abreviaturas	xii
	RESUMO	xv
	ABSTRACT	xvi
1	INTRODUÇÃO	1
2	FUNDAMENTAÇÃO TEÓRICA	8
	2.1 Caracterização dos acidentes de transporte	8
	2.1.1 Fonte de dados	11
	2.1.1.1 Relacionamento de banco de dados	14
	2.2 Informação e informática em saúde	18
	2.2.1 Informação em saúde	18
	2.2.2 Informática em saúde	21
	2.3 Descoberta conhecimento em bases de dados	24
	2.3.1 Tarefas da mineração dos dados	28
	2.3.1.1 Regras de associação	30
	2.3.2 Técnica da mineração dos dados	33
	2.3.3 Ferramentas utilizadas em mineração dos dados	38
	2.3.4 Mineração dos dados no campo da saúde	40
3	OBJETIVO	42
	3.1 Objetivo geral	42
	3.1.1 Objetivo específicos	42
4	MÉTODO	43
	4.1 Tipo de estudo	43
	4.2 Local de estudo	43
	4.3 Fonte de dados, período e considerações éticas	44
	4.4 Processamento e análise dos dados	44

4.4.1	Relacionamento dos banco de dados	47
4.4.2	Mineração dos dados e análise	51
4.5	Limitação do estudo	52
5	RESULTADOS	53
5.1	Aplicação da mineração dos dados	53
5.1.1	Preparação dos dados	53
5.1.2	Entrada de dados na ferramenta	58
5.1.3	Análise preliminar dos dados	61
5.1.4	Aplicação do algoritmo <i>Apriori</i>	64
6	DISCUSSÃO	71
6.1	Aplicação da mineração dos dados	71
6.1.1	Preparação dos dados	71
6.1.2	Entrada de dados na ferramenta	72
6.1.3	Análise preliminar dos dados	73
6.1.4	Aplicação do algoritmo <i>Apriori</i>	75
7	CONCLUSÕES	79
8	CONSIDERAÇÕES FINAIS	80
9	REFERÊNCIAS BIBLIOGRÁFICA	81
	ANEXOS	97
	APÊNDICE	103

LISTA DE FIGURAS

Figuras		Pág.
Figura 1.	Etapas do processo do KDD	26
Figura 2.	Etapas do processo de Mineração de Dados	28
Figura 3.	Número exclusão e registros selecionados dos bancos de dados da SEJUSP, SIH/SUS e SIM. Cuiabá-MT, 2006	48
Figura 4.	Arquivo com cabeçalho e seus atributos no formato exigido pela ferramenta WEKA.	59
Figura 5.	Tela do pré-processamento do <i>software</i> WEKA com atributos ACIDENTECUIABA.arff	60
Figura 6.	Tela de visualização gráfica das variáveis	60
Figura 7.	Valores dos parâmetros selecionados	64
Figura 8.	Resultados dos <i>Itemsets</i> gerados pelo algoritmo <i>APriori</i>	66
Figura 9.	Resultados das regras associativas gerados pelo algoritmo <i>APriori</i>	67

LISTA DE TABELAS E QUADROS

Tabelas		Pág.
Tabela 1.	Número de pares verdadeiros obtidos por meio do relacionamento probabilístico	50
Tabela 2.	Variáveis, descrição e suas categorias	55
Tabela 3.	Categorias da variável idade e dias de internação	56
Tabela 4.	Frequência e proporção do conjunto de dados demográficos das vítimas de acidente de transporte, segundo variáveis selecionadas, Cuiabá / MT, 2006	61
Tabela 5	Frequência e proporção do conjunto de dados ocorrência e das vítimas de acidente de transporte, segundo variáveis selecionadas, Cuiabá / MT, 2006	62
Tabela 6	Frequência e proporção do conjunto de dados das vítimas de acidente de transporte, segundo variáveis assistência e evolução, Cuiabá / MT, 2006	63
Tabela 7	Confiança, suporte mínimos estipulados e número regras geradas pelo algoritmo <i>APriori</i>	64
Quadros		
Quadro 1.	Variáveis selecionadas nos bancos de dados da SEJUSP, SIH/SUS e SIM	46
Quadro 2.	Variáveis de estudo e suas categorias	54
Quadro 3.	Regras associativas e seus parâmetros o <i>Interest</i> , suporte e confiança	68

LISTA DE ABREVIATURAS E SÍMBOLOS

AG	Algoritmos Genéticos
AIDS	<i>Acquired Immune Deficiency Syndrome</i>
AIH	Autorização de internação hospitalar
BI	<i>Business Intelligence</i>
BO	Boletim de Ocorrência
CEPROMAT	Centro de Processamento de Dados do Estado de Mato Grosso
CID	Classificação Internacional de Doenças
CID - 10	Classificação Internacional de Doenças- Décima Revisão
CNES	Cadastro Nacional dos Estabelecimentos de Saúde
CNPQ	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONF	Confiança
CPF	Cadastro de Pessoas Físicas
CTB	Código de Trânsito Brasileiro
DATASUS	Departamento de Informação e Informática do Sistema Único de Saúde
DETRAN	Departamento Estadual de Transito
DM	<i>Data Mining</i>
DN	Data de Nascimento
DO	Declaração de óbito
EUA	<i>Estados Unidos da América</i>
EX	Exemplo
HCANCERMT	Hospital do Câncer de Mato Grosso
HGU	Hospital Geral Universitário
HIV	Vírus da Imunodeficiência Humana
HJDCBA	Hospital Jardim Cuiabá
HMBOMJESUS	Hospital Municipal Bom Jesus
HPSMCBA	Hospital e Pronto Socorro Municipal de Cuiabá

HSTAHELENA	Hospital Santa Helena
HSTAROSA	Hospital Santa Rosa
HUJM	Hospital Universitário Julio Muller
IBGE	Instituto Brasileiro de Geografia e Estatística
IBM	<i>INTL BUSINESS MACH</i>
IPDU	Instituto de Pesquisa e Desenvolvimento Urbano
IPEA	Instituto de Pesquisa Econômica Aplicada
MD	Mineração de Dados
Medline	National Library of Medicine
MS	Ministério da Saúde
MT	Mato Grosso
OMS	Organização Mundial de Saúde
RNA	Rede Neural Artificial
SAMU	Serviço de Atendimento Móvel de Urgência
SAS	<i>Statistical Analysis System</i>
SBIS	Sociedade Brasileira de Informática
SCIELO	<i>Scientific Eletronic Library online–Brasil</i>
SEJUSP	Secretaria de Estado de Justiça e Segurança Pública
SEPLAN	Secretaria de Planejamento
SES	Secretaria de Estado de Saúde
SES/MT	Secretaria de Estado de Saúde de Mato Grosso
SGBD	Gerenciadores de Banco de Dados
SIG-MT	Sistema de Informações Gerenciais
SIH/SUS	Sistema de Informações Hospitalares do Sistema Único de Saúde
SIM	Sistema de Informações sobre Mortalidade
SINAN	Sistema de Informações de Agravos de Notificação
SINF	Sem Informação (campo em branco)
SIOPS	Sistema de Informações sobre Orçamentos Públicos em Saúde
SIS	Sistema de Informação de Saúde
SMS	Secretaria Municipal de Saúde

SOTRAUMA	Hospital Só Trauma
SPSS	<i>Statistical Package for the Social Sciences</i> for Windows
SROP	Sistema de Registro de Ocorrências Policiais
STACASA	Hospital Santa Casa
SUP	Suporte
SUS	Sistema Único de Saúde
UF	Unidade Federada
UFMT	Universidade Federal de Mato grosso
UNIFESP	Universidade Federal de São Paulo
WEKA	<i>Waikato Environment for Knowledge Analysis</i>):

RESUMO

Galvão ND. **Aplicação da Mineração de Dados em Bancos da Segurança e Saúde Pública em Acidentes de Transporte**. São Paulo; 2009. (Tese de Doutorado – Departamento de Enfermagem da Universidade Federal de São Paulo – DENF/UNIFESP).

Introdução: Com a evolução científica e tecnológica, grandes volumes de dados têm sido coletados e armazenados. Para investigar tais bases de dados, a informática em saúde apropria-se de metodologias como a *Knowledge Discovery in Databases* (KDD), ou seja, descoberta de conhecimento das bases de dados. A mineração de dados – ou *Data Mining*, umas das etapas do KDD – permite investigar os dados à procura de padrões, muitas vezes não visíveis pela simples observância. **Objetivo:** Identificar, por meio da aplicação da tecnologia de mineração de dados, regras sobre acidentes de transporte ocorridos no município de Cuiabá, Mato Grosso, no ano de 2006, a partir dos dados da Secretaria de Justiça Segurança Pública, do Sistema Único de Saúde municipal de Cuiabá (morbidade e mortalidade). **Método:** Por meio de estudo observacional, transversal, retrospectivo e exploratório das bases de dados da segurança e saúde pública do município de Cuiabá / MT, do ano de 2006, foram relacionados os três bancos pelo método probabilístico, por meio do *software* livre Reclink. Foram obtidos 139 pares verdadeiros de vítimas de acidentes de transporte terrestre. Neste banco relacionado, foi aplicada a tecnologia de minerar dados, por meio do algoritmo *APriori*, o *software* utilizado foi WEKA, também de domínio livre. **Resultados:** Uma análise preliminar na fase de pré-processamento da ferramenta WEKA mostrou que, das 139 vítimas de acidentes, 80,6% eram do sexo masculino, na faixa etária de 20-29 anos (41,7%). O tipo de vítima que predominou foi condutor do veículo (35,3%), o meio de transporte da vítima ou tipo de veículo da vítima foi a motocicleta (33,1%). A colisão foi o principal tipo (51,8%) de acidente verificado pela análise. A maioria das vítimas teve assistência médica (87,1%), sendo o Hospital Pronto Socorro Municipal de Cuiabá (HPSMCBA) o local que mais recebeu vítimas desse conjunto de dados (36,7%); em média, cada vítima ficou hospitalizada cinco dias. Com aplicação do algoritmo *APriori*, foram geradas cinquenta melhores regras, sendo que seis indicaram um conhecimento compreensível e útil para caracterização das vítimas de acidentes em Cuiabá. **Conclusão:** Com base nestes resultados, programas de ensino e prevenção podem ser estabelecidos e, portanto, vale considerar que tecnologia de mineração de dados como uma poderosa ferramenta na análise de dados secundários, subsidiando o processo de tomada de decisões com a extração de conhecimento útil de bases de dados provenientes dos sistemas de informações da saúde e da segurança pública.

Palavras-chave: mineração de dados, acidente de transporte, informática em saúde

ABSTRACT

Galvão ND. **Use of Data Mining in the Health Banks and Public Security in Traffic Accidents.** São Paulo; 2009. (Doctoral Thesis – Nursing Department of the Federal University of São Paulo – DENF/UNIFESP).

Introduction: With the scientific and technological evolution a large number of data have been collected and stored. In order to investigate such databases, the health informatics uses the *Knowledge Discovery in Databases* (KDD) methodologies, that is, knowledge discovery of the databases. *Data Mining*, one of the phases of KDD permits the investigation of data in search for standards, often not visible just by simple observation. **Aim:** To identify, through the use of the data mining technology, rules about traffic accidents making use of data of the Justice Secretariat and Public Security, of the Unified Health System municipal of Cuiabá (morbidity and mortality). **Method:** An exploratory, retrospective, observational, cross-sectional study of the databases of security and public health of the municipality of Cuiabá-MT, in 2006 was used. The three banks were related using the probabilistic method, through the free *software* RecLink. A hundred and thirty-nine (139) true pairs of road accident victims were obtained. In this related bank the mining data technology was applied, using the *APriori* algorithm, the *software* used was WEKA, also of a free domain. **Results:** A preliminary analysis in the pre-processing phase of the WEKA tool, showed that of the 139 victims of accidents, 80,6% were male, between 20-29 years of age (41,7%). Most of the victims were drivers (35,3%), the means of transportation used by the victim was the motorcycle (33,1%). Collision was the main cause of accident (51,8%) verified by the analysis. Most of the victims received medical assistance (87,1%), and the Municipal Emergency Hospital of Cuiabá (HPSMCBA) received most of the victims of this set of data (36,7%); in average each victim remained in hospital for five days. With the application of the *APriori algorithm*, fifty best rules were created, six of them, indicated a useful and comprehensible knowledge to characterize the victims of accidents in Cuiabá. **Conclusion:** Based on these results, teaching and prevention programs can be established and so, it is worth considering the data mining technology as a powerful tool in the analysis of secondary data, helping the decision-making process with the extraction of useful knowledge of databases originated from the health information systems and public security.

Key-words: data mining, traffic accidents, Health Informatics.

1 INTRODUÇÃO

Nas últimas décadas, em que a maioria das operações e atividades das instituições privadas e públicas é registrada computacionalmente e acumula-se em grandes bases de dados, a mineração de dados – *Data Mining* (DM) – é uma das alternativas mais eficazes para extrair conhecimento a partir de grandes volumes de dados, descobrindo relações ocultas, padrões e gerando regras para prever e correlacionar dados, que podem ajudar as instituições nas tomadas de decisões mais rápidas ou até mesmo a atingir um maior grau de confiança ¹.

Hoje, a informação e o conhecimento são prerrogativas legais, estratégicas e imprescindíveis na busca de maior autonomia nas ações das empresas, de saúde, controle social e na tomada de decisão com prazos cada vez mais curtos. Por isso, diversas empresas nacionais e internacionais de produção, consumo, mercado financeiro, instituições de ensino e bibliotecas já adotaram nas suas rotinas a mineração de dados para monitorar arrecadações, consumo de clientes, prevenir fraudes e previsão de riscos do mercado, dentre outras ¹⁻²⁻³⁻⁴. No setor saúde, principalmente no público, a aplicação está sendo aceita como uma forma de agilizar a busca de conhecimento. Além do mais, a utilização da mineração de dados nos grandes bancos de dados hospitalares ou até mesmo nos sistemas de informação de saúde pública contribui para descobrir relacionamentos para que, desta forma, possa ser feita uma previsão de tendências futuras baseadas no passado, caracterizando melhor o paciente que busca assistência, identificando terapias médicas de sucesso para diferentes doenças e demonstrando padrões de novos agravos.

Contudo, há uma apreensão por parte de vários gestores e profissionais de saúde em compreender os dados e em utilizar a informação e conhecimento das bases de dados da saúde para promover a gestão da informação e qualidade de cuidados ⁵⁻⁶. Isso provavelmente é decorrência de um ritmo alucinante de geração de dados ¹, que produz uma incapacidade natural no ser

INTRODUÇÃO

humano de explorar, extrair e interpretar estes mesmos dados para obter conhecimento dessas bases.

Por outro lado, os profissionais de saúde, enquanto executores de profissões destinadas à assistência à saúde e ao cuidado do ser humano no processo saúde-doença, defrontam-se em sua vivência diária com grande número de informações, referentes a todas as medidas administrativas necessárias para garantia do acesso, da oferta, da equidade, universalidade e controle social nos serviços de saúde.

Nesse sentido, a informática e as tecnologias voltadas para coleta, armazenamento e disponibilização de dados vêm evoluindo e disponibilizando técnicas, métodos e ferramentas computacionais automatizados, capazes de auxiliar na extração de informações úteis contidas nesses grandes volumes de dados complexos⁶⁻⁷.

O campo de informática em saúde tem apresentado um crescimento extraordinário na última década. Segundo Marin et al.⁸, a informática é uma ciência que tem finalidade de tratar do processamento automático das informações, colaborando na área de saúde por meio da utilização de recursos tecnológicos para o armazenamento e a manipulação dos dados nos sistemas de informações.

A utilização da informática em saúde surgiu justamente com o propósito de facilitar o acesso à informação, auxiliando assim, a prática nos serviços de saúde. A informação em saúde, no entanto, é extremamente complexa e pouco estruturada e, durante muito tempo, a melhor forma que se encontrou para representar a informação foi por intermédio do simples armazenamento de documentos descritos em texto livre.

Porém, os computadores precisam de dados estruturados para poderem processar eficientemente a informação. Assim, foram criadas complexas bases de dados em saúde. Mas, atualmente, não basta mais ter simplesmente uma grande base de documentos ou fichas. É importante também, a partir de um conjunto de fichas, ser capaz de perguntar e oferecer respostas a estas perguntas

INTRODUÇÃO

que os computadores podem efetivamente auxiliar na prática de assistência à saúde.

Segundo Moraes ⁹, as tecnologias oriundas da eletrônica e da informática, com relação às informações, abrem possibilidades imensas para seu uso e sua aplicação na área da saúde, tanto no uso da telemedicina como na disseminação da informação para a sociedade, mas principalmente, com a perspectiva de difusão de experiências locais por intermédio da Internet, os *websites* de cada município, estado e/ou instituição.

As tecnologias de informação garantem a ampliação e consolidação não só de redes invisíveis de olhares, mas também de vozes e imagens que ligam, interligam, separam ideias e visões de mundo, em todo o globo terrestre. Com os meios de comunicação instantâneos, tais recursos fazem parte da vida cotidiana de indivíduos ou da coletividade.

Segundo Halamka ¹⁰, a tecnologia de informação em saúde possui pontos positivos quando implantados nos serviços de saúde, tais como:

1. o aumento da adesão aos protocolos clínicos;
2. incremento na capacidade de executar vigilância e monitorar as condições da doença;
3. redução de erros nas medicações e custos de diagnósticos; e
4. maior aproveitamento do tempo dos profissionais de saúde.

Além disso, o autor ressalta a utilização em grande escala da tecnologia e informação em saúde nos países, tais como Reino Unido, Canadá e Suécia, apesar das limitações de retroalimentação que ocorrem dos bancos de dados entre as unidades de serviços.

Em uma revisão sistemática, Chaudhry et al. ¹¹, demonstraram que a tecnologia e informação em saúde melhora significativamente a qualidade e

INTRODUÇÃO

eficiência dos serviços médicos. Esta revisão contemplou 257 estudos que atenderam aos critérios de inclusão. A maioria dos estudos tratou dos sistemas de apoio à decisão ou registros eletrônicos de saúde. Sobre os achados da qualidade, os autores enfatizam que as instituições que utilizam a tecnologia e informação tiveram crescente aderência aos protocolos sobre cuidados, aumento na vigilância e monitoramento ao paciente e diminuição nos erros de medicação. Já os efeitos da tecnologia e informação sobre a eficiência foram observados por meio de redução das taxas de utilização dos serviços médicos.

O início da utilização dos recursos da informática em larga escala no Brasil remonta à década de 1970. Desde então, um crescente contingente de serviços de saúde pública tem-se beneficiado de tais recursos. Com avanço da informática nas últimas décadas, cada vez mais a área da saúde tem conseguido obter melhoria, rapidez na ação e na resolução de problemas básicos. Porém, a informatização dos dados foi surgindo a partir de planilhas e relatórios, até estes serem substituídos por Sistemas de Informação informatizados.

Os sistemas de informação têm sido desenvolvidos para otimizar o fluxo de informação relevante no âmbito de uma organização, desencadeando um processo de conhecimento e de tomada de decisão e intervenção ¹².

Os Sistemas de Informação em Saúde (SIS) agregam milhões de dados e permitem inúmeras possibilidades de uso. Muitos sistemas consistem-se de bases complexas que dificultam o manuseio dos dados e informações pelos profissionais de saúde e ainda, a interação entre os sistemas é muito difícil, pois cada sistema possui chaves e linguagem diferentes. Ou seja, o interfaceamento de sistemas em saúde é ainda um grande desafio na área.

Entretanto, observa-se uma preocupação crescente em agilizar, ampliar a função destes sistemas e apropriar-se do potencial de variáveis e informações disponibilizadas para verificar as definições mais acertadas na priorização de ações no setor saúde.

Muitas vezes, a falta de clareza quanto à informação realmente necessária faz com que sejam coletados dados que não terão qualquer utilidade,

INTRODUÇÃO

sobrecarregando sistemas. Grande quantidade de dados parece dar a ilusória sensação de que a situação está sob controle ou de que se tem tudo o que é necessário para conhecer um problema.

Os dados nos sistemas computadorizados, via de regra, não se comunicam com os demais sistemas e entre as diferentes instituições, os quais apresentam uma limitação e geram custos para os serviços de saúde. Em Massachusetts, Estados Unidos, aproximadamente 15% de testes de diagnóstico são repetidos, em um custo estimado de US\$ 4.5 bilhões por ano, devido a não retroalimentação de dados entre instituições e também a não comunicação entre os diversos sistemas ¹⁰.

Há quase vinte anos, vem-se enfatizando que os dados e a informação em saúde devem ter acesso amplo, rápido e fácil; ou seja, o público em geral deve ter acesso, de modo a assegurar a transparência e democratização da administração ¹³. Na verdade, na concepção do Sistema Único de Saúde (SUS), todos os níveis de decisão devem agregar banco de dados que resultará em um banco central, que permitirá instrumentar um sistema de planejamento e de acompanhamento nos cinco níveis de gestão, a saber: unidade de saúde, município, distrito ou região, estado e país.

Esta realidade também acontece nos dados e informações sobre acidentes de transporte terrestre que se encontram fragmentados nos mais diversos sistemas de informações do setor saúde e da segurança pública. Isto, conseqüentemente, dificulta a tomada de decisão pelos profissionais, nas mais distintas instituições envolvidas neste problema.

A morbidade e mortalidade crescentes por este agravo têm gerado preocupação constante na sociedade moderna. Para o sistema de saúde, as conseqüências dos acidentes de transporte terrestre, dentre outros aspectos, evidenciam-se no aumento de gastos com emergência, assistência e reabilitação.

Desta forma, as unidades de serviços de saúde, antes muito mais orientadas para enfermidades de origem biomédicas, são hoje chamadas para dar

INTRODUÇÃO

respostas às vítimas de lesões e traumas físicos e emocionais, devendo equipar-se para isso ¹⁴.

A mortalidade por causas externas, no entanto, somente representa a fração mais visível e mais trágica do problema, sem revelar o que esses agravos significam em termos de morbidade, não expressando o número de feridos, de inválidos ou de incapacitados, temporária ou permanentemente. A crescente importância dos adolescentes e jovens, como grupos e como vítimas da violência, relaciona-se com as mudanças cada vez mais rápidas que se processam nos contextos social, econômico e cultural.

Embora as causas externas tenham em sua gênese fatores não diretamente relacionados ao setor saúde, é sobre esse setor que vai recair o maior ônus das consequências dos acidentes e das violências. Para o SUS, isso representa minimamente, maior demanda por assistência geral e especializada, comprometimento crescente de recursos (geralmente escassos) e dificuldade para suprir outras necessidades de saúde, em função dessa demanda. Além disso, estas causas geram ônus para a Previdência Social, na forma de aposentadorias precoces, pensões e auxílios-doença.

No estado de Mato Grosso, a mortalidade por causas externas representa 1/5 das mortes e tem sido a segunda causa de óbito há quase uma década, e a primeira entre a população masculina com menos de 50 anos. Representa também mais de 5,0% de todas as internações do SUS no estado e quase 7,0% dos gastos hospitalares, nada se conhecendo sobre os custos sociais e econômicos relacionados às vítimas fatais e não fatais desses agravos, neste estado ¹⁵.

A relevância dos acidentes e violências em Cuiabá, capital do estado de Mato Grosso, é revelada pelo quadro de mortalidade por causas externas. No período de 1991 a 2000, entre as capitais brasileiras, Cuiabá apresentou um dos maiores crescimentos das taxas de mortalidade por acidentes de transporte ¹⁶ e o maior incremento da taxa de mortalidade por armas de fogo (690,0%) ¹⁷.

INTRODUÇÃO

Em 2005, entre as capitais brasileiras, Cuiabá apresentou a sexta maior taxa de mortalidade por causas externas na população geral (90,9/100.000 hab), a quinta em acidentes de transporte e a sexta em homicídios. Os dados de morbidade hospitalar para Cuiabá revelam-se acima dos índices do Brasil, da região Centro-Oeste e do estado de Mato Grosso ¹⁸.

Neste sentido, iniciativas visando ao maior conhecimento do problema e à proposição de intervenções, são oportunas e relevantes para o SUS e mesmo para a sociedade.

Por esta consideração, o presente trabalho teve como objetivo identificar nos conjuntos de dados da Secretaria de Estado de Justiça e Segurança Pública (SESJUSP), Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS) e Sistema sobre Informação de Mortalidade (SIM) relacionados entre si, e por meio da tecnologia de mineração de dados, identificar nas regras de associação os padrões para acidentes de transporte ocorridos em Cuiabá no ano de 2006.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Caracterização dos acidentes de transporte

Os acidentes configuram um conjunto de agravos à saúde que podem ou não levar ao óbito, nos quais se incluem as causas ditas acidentais, ou seja, em razão do trânsito, trabalho, quedas, envenenamentos, afogamentos e outros tipos de acidentes.

Atualmente, esse conjunto de eventos consta na Classificação Internacional de Doenças, 10ª revisão – CID 10¹⁹, sob a denominação de causas externas e com os seguintes códigos:

- Acidentes: acidentes de transporte (V 00 – V 99) e demais acidentes (W 00 –X 59).

Quanto à natureza da lesão, tais eventos e/ou agravos englobam todos os tipos de lesões e envenenamentos, como ferimentos, fraturas, queimaduras, intoxicações, afogamentos, entre outros.

A Política Nacional de Redução da Morbimortalidade por Acidentes e Violências²⁰ define acidente como evento não intencional e evitável, causador de lesões físicas e/ou emocionais no âmbito doméstico ou nos outros ambientes sociais, como o do trabalho, do trânsito, da escola, de esportes e o de lazer.

Os acidentes de transporte terrestre são eventos não intencionais e com uma magnitude importante em todo o mundo, expressa no grande número de mortes, incapacidades e sequelas psicológicas, sendo um dos fatores negativos da transição epidemiológica que o Brasil tem experimentado nas últimas décadas²¹⁻²².

Em 2000, em todo mundo, mais de 1,2 milhão de pessoas morreram em consequência de acidentes de trânsito, a maioria adultos jovens.

FUNDAMENTAÇÃO TEÓRICA

Cerca de 90% dessas mortes ocorreram em países subdesenvolvidos e em desenvolvimento. O perfil das vítimas nesses países indica uma alta proporção de mortes entre pessoas mais vulneráveis, como pedestres, ciclistas, motociclistas e passageiros de ônibus e caminhão ²³.

O Brasil está entre os países que tem registrado as maiores frequências de indivíduos feridos por acidentes de trânsito, com taxas que, em 2002, alcançaram 219,5 vítimas por 100 mil habitantes ²⁴ e de 18,7 mortes por 100 mil habitantes ²⁵. Ao lado da grande importância social desse agravo, estima-se que o impacto econômico para o país seja de aproximadamente 5,3 bilhões de reais, o que equivale a 1% do Produto Interno Bruto (PIB) nacional por ano. Vale destacar que estes custos somente foram dos acidentes em aglomerações urbanas ²⁶.

A mortalidade por causas externas foi a terceira principal causa conhecida de óbitos no Brasil em 2001. Neste grupo, ocorreram 31 mil óbitos por acidentes de transporte terrestre, o que representou 25,7% do total de óbitos por causas externas. Os adultos jovens do sexo masculino constituíram o grupo mais atingido por acidentes de transporte terrestre ²⁵. Pode-se dizer que no conjunto de causas externas, os acidentes de transportes destacaram-se em termos de magnitude, tanto de mortes quanto de feridos.

O novo Código de Trânsito Brasileiro (CTB), instituído pela lei número 9.503 ²⁷, privilegia as questões de segurança e de prevenção da vida. Uma de suas características é conter expressivo conjunto de medidas de prevenção, não sendo, por conseguinte, um instrumento apenas punitivo. A sua implantação configura, assim, mecanismo legal e eficaz para diminuição dos principais fatores de risco, envolvendo condutor, pedestre, veículos e via pública ^{28, 20}. O impacto da implantação do novo CTB tem sido estudado há cerca de uma década, pelos autores Bastos et al. ²⁹ e Liberatti et al. ³⁰ que revelam uma redução dos feridos e das vítimas fatais de acidentes na região sul do Brasil.

Porém, a mortalidade por acidente de transporte terrestre é um indicador que gera discussão no meio dos gestores do trânsito. Há quase dez anos

desde a criação do atual CTB, ainda se constata uma elevada mortalidade. Em 2006, os brasileiros morreram por acidentes de transporte terrestre com taxa de 39,4 por 100 mil habitantes. No mesmo ano, o estado de Mato Grosso, apresentou um risco 56,8 por 100 mil habitantes, já a capital registrou uma taxa menor de 25,2 por 100 mil habitantes ²⁵.

Em relação aos acidentes de transporte, Cuiabá apresentou um dos maiores crescimentos das taxas de mortalidade no período de 1991 a 2000 ¹⁶.

Referente à morbidade hospitalar, dados do SIH/SUS mostram que, em 2002, das internações ocorridas no Brasil, cerca de 8%, excluindo-se os partos (capítulo XV, CID 10), foram por lesões e traumatismos decorrentes de causas externas (capítulo XIX e capítulo XX, CID 10) ³¹.

No Brasil, referentes às internações ocorridas no SUS, em 2002, os principais tipos de causas externas foram: quedas (34,3%), acidentes de transporte (19,6%) e tentativas de homicídios (9,9%) ³¹. Em 2000, as quedas ocuparam o primeiro lugar entre os motivos de internações, com 279.336 pessoas hospitalizadas, representando 42,8% do total; 58,5% dessas internações foram decorrentes de fraturas, especialmente em membros superiores (29,0%), seguidas dos membros inferiores (24,7%). O sexo masculino apresentou maior número de internações com diagnóstico de fraturas, porém especificamente em relação às fraturas do fêmur, o número de mulheres hospitalizadas aproxima-se do número de homens ³².

Neste mesmo ano, os acidentes de transportes foram responsáveis por um número considerável de hospitalizações no país (118.623), representando 18,2% do total de internações, sendo que os atropelamentos determinaram 39,5% ³¹. No Rio de Janeiro, Deslandes & Silva ³³, estudando os atendimentos de urgência e emergência em dois hospitais, identificaram que 11,8% dos atendimentos foram provenientes de acidentes de trânsito. Os dados de morbidade hospitalar para Cuiabá revelam-se também acima dos índices do Brasil, da região Centro-Oeste e do estado de Mato Grosso ³¹.

Os acidentes de trânsito tornaram-se um importante problema de

FUNDAMENTAÇÃO TEÓRICA

saúde pública e sobrecarregam o setor, em função dos elevados percentuais de internação e dos altos custos hospitalares, além de gerarem problemas para a sociedade, tais como perdas materiais, despesas previdenciárias e grande sofrimento para as vítimas e seus familiares.

Estudo de Bastos et al.²⁹ constatou redução de 12,3% e 18,5% do número de acidentes e de vítimas, em períodos de três meses cada, anteriores e posteriores à implantação do novo CTB em Londrina / PR.

As morbidades de acidentes de trânsito são pouco estudadas, embora representem importante parcela da demanda por atendimento de emergência nos hospitais. Em artigo de revisão sobre acidentes de trânsito nacionais e internacionais, observou-se haver o dobro de estudos sobre mortalidade em relação aos estudos sobre morbidade por acidentes de trânsito³⁴.

2.1.1 Fonte de dados

Em relação aos acidentes, são várias as fontes a partir das quais se pode investigá-los, cada uma constituída de modo a satisfazer as necessidades institucionais que as geram. Além disso, sofrem diretamente a influência das limitações características dos sistemas de notificação, às vezes difíceis de serem compatibilizados. Desse modo, os resultados das investigações são divergentes, dependendo da fonte consultada, ocasionando distorções e erros interpretativos e, conseqüentemente, dificuldade na tomada de decisões³⁵.

As principais fontes oficiais de informação na saúde para o estudo da morbidade e da mortalidade dos acidentes são: o Sistema de Informações Hospitalares do SUS (SIH/SUS) e o Sistema de Informação sobre Mortalidade (SIM).

O Sistema de Informações Hospitalares (SIH/SUS), que abarca informações das instituições públicas e conveniadas ou contratadas pelo SUS, engloba cerca de 80% da assistência hospitalar do país e seus dados dizem respeito

FUNDAMENTAÇÃO TEÓRICA

à aproximadamente 13 milhões de internações/ano. Nesse aspecto, é importante salientar que, até 1997, os dados de acidentes informavam somente a natureza da lesão que levou à internação, sem quaisquer esclarecimentos sobre as circunstâncias que redundaram naquela lesão. A partir de 1998, em decorrência da Portaria Ministerial n. 142 de 1997 ³⁶, os tipos de causas externas geradoras das lesões que ocasionaram a internação também passaram a ser codificados. Assinala-se que não existem sistemas de informação epidemiológica relacionados aos atendimentos em prontos-socorros e ambulatórios, nem estudos mais apurados relativos a acidentes e violências.

Salienta-se como aspecto positivo do SIH/SUS a agilidade do sistema para dar conhecimento do perfil dos atendimentos da rede hospitalar. Como aspectos negativos ressaltam-se: problemas de qualidade dos dados que aportam, em decorrência de fraudes e manipulações pelos prestadores de serviços; não identificação de reinternação e transferências de outros hospitais, o que leva eventualmente a duplas ou triplas contagens de um mesmo paciente e evidente precariedade das informações anotadas nos prontuários médicos ³⁷.

O Sistema de Informação sobre Mortalidade é atualmente gerenciado pela Secretaria de Vigilância em Saúde do Ministério da Saúde, foi desenvolvido e implantado no Brasil em 1975. Embora este sistema abranja atualmente mais de 900 mil óbitos/ano, a cobertura correspondente ainda não é completa em algumas áreas, principalmente nas regiões Norte e Nordeste. Quanto à qualidade das informações relativas a acidentes e a violências, ainda há deficiências. Os dados de óbitos por causas externas originam-se, em sua maioria, das Declarações de Óbitos (DO) preenchidas nos Institutos de Medicina Legal. A qualidade das informações é ainda discutível, na medida em que algumas vezes, não existe detalhamento quanto ao tipo ou intencionalidade da causa externa responsável pelas lesões que provocaram o óbito. Nesses casos, sabe-se apenas que se trata de uma morte decorrente de causas externas. Tal ocorrência vem sendo verificada em cerca de 10% do total de mortes por acidentes no país, alcançando em algumas áreas, valores bem mais elevados. Em determinadas

FUNDAMENTAÇÃO TEÓRICA

localidades, essa distorção tem sido sanada ou minimizada com o auxílio de informações de outras fontes, tais como consultas ao prontuário hospitalar, laudos de necropsia ou utilização de noticiários de jornais ²⁰.

Alguns dos aspectos positivos do SIM são a universalidade do sistema e a padronização das informações, principalmente aquelas referentes à causa básica, o que permite a comparabilidade dos dados de diferentes localidades, possibilitando maior confiabilidade ao mesmo. Salientam-se como aspectos negativos deste sistema o acentuado número de óbitos por Sintomas e Sinais Mal Definidos, o preenchimento inadequado de algumas variáveis da DO e a subnotificação ^{37, 20}.

A Secretaria de Estado de Justiça e Segurança Pública (SEJUSP) é ainda uma fonte de dados imprescindível sobre os acidentes de transporte, uma vez que coleta, processa e armazena os dados dos atendimentos móveis de urgência e emergência realizados pelo Corpo de Bombeiros e dos do Serviço de Atendimento Móvel de Urgência (SAMU) até meados de 2007. As ocorrências da Polícia Militar registradas no Boletim de Ocorrência (BO), estão armazenadas no Sistema de Informações Gerenciais, o SIG-MT, adquirido recentemente pelo estado de Mato Grosso, sendo caracterizado como uma ferramenta de consulta de dados dos sistemas transacionais. A tecnologia utilizada, *Business Intelligence* (BI), consiste em filtrar dados com o objetivo de retirar informações úteis e relevantes para a tomada de decisão e alcance de resultados, além de permitir que as consultas sejam rápidas. A ferramenta ainda permite realizar correlações de variáveis, elaborar gráficos, fazer consultas via Internet, reduzindo assim custos e esforços, possibilitando a padronização das informações e conseqüentemente a melhoria na tomada de decisão por meio de informações mais rápidas e precisas. Vale destacar que proporciona melhoria na adaptação da organização para enfrentar os acontecimentos não previstos, além de permitir ao usuário a inteira padronização e personalização de seus relatórios. A gestão deste sistema é realizada pelo Centro de Processamento de Dados do Estado de Mato Grosso (CEPROMAT).

A coleta de dados é realizada pelos militares (Polícia e/ou

Bombeiro) que registram dados do fato acontecido em uma ficha com maior número de informações possíveis; depois, esse policial encaminha este formulário a uma central de registros de ocorrências ou a algum batalhão e, então, os dados dessa ficha são inseridos em um sistema chamado Sistema de Registro de Ocorrências Policiais (SROP), que faz uma replicação desses dados nos servidores da SEJUSP que se localizam fisicamente no CEPROMAT. Deste modo, as informações se juntam a uma base de dados de ocorrências policiais para que possam mais tarde ser acessadas pelos usuários do sistema, para somente então serem analisadas.

Esses bancos de dados, tanto da saúde como da SEJUSP, são importantes para o planejamento e gestão; porém, não possuem nenhum tipo de relacionamento, o que dificulta a verificação da trajetória do paciente fatal ou não fatal e também a tomada de decisão intersetorial.

Entretanto, o interesse em relacionar registros em diferentes bases de dados vem aumentando nas últimas décadas, juntamente com a crescente disponibilidade de grandes bases de dados em saúde informatizadas.

2.1.1.1 Relacionamento de banco de dados

Relacionar registros em diferentes bases de dados é uma tarefa trivial nos casos onde os registros de cada base incluem um atributo ou variável comum, pouco sujeito a erros e que permita a identificação de cada ocorrência de forma unívoca ³⁸. Porém, nem todas as bases de dados, por exemplo da saúde pública – SIM e SIH/SUS –, possuem este identificador unívoco comum.

Este processo de relacionamento pode ser realizado por meio de dois métodos: o determinístico e o probabilístico. Denomina-se relacionamento determinístico quando a busca é feita por uma concordância exata entre uma ou mais variáveis existentes em um ou mais arquivos e quando existe um código ou identificador unívoco comum entre as bases, como por exemplo, o Cadastro de

Pessoas Físicas (CPF). Já o relacionamento probabilístico de bases de dados, pode ser definido como um processo de pareamento de duas (ou mais) bases, que utiliza probabilidades de concordância e discordância entre um conjunto de variáveis comuns às duas bases ³⁹. Os pares, usualmente, são definidos a partir das variáveis nome, nome da mãe, data de nascimento. Informações como renda, educação, entre outras, podem ser utilizadas, na dependência da qualidade destes campos ⁴⁰. O principal objetivo do relacionamento probabilístico de base de dados é encontrar pares de registros que se referem a uma mesma pessoa.

Segundo Camargo Jr. & Coeli ⁴¹, o relacionamento probabilístico baseia-se em um conjunto de processos: (1) a utilização de rotinas para a padronização dos campos comuns a serem empregados no relacionamento (ex: quebra do campo nome em seus componentes e a formatação de campos data); (2) a blocagem (*blocking*), que consiste na criação de blocos lógicos de registros dentro dos arquivos a serem relacionados, permitindo que a comparação entre registros se faça de uma forma mais otimizada; (3) a aplicação de algoritmos para a comparação aproximada de cadeias de caracteres, que levam em consideração possíveis erros fonéticos e de digitação (ex. Manoel e Manuel seriam reconhecidos como iguais); (4) o cálculo de escores, que sumarizam o grau de concordância global entre registros de um mesmo par; (5) a definição de limiares para a classificação dos pares de registros relacionados em pares verdadeiros, não pares e pares duvidosos; (6) a revisão manual dos pares duvidosos visando à classificação dos mesmos como pares verdadeiros ou não pares; (7) a verificação de duplicidade de registros em arquivos.

No método probabilístico, utiliza-se o sistema fonético - código *soundex*, para minimizar erros com eliminação de vogais e substituição de consoantes com sons similares da variável nome, a fim encontrar maior número de pares verdadeiros. Este código é formado por quatro dígitos, sendo o primeiro representado pela primeira letra da palavra a ser codificada enquanto os demais são dígitos numéricos codificados segundo regras de eliminação de vogais, substituição

FUNDAMENTAÇÃO TEÓRICA

de consoantes com sons similares por um código numérico comum. Por exemplo, o soundex da palavra Pedro é P360, enquanto o da palavra Souza é S200.

No sistema fonético (soundex) relativo ao primeiro e último nome, a primeira sílaba é modificada segundo as seguintes transformações descritas no manual do RecLink III ⁴¹:

- Primeira letra W e segunda A – Primeira letra passa a V;
- Primeira letra H – Deleta primeira letra;
- Primeira letra K e segunda A, O ou U – Primeira letra passa a C;
- Primeira letra Y – Primeira letra passa a I;
- Primeira letra C e segunda E ou I – Primeira letra passa a S;
- Primeira letra G e segunda E ou I – Primeira letra passa a J;

Vale ressaltar que o pareamento de registros é baseado na construção de escores para os diferentes pares possíveis de serem obtidos, a partir de uma determinada estratégia de blocagem empregada. O escore final de cada par é construído a partir da soma dos escores ponderados de cada campo empregado no processo de pareamento (nome, último nome, sexo e data de nascimento, por exemplo), permitindo, desta maneira, que cada campo contribua de forma diferenciada para o escore total do par. A contribuição diferenciada é recomendável, pois os campos apresentam diferente poder discriminatório e ao mesmo tempo, apresentam uma maior ou menor probabilidade de terem seus conteúdos registrados de forma incorreta. Por exemplo, o campo sexo apresenta um baixo poder discriminatório, mas o seu registro é geralmente feito de forma correta. Já o campo último nome, apesar de apresentar um bom poder discriminatório, é mais sujeito a erros de registro.

O grau de concordância e de discordância, para o cálculo dos escores, é obtido, respectivamente, pelas seguintes equações:

$$\text{Equação 1: } \log_2 \left(\frac{m_i}{u_i} \right) ; e$$

$$\text{Equação 2: } \log_2 \left[\frac{(1-m_i)}{(1-u_i)} \right]$$

Onde m_i é a probabilidade dos campos concordarem, dado que se trata de par verdadeiro (equivalente à sensibilidade); u_i é a probabilidade dos campos concordarem, dado que se trata de par falso (1-especificidade), $(1-m_i)$ refere-se ao complementar da sensibilidade e, finalmente, $(1-u_i)$ equivale à especificidade. Estes conceitos são muito utilizados entre os epidemiologistas na avaliação da acurácia de testes diagnósticos. Neste estudo, para a composição do escore total de cada par, foram utilizados valores sugeridos por Camargo Jr. & Coeli⁴¹, presentes também nos *links* <http://www.mediafire.com/?sharekey=bc99288de55df934d2db6fb9a8902bda> ou www.iesc.ufrj.br, para vários campos na etapa de pareamento. Esses campos, com o algoritmo de comparação, bem como os valores de m_i e u_i são utilizados para o cálculo dos escores.

Exemplificando o cálculo do escore total na situação de concordância completa do primeiro nome, último nome, inicial(is) do(s) nome(s) do meio e data de nascimento, o escore máximo será obtido por meio da soma apenas dos fatores de ponderação de concordância pela equação 2.

Portanto, nesse caso teremos a soma que resultará em aproximadamente 25,3:

$\log_2 \left(\frac{99}{2} \right) + \log_2 \left(\frac{99}{3} \right) + \log_2 \left(\frac{89}{3} \right) + \log_2 \left(\frac{91}{10} \right) + \log_2 \left(\frac{94}{23} \right) + \log_2 \left(\frac{92}{4} \right)$, cujas parcelas correspondem, respectivamente, aos escores dos campos: primeiro nome, último nome, inicial(is) do(s) nome(s) do meio, dia, mês e ano de nascimento.

Estas bases dados relacionadas estão sendo empregadas na saúde pública para monitorar a ocorrência de evento de interesse ou com objetivo de ampliar a quantidade de informação a ser obtida por cada unidade de estudo a partir da combinação de bases qualitativamente distintas, por exemplo, a mortalidade materna, neonatal e infantil⁴²⁻⁴³⁻⁴⁴⁻⁴⁵, a subnotificação de casos de AIDS⁴⁶⁻⁴⁷, as causas mal definidas⁴⁸ e diabetes *mellitus*⁴⁹⁻⁵⁰.

Porém, ainda que essas condições sejam satisfeitas para relacionamento probabilístico, existe limitação do método, decorrente das coincidências de identidades de indivíduos: mesmo nome, sexo e data de nascimento ^{38,41}.

Leles ⁵¹ destaca também problemas que são verificados em sistemas automáticos de relacionamento probabilístico, que são os da identificação inequívoca do par, ocasionada por preenchimentos diferentes de mesmos campos ou ausência de alguns campos que participam da identificação do campo-chave. Consequentemente, muitas vezes isto resulta em falha na determinação do grau de concordância e discordância nos processos de blocagem e o pareamento de registros torna-se lento e ineficiente, principalmente quando grandes bases de dados (na ordem de milhões de registros) estão envolvidas.

No entanto, mesmo com problemas, esta forma de relacionamento está se expandindo cada vez mais na área da saúde, tendo em vista as diversas bases de dados desenvolvidas sem nenhuma integração, já que unir registros de duas ou mais bases para criar um novo registro em uma nova base de dados, tem tido progressos em razão dos avanços que houve na área da informática em saúde ^{38, 41, 51}.

2.2 Informação e a informática em saúde

2.2.1 Informação em saúde

Segundo Moraes (2002, p. 11): ⁹

(...) “a informação em saúde deve ser trabalhada no sentido de reforçar os direitos humanos, de contribuir para eliminação da miséria e das desigualdades sociais e ao mesmo tempo subsidiar o processo decisório na área da

FUNDAMENTAÇÃO TEÓRICA

saúde, em prol da efetividade e qualidade dos serviços de saúde, em respeito à singularidade de cada indivíduo ao contexto de cada população”.

A autora busca o significado das informações em saúde na sociedade atual, enquanto um instrumento potente e potencial a serviço de um conhecimento. Para ter essa informação adequada, precisa-se de dados. Alguns autores⁵²⁻⁵³ reforçam que os dados se diferenciam de informação, isto é: os dados são a “matéria-prima” da informação, ou seja, são valores ainda não trabalhados ou, ainda, uma descrição limitada do real, desvinculada de referencial, um explicativo e difícil de ser utilizada por ser ininteligível⁵⁴.

Para Mota & Carvalho⁵⁵, o dado tende a ser unidimensional e pode ser a simples enumeração de eventos ou de características quantitativas ou ainda, a representação numérica das propriedades dos eventos que necessitam ser consolidadas e analisadas.

A informação refere-se a uma descrição completa de uma situação real associada a um referencial explicativo sistemático, que atua como uma ponte entre os fatos reais ou ideias de algumas pessoas e as ideias ou conhecimento de outras, constituindo uma representação simbólica de fatos ou ideias potencialmente capazes de alterar um estado de conhecimento⁵².

Os dados são elementos brutos ou “crus”, não interpretados, que estão sem significado e desvinculados da realidade, constituindo a matéria-prima para a informação. Já a informação é o dado com significado, dotado de relevância e propósito. Portanto, dados sem qualidade levam a informações e decisões da mesma natureza⁵⁶⁻⁵⁷.

Além do mais, os dados devem ser processados e este processamento ocorre dentro de uma estrutura organizada que dê significado ao contexto. Após o processamento, tem-se uma coleção de dados armazenados, organizados e estruturados, que são definidos como base de dados. A base de dados organizada e compartilhada facilita a recuperação da informação útil com maior agilidade para o processo de tomada de decisão⁵⁷.

Tendo em mãos esses conceitos, pode-se dizer que o conhecimento é o processo de transformação realizado por meio de interpretação e análise dos dados e informações, ou seja, insumos provenientes das diversas fontes, os quais são combinados na síntese de um produto final, o conhecimento. Por outro lado, o conhecimento não é estático, ele se modifica quando há uma interação com o ambiente e este processo é denominado aprendizado^{52,12}.

Esta distinção entre dados, informação e conhecimento, reforça a ideia de que a forma, como as pessoas ou até mesmo os profissionais da área da saúde, associa-se a estes fluxos informacionais, e não se resume apenas à oferta de informação⁵⁸. Ainda para Sousa⁵⁸, a informação integra o processo de produção de conhecimento, após ser assimilada, interiorizada e processada.

É por meio desse conhecimento e/ou aprendizado que a informação pode contribuir para assessoramento das decisões e compreensão mais efetiva da situação problema na área da saúde. E assim, a lógica da rapidez na busca de mais informação em um menor tempo reduz a absorção das informações e a tecnologia da informática passa a ser utilizada para racionalizar e até para melhorar a qualidade de vida dos profissionais de saúde e dos pacientes^{12, 9, 55}.

O conhecimento extraído das informações deve demonstrar características tais como: ser compreensível pelos usuários, interessante, útil, novo e ser correto para tomada de decisão^{59, 56}.

Os dados, informações e conhecimento são elementos essenciais à tomada de decisão, uma vez que determinam uma cadeia de agregação de valor, quando partilhados por meio de um sistema de comunicação bem estruturado, que envolva a gestão e instalação de infraestruturas tecnológicas adequadas⁵⁶.

Com o desenvolvimento de sistemas de informação computadorizados na saúde a partir da década de 1970, as pesquisas em grandes volumes de dados numa perspectiva descritiva e preditiva vêm-se norteando cada vez mais; no entanto, a facilidade de armazenar os dados traz uma apreensão para vários gestores e pesquisadores, que fazem questionamentos a respeito dos dados de diversas doenças e agravos relevantes a saúde.

Segundo Safran & Perreault ⁶⁰, o sistema de informação (SI) usado dentro de uma organização de assistência à saúde facilita a comunicação, integra a informação, documenta as intervenções do cuidado, garante o registro e garante o apoio nas demais funções da instituição.

Por outro lado, há uma grande preocupação dos pesquisadores quanto à utilização desses dados. Pelletier & Diers ⁶¹ destacam que os profissionais de saúde nem sempre se mostraram aptos para compreender os dados e utilizar a informação disponível para promover a gestão da informação e qualidade de cuidados.

No Brasil, a informática médica ou informática em saúde, como denominada atualmente, está sendo expandida como uma ciência aplicada na prática em busca das soluções dos grandes volumes de dados complexos em muitas universidades e instituições governamentais e não-governamentais ⁶².

Nakayama ⁶³ ressaltou em seu estudo que a informática em saúde tem como objetivo estabelecer um sistema para facilitar o fluxo, a circulação e aplicação do uso da informação em saúde.

2.2.2 Informática em saúde

As expressões “informática em saúde” e “informática médica” têm sido usadas como sinônimos, dado que as atividades relativas à saúde abrangem não só a medicina, mas também as demais áreas das ciências biológicas e da saúde. Pode-se dizer que informática em saúde é o estudo e uso de computadores e sistemas de comunicação e informação na assistência médica, ensino e pesquisa na área da saúde ⁶⁴.

Os autores Greenes & Shortliffe ⁶⁵ destacam que informática médica é a área que se relaciona com a cognição, o processamento de informação e comunicação de tarefas da prática médica, educação e pesquisa, incluindo a ciência

FUNDAMENTAÇÃO TEÓRICA

da informação e a tecnologia que apoia tais atividades. Já para Coiera ⁶⁶, a informática médica é o estudo de como o conhecimento médico é criado, modelado, compartilhado e aplicado.

Segundo Degoulet & Fieschi ⁶⁷, os médicos recorrem à informática para auxiliar atividades de coleta e organização de dados. Quando aplicadas na área médica, estas atividades são classificadas como sendo de “informática médica” e têm por objetivo “gerenciar recursos, dispositivos e métodos de otimização do armazenamento, recuperação e gerenciamento de informações biomédicas” ⁶⁸. Bastante difundida no exterior, a informática médica também é estudada no Brasil, tendo sido tratada, mais genericamente, como informática em saúde ⁶⁹.

A palavra médica deve-se à origem na área da medicina, mas ela refere-se a todos os profissionais dos serviços de saúde; deste modo, os autores ⁷⁰ do livro *Introduction to nursing informatics – Introdução à informática em enfermagem* – definiram informática em saúde como “o uso da tecnologia de informação no processo de tomada de decisão para o cuidado de pacientes, realizado por profissionais de saúde (p. 5)”

Gunther Eysenbach ⁷¹, pesquisador da University of Heidelberg da Alemanha, também aborda o crescimento da informática em saúde e que esta área está interligada com outras áreas, tais como: cuidado assistencial, saúde pública, promoção à saúde, educação em saúde, bibliotecas, ciência da comunicação, e talvez seja o campo que mais desenvolve em diversas áreas do campo médico.

A Sociedade Brasileira de Informática em Saúde (SBIS) traz definições em sua *home page* (www.sbis.org.br) ⁷². A primeira definição é dada por Blois & Shortliffe (1990) como "um campo de rápido desenvolvimento científico que lida com armazenamento, recuperação e uso da informação, dados e conhecimentos biomédicos para a resolução de problemas e tomada de decisão".

Outra definição disponibilizada afirma que: "A Saúde é uma das áreas onde há maior necessidade de informação para a tomada de decisões. A

FUNDAMENTAÇÃO TEÓRICA

Informática Médica é o campo científico que lida com recursos, dispositivos e métodos para otimizar o armazenamento, recuperação e gerenciamento de informações biomédicas. O crescimento da Informática Médica como uma disciplina deve-se, em grande parte aos avanços nas tecnologias de computação e comunicação, à crescente convicção de que o conhecimento médico e as informações sobre os pacientes são ingerenciáveis por métodos tradicionais baseados em papel, e devido à certeza de que os processos de acesso ao conhecimento e tomada de decisão desempenham papel central na medicina moderna" (acessível em: <http://www.thenewmedicine.med.br/sobre.htm>) .

Além do mais, o SBIS define áreas de atuação da informática em saúde tais como: Sistemas de Informação em Saúde; Prontuário Eletrônico do Paciente; Telemedicina; Sistemas de Apoio à Decisão; Processamento de Sinais Biológicos; Processamento de Imagens Médicas; Internet em Saúde e Padronização da Informação em Saúde.

Segundo Sigulem (1997 p. 12) ⁷³: "Informática médica é uma ciência que, a exemplo de outras disciplinas, como a biologia molecular ou a neurociência, tem raízes na história e nas ideias da teoria da informação. É caracterizada por seu objeto (medicina) e seus métodos (os de gerenciamento de informação). Informática médica evoca outras disciplinas, como a matemática, a estatística, a linguística e a ciência da cognição ou filosofia. É bem adequada à abordagem experimental: sugestão de hipótese; modelagem; experimentação, frequentemente na forma de desenvolvimento ou implantação de programas ou protótipos de sistemas de informação; avaliação; validação; e, por fim, generalização do processo."

Outros autores defendem que a área de informática em saúde é uma área interdisciplinar entre as áreas de ciências da computação e áreas da ciência da saúde. Da ciência da computação, a informática em saúde trouxe consigo o desenvolvimento de sistemas computacionais úteis e relativamente novos. Já das ciências da saúde, herdou avaliação empírica ou clínica, bem como o uso de bases de dados bibliográficas para gerar uma pesquisa válida ⁶⁴.

Assim, a informática em saúde expande-se em razão da crescente informatização na área da saúde e da necessidade de recursos de armazenamento e análise dos dados que, nos últimos anos, graças ao desenvolvimento científico, tem apresentado um aumento considerável de volume nas instituições. Manipular esses dados e, a partir deles, extrair informações que auxiliem nos processos decisórios, é cada vez mais indispensável para os serviços de saúde que desejam realizar um planejamento adequado. Entretanto, mais uma vez, vale enfatizar que essas informações normalmente não são fáceis de serem conseguidas.

A dificuldade em obter conhecimento útil de grandes volumes de dados faz com que haja a necessidade de obter-se meios para o devido tratamento e extração de informações que possam ser úteis para os serviços de saúde. Essa necessidade vem fazendo com que técnicas e ferramentas sejam criadas e aprimoradas, com o intuito de facilitar essa tarefa.

Nesse sentido, a informática em saúde apropria-se de diversas metodologias da ciência da computação para realizar estudos na saúde, dentre elas a metodologia *Knowledge Discovery in Databases* (KDD), ou seja, extração de conhecimento das bases de dados.

2.3 Descoberta de conhecimento em bases de dados

Para atender a este novo contexto, a informática em saúde apropria-se de metodologias da ciência da computação para realizar estudos na saúde, dentre elas, a metodologia *Knowledge Discovery in Databases* (KDD), ou seja, descoberta de conhecimento das bases de dados. A metodologia KDD pode ser definida como o processo de extração de informação a partir de dados registrados numa base de dados, um conhecimento implícito, previamente desconhecido, potencialmente útil e compreensível^{7, 2, 74}.

Normalmente, a expressão Mineração de Dados (*Data Mining*) surge como um sinônimo de KDD. No entanto, é apenas uma das etapas da descoberta de conhecimento em bases de dados no processo global KDD e consiste na utilização de métodos específicos para produzir padrões particulares e com valor sobre os dados em análise.

O conhecimento que se consegue adquirir por meio da Mineração de Dados (MD) tem-se mostrado bastante útil nas mais diversas áreas como a medicina, finanças, comércio, marketing, telecomunicações, meteorologia, entre outras ^{2, 5}.

Para Fayyad et al. ⁵⁹, a *Data Mining* (DM) não é um processo trivial e consiste na habilidade de identificar, nos dados, os padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis, envolvendo métodos estatísticos, ferramentas de visualização e técnicas de inteligência artificial. Estas características utilizadas pelos autores – tais como dados, padrões válidos, novos e potencialmente úteis – possuem definições próprias que contribuem para a compreensão da técnica DM.

Os dados são elementos brutos ou “crus”, não interpretados, que estão sem significado e desvinculados da realidade, constituindo a matéria-prima para a informação. Este conjunto de informações é armazenado em Sistemas Gerenciadores de Banco de Dados (SGBD), em vários registros, com os mesmos atributos e representam um tipo de coleção ^{59, 57-56}.

O padrão é o grupo de itens, atributos ou variáveis que tem maior incidência em um conjunto de dados. Padrões Válidos são aqueles considerados válidos e interessantes ao objetivo traçado. Conhecimento novo adquirido representa todo aquele que não estava previsto ou não poderia ser deduzido por meio de hipótese. Já o padrão potencialmente útil são aqueles relevantes e que possuem alguma utilidade para a tomada de decisão; deve ser entendido pelos seres humanos e acrescentar conhecimento útil para adotar medidas decisivas ⁵⁹.

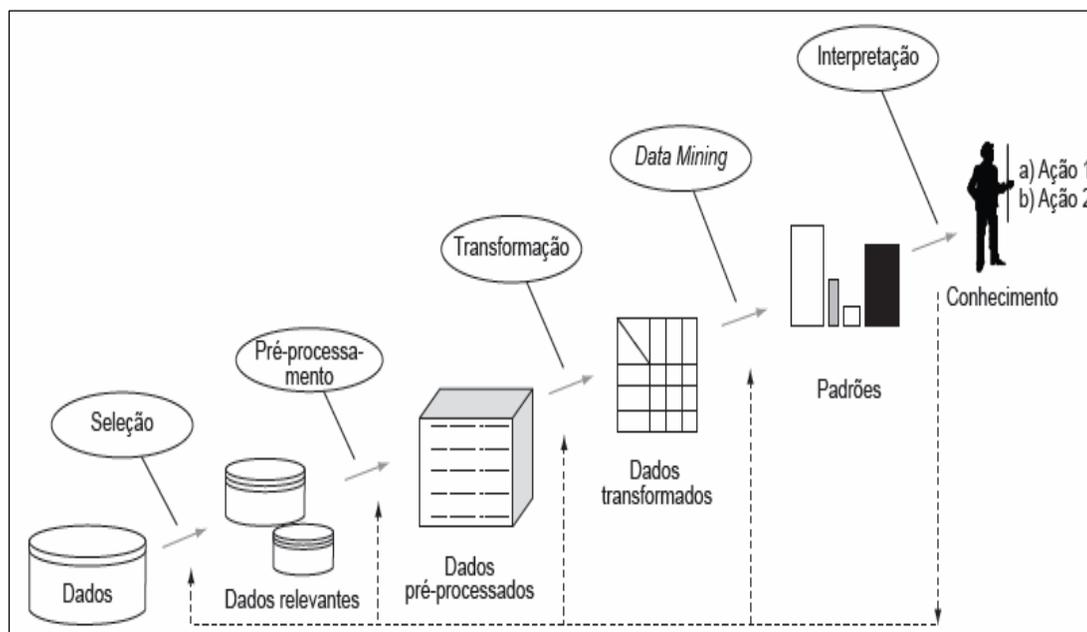
O processo da KDD utiliza conceitos de base de dados, métodos estatísticos, ferramentas de visualização e técnicas de inteligência artificial (IA),

dividindo-se nas etapas de seleção, pré-processamento, transformação, mineração de dados (MD) e avaliação/interpretação^{59, 75, 2, 74} (Figura 1).

Dentre essas etapas, a mais importante é a mineração de dados, foco de inúmeros trabalhos e responsável pela automação da maioria dos métodos de aprendizado de máquina na realização de tarefas especializadas. Conforme já mencionado, este processo descobre informações relevantes, como padrões, associações, mudanças, anomalias e estruturas em grande quantidade de dados armazenados em banco de dados.

A sua aplicação torna possível comprovar o pressuposto da transformação de dados em informação e, posteriormente, em conhecimento, o que faz com que a técnica seja imprescindível para o processo de tomada de decisão. Vários estudos utilizaram este método para demonstrar a informação e o conhecimento a partir de bases de dados^{7, 76, 75, 5}.

Figura 1 – Etapas do processo do KDD



Fonte: Adaptado de Fayyad et al 1996⁵⁹.

As várias etapas envolvidas na MD é ilustrada por Quoniam et al.⁷ sendo que a primeira etapa é a definição clara do problema; a segunda é a seleção a fim de identificar todas as fontes internas e externas de dados para aplicação da MD, que contemple o problema; a penúltima etapa corresponde à preparação dos dados, que inclui o pré-processamento e reformatação destes e significa praticamente 60% do trabalho de MD. Na última etapa, ou seja, na análise dos resultados obtidos do processo de MD, dois aspectos fundamentais devem ser considerados: (a) informar as novas descobertas e apresentá-las de maneira que possam ser exploradas e (b) analisar o conjunto de dados formado a partir da MD, com ajuda de especialista da área a fim de extrair, ao máximo, os dados e a consequente assimilação do conhecimento para a tomada de decisão (Figura 2).

Vale ressaltar que na terceira etapa é crucial, para assegurar a qualidade final dos resultados, a utilização de *softwares* para agregar, efetuar conversão e filtrar variáveis das bases de dados com o objetivo de obter um formato que possa realizar a exportação de dados e base de dados relacionáveis.

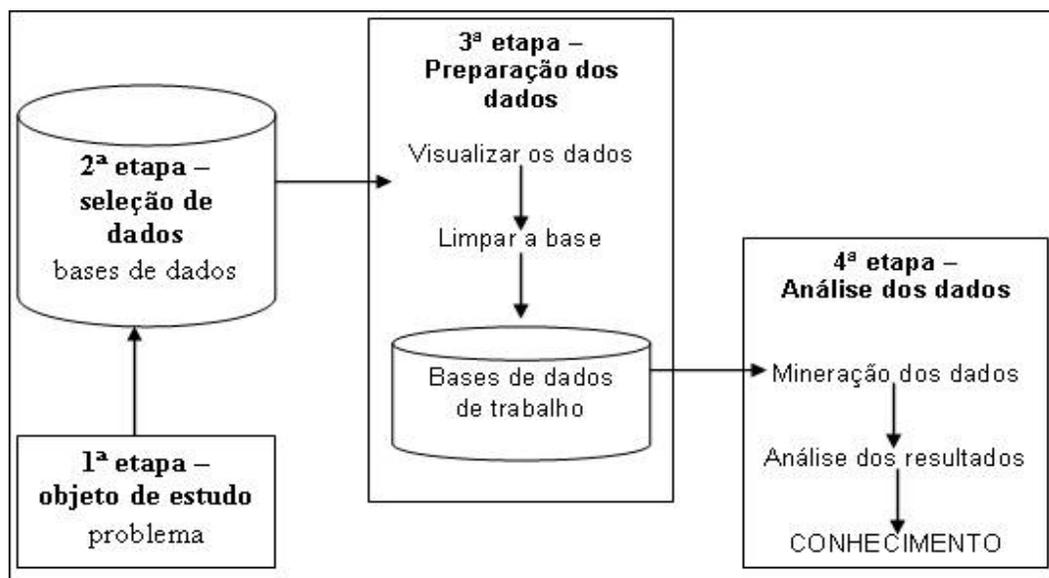
A análise do conjunto de dados, formado por meio da mineração de diversas bases de dados, pode apresentar as principais tarefas que visam ao suporte da decisão. Por intermédio de algoritmos e da estatística, pode-se encontrar um relacionamento entre os atributos e uma classe, de modo que a relação possa prever a função não conhecida da classe, ou seja, um exemplo novo e desconhecido. Além disso, por meio da mineração também se encontram as tendências que possam servir para compreender os padrões de comportamento dos dados^{76, 5}.

Segundo Steiner et al.⁵, a descoberta do conhecimento deve apresentar as seguintes características: ser eficiente (acurado), genérico (aplicável a vários tipos de dados) e flexível (facilmente modificável).

O processo de desenvolvimento de MD envolve as tarefas, métodos e algoritmos para que possa ser realizada a extração de novos conhecimentos¹. Entre as várias tarefas de MD, destacam-se algumas que são as

mais utilizadas: associação, classificação, regressão, clusterização e sumarização ¹⁻
2, 77-78

Figura 2 – Etapas do processo de Mineração de Dados



Fonte: Adaptado de Quoniam et al 2001.

2.3.1 Tarefas da mineração de dados

Na mineração de dados, são definidas as tarefas e os algoritmos que serão utilizados conforme os objetivos do estudo, a fim de obter resposta do problema. As tarefas possíveis de um algoritmo de extração de padrões podem ser agrupadas em atividades preditivas e descritivas ⁷⁹⁻⁸⁰⁻⁸¹. Atividades de predição ou Mineração de Dados preditiva consistem na generalização ou na aplicação de experiências passadas com respostas conhecidas, em uma linguagem capaz de reconhecer a classe de um novo exemplo.

Os dois principais tipos de tarefas para predição são: classificação e regressão. A classificação consiste na predição de uma variável categórica, ou seja, descobrir uma função que mapeie um conjunto de registros em um conjunto de

variáveis predefinidas, denominadas classes. Tal função descoberta pode ser aplicada em novos registros, de forma a prever a classe em que tais registros se enquadram. Vários algoritmos são aplicados na tarefa de classificação, mas os que mais se destacam são as Redes Neurais Back-Propagation, Classificadores Bayesianos, C4.5, (por extenso) KNN (por extenso) e Algoritmos Genéticos ^{75, 2, 5, 74}.

Na regressão, busca-se por funções lineares ou não, sendo que a variável a ser predita consiste em um atributo numérico (contínuo) presente em banco de dados com valores reais ⁸²⁻⁸³. Para implementar a tarefa de regressão, utilizam-se métodos da estatística e de Redes Neurais.

Os algoritmos são atividades de descrição ou Mineração de Dados descritivos, que por sua vez, consistem na identificação de comportamentos intrínsecos do conjunto de dados, sendo que estes não possuem uma classe especificada. Algumas das tarefas de descrição são as chamadas *clustering*, regras de associação e sumarização ^{79, 81}.

A tarefa de clusterização é utilizada para separar os registros de uma base de dados em subconjuntos ou *clusters*, de tal forma que os elementos de um *cluster* compartilhem propriedades comuns, que servem para distinguir os elementos em outros clusters, tendo como objetivo maximizar similaridade intracluster e minimizar similaridade intercluster. Diferente da tarefa de classificação, em que as variáveis são predefinidas, a clusterização precisa automaticamente identificar os grupos de dados, aos quais os pesquisadores deverão atribuir as variáveis ⁵⁹. Os algoritmos mais utilizados nessa tarefa são os K-Means, KModes, K-Prototypes, K-Medoids, Kohonen, dentre outros ².

A sumarização procura identificar e indicar características comuns entre conjunto de dados. Essa tarefa é aplicada nos agrupamentos obtidos na clusterização, sendo a Lógica Indutiva e Algoritmos Genéticos exemplos de técnicas que podem implementar a sumarização ².

A tarefa de associação consiste em identificar e descrever associações entre variáveis no mesmo item ou associações entre itens diferentes, que ocorram simultaneamente, de forma frequente em banco de dados. É também

comum a procura de associações entre itens durante um intervalo temporal ^{75, 74}. Portanto, os algoritmos *Apriori*, GSP, DHP, entre outros, implementam a tarefa de descoberta de associações.

2.3.1.1 Regras de associação

Regra de associação é uma técnica de mineração de dados utilizada para encontrar relacionamentos ou padrões frequentes em conjuntos de dados, é descrita em forma de regras do tipo “**Se X então Y**”, ou “**X → Y**”, (X implica em Y), em que X e Y são conjuntos de itens da base de dados; X é o antecedente da regra (lado esquerdo) e Y é o conseqüente da regra (lado direito) e pode envolver qualquer número de itens em cada lado da regra. Vale destacar, que a interseção ($X \cap Y = \emptyset$) vazia entre antecedente (X) e conseqüente (Y) da regra assegura que não sejam extraídas regras óbvias as quais indiquem que um item esteja associado a ele próprio ² (negrito nosso).

Um exemplo prático é afirmar que "30% dos registros que contêm X também contêm Y; 2% dos registros contêm ambos". O significado desta regra é que as transações da base que contêm X tendem a conter Y. Destaca-se que a transação é o nome atribuído ao elemento de ligação existente em cada ocorrência de itens no banco de dados.

A regra de associação possui dois parâmetros básicos: o suporte e a confiança. Estes parâmetros limitam a quantidade de regras que serão extraídas e descrevem a qualidade delas. Considerando que os conjuntos de itens X e Y estão sendo analisados, o **suporte** é definido como a fração de registros que satisfaz a união dos itens no conseqüente (Y) e no antecedente (X), correspondendo à significância estatística da regra ⁸⁴.

A **confiança** é expressa pelo percentual de registros que satisfaz o antecedente (X) e o conseqüente (Y), medindo a força da regra ou sua precisão ⁸⁴.

No exemplo anteriormente citado, 30% é o fator de confiança e 2% é o suporte da regra. Berry & Linoff ⁸⁵ definem a confiança como a frequência com que o relacionamento mantém-se verdadeiro na amostra de treinamento e o suporte como a frequência com que a combinação acontece. Assim, uma associação pode manter-se 100% do tempo e ter a mais alta confiança, porém pode ser de pouca utilidade se a combinação ocorrer raramente.

Para Agrawal et al. ⁸⁴, o problema das regras de associação é encontrar todas as que possuem o suporte e a confiança acima de um determinado valor mínimo, pois, na prática os usuários normalmente estão interessados somente num subconjunto de associações.

Segundo Goldshmidt & Passos ², a medida de confiança procura expressar a qualidade de uma regra, indicando o quanto a ocorrência do antecedente (X) da regra pode assegurar a ocorrência do conseqüente (Y) desta regra.

Segundo Brusso ⁸⁶, objetivo de um algoritmo para descoberta de regras de associação é identificar todas aquelas que tenham suporte (Sup) e confiança (Conf) maiores do que os valores mínimos estipulados, onde o suporte é um número mínimo de ocorrências e a confiança é o percentual das transações que satisfazem X e Y. Assim, $Sup \geq Sup_{\text{mínimo}}$ e $Conf \geq Conf_{\text{mínima}}$ devem ser satisfeitos.

Sendo:

$$\text{Equação 3: } \text{Suporte } X \cup Y = \frac{\text{número de registro com } X \cup Y}{\text{número de total Transações}}$$

e

$$\text{Equação 4: } \text{Confiança } X \Rightarrow Y = \frac{\text{número de Transações que suportam } X \cup Y}{\text{número Transações que suportam } X}$$

A medida de interesse, também conhecida como *interest*, é uma das mais utilizadas para avaliar dependências. Dada uma regra de associação $X \rightarrow Y$, esta medida indica o quanto mais freqüente torna-se Y quando X ocorre. Seja D

uma base de dados de transações, Seja $X \rightarrow Y$ uma regra de associação obtida a partir de D. O valor do *interest*, para $X \rightarrow Y$ é computado por:

$$\text{Equação 5: } \text{Interest } X \Rightarrow Y = \frac{\text{Confiança } X \Rightarrow Y}{\text{Suporte}(Y)}$$

Se *interest*, $X \rightarrow Y = 1$, então X e Y são independentes, Se *interest*, $X \rightarrow Y > 1$, então X e Y, então são positivamente dependentes. Se *interest* $X \rightarrow Y < 1$, X e Y são negativamente dependentes. Esta medida varia entre 0 e ∞ e possui interpretação bastante simples: quanto maior o valor do *interest*, mais interessante é a regra ⁸⁷.

Um dos algoritmos mais referenciados para este método é o *Apriori*, nas diversas variações, tais como o *Apriori*Tid, DHP e *Partition*. O *Apriori* que se baseia no princípio da antimonotonicidade do suporte, ou seja, um $(K-1)$ -*itemset*, somente pode ser frequente se todos os seus $(K-1)$ -*itemsets* forem frequentes. Denomina-se $(K-1)$ *itemsets* a todo conjunto de itens com exatamente K elementos. Sendo assim, a combinação de *itemsets* para gerar um novo *itemset* somente ocorre quando estes forem frequentes.

Duas operações são necessárias no algoritmo para gerar regras de associação:

- (a) Iteração: encontrar todos os conjuntos de itens frequentes, ou seja, identificar os $(K-1)$ -*itemset* frequentes (satisfazem à condição de suporte maior que mínimo) e por sua vez, combinar os $(K-1)$ -*itemsets* frequentes para gerar os 2-*itemsets* e assim sucessivamente, até K -*itemset* do i -ésimo nível ou suporte maior que mínimo);
- (b) a partir do conjunto de itens frequentes, gerar regras de associação cuja confiança seja superior à confiança mínima definida pelo pesquisador.

Segundo Goldshmidt & Passos ², a etapa (a) demanda maior custo computacional e a etapa (b) é mais imediata.

2.3.2 Técnicas de mineração de dados

As técnicas existem independentemente do contexto de mineração de dados, uma vez que, aplicados na descoberta de conhecimento ou na KDD, produzem resultados supostamente bons. São várias técnicas existentes e o objetivo aqui não é esgotar o assunto e sim identificar os mais utilizados, que são: Rede Neurais, Árvore de Decisão, Algoritmos Genéticos, Lógica Nebulosa (*Fuzzy logic*) e Estatística ⁵.

A Rede Neural Artificial (RNA) é uma técnica computacional que constrói modelo matemático inspirado em cérebro humano, com capacidade de aprendizado, generalização, associação e abstração, constituídos por sistemas paralelos distribuídos em compostos de unidades simples de processamento ^{2, 88, 74, 81, 89}. Inicialmente, a estratégia era de simular os processos mentais humanos para reconhecimento de imagens e sons, mas se tornou instrumento tecnológico eficiente para inúmeras tarefas ^{88, 74, 89}.

Penna ⁸⁸ definiu que a rede neural é composta por neurônios ou nós e suas conexões podem ser classificadas como entradas e saídas em relação a cada neurônio individual. O nó ou neurônio corresponde à etapa que executa o processamento matemático. Esse processamento consiste em dois passos: o primeiro é a soma ponderada das entradas ($\sum W_i X_{ii}$) e o segundo, a aplicação de uma função de ativação a essa soma, gerando a saída do neurônio, que pode se constituir em entrada para outros neurônios. Essa função é, em geral, a função logística ou a tangente hiperbólica. Essas funções têm forma sigmoidal, apresentando variações muito pequenas para valores de x extremos, o que simula a saturação de um neurônio biológico quando os estímulos de entrada são muito grandes.

Já Kovács ⁹⁰ & Zanetti et al. ⁸⁹ caracterizaram as unidades de processamento por uma ou diversas camadas interligadas por um grande número de

conexões. Na maioria dos modelos, essas conexões estão associadas a pesos, que, após o processo de aprendizagem, armazenam o conhecimento adquirido pela rede.

As RNAs têm sido utilizadas com sucesso para modelar relações envolvendo séries temporais complexas em várias áreas do conhecimento ^{90-91, 89}. Para Sudheer et al. ⁹¹, a maior vantagem das RNAs sobre os técnicas convencionais é que estas não requerem informação detalhada sobre os processos físicos do sistema a ser modelado, sendo este descrito explicitamente na forma matemática (modelo de entrada-saída).

A partir de repetidas apresentações dos dados à rede, a RNA aprende padrões. Nesse sentido, procura relacionamentos e constrói modelos automaticamente. Segundo Baesens et al. ⁹², as RNAs possuem uma alta taxa de acurácia preditiva e são bem robustas.

No campo da saúde, vários pesquisadores apropriaram-se dessa técnica e os estudos são cada vez mais desenvolvidos para investigar problemas de saúde tais como: cirurgia refrativa, glaucoma, acessibilidade dos serviços de saúde, cólera e conjuntivite alérgica ^{92-93, 6, 88, 94}.

A Árvore de Decisão é um modelo representado graficamente por nós e galhos, parecido com uma árvore, mas no sentido invertido ⁹⁵⁻⁹⁶. É um modelo de conhecimento que em cada nó (galho) interno da árvore representa-se uma decisão sobre uma variável que determina como os dados estão particionados por uma série de galhos (nós filhos). Com isso, descreve-se uma associação entre o atributo e variável alvo, ou seja, associação de cada galho com outro(s) galho(s) – filho gerado ^{2, 74, 81}.

Segundo Meira et al. ⁹⁶, o nó raiz é o primeiro nó da árvore, no topo da estrutura. Os nós internos, incluindo o nó raiz, são nós de decisão. Cada um contém um teste sobre uma variável independente e os resultados deste teste formam os ramos da árvore. Os nós folhas, nas extremidades da árvore, representam valores de predição para a variável dependente ou distribuições de probabilidade desses valores. As árvores de decisão são também chamadas de

árvores de classificação ou de regressão, caso a variável dependente seja categórica ou numérica, respectivamente.

A finalidade da indução de uma Árvore de Decisão é produzir um modelo de predição preciso ou descobrir a estrutura preditiva do problema. No último caso, a intenção é compreender quais variáveis e interações dessas variáveis conduzem o fenômeno estudado. Esses dois propósitos não são excludentes, podendo aparecer juntos em um mesmo estudo ^{97, 96}.

As Árvores de Decisão são de interesse especial para a mineração de dados ou descoberta de conhecimento em bases de dados ⁵⁹, pois utilizam representações simbólicas e interpretáveis. Soluções simbólicas permitem a compreensão das fronteiras de decisão que existem nos dados e também da lógica implícita neles ⁹⁷. A multicolinearidade entre as variáveis independentes não afeta o desempenho das árvores de decisão, diferentemente das técnicas de regressão ⁸³. Diversas variáveis, numéricas ou categóricas, podem ser analisadas ao mesmo tempo, sendo que o próprio algoritmo de indução se encarrega de selecionar as de maior importância.

Algumas pesquisas recentes na saúde têm utilizado a indução de Árvore de Decisão para prever e obter conhecimento, como estratégias auxiliares para graduação dos tumores astrocíticos, segundo os critérios histopatológicos estabelecidos pela OMS ⁹⁸ e a prevalência pontual de depressão na esquizofrenia, investigando os fatores associados e sua relação com a qualidade de vida ⁹⁹.

Já os Algoritmos Genéticos (AGs) formulam estratégias de otimização algorítmica inspiradas nos princípios observados na evolução natural e na genética para solução de problemas que, neste caso, podem apresentar diversos parâmetros ou características os quais necessitam ser combinados para encontrar a melhor solução. Os algoritmos genéticos usam os operadores de seleção, cruzamento e mutação para desenvolver sucessivas gerações de soluções. Com a evolução do algoritmo, somente as soluções com maior poder de previsão sobrevivem, até convergirem numa solução ideal ^{79, 2}.

Para Leite et al.¹⁰⁰, os AGs baseiam-se, inicialmente, na geração de uma população formada por um conjunto aleatório de indivíduos que podem ser vistos como possíveis soluções para um problema. Durante o processo evolutivo, esta população é avaliada, atribuindo-se um índice para cada indivíduo, refletindo sua habilidade de adaptação a um determinado ambiente. Uma porcentagem dos índices mais adaptados é mantida, enquanto os outros são descartados. Os membros selecionados podem sofrer modificações em suas características fundamentais por meio de mutações e cruzamentos genéticos, gerando descendentes pelo processo chamado de reprodução, o qual é repetido até que um conjunto de soluções satisfatórias seja encontrado.

Pesquisas investigando a otimização, configuração, operação e apoio à decisão dos sistemas médicos emergenciais em rodovias, utilizando os AGs por intermédio dos modelos de hipercubo, mostraram que as medidas de desempenho do sistema podem ser melhoradas somente modificando as áreas de atuação de cada ambulância. Isto é, sem ter de reposicionar as ambulâncias¹⁰¹⁻¹⁰²⁻¹⁰³⁻¹⁰⁴.

Outra técnica muito utilizada é a Lógica Nebulosa (*Fuzzy logic*), uma teoria matemática que permite uma modelagem do modo aproximado de raciocínio, imitando a habilidade humana de tomar decisões em ambientes de incertezas e imprecisão. Com isso, podem-se construir sistemas inteligentes de controle e suporte à decisão que lidem com informações imprecisas e subjetivas².

Segundo Sousa et al.¹⁰⁵, a lógica *fuzzy* pode ser utilizada principalmente de duas formas: 1) representar a extensão da lógica clássica para uma forma mais flexível, com objetivo de formalizar conceitos imprecisos; 2) verificar onde se aplicam conjuntos *fuzzy* a diversas teorias e tecnologias para processar informações imprecisas, por exemplo, em processos de tomada de decisão.

A teoria dos conjuntos *fuzzy* utiliza expressões linguísticas como alto, leve, rápido e infeccioso na sua aplicação; definem-se conjuntos aos quais coisas (objetos de estudo) são alocadas com diferentes graus de pertinência entre o

falso (zero) e o verdadeiro (um), o que rompe os limites da rígida dicotomia da lógica clássica ^{106, 105}.

Para Sousa et al. ¹⁰⁵, o campo das ciências da saúde possui níveis de incerteza e imprecisão mais diversos, e com frequência, o processo de tomada de decisão acaba apoiando-se em conceitos vagos estranhos à lógica clássica e em parâmetros de natureza subjetiva. Nesse sentido, esta técnica encontra-se em ascensão no campo da saúde com diversos estudos publicados ¹⁰⁷,
106, 108-109-110-111

Por fim, a estatística, uma das técnicas mais tradicionais, fornece modelos para análise e interpretação de dados. Os modelos mais utilizados são Redes Bayesianas, Análise Discriminante, Análise Exploratória de Dados, dentre outros.

Para Roazzi et al. ¹¹², um dos princípios estatísticos de base concerne à maneira pela qual é possível se estimar a probabilidade de um evento a partir de dois tipos de conhecimento: não somente o que se sabe sobre tal evento antes que o mesmo se verifique, como também eventuais informações que podem ser obtidas em seguida. Isto é, tal princípio, formulado pelo matemático inglês Thomas Bayes (1702-1761), possibilita calcular a probabilidade de uma hipótese fundamentando-se na probabilidade *a priori* e em eventuais novas evidências relevantes.

Para Abbott & Lee ⁷⁴, as Redes Bayesianas emergiram em anos recentes como uma poderosa tecnologia de mineração de dados, estas redes fornecem representações gráficas de distribuições probabilísticas derivadas da contagem da ocorrência dos dados num dado conjunto, representando um relacionamento de variáveis. Consistem em grafos acíclicos dirigidos, onde os nós representam os atributos e as ligações representam as dependências probabilísticas entre os atributos. Associada a cada nó (atributo) da rede, existe uma Tabela de Probabilidades Condicionais, esta descreve a relação entre o nó e os seus nós “pai”.

Por fim, as Redes Bayesianas estão amplamente difundidas nos estudos das áreas da saúde. Diversos estudos têm empregado a técnica para obter distribuições probabilísticas de vários problemas de saúde ¹¹³⁻¹¹⁴⁻¹¹⁵⁻¹¹⁶.

2.3.3 Ferramentas utilizadas em mineração de dados

Existem diversas ferramentas disponíveis no mercado que realizam várias tarefas de mineração ². A escolha pode ser feita conforme o problema e o resultado pretendido. Algumas delas são:

- *Waikato Environment for Knowledge Analysis (WEKA)*: um *software* de domínio público desenvolvido pela Universidade de Waikato, na cidade de Hamilton em Nova Zelândia, que implementa uma série de algoritmos de Data mining ^{2, 117};
- *Intelligent Miner*: desenvolvido pela IBM, também é uma ferramenta de mineração;
- *Oracle Data Miner*: é uma ferramenta de mineração que possui interligação com o banco de dados Oracle Enterprise;
- *Enterprise Miner*: nova versão do SAS Enterprise Miner, usado na área de negócios, marketing e inteligência competitiva;
- *Statistica Data Miner*: este *software* acrescenta as ferramentas de mineração aos tradicionais pacotes de aplicações estatística;
- *Tanagra*: um *software* de código aberto para fins acadêmicos e para pesquisa, desenvolvido por professor da Universidade de Lyon 2, na França, que implementa vários métodos de extração a partir de análise exploratória de dados, estatística e aprendizagem automática dos dados ¹¹⁸.

Neste trabalho, foi selecionada a ferramenta WEKA que é um *software* implementado em JAVA, segundo o paradigma de orientação a objetos, composto por uma série de algoritmos de aprendizagem para solucionar problemas de mineração de dados. Os algoritmos podem ser aplicados diretamente a uma série de dados.

São distinguidas três tarefas na mineração de dados; cada uma é adequada para atingir um objetivo, ou seja, possui melhores soluções para determinados problemas. São elas: classificação, clusterização e associação. Além do mais, a WEKA implementa as atividades de pré-processamento, seleção de atributos e visualização.

O pré-processamento dos dados usa-se um algoritmo de filtragem. Esses filtros podem ser utilizados para transformar os dados (por exemplo, transformar em atributos numéricos em discretos), e tornar possível a remoção de instâncias e atributos de acordo com critérios específicos.

A seleção de atributos envolve pesquisar através de todas as combinações possíveis de atributos de dados para encontrar o subconjunto de atributos que funciona melhor para uma previsão. Para isso, dois objetos devem ser criados: um atributo avaliador e uma técnica de pesquisa. O avaliador determina qual técnica é utilizado para atribuir um valor a cada subconjunto de atributos. A técnica de pesquisa permite determinar qual estilo de pesquisa será realizada.

A visualização mostra uma parcela matriz de pontos, onde espalhe parcelas individuais podem ser selecionadas e expandidas, e ainda analisados utilizando vários operadores seleção.

O painel de classificação permite ao usuário aplicar tarefas de classificação para os dados resultantes. Dentre as técnicas implementadas pode-se listar: árvore de decisão induzida, regras de aprendizagem, naive Bayes, tabelas de decisão, regressão local de pesos, aprendizado baseado em instância, regressão lógica, perceptron, perceptron multicamada, comitê de perceptrons e SVM.

No painel de cluster que dá acesso as opções de agrupamento. Os principais algoritmos de agrupamento implementados no WEKA são: EM, DBScan, XMeans, SimpleKMeans, Cobweb, OPITICs.

A tarefa de associação fornece acesso as regra de associativas na tentativa de identificar todas as inter-relações entre atributos importantes nos dados, por meio dos algoritmos Apriori e Tertius.

2.3.4 Mineração de dados no campo da saúde

Neste contexto, diversos estudos têm utilizado a MD para automatizar a tarefa de extrair conhecimento útil a partir de grandes volumes de dados. Nas áreas de ciências biológicas e da saúde, vários pesquisadores estão se apropriando dessa tecnologia, em razão do crescimento da área de informática em saúde, que possui uma grande interface com a ciência da computação.

Nos mais diversos países, os estudos com MD estão crescendo gradativamente na área de ciências biológicas, como na subárea da genética que em vários estudos têm utilizado a MD, em razão de suas grandes bases de dados que são analisadas ^{78, 6, 119, 120}.

Já na área da ciência da saúde, pode-se citar alguns estudos, tais como um financiado pela Organização Mundial de Saúde (OMS), que aplicou MD e regressão linear para investigar o relacionamento dos diferentes grupos de fatores – sociais, econômicos e mão-de-obra nos cuidados de saúde – com a taxa de prevalência de HIV/AIDS, usando grandes e diversas bases de dados. Este estudo concluiu que existe a necessidade de investir em mão de obra para promover um apoio adicional na educação de enfermeiros para os programas de HIV/AIDS e, com isso, obter um impacto positivo na taxa de prevalência ¹²¹.

Outro estudo realizado nos Estados Unidos por pesquisadores da Universidade da Johns Hopkins demonstrou que a tecnologia da MD poderia ser utilizada em meta-análise para analisar acidentes de trânsito de pessoas idosas ⁷⁵.

Na Alemanha, pesquisadores utilizaram a tecnologia MD e a tarefa de clusterização, para prever o risco de doenças cardiovasculares em pacientes em hemodiálise. Eles obtiveram um excelente sucesso com *clusters* para prever o risco individual dos pacientes em hemodiálise e, ainda, validaram esse método com um apoio da tomada de decisão para essa especialidade e para outras esferas da medicina ¹²².

Kaur & Wasan ⁸¹, na Índia, investigaram crianças com diabetes *mellitus* e diabetes *insipidus*, por meio da aplicação da MD e da tarefa de classificação. A base de dados tinha atributos do exame físico e dos exames bioquímicos laboratoriais. Com atributos condicionais na base de dados disponíveis de cada criança, a identificação de novos pacientes foi mais rápida para iniciar o tratamento, reduzindo assim sofrimento das crianças e as sequelas do diabetes.

Em Criciúma, Santa Catarina, foi realizado um estudo sobre alergia e rinite em escolares, com objetivo de verificar a prevalência destas causas numa base de dados existentes, aplicando a MD e utilizando um *software* Orion Data Mining, ferramenta gratuita desenvolvida pela Universidade do Extremo Sul Catarinense –UNESC ¹²³⁻¹²⁴.

Em síntese, os exemplos dos estudos apresentados confirmam evidências da importância da MD no campo da saúde, fornecendo recursos para realização de tomada de decisão criteriosa, promovendo melhores condições para planejamento da utilização e implantação dos recursos de saúde no país.

3 OBJETIVO

3.1 Objetivo Geral

Identificar, por meio da aplicação da tecnologia de mineração de dados, padrões significativos sobre acidentes de transporte, a partir dos dados da Secretaria de Justiça Segurança Pública, do Sistema de Informação Hospitalar do Sistema Único de Saúde e do Sistema de Informação sobre Mortalidade, ocorridos no município de Cuiabá, Mato Grosso, no ano de 2006

3.1.1 Objetivos específicos

Aplicar a tecnologia de mineração de dados no banco relacionado de vítimas de acidentes de transporte de Cuiabá, Mato Grosso no ano de 2006;

Identificar regras de associação no banco de dados de vítimas de acidentes de transporte em Cuiabá, Mato Grosso no ano de 2006;

Destacar as melhores regras associativas segundo a medida de interesse - *interest*.

4 MÉTODOS

4.1 Tipo de estudo

Estudo observacional, transversal, retrospectivo e exploratório das bases de dados secundárias da segurança e saúde pública.

4.2 Local de estudo

O município selecionado foi Cuiabá, sendo que a análise de suas bases de dados foi feita por conveniência e também por facilidade em localizar os dados secundários nos serviços de saúde e segurança pública.

O município de Cuiabá está localizado na mesorregião centro-sul mato-grossense, microrregião Cuiabá, é a capital do estado de Mato Grosso e foi o primeiro município de Mato Grosso criado por carta régia, em 8 de abril de 1719. De Cuiabá, derivam todos os municípios que compõem atualmente os estados de Mato Grosso, Mato Grosso do Sul, e Rondônia. Possui uma extensão territorial de 3.984,9 km² e uma população, segundo o censo do IBGE de 2000, de 482.498 habitantes e, na contagem de 2007, de 526.830 habitantes ¹²⁵⁻¹²⁶. Limita-se com os municípios de Chapada dos Guimarães, Rosário Oeste, Campo Verde, Santo Antonio do Leverger, Várzea Grande e Acorizal. O município possui uma malha viária importante sendo que, das cinco rodovias federais do estado, três cortam Cuiabá (BR 364, BR 163 e BR 070). Para o Departamento Estadual de Trânsito (DETRAN), a frota de veículo é de aproximadamente 280 mil veículos ¹²⁷. As principais atividades econômicas são: indústria, comércio, pecuária e pesca, com destaque especial para o turismo. Para administração pública, a cidade de Cuiabá está dividida em quatro zonas: norte, sul,

leste e oeste, conforme mostra o Anexo 1.

4.3 Fontes de dados, período e considerações éticas

As fontes de dados que fizeram parte do estudo foram os dados da Secretaria Estadual de Justiça e Segurança Pública, disponibilizados por meio da Coordenadoria de Tecnologia da Informação, que forneceu banco de dados das vítimas de acidentes de transporte terrestre atendidas pelas viaturas do Corpo de Bombeiros Militar e SAMU com Boletim de Ocorrência (BO).

Complementando o banco de dados da área de segurança pública, existem os sistemas de informação da saúde que contém os registros de acidentes em qualquer faixa etária e sexo, sendo, portanto, os sistemas que registram as internações no SUS: Sistema de Informações hospitalares do Sistema Único de Saúde (SIH/SUS), registro dos óbitos e o Sistema de Informação sobre Mortalidade (SIM).

Os dados foram coletados no período de 1º de janeiro a 31 de dezembro do ano de 2006. Os bancos de dados do ano de 2006, as fichas de coleta ou instrumento de coleta dos mesmos foram cedidas pela Secretaria Municipal de Saúde Cuiabá / MT e Secretaria de Estado de Justiça e Segurança Pública (SEJUSP), por meio de Autorizações dos Gestores (Anexo 2). O estudo foi submetido e aprovado pelo Comitê de Ética da UNIFESP sob número de protocolo 1595/07 (Anexo 3).

4.4 Processamento e análise dos dados

A parte de processamento dos dados foi dividida em duas partes: a

MÉTODO

primeira contendo o relacionamento dos bancos de dados e, a segunda, a mineração dos dados propriamente dita, sendo que algumas etapas como seleção, limpeza do banco foram antecipadas para que obtivesse um relacionamento de dados mais otimizado.

Os bancos de dados recebidos do SEJUSP, SIH/SUS e SIM estavam em uma planilha do Excel, Microsoft Office Excel 2003, com 2.547, 38.273 e 3.198 registros, respectivamente. Vale ressaltar que, no banco da SEJUSP, tinham dezessete variáveis, no SIH/SUS, havia noventa e quatro e, no SIM, o número de variáveis era de sessenta e dois. Para operação de processamento, foram selecionadas apenas variáveis de cada banco relacionadas com acidente de transporte e a identificação da vítima / paciente / falecido, sendo onze, quinze e vinte e um no da SEJUSP, SIH/SUS e SIM, respectivamente (Quadro 1).

Após esta seleção, realizou-se uma checagem manual dos bancos para averiguar as variáveis existentes e o preenchimento dos atributos-chave: nome completo (nome e sobrenome), data de nascimento (DN) correta no formato dia/ mês/ano ou ano/mês/dia.

O critério de seleção dos registros nos bancos de dados considerou todos os registros que tinham os atributos-chave presentes, ou seja:

- nome completo – nome e sobrenome - da vítima / paciente internado / falecido;
- nome completo da mãe - nome e sobrenome - da vítima / paciente internado / falecido;
- data de nascimento – dia, mês e ano ou ano, mês e dia.

Os registros que não tinham estes dados completos foram excluídos do banco.

MÉTODO

Quadro 1 - Variáveis selecionadas nos bancos de dados da SEJUSP, SIH/SUS e SIM

Variáveis selecionadas			
Item	SEJUSP	SIH/SUS	SIM
1	Nome da vítima	Nome da paciente	Nome da falecido
2	Nome da mãe	Nome do responsável	Nome da mãe
3	Nome do pai	Sexo	Nome do pai
4	Número ocorrência	Data de nascimento	Sexo
5	Data da ocorrência	Logradouro (residência)	Data de nascimento
6	Sexo	Número	Endereço (residência)
7	Data de nascimento	Complemento (residência)	Número (residência)
8	Bairro da ocorrência	Bairro (residência)	Complemento (residência)
9	Tipo de acidente	Município (residência)	Bairro (residência)
10	Meio de transporte da vítima	UF (residência)	Município (residência)
11	Tipo de vítima	Diagnóstico secundário	UF (residência)
12		Data de internação	Endereço (ocorrência)
13		Data de saída	Número (Ocorrência)
14		Número CNES (local da internação)	Complemento (Ocorrência)
15		Motivo da cobrança da AIH	Bairro (Ocorrência)
16			Município (Ocorrência)
17			UF (Ocorrência)
18			Causa básica
19			Local da ocorrência do óbito
20			Código do estabelecimento que prestou assistência
21			Recebeu assistência médica

Na etapa de limpeza, o banco da SEJUSP foi o que apresentou o maior número de erros na data de nascimento, sendo, portanto, excluídos registros que se apresentavam como, por exemplo: 01/01/0001 (1066), 24/4/0132 (1),

MÉTODO

14/10/8199 (1), 30/03/2012 (1), 20/08/2018 (1), 22/09/2025 (1), 15/03/2026 (1), 06/01/2030 (1), 09/09/2033 (1), 09/09/2033(1), 23/04/2038 (1), 01/05/2039 (1), 06/06/2039 (1). Assim, havia 1.078 registros e mais 59 duplicidades (total de 1137) que foram excluídas, finalizando, assim, um total de 1411 registros selecionados para estudo. Para excluir os nomes das vítimas que estavam duplicados, foram analisados quatro variáveis: nome da vítima, DN, data de ocorrência e nome da mãe.

No SIH/SUS, não tiveram campos em branco e erro nas DN. Desta forma, todos os registros foram selecionados (38.273). Já no SIM, houve exclusão de 97 registros com campos em branco DN, perfazendo, assim, 3.101 óbitos selecionados, conforme mostra a Figura 3.

4.4.1 Relacionamento dos bancos de dados

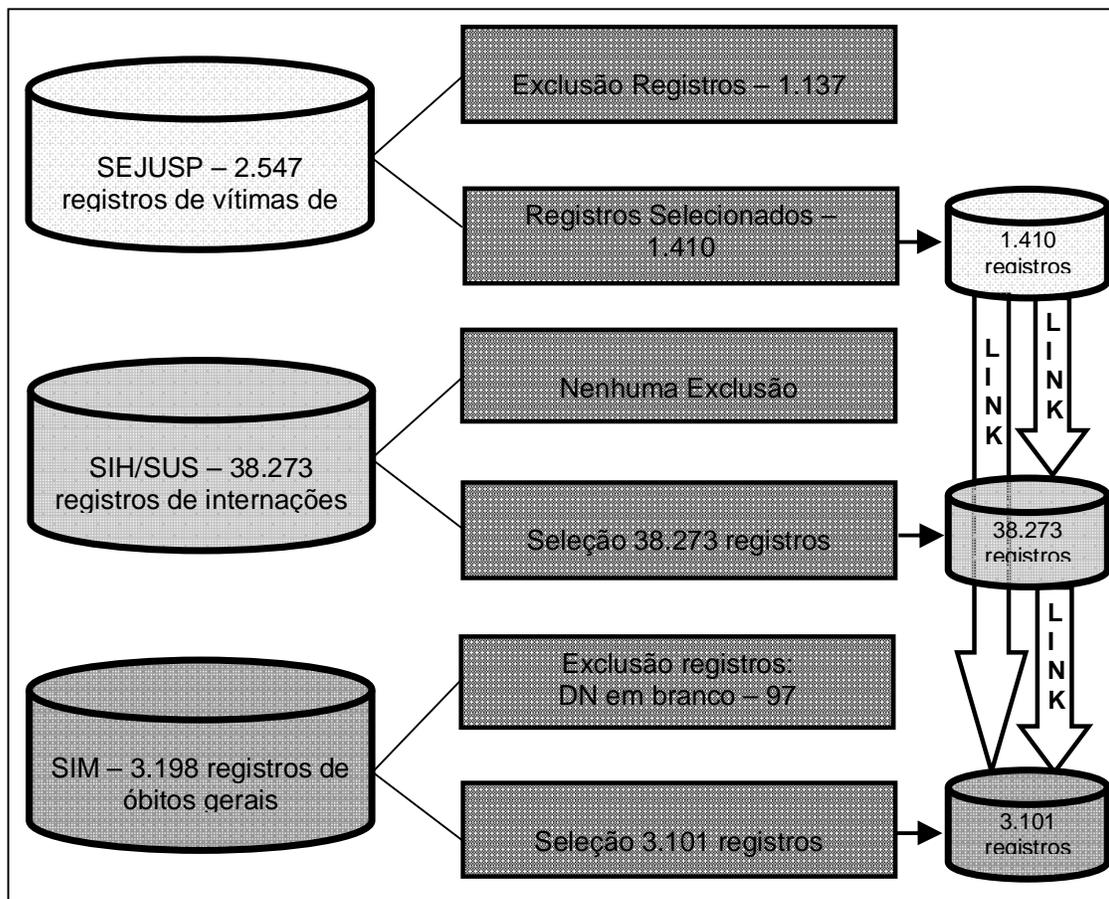
Foi utilizado o método probabilístico de relacionamento de banco de dados para concatenar os dados do Sistema de Informação do SUS com os dados da Secretaria Estadual de Justiça e Segurança Pública (SEJUSP) para extração de informações relativas a indivíduos comuns, que foram vítimas de acidente de transporte no município de Cuiabá / MT no ano de 2006. Esses arquivos continham informações que permitiram identificar as vítimas de acidente de transporte (como nome, data de nascimento, nome da mãe e outras variáveis que descrevem o acidente) para fins de relacionamento dessas bases de dados. Foi resguardado o sigilo do nome de cada pessoa identificada no relacionamento dos dados e, posteriormente nas análises, o nome foi alterado para um código decimal.

Os três arquivos foram transformados em arquivo dBase IV, ou seja, extensão *dbf*, para entrada dos dados no *software* RecLink III. Esta ferramenta é gratuita e está sendo disponibilizada pelos pesquisadores do Instituto de Estudos em Saúde Coletiva da Universidade Federal do Rio de Janeiro, nos seguintes

MÉTODO

endereços <http://www.mediafire.com/?sharekey=bc99288de55df934d2db6fb9a8902bda> ou www.iesc.ufrj.br.

Figura 3 – Número exclusão e registros selecionados dos bancos de dados da SEJUSP, SIH/SUS e SIM. Cuiabá / MT, 2006



O método probabilístico de relacionamento de banco de dados realizado por este *software* baseia-se em processos: a padronização de registros, a criação de variável padrão e o pareamento de registros ³⁸.

Realizaram-se três relacionamentos separadamente: (a) SEJUSP vs. SIH/SUS, (b) SEJUSP vs. SIM e, (c) com os registros do primeiro relacionamento SEJUSP/SIH vs. SIM. As variáveis selecionadas e comuns nos três arquivos de

MÉTODO

dados para serem chaves de pareamento foram: nome do paciente ou da vítima ou do falecido e data de nascimento (dia, mês e ano).

A padronização dos três arquivos de dados foi a primeira etapa a ser realizada e envolveu a preparação dos campos de dados, onde buscou-se minimizar a ocorrência de erros durante o processo de pareamento de registros. Este processo foi fundamental para os campos não estruturados, como por exemplo, o campo nome, que se caracteriza pela entrada relativamente livre de controles. Esta variável, nome do paciente, foi transformada em todos os caracteres alfabéticos da forma minúscula para a maiúscula, eliminando-se também caracteres de pontuação erroneamente digitados, espaços em branco no início do campo e das preposições "de", "da", "do", "dos", "das". A data de nascimento do banco do SIM foi transformada no formato americano, ou seja, *yyyymmdd*; nos bancos do SEJUSP e no SIH/SUS, as mesmas já se encontravam nesta forma.

Em seguida, realizou-se a subdivisão do nome, com o objetivo de minimizar a perda de pares verdadeiros, tendo em vista que os nomes são mais sujeitos a erros de registro. A variável nome do paciente ou da vítima ou do falecido foi subdividida em seis campos, como por exemplo: Pedro Neves Souza Filho, o campo:

- FNOMEF: constou o primeiro nome ou prenome = Pedro;
- FNOMEU: último nome ou sobrenome = Souza;
- FNOMEI: as iniciais do nome do nome composto ou primeiro sobrenome = N;
- FNOMEA: o agnome = Filho;
- Os dois últimos campos criados (PBLOCO e UBLOCO) foram dedicados ao Sistema fonético.

As chaves dos campos utilizadas no passo pareamento foram: (1) código fonético (*Soundex*) do primeiro nome formatado para blocos – Pbloco; (2)

MÉTODO

código fonético (*Soundex*) do último nome formatado – Ubloco e (3) ano de nascimento – Anonas.

Os pares óbitos, por meio do pareamento de registros, foram classificados de acordo com a definição de escores inferior e superior. Os pares que atingiram valor de escore acima do limiar superior foram considerados verdadeiros, os que atingiram escore abaixo ao limiar inferior foram considerados falsos e os que tiveram valores intermediários foram considerados duvidosos e, portanto, submetidos a uma checagem manual. Na análise manual dos pares duvidosos, verificou-se além dos campos-chave (nome, DN), os campos do sexo, endereço residencial e o nome da mãe. Foram encontrados 139 pares verdadeiros nos três relacionamentos (Tabela 1).

Tabela 1 - Número de pares verdadeiros obtidos por meio do relacionamento probabilístico

Relacionamento	Pares verdadeiros
SEJUSP VS. SIH/SUS	111
SEJUSP VS. SIM	25
SEJUSP/SIH vs. SIM	03
Total de registro	139

O último processo foi a junção dos pares verdadeiros em um único arquivo. Nessa operação, utilizou-se a planilha o Excel, Microsoft Office Excel 2003. As variáveis triplicadas – como, por exemplo, nome, data de nascimento e sexo – foram excluídas e foi selecionada apenas a variável que estava mais completa.

4.4.2 Mineração dos dados e análise

A ferramenta utilizada para extração de conhecimento do banco de dados de acidentes de Cuiabá foi a WEKA – *Waikato Environment for Knowledge Analysis*. Os motivos da escolha foram:

1. É uma ferramenta desenvolvida na linguagem JAVA, que tem como característica principal a portabilidade (facilidade de ser executada em várias plataformas de Sistema Operacional);
2. Tem o código fonte aberto;
3. É de fácil acesso pela internet, disponível em <http://www.cs.waikato.ac.nz/ml/weka/>.
4. Estudos em bioinformática têm utilizado o *software*, obtendo ótimos resultados ^{117, 128-129, 80}.

Outro *software* utilizado foi *Statistical Package for the Social Sciences* for Windows (SPSS) versão 15.0 para calcular o quartil dos dias de internação e alterar o formato das datas de nascimento, dia da ocorrência do acidente, entrada e saída da internação. Vale esclarecer que estas datas estavam no formato americano (yyyymmdd) e para realização do relacionamento e foram transformadas em ddmmyyyy, pela função *date and time wizard* deste pacote estatístico.

A análise preliminar dos dados foi realizada na parte de visualização da ferramenta (pré-processamento), onde foi obtida a frequência absoluta de cada variável e o cálculo das frequências relativas das variáveis nominais. Com relação às variáveis numéricas (idade e dias de internação), a ferramenta WEKA calculou as medidas de tendência central (média) e de dispersão (desvio padrão, valor mínimo e máximo).

Para verificar as regras e padrões de associação, foi selecionado o algoritmo *Apriori* que faz uma varredura no conjunto de dados, procurando por subconjuntos que tenham relacionamentos que sejam frequentes ². Para

MÉTODO

processamento do algoritmo utilizou-se vários valores dos parâmetros confiança e suporte, com auxílio de um pesquisador / especialista de área de acidentes de trânsito e por meio da medida de interesse *interest* escolher as regras dos os melhores modelos associação com os seguintes parâmetros:

1. Confiança 90% e suporte 10%;
2. Confiança 80% e suporte 20%;
3. Confiança 70% e suporte 30%; e
4. Confiança 70% e suporte 40%;

Como citado anteriormente na fundamentação teórica, o objetivo do algoritmo de associação é identificar regras com suporte, confiança maiores que os valores mínimos estipulados e medidas de interesse maior 1; assim as regras que apresentaram um dos parâmetros, principalmente *interest* menor que 1 foram desconsiderados.

4.5 Limitação do estudo

A qualidade dos dados foi uma das limitações encontradas neste trabalho. Isto ocorreu em razão dos erros e do preenchimento incorreto de algumas variáveis da AIH, da DO e também do relatório de ocorrência do atendimento móvel de urgência da SEJUSP, fatores que dificultaram a interpretação dos resultados.

Os dados inconsistentes denotam a falta de comprometimento dos responsáveis com o correto preenchimento dos instrumentos de coleta e com a alimentação do banco de dados e que, por consequência, prejudicam a interpretação dos dados. Se não fosse assim, o sistema poderia trazer resultados mais precisos e uma melhor qualidade das informações.

A falta de integração entre as bases de dados selecionadas impossibilitou que o estudo se estendesse a outros anos e a outras esferas

MÉTODOS

geográficas, que possibilitou apenas utilizar o método probabilístico de relacionamento dos registros que possui grandes limitações para encontrar os pares verdadeiros.

O estudo também está limitado a três bases de dados, não tendo, dados dos pacientes que foram atendidos no ambulatório dos serviços de saúde do SUS de Cuiabá, sendo uma das possíveis causas da redução dos pares verdadeiros, pois alguns estudos apontam que grande parte das vítimas acidentadas possui ferimentos leves que não necessitam de internação.

5 RESULTADOS

5.1 Aplicação da mineração dos dados

5.1.1 Preparação dos dados

A aplicação da técnica de mineração de dados é dividida em etapas, como descrito anteriormente. As primeiras etapas (seleção dos dados e limpeza do banco) foram realizadas no processo de relacionamento dos dados, também já descrito. A etapa seguinte, de pré-processamento ou reformatação dos dados, foi realizada após junção do banco de dados.

Também, vale reforçar que, no banco de dados, algumas variáveis como as datas (nascimento, datas de ocorrência, internação, saída do hospital) que estavam no formato americano, *yyyymmdd*, foram transformadas em *ddmmyyyy*, pela função *date and time wizard* do *software* SPSS versão 15.

Com a junção dos bancos, foram obtidos atributos ou variáveis que puderam ser agregados e, com isso, houve a transformação de variáveis tais como: assistência médica, local da assistência e evolução (Quadro 2). Ainda foi reformatada a variável data de ocorrência, que contém informações de dia, mês e ano. Para este estudo, foi considerado somente o mês de ocorrência (Quadro 2).

As variáveis bairro de residência e de ocorrência foram categorizados em zonas norte, sul, leste e oeste (Quadro 2). A base cartográfica de Cuiabá foi utilizada do Instituto de Pesquisa e Desenvolvimento Urbano – IPDU, que é o órgão de planejamento do Sistema Municipal de Desenvolvimento Urbano de Cuiabá (SMDU), disponível na *homepage* da Prefeitura de Cuiabá – <http://www.cuiaba.mt.gov.br/orgaos/ipdu/mapas.jsp>¹³⁰, conforme relação dos bairros citados no Anexo 4.

RESULTADOS

Quadro 2 – Variáveis de estudo e suas categorias

Variáveis precursoras	Variáveis novas	Categorias
Diferença entre data de ocorrência do acidente e a DN	Idade em 2006	Número arábico
Data de ocorrência (dia, mês, ano)	Mês de ocorrência	Jan, fev, abr, mai, jun, jul, ago, set, out, nov, dez
Diferença entre a data de saída e a data da internação	Dias de internação	Número arábico
Bairro de residência	Zona de residência	Norte, sul, leste e oeste
Bairro de ocorrência	Zona de ocorrência	Norte, sul, leste e oeste
Dias de internação, local da ocorrência do óbito, falecido recebeu assistência médica	Assistência Médica	SIM, Não
Assistência médica, número do CNES do hospital, código do estabelecimento do óbito	Local de assistência	Hospital do Câncer de Mato Grosso (HCANCERMT), Hospital Geral Universitário (HGU), Hospital Universitário Julio Muller (HUJM), Hospital Jardim Cuiabá (HJDCBA), Hospital Municipal Bom Jesus (HMBOMJESUS), Hospital e Pronto Socorro Municipal de Cuiabá (HPSMCBA), Hospital Santa Helena (HSTAHELENA), Hospital Santa Rosa (HSTAROSA), Hospital Só Trauma (SOTRAUMA), Hospital Santa Casa (STACASA), NENHUM
Motivo da cobrança da AIH e nome do falecido no banco do SIM	Evolução em 2006	Alta, óbito

As demais variáveis foram mantidas no formato e também com suas categorias originais, conforme os bancos precursores. O banco de dados de acidentes de transporte foi formado com 13 variáveis, 2 numéricas (idade e dias de internação) e as demais (11) nominais (Tabela 2).

RESULTADOS

Tabela 2 – Variáveis, descrição e suas categorias

Variável	Descrição	Categorias
COD_VIT	Código da vítima	NÚMERO ARÁBICO
SEXO	Sexo	M, F
IDADE	Idade em 2006	NÚMERO ARÁBICO
MES_OC	Mês de ocorrência	JAN, FEV, ABR, MAI, JUN, JUL, AGO, SET, OUT, NOV, DEZ
TP_ACID	Tipo de acidente	ABALROAMENTO, ATROPELAMENTO, CAPOTAMENTO, CHOQUE, COLISÃO, QUEDA, SINF
TRANSP_VIT	Meio de transporte da vítima	A PÉ, VEICULO, MOTOCICLETA, AUTOMÓVEL, BICICLETA, SINF
TP_VIT	Tipo de vítima	CONDUTOR, PASSAGEIRO, PEDESTRE, IGNORADO
DIA_INTER	Dias de internação	NÚMERO ARÁBICO
ZONA_RES	Zona de residência	NORTE, SUL, LESTE, OESTE, SINF
ZONA-OC	Zona de ocorrência	NORTE, SUL, LESTE, OESTE, SINF
ASSIST_MED	Assistência médica	SIM, NÃO
LC_ASSIST	Local de assistência	HCANCERMT, HGU, HUJM, HJDCBA, HMBOMJESUS, HPSMCBA, HSTAHELENA, HSTAROSA, SOTRAUMA, STACASA, NENHUM
EVOL	Evolução	ALTA, ÓBITO

As duas variáveis numéricas criadas, idade e dias de internação, também foram categorizadas, pois o algoritmo *Apriori* somente aceita variáveis nominais. A idade foi categorizada em seis classes com intervalos de 10,^{18, 131-132}, com exceção da última classe de 60 anos e mais, a qual representa a faixa dos idosos. A variável dias de internação foi categorizada segundo os quartis (Tabela 3).

RESULTADOS

Tabela 3 - Categorias da variável idade e dias de internação

Variável	Descrição	Categorias
FX_ET	Faixa etária	1) 0-9; 2) 10-19; 3) 20-29; 4) 30-39; 5) 40-49; 6) 50-59; 6) 60 E MAIS
CAT_DIAS_ INTER	Dias de internação (categorizada)	0) sem internação; 1) 1 a 2 dias 2) 3 a 5 dias; 3) 6 dias e mais de internação

Vale destacar as definições das categorias das variáveis: tipo de acidente, meio de transporte da vítima e tipo de vítima para a SEJUSP.

Tipo de acidente:

- Abalroamento: acidente em que os veículos colidem lateral ou transversalmente, estando os mesmos trafegando pela mesma via, podendo ser no mesmo sentido ou em sentidos contrários.
- Atropelamento: Acidente em que o pedestre sofre impacto de um veículo.
- Capotamento: acidente em que o veículo gira sobre si mesmo, em qualquer sentido, chegando a ficar com as rodas para cima, imobilizando-se em qualquer posição.
- Choque: acidente em que há impacto de um veículo contra qualquer objeto fixo (muro, árvore, poste) ou móvel, mas sem movimento (veículo parado).
- Colisão: acidente em que um veículo em movimento sofre o impacto de outro veículo, também em movimento.
- Queda: quando ocorre os acidentes de transporte envolvendo veículo em movimento, tais como: queda de um objeto sobre ou no interior do veículo ou projeção acidental.
- Sinf: campo em branco.

Meio de transporte da vítima:

- A pé: a pessoa que se locomove, por meio dos seus membros inferiores, envolvida em acidente de transporte.
- Motocicleta: veículo a motor de duas rodas com um ou dois assentos para os passageiros e, algumas vezes, uma terceira roda para manter um “*side-car*”. Inclui bicicleta motorizada, motoneta, patinete motorizado.
- Automóvel: veículo de quatro rodas projetado essencialmente para transportar até dez pessoas.
- Bicicleta: veículo de transporte terrestre movido apenas por meio de pedais. Inclui triciclo e velocípede.
- Veículo: qualquer outro tipo de veículo não contemplado nas categorias acima. Inclui veículo de tração animal, animal montado, ambulância, kômbi, trem, metrô, van, perua e veículos pesados como caminhão, carreta e trator.
- Sinf: campo em branco.

Tipo de vítima:

- Condutor: Pessoa que conduz um veículo de transporte ou tem intenção de manobrá-lo.
- Passageiro: ocupante de um veículo, que não o condutor e que esteja sentado no banco da frente ou de trás.
- Pedestre: toda pessoa envolvida em um acidente, mas que no momento em que o mesmo ocorreu, não estava viajando no interior de ou sobre um meio de transporte, ou seja, estava a pé.
- Sinf: campo em branco

5.1.2 Entrada de dados na ferramenta

Após a formatação das variáveis, preparou-se o banco para entrada de dados na ferramenta selecionada, o *software* Weka versão 3.6.0. Este *software* lê um formato de arquivo texto padronizado com extensão **.arff**. Para importar os dados do banco de acidentes de transporte de Cuiabá, inicialmente armazenados em uma planilha do Microsoft Office Excel 2003, foi necessário transformá-la em arquivo onde os dados foram separados por vírgula, ou seja, arquivo **.csv**.

Em seguida, abriu-se um arquivo **.csv** no bloco de notas, versão 5.1, para transformá-lo na extensão **.arff**. Já com arquivo aberto, a primeira linha representa o cabeçalho das colunas, ou seja, os nomes dos atributos que estão separados por ponto e vírgula. As linhas seguintes representam os registros onde cada atributo do registro está separado por ponto e vírgula. Como o *software* WEKA somente aceita os registros separados por vírgula, então foi utilizado o recurso do menu, localizar e substituir o ponto e vírgula {;} por vírgula {,}.

Vale destacar que, no cabeçalho do arquivo (Figura 4) com extensão **.arff**, foi acrescentado o seguinte:

- na primeira linha, o nome do conjunto de dados atribuído pelo comando `@relation nome_do_conjunto_de_dados- @Relation ACIDENTECUIABA;`
- em seguida, a relação dos atributos, onde foi inserido o nome do atributo, o tipo e/ou suas categorias, definido por `@attribute nome_do_atributo tipo ou {categorias}`, por exemplo: `@attribute TP_ACID{ABALROAMENTO, ATROPELAMENTO, CAPOTAMENTO, CHOQUE, COLISÃO, QUEDA, SINF};`

RESULTADOS

- na seção dos dados, foi inserido o comando @data e, nas linhas seguintes, os registros foram representados, sendo um em cada linha;
- o último passo foi gravar o arquivo na extensão .arff, recebendo o nome - ACIDENTECUIABA.arff.

Figura 4 - Arquivo com cabeçalho e seus atributos no formato exigido pela ferramenta WEKA

```

@Relation ACIDENTECUIABA
@attribute COD_VIT real
@attribute SEXO{M, F}
@attribute IDADE real
@attribute Fx_ET{1, 2, 3, 4, 5, 6}
@attribute MES_OC{JAN, FEV, MAR, ABR, MAI, JUN, JUL, AGO, SET, OUT, NOV, DEZ}
@attribute ZONA_RES{NORTE, SUL, LESTE, OESTE, SINF}
@attribute ZONA_OC{NORTE, SUL, LESTE, OESTE, SINF}
@attribute TP_ACI{ABALOAMENTO, ATROPELAMENTO, CAROTAMENTO, CHOQUE, COLISÃO, QUEDA, SINF}
@attribute TP_VIT{CONDUTOR, PASSAGEIRO, PEDESTRE, IGNORADO}
@attribute TRANSP_VIT{APÉ, VEICULO, MOTOCICLETA, AUTOMÓVEL, BICICLETA, SINF}
@attribute ASSIST_MED{SIM, NÃO}
@attribute DIA_INTER real
@attribute EVOL{ALTA, OBITO}
@attribute CAT_DIAS_INTER{0, 1, 2, 3}

@DATA
1, M, 65, 6, AGO, LESTE, SINF, ATROPELAMENTO, PEDESTRE, APÉ, SIM, HPSMCBA, 3, ALTA, 1
2, F, 26, 3, JUN, LESTE, SINF, ATROPELAMENTO, PEDESTRE, APÉ, SIM, STACASA, 2, ALTA, 1
6, M, 21, 3, MAI, LESTE, LESTE, SINF, CONDUTOR, VEICULO, SIM, HMBOMJESUS, 3, ALTA, 2
9, M, 27, 3, MAI, SINF, LESTE, COLISÃO, IGNORADO, SINF, SIM, HMBOMJESUS, 3, ALTA, 2
11, F, 72, 6, NOV, LESTE, SINF, ATROPELAMENTO, PEDESTRE, APÉ, SIM, SOTRAUMA, 2, ALTA, 1
12, M, 31, 4, NOV, NORTE, NORTE, COLISÃO, PASSAGEIRO, VEICULO, SIM, HMBOMJESUS, 3, ALTA, 2
13, F, 34, 4, MAR, OESTE, LESTE, QUEDA, CONDUTOR, VEICULO, SIM, HPSMCBA, 7, ALTA, 3
14, M, 39, 4, JUN, SUL, SINF, PASSAGEIRO, MOTOCICLETA, SIM, HSTAELENA, 13, ALTA, 3
17, M, 63, 6, AGO, LESTE, LESTE, SINF, CONDUTOR, MOTOCICLETA, SIM, HPSMCBA, 21, ALTA, 3
18, M, 63, 6, MAI, LESTE, SINF, SINF, IGNORADO, SINF, SIM, HPSMCBA, 3, ALTA, 2
20, M, 29, 3, JUL, SINF, SINF, SINF, CONDUTOR, VEICULO, SIM, HPSMCBA, 8, ALTA, 3
21, M, 18, 2, JUN, SUL, SUL, COLISÃO, IGNORADO, SINF, SIM, HPSMCBA, 3, ALTA, 2
22, M, 18, 2, SET, LESTE, NORTE, COLISÃO, CONDUTOR, MOTOCICLETA, SIM, HPSMCBA, 2, ALTA, 1
23, M, 54, 3, ABR, LESTE, OESTE, COLISÃO, PASSAGEIRO, MOTOCICLETA, SIM, HPSMCBA, 3, OBITO, 2
24, M, 16, 2, SET, NORTE, NORTE, QUEDA, PASSAGEIRO, VEICULO, SIM, HPSMCBA, 29, ALTA, 3
25, M, 27, 3, FEV, SUL, SUL, QUEDA, PASSAGEIRO, MOTOCICLETA, SIM, HPSMCBA, 3, ALTA, 2
26, M, 29, 3, AGO, OESTE, SINF, SINF, PASSAGEIRO, VEICULO, SIM, STACASA, 11, OBITO, 3
27, M, 39, 4, AGO, OESTE, LESTE, COLISÃO, CONDUTOR, MOTOCICLETA, SIM, HPSMCBA, 3, ALTA, 2
28, M, 18, 2, JUN, OESTE, SINF, COLISÃO, IGNORADO, SINF, SIM, STACASA, 3, ALTA, 2
29, M, 28, 3, ABR, SUL, SUL, COLISÃO, PASSAGEIRO, MOTOCICLETA, SIM, HSTAELENA, 1, ALTA, 1
30, M, 56, 5, JUN, OESTE, OESTE, SINF, IGNORADO, MOTOCICLETA, SIM, HPSMCBA, 2, ALTA, 1
31, M, 29, 3, JUN, SUL, OESTE, SINF, PASSAGEIRO, MOTOCICLETA, SIM, SOTRAUMA, 1, ALTA, 1
34, M, 26, 3, MAR, SINF, SINF, COLISÃO, IGNORADO, MOTOCICLETA, SIM, HMBOMJESUS, 1, ALTA, 1
35, M, 59, 5, SET, OESTE, OESTE, SINF, PASSAGEIRO, VEICULO, SIM, STACASA, 2, ALTA, 1
37, M, 29, 3, ABR, OESTE, OESTE, COLISÃO, IGNORADO, SINF, SIM, HMBOMJESUS, 3, ALTA, 2
41, M, 45, 5, SET, SUL, NORTE, SINF, PASSAGEIRO, VEICULO, SIM, HMBOMJESUS, 3, ALTA, 2
43, M, 29, 3, SET, SUL, OESTE, CHOQUE, IGNORADO, MOTOCICLETA, SIM, HPSMCBA, 1, ALTA, 1
44, M, 38, 4, SET, OESTE, OESTE, ATROPELAMENTO, PEDESTRE, APÉ, SIM, SOTRAUMA, 2, ALTA, 1
45, M, 64, 6, OUT, NORTE, SINF, COLISÃO, CONDUTOR, VEICULO, SIM, HSTAELENA, 1, ALTA, 1
46, M, 46, 5, MAR, NORTE, NORTE, COLISÃO, IGNORADO, MOTOCICLETA, SIM, HMBOMJESUS, 4, ALTA, 2
50, M, 23, 3, ABR, NORTE, NORTE, COLISÃO, PASSAGEIRO, MOTOCICLETA, SIM, HMBOMJESUS, 3, ALTA, 2
52, F, 25, 3, MAI, SUL, SINF, COLISÃO, IGNORADO, VEICULO, SIM, HPSMCBA, 2, ALTA, 1

```

No módulo Explorer da ferramenta WEKA, o arquivo com a extensão .arff foi aberto. Após o arquivo .arff ser carregado, a ferramenta proporciona uma parte de visualização das variáveis em forma de gráficos, na opção de pré-processamento (Figura 5 e 6), ou seja, descrição das variáveis em frequência absoluta.

RESULTADOS

Figura 5 - Tela do pré-processamento do software WEKA com atributos ACIDENTECUIABA.arff

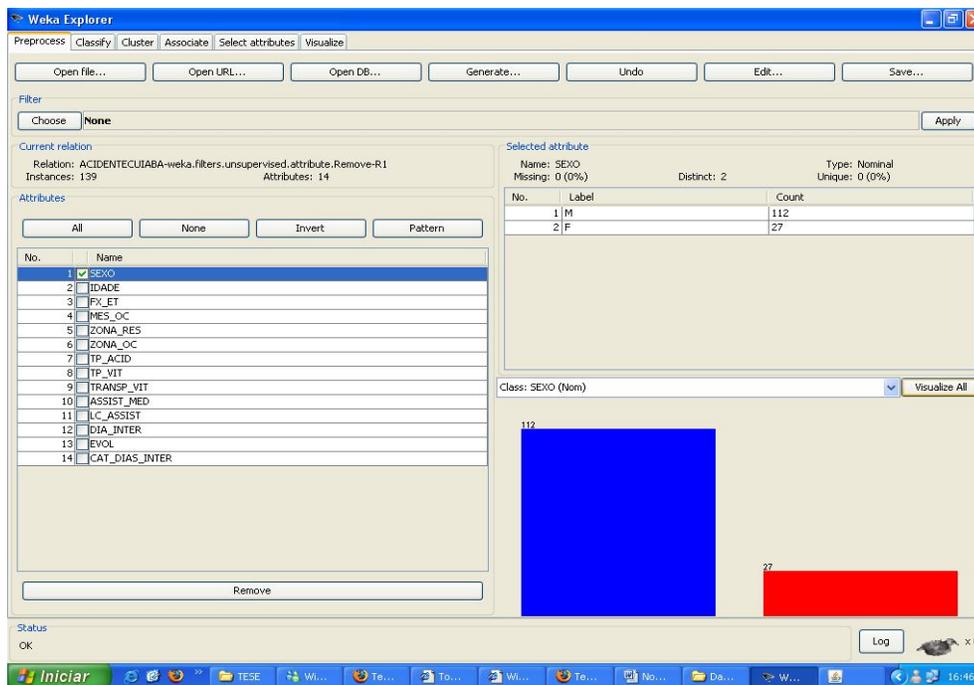
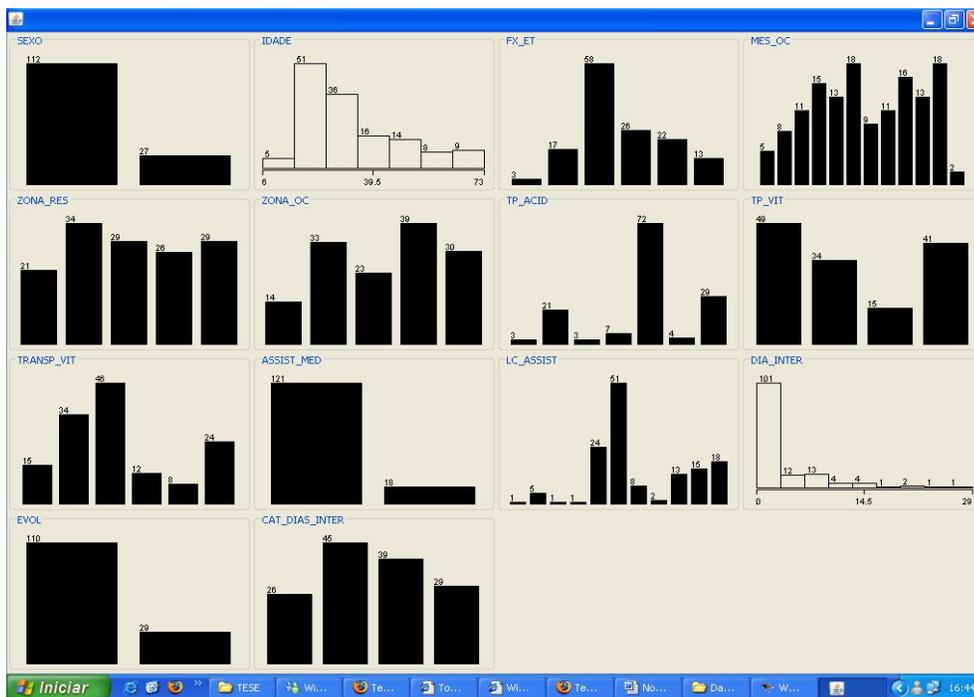


Figura 6 - Tela de visualização gráfica das variáveis



RESULTADOS

5.1.3 Análise preliminar dos dados

Por meio da distribuição de frequência, foi observada a proporção dos indivíduos em cada categoria, conforme mostra a Tabela 4 e 5.

Na Tabela 4, dos 139 pares de vítimas de acidente de transporte formados pelos bancos de dados SEJUSP, SIH/SUS e SIM do município de Cuiabá do ano de 2006, observou-se que as vítimas foram do sexo masculino 80,6%, pertencentes a faixa etária de 20-29 anos (41,7%), com a idade mínima de 6 e a máxima de 72 anos, com uma média de 33 anos, a mediana de 28 e um desvio padrão de 15 anos. Vale ressaltar, que a idade possui uma distribuição simétrica dos dados, ou seja, valores concentrados numa classe central com as frequências diminuindo à medida que se afasta nos dois sentidos dessa classe. Quanto à zona de residência, destacou-se a zona sul (24,5%) e leste (20,9%).

Tabela 4 - Frequência e proporção do conjunto de dados demográficos das vítimas de acidente de transporte, segundo variáveis selecionadas, Cuiabá / MT, 2006

Variável	Categoria	Acidente de transporte	
		Frequência	%
Sexo	Masculino	112	80,6
	Feminino	27	19,4
Faixa Etária (anos)	0-9	3	2,2
	10-19	17	12,2
	20-29	58	41,7
	30-39	26	18,7
	40-49	14	10,1
	50-59	8	5,8
	60 +	13	9,4
Zona de Residência	Leste	29	20,9
	Norte	21	15,1
	Oeste	26	18,6
	Sul	34	24,5
	SINF	29	20,9

RESULTADOS

Tabela 5 - Frequência e proporção do conjunto de dados ocorrência e das vítimas de acidente de transporte, segundo variáveis selecionadas, Cuiabá / MT, 2006

Variável	Categoria	Acidente de transporte	
		Frequência	%
Mês de Ocorrência	Janeiro	5	3,6
	Fevereiro	8	5,8
	Março	11	7,9
	Abril	15	10,8
	Mai	13	9,4
	Junho	18	12,9
	Julho	9	6,5
	Agosto	11	7,9
	Setembro	16	11,5
	Outubro	13	9,4
	Novembro	18	12,9
	Dezembro	2	1,4
Zona de Ocorrência	Leste	23	16,5
	Norte	14	10,1
	Oeste	39	28,1
	Sul	33	23,7
	SINF	30	21,6
Tipo de Vítima	Condutor	49	35,6
	Passageiro	34	24,5
	Pedestre	15	10,8
	SINF	41	29,5
Meio de Transporte da Vítima	A pé	15	10,8
	Automóvel	12	8,6
	Bicicleta	8	5,8
	Motocicleta	46	33,1
	Veículo	34	24,5
	SINF	24	17,3
Tipo de Acidente	Abalroamento	3	2,2
	Atropelamento	21	15,1
	Capotamento	3	2,2
	Choque	7	5,0
	Colisão	72	51,8
	Queda	4	20,9
	SINF	29	20,9

No conjunto de dados das vítimas relacionados, a ocorrência desses acidentes de transporte aconteceram em torno de 48,1% nos quatro meses

RESULTADOS

do ano de 2006, distribuídos da seguinte forma: em abril, 10,8%; junho com 12,9%; setembro com 11,5% e, em novembro, a ocorrência foi de 12,9% (Tabela 5). Vale ressaltar que, neste conjunto, a zona oeste da cidade apresentou 28,1% das ocorrências. A principal vítima foi o condutor do veículo (35,3%), o meio de transporte da vítima ou tipo de veículo que predominou foi a motocicleta (33,1%), e o tipo de acidente que se destacou foi a colisão (51,8%) (Tabela 5) .

Em relação à assistência das vítimas, a maioria teve assistência médica (87,1%), sendo o Hospital Pronto Socorro Municipal de Cuiabá (HPSMCBA), o que mais recebeu vítimas desse conjunto de dados (36,7%). O período de internação, em média, para cada vítima foi de aproximadamente cinco dias, tendo sido no mínimo um e, no máximo, vinte e nove dias (Tabela 6).

Tabela 6 - Frequência e proporção do conjunto de dados das vítimas de acidente de transporte, segundo variáveis assistência e evolução, Cuiabá / MT, 2006

Variável	Categoria	Acidente de transporte	
		Frequência	%
Assistência Médica	Sim	121	87,1
	Não	18	12,9
Local da Assistência	HCANCERMT	1	0,7
	HGU	5	3,6
	HUJM	1	0,7
	HJDCBA	1	0,7
	HMBOMJESUS	24	17,3
	HPSMCBA	51	36,7
	HSTAHELENA	8	5,8
	HSTAROSA	2	1,4
	SOTRAUMA	13	9,4
	STACASA	15	10,8
	NENHUM	18	12,9
Dias de Internação	1º quartil (sem internação)	26	18,7
	2º quartil (1 a 2 dias)	45	32,4
	3º quartil (3 a 5 dias)	39	28,1
	4º quartil (6 dias e mais)	29	20,9
Evolução	Alta	110	79,1
	Óbito	29	20,9

RESULTADOS

Destaca-se, que a variável – dias de internação – teve uma distribuição dos dados assimétrica à direita, com uma concentração de valores à esquerda, com as freqüência decrescendo à direita. Os dias de internação foram categorizados segundo os quartis, assim distribuídos: no primeiro quartil, 18,7% dos pacientes que não foram internados e 32,4%, 28,1% e 20,9% nos quartis dois, três, quatro, respectivamente. Das 139 vítimas, 110 (79,1%) tiveram evolução para alta no ano de 2006 (Tabela 6).

5.1.4 Aplicação do algoritmo *APriori*

Após a análise de frequências das variáveis, selecionou-se o método de associação (*Associate*) na ferramenta WEKA, que apresenta quatro algoritmos para obter regras de associação, a saber: *APriori*, *Hotspot*, *PredictiveAPriori* e *Tertius*. Porém, neste estudo, optou-se por selecionar o algoritmo *APriori*. Como não encontrou na literatura um padrão dos parâmetros para este estudo, optou-se simular a várias situações com valores diferentes de confiança mínima e suporte mínimo (Tabela 7). Que verificou um decréscimo acentuado no número de regras com aumento do suporte e redução da confiança.

Tabela 7 – Confiança, suporte mínimos estipulados e número regras geradas pelo algoritmo *APriori*.

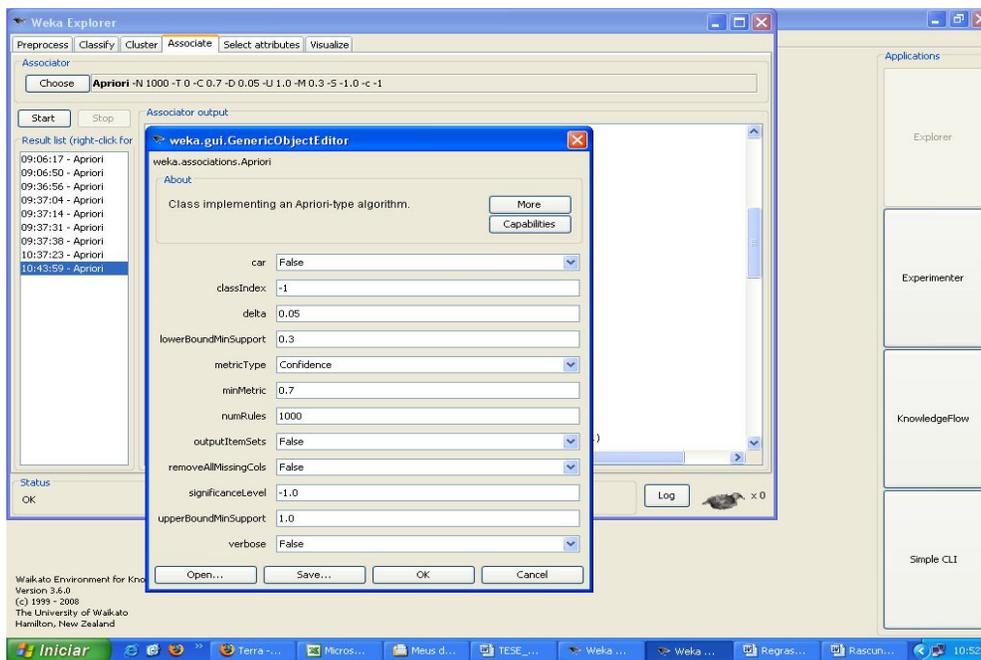
Confiança mínima	Suporte mínimo	Número regras geradas
90	10	490
80	20	153
70	30	50
70	40	18

RESULTADOS

Com auxílio da medida de interesse *lift*, também conhecida como *interest*, de um pesquisador / especialista da área de acidentes de transporte que colaborou para escolha das melhores regras geradas pelo algoritmo Apriori.

Nas regras geradas pelo algoritmo de associação das vítimas de acidentes de transporte a partir dos dados relacionados da Secretaria de Estado Justiça e Segurança Pública, do Sistema de Informação Hospitalar do Sistema Único de Saúde e do Sistema de Informação sobre Mortalidade (Apêndice 1) com os parâmetros estipulados, observou-se muitas regras redundantes, por meio da medida de interesse e com auxílio de um especialista da área escolheu as regras geradas pelo modelo que continha os parâmetros: confiança 70% e suporte 30% (Figura 7), destacando que estas representavam melhor o contexto do município, apesar das limitações do dados dos sistemas de informação.

Figura 7 - Valores dos parâmetros selecionados

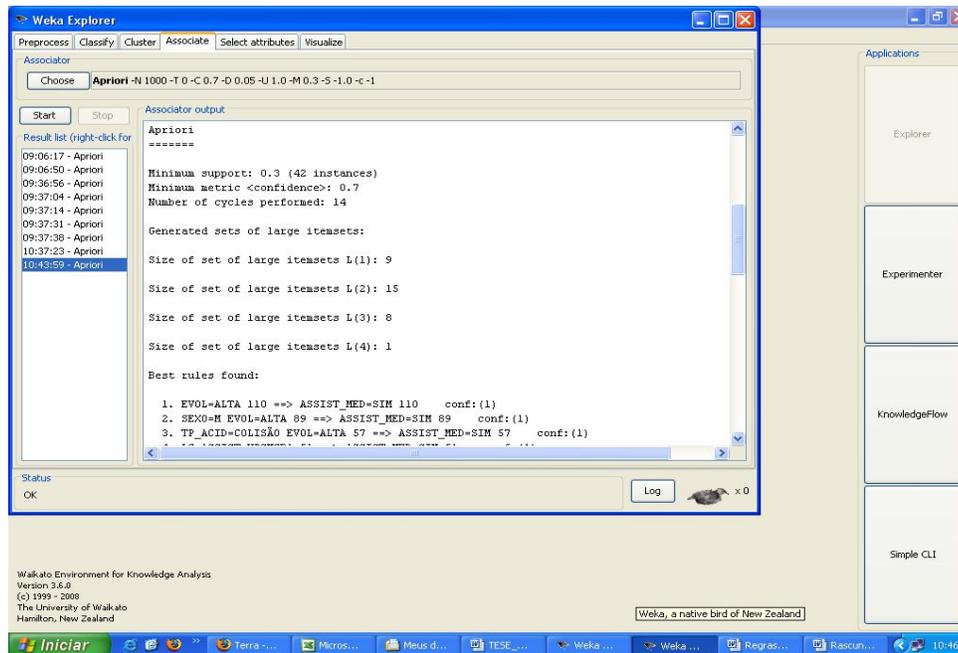


Na Figura 8 mostra os conjuntos gerados, que foram quatro *itemsets*, sendo o primeiro com nove atributos. Estes se combinaram para gerar o

RESULTADOS

segundo *itemsets*, que teve quinze subconjuntos e cada uma com dois atributos. Já com a combinação do segundo *itemsets*, foi gerado o terceiro *itemsets* com oito subconjuntos com três variáveis.

Figura 8 - Resultados dos *Itemsets* gerados pelo algoritmo *APriori*



O último *itemsets* teve um único subconjunto com quatro variáveis - SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM EVOL=ALTA 50, sendo que dos 139 vítimas de acidentes 35,9% foram do sexo masculino, o tipo de acidente foi por colisão, tiveram assistência médica e evoluíram para alta.

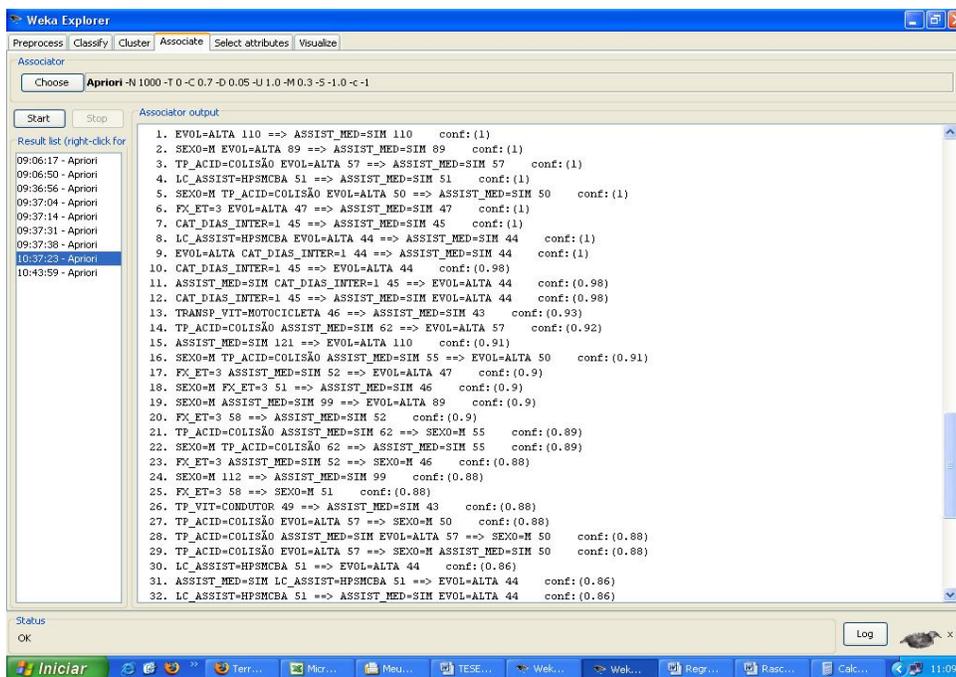
As regras associativas foram apresentadas no formato conforme mostrado na Figura 9.

No Quadro 3 mostra as cinquenta regras geradas pelo algoritmo de associação das vítimas de acidentes de transporte a partir dos dados relacionados da Secretaria de Estado Justiça e Segurança Pública, do Sistema de Informação Hospitalar do Sistema Único de Saúde e do Sistema de Informação sobre Mortalidade, foi verificado que, em aproximadamente quarenta regras

RESULTADOS

geradas, os atributos ou variáveis referentes à assistência médica hospitalar (sim) e evolução (alta) estavam presentes, como antecedente ou consequente. Outras duas variáveis frequentes nas regras geradas foram o tipo de acidente por colisão e o sexo masculino.

Figura 9 - Resultados das regras associativas gerados pelo algoritmo *APriori*



Quanto às regras, pode-se observar que todas apresentaram a medida de interesse maior 1, demonstrando uma associação positiva. Com esta condição, o especialista analisou as regras e com seu conhecimento da área, eliminou as regras redundantes, selecionou as melhores regras que fossem úteis para o contexto do município de Cuiabá. Das cinquenta, 6 regras foram selecionadas que estão em negrito no Quadro 3.

A poda das regras demonstrou que várias regras apresentaram redundância, como por exemplo: a regra de número 5 igual a 13, 28, 42 e 43; a de número 3 foi igual a 14 e 48, a de número 8 foi igual a 31 e 32; a de número 18 igual a 23 e 46; a de número 21 igual a 22 e a de número 39 foi igual a 45 e 50.

RESULTADOS

Quadro 3 - Regras associativas e seus parâmetros: *Interest*, suporte e confiança

Regras X ==>Y	Conf. (%)	Sup. XY (%)	<i>interest</i>
1. EVOL=ALTA 110 ==> ASSIST_MED=SIM 110	100,0	79,1	1,3
2. SEXO=M EVOL=ALTA 89 ==> ASSIST_MED=SIM 89	100,0	64,0	1,6
3. TP_ACID=COLISÃO EVOL=ALTA 57 ==> ASSIST_MED=SIM 57	100,0	41,0	2,4
4. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM 51	100,0	36,7	2,7
5. SEXO=M TP_ACID=COLISÃO EVOL=ALTA 50 ==> ASSIST_MED=SIM 50	100,0	36,0	2,8
6. FX_ET=3 EVOL=ALTA 47 ==> ASSIST_MED=SIM 47	100,0	33,8	3,0
7. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM 45	100,0	32,4	3,1
8. LC_ASSIST=HPSMCBA EVOL=ALTA 44 ==> ASSIST_MED=SIM 44	100,0	31,7	3,2
9. EVOL=ALTA CAT_DIAS_INTER=1 44 ==> ASSIST_MED=SIM 44	100,0	31,7	3,2
10. CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44	98,0	31,7	3,1
11. ASSIST_MED=SIM CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44	98,0	31,7	3,1
12. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM EVOL=ALTA 44	98,0	31,7	3,1
13. TRANSP_VIT=MOTOCICLETA 46 ==> ASSIST_MED=SIM 43	93,0	30,9	3,0
14. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> EVOL=ALTA 57	92,0	41,0	2,2
15. ASSIST_MED=SIM 121 ==> EVOL=ALTA 110	91,0	79,1	1,1
16. SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM 55 ==> EVOL=ALTA 50	91,0	36,0	2,5
17. FX_ET=3 ASSIST_MED=SIM 52 ==> EVOL=ALTA 47	90,0	33,8	2,7
18. SEXO=M FX_ET=3 51 ==> ASSIST_MED=SIM 46	90,0	33,1	2,7
19. SEXO=M ASSIST_MED=SIM 99 ==> EVOL=ALTA 89	90,0	64,0	1,4
20. FX_ET=3 58 ==> ASSIST_MED=SIM 52	90,0	37,4	2,4
21. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> SEXO=M 55	89,0	39,6	2,2
22. SEXO=M TP_ACID=COLISÃO 62 ==> ASSIST_MED=SIM 55	89,0	39,6	2,2
23. FX_ET=3 ASSIST_MED=SIM 52 ==> SEXO=M 46	88,0	33,1	2,7
24. SEXO=M 112 ==> ASSIST_MED=SIM 99	88,0	71,2	1,2
25. FX_ET=3 58 ==> SEXO=M 51	88,0	36,7	2,4

RESULTADOS

Continuação do Quadro 3			
Regras X ==>Y	Conf. (%)	Sup. XY (%)	interest
26. TP_VIT=CONDUTOR 49 ==> ASSIST_MED=SIM 43	88,0	30,9	2,8
27. TP_ACID=COLISÃO EVOL=ALTA 57 ==> SEXO=M 50	88,0	36,0	2,4
28. TP_ACID=COLISÃO ASSIST_MED=SIM EVOL=ALTA 57 ==> SEXO=M 50	88,0	36,0	2,4
29. TP_ACID=COLISÃO EVOL=ALTA 57 ==> SEXO=M ASSIST_MED=SIM 50	88,0	36,0	2,4
30. LC_ASSIST=HPSMCBA 51 ==> EVOL=ALTA 44	86,0	31,7	2,7
31. ASSIST_MED=SIM LC_ASSIST=HPSMCBA 51 ==> EVOL=ALTA 44	86,0	31,7	2,7
32. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM EVOL=ALTA 44	86,0	31,7	2,7
33. TP_ACID=COLISÃO 72 ==> SEXO=M 62	86,0	44,6	1,9
34. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM 62	86,0	44,6	1,9
35. ASSIST_MED=SIM 121 ==> SEXO=M 99	82,0	71,2	1,2
36. FX_ET=3 58 ==> EVOL=ALTA 47	81,0	33,8	2,4
37. FX_ET=3 58 ==> ASSIST_MED=SIM EVOL=ALTA 47	81,0	33,8	2,4
38. EVOL=ALTA 110 ==> SEXO=M 89	81,0	64,0	1,3
39. ASSIST_MED=SIM EVOL=ALTA 110 ==> SEXO=M 89	81,0	64,0	1,3
40. EVOL=ALTA 110 ==> SEXO=M ASSIST_MED=SIM 89	81,0	64,0	1,3
41. SEXO=M TP_ACID=COLISÃO 62 ==> EVOL=ALTA 50	81,0	36,0	2,3
42. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> SEXO=M EVOL=ALTA 50	81,0	36,0	2,3
43. SEXO=M TP_ACID=COLISÃO 62 ==> ASSIST_MED=SIM EVOL=ALTA 50	81,0	36,0	2,3
44. SEXO=M 112 ==> EVOL=ALTA 89	79,0	64,0	1,2
45. SEXO=M 112 ==> ASSIST_MED=SIM EVOL=ALTA 89	79,0	64,0	1,2
46. FX_ET=3 58 ==> SEXO=M ASSIST_MED=SIM 46	79,0	33,1	2,4
47. TP_ACID=COLISÃO 72 ==> EVOL=ALTA 57	79,0	41,0	1,9
48. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM EVOL=ALTA 57	79,0	41,0	1,9
49. TP_ACID=COLISÃO 72 ==> SEXO=M ASSIST_MED=SIM 55	76,0	39,6	1,9
50. ASSIST_MED=SIM 121 ==> SEXO=M EVOL=ALTA 89	74,0	64,0	1,2

RESULTADOS

A primeira regra selecionada foi a de número 5 - SEXO=M TP_ACID=COLISÃO EVOL=ALTA 50 ==> ASSIST_MED=SIM 50 – a medida de interesse foi maior que 1 (2,8) demonstrando uma associação positiva com uma confiança de 100,0% (0,92), ou seja, maior que a mínima estipulada. Ainda, observa-se que 100,0% das vítimas foram do sexo masculino, o tipo de acidente, mais especificamente, a colisão – tiveram assistência médica e evoluíram para alta.

O suporte de 36,0% também é maior que o mínimo estipulado; das 139 vítimas de acidentes de transporte terrestre ocorridos no município, 36,0% por colisão do sexo masculino, tiveram assistência médica e evoluíram para alta.

A segunda regra selecionada foi a de número 8, (LC_ASSIST=HPSMCBA EVOL=ALTA 44 ==> ASSIST_MED=SIM 44), demonstra que 44 vítimas de acidentes tiveram assistência médica hospitalar e foram internadas no hospital e Pronto Socorro Municipal de Cuiabá (HPSMCBA) e 100,0% destas evoluíram para alta. Do total das vítimas de acidentes de transporte terrestre, 31,7% tiveram assistência médica hospitalar foram internadas no HPSMCBA que, por sua vez, também evoluíram para alta. A outra regra selecionada foi de número 11 (ASSIST_MED=SIM CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44), indicando que das 45 vítimas, 98,0% tiveram assistência médica, foram internadas e permaneceram 1 a 2 dias hospitalizadas e evoluíram para alta.

Na quarta regra selecionada, foi a de número 13, verificou-se nesta regra o tipo de transporte da vítima, que foi a motocicleta e que 93,0% destes tiveram assistência médica. A penúltima regra selecionada foi a de número 26, apresentou o tipo de vítima, o condutor do veículo, que em 88,0% dos acidentes tiveram assistência médica. As regras de número 13 e 26 apresentaram um suporte no limite do esperado, mas por sua vez indicaram um grau de interesse de 3,0 e 28, respectivamente.

A última regra selecionada foi a de número 18, apresentou um grau de interesse de 2,7. Das 51 vítimas de acidentes de trânsito (90,0%) foram do sexo masculino na faixa etária de 20 a 29 anos e tiveram assistência médica.

6 DISCUSSÃO

6.1 Aplicação da mineração dos dados

6.1.1 Preparação dos dados

Neste estudo, a aplicação da técnica de mineração de dados foi apresentada por meio de um banco de dados que possuía grande volume. Vale enfatizar que, com a realização das etapas de seleção de variáveis, estabelecimento do objetivo do estudo e procedida a limpeza do banco, obteve-se uma redução significativa de variáveis e de registro. Porém, são etapas previstas na descoberta de conhecimento em bases de dados e abordadas por diversos autores ^{59, 7, 2, 4, 1}. Para Steiner ⁵ e Quoniam et al. ⁷, os ruídos da base de dados a ser minerada, deveriam ser eliminados, pois podem interferir nos padrões novos e válidos adquiridos e também no conhecimento potencialmente útil e compreensível.

Além do mais, a reformatação e a transformação variáveis foram necessárias para que ficassem em um formato que facilitasse a entrada dos dados na ferramenta e realizasse a aplicação do algoritmo selecionado. Estas etapas foram chamadas por Goldschmidt & Passos ² de codificação e enriquecimento dos dados. Eles ressaltaram que a função de enriquecimento consiste em conseguir, de alguma forma, agregar registros existentes para obter mais informações para o processo de descoberta de conhecimento. Esta função foi realizada após a junção dos bancos, onde existiam diversas variáveis que poderiam ser agregadas para gerar um novo atributo, conforme foi mostrado no Quadro 2. Com essas etapas concluídas, o banco de dados relacionado estava preparado para a próxima operação, a de entrada desses dados no *software* selecionado para minerá-los.

6.1.2 Entrada de dados na ferramenta

A entrada dos dados na ferramenta selecionada, WEKA, não é muito operacional, tendo em vista que estes estavam em uma planilha do Microsoft Office Excel 2003 e foi necessário realizar diversas operações para obter o arquivo com extensão arff.

A ferramenta TANAGRA,¹¹⁸ recentemente lançada, gratuita, que minera os dados e está sendo implementada, possui diversas alternativas para entrada, inclusive no formato (xls). Contudo, a ferramenta WEKA, já está consolidada no meio educacional e nas pesquisas científicas.

Em pesquisa realizada recentemente nas bases de dados SCIELO (*Scientific Eletronic Library online* – Brasil), foram obtidos dois artigos e, na base literária MEDLINE, da *National Library of Medicine* (Medline/Pubmed), foram encontrados trinta e seis artigos publicados nas mais diversas áreas da ciência. Para exemplificar, alguns desses artigos recentemente publicados nessas bases de dados são: (a) integração múltipla do classificador para predizer as classes estruturais da proteína⁸⁰; (b) estudo para predizer polimorfismos funcionais e reguladores¹²⁹; (c) desenvolvimento de uma árvore de decisão para determinar a densidade óssea em doenças inflamatórias usando o *software* WEKA¹³³; (d) determinação dos fatores que influenciam no sucesso da imunização na taxa de leucócitos¹³⁴; (e) extração de regras de classificação a partir de redes neurais, para auxílio à tomada de decisão na concessão de crédito bancário¹³⁵; (f) identificação de resultados da intervenção educacional em diabetes tipo 2¹³⁶; (g) demonstração do valor da relação do citrato/cálcio na urina na avaliação do risco de urolitíase¹³⁷ e, finalmente, (h) desenvolvimento ambiental para predizer a modelagem em alimentos

6.1.3 Análise preliminar dos dados

Sabe-se que a razão de sexo (M:F) é maior para este agravo e, neste estudo, encontrou-se uma razão de quatro homens vítimas de acidente de transporte para uma mulher. Os homens jovens da faixa etária (20-39 anos) também foram os mais comprometidos. Corroborando com os resultados deste estudo, pesquisas anteriores na Grande Cuiabá ^{132,139} e em Cuiabá, feitas nas unidades de Urgência e Emergência¹⁸ mostraram que os homens também foram as maiores vítimas de acidentes de transporte.

Os motociclistas foram as principais vítimas identificadas dos acidentes de transporte, sendo os adultos jovens, os mais atingidos. Resultados coincidentes foram observados em Alta Floresta¹³¹, município do interior do estado de Mato Grosso. Esses dados puderam ser explicados pelos seguintes fatores: a) motociclistas têm maior exposição corpórea; b) a dificuldade de visualização da moto, por outros motoristas é maior; c) entre os motociclistas, é maior a prevalência de comportamento inadequado no trânsito ¹⁴⁰.

Sabe-se que os acidentes envolvendo motocicletas são um problema emergente em Cuiabá em razão do crescente uso deste veículo não somente para lazer, mas principalmente para o trabalho nas atividades de serviços de “motoboy” e “mototáxi”. Estes fatos foram comprovados por dados do DETRAN-MT, publicados no anuário estatístico do estado de Mato Grosso, em 1999 em Cuiabá, onde as motocicletas representavam cerca de 10,0% da frota total de veículos. Desde 2005, esse percentual vem crescendo para 18,0%, representando um aumento de 80,0% ^{141,127}. Na atual década também foi registrado o crescimento desta frota, principalmente nas áreas urbanas brasileiras ²⁰. Além desse aumento efetivo na circulação, a proposta de tais serviços é a rapidez no atendimento ao cliente, o que pode contribuir para o maior número de acidentes com este meio de transporte.

DISCUSSÃO

O acidente por colisão foi o mais frequente, conforme demonstrado acima. Em geral, isso acontece após perda de controle do veículo pelo condutor, sendo a alta velocidade a possível causa, em razão da elevada energia transferida ao corpo da vítima no momento do impacto, o que favoreceu ainda mais a ocorrência de desfechos mais graves. Além da velocidade, o tamanho (massa) do veículo envolvido no acidente também foi responsável por parcela significativa do valor dessa energia transferida, sendo, portanto, esperados maiores percentuais de pessoas traumatizadas em confrontos com veículos pesados (caminhão ou ônibus)¹⁴². Em estudo realizado no Distrito Federal no ano de 2005, a colisão foi responsável pela metade das vítimas de acidente de trânsito naquele período¹⁴³.

Outro evento importante a ser ressaltado é a legislação de trânsito, pois no decorrer de dezesseis anos da implantação do novo Código de Trânsito Brasileiro (CTB), ainda se registram acidentes com alta velocidade, em razão da não utilização de equipamentos de segurança e do consumo de bebidas alcoólicas. Souza et al.¹⁴⁴ chamaram a atenção para a continuidade da implantação do CTB, promovendo aprofundamento do processo educativo dos condutores e pedestres, com uma fiscalização rigorosa do consumo de bebidas alcoólicas, de cintos de segurança e do respeito aos limites de velocidade.

Sabe-se que as ocorrências de acidentes com vítimas tiveram um registro de Boletim de Ocorrência (BO). Tais vítimas que se encontravam com sinais vitais ativos foram atendidas e encaminhadas pelo Corpo de Bombeiro ou Serviço de Atendimento Móvel de Urgência (SAMU) à assistência médica hospitalar. O local de assistência que mais recebeu esses pacientes foi o Hospital Pronto Socorro Municipal de Cuiabá, que é referência para as pessoas traumatizadas do município. A letalidade daquelas que tiveram assistência médica hospitalar foi aproximadamente de 10,0%, mesmo com atendimento móvel de urgência e emergência (pré-hospitalar). Este indicador é preocupante, pois num estudo realizado nas capitais das Regiões Norte e Centro-Oeste por Duarte et al.²², a capital do estado de Mato Grosso teve um forte incremento do risco de morte por acidente de transporte terrestre após o período pós-implantação do CTB.

Malvestio & Sousa ¹⁴⁵ destacaram em uma publicação que, na fase pré-hospitalar, não se reverte um quadro extremamente grave, mas a rapidez de chegada à cena e ao hospital, bem como as intervenções iniciais apropriadas, previne o agravamento do quadro e o surgimento de novas lesões, melhorando as condições para alguns casos e até atrasando resultados fatais, dando à vítima a chance de chegar ao tratamento definitivo e beneficiar-se dele.

Em relação à média de dias de internação das vítimas de trânsito em Cuiabá / MT, verificou-se que foi de cinco dias, o que está abaixo da média encontrada em um estudo realizado no Hospital Universitário Cajuru, no município de Curitiba – PR, que foi de seis dias de internação ¹⁴⁶. Os autores consideraram duas questões quanto à internação: a primeira diz respeito à média de internação e alta, já que a falta de leitos é um problema comum nos hospitais públicos brasileiros e, a segunda, está diretamente relacionada com os altos custos do tratamento hospitalar para as vítimas de acidente de trânsito. Em outra pesquisa realizada na grande Cuiabá, os autores encontraram um tempo médio de internação, durante o período de 1998 a 2005, de 6,9 dias de internação ¹³².

Por fim, os dados da análise preliminar retratam peculiaridades das vítimas de acidentes de transporte no município de Cuiabá e apontam para a necessidade de intervenções específicas para redução desses eventos. Nesse sentido, é necessário reavaliar as políticas públicas de transporte, o plano diretor da cidade e o estatuto da cidade.

6.1.4 Aplicação do algoritmo *APriori*

Uma importante dificuldade na análise desse estudo ocorre em consequência do processo de comparação dos dados referentes a acidentes de transporte terrestre, pela técnica de mineração de dados e aplicação do algoritmo, que possui uma escassez de publicações específicas sobre o tema. Grande parte

DISCUSSÃO

das produções científicas nacionais e internacionais referentes a causas externas está relacionada à mortalidade ou morbidade hospitalar ¹⁴⁷⁻¹⁴⁸. Até o presente momento, apenas uma publicação foi encontrada nas bases de dados literárias indexadas sobre a mineração de dados de acidentes de trânsito ⁷⁵.

Reconhece-se ainda, que este estudo oferece somente uma perspectiva pontual da realidade, por meio da técnica de mineração de dados dos acidentes em Cuiabá. Porém, apesar desta limitação e restrição a apenas uma área e um período, é possível constatar características dos pares relacionados das vítimas nas três bases de dados selecionadas. Destaca-se que maioria das vítimas de trânsito são atendidas no ambulatório ^{18, 131}, por isso a redução dos pares verdadeiros neste estudo. Como Sistema de Informação Ambulatorial não fez parte deste estudo por não ter atributos chaves como nome do paciente, nome da mãe e idade no seu banco de dados.

Quanto aos parâmetros estipulados na ferramenta, vale ressaltar que estes foram feitos baseando-se em dados e informações sobre o uso desta tecnologia, o que pode estabelecer também um grau de limitação no estudo, uma vez que é evidente a escassez de publicações sobre análise dos dados por meio desta técnica. A maioria das regras geradas teve uma alta confiança, indicando a ótima precisão destas. Porém, autores como Berry & Linoff ⁸⁵ já destacavam que uma associação pode ter a mais alta confiança, todavia pode ser pouco útil, se a combinação ocorrer raramente.

Por outro lado, Agrawal et al. ⁸⁴, ressaltavam que no estabelecimento das regras de associação, dificilmente encontra-se todas as regras com suporte e a confiança acima de um determinado valor mínimo, pois na prática, os usuários desta tarefa de mineração normalmente estão interessados somente em um subconjunto de associações.

Entretanto, neste estudo considera-se seis regras do modelo com confiança 70,0% e suporte 30,0%, estando as mesmas com parâmetros acima do selecionado, com *Interest* maior 1. O especialista analisou e descreveu as seis regras que apresentaram um conhecimento útil e interessante para contexto.

DISCUSSÃO

O conhecimento obtido da regras foi compreensível e em consonância com os resultados e pesquisas anteriores em Cuiabá e na Grande Cuiabá ^{18, 139}. Porém, vale ressaltar que estas pesquisas não utilizaram a técnica mineração de dados para analisar os dados.

Ainda sobre os parâmetros mínimos estipulados, Dunham ¹⁴⁹, destacou a prática comum de estabelecer valores de suporte e confiança mais baixos, pois quando seus valores são altos, podem-se obter regras muito triviais, não possibilitando a aquisição de novos conhecimentos.

Além do mais, o número de transações relativamente pequeno que traz uma grande limitação para estudo. Outra restrição foi às variáveis que ocorrem na maioria das regras que podem mascarar ou modificar os resultados. Porém, das regras selecionadas pelo especialista a maioria teve uma medida de interesse maior 2, que garante que as regras possui uma associação positiva, e não seja um fato aleatório.

No entanto, resultado das regras muito se assemelha ao encontrado em outros estudos realizados por outros métodos e técnicas de análise em Cuiabá / MT ¹⁸, na Grande Cuiabá ^{132, 139}, Alta Floresta / MT¹³¹ e no Distrito Federal ¹⁴³.

Pode-se salientar que as regras geradas, apesar de triviais, demonstram que o tipo de acidente por colisão faz vítimas e vítimas fatais, ilustrando que todos os tipos de veículos não são armaduras invulneráveis e que os condutores necessitam de preparo para sua condução. Pode-se inferir que a exposição ao risco está firmemente ligada ao comportamento, ao grau de prudência, à maneira de agir e à maneira de sentir, que em síntese, estabelecem a maneira de conduzir. Frente à grande dificuldade em se intervir no aumento da população e no crescimento da frota de veículos, propõe-se uma atitude mais efetiva e agressiva para educar e prevenir a população.

Nas regras apresentadas, destaca-se também que os homens foram as principais vítimas de acidentes uma faixa etária jovem, mesmo que uma grande proporção tenha recebido assistência médica e evoluído para alta. Sabe-se,

DISCUSSÃO

porém, que essa situação interfere na economia familiar, uma vez que a faixa etária mais acometida foi a dos indivíduos economicamente ativos, além disso, inclui-se o prejuízo social imensurável que o trauma traz. Não há como avaliar quantitativamente o impacto social das sequelas que o trauma causa para a população, por outro lado, há preocupação com os custos financeiros do seu atendimento (assistência médica), acompanhamento e reabilitação.

Segundo pesquisa realizada pelo Instituto de Pesquisa Aplicada (IPEA) entre os anos de 2001 e 2003, os custos dos acidentes de trânsito em aglomerações urbanas produziram perdas anuais da ordem de 5,3 bilhões de reais²⁶.

Outra pesquisa do mesmo instituto, em 2006, demonstrou que os impactos sociais e econômicos dos acidentes de trânsito nas rodovias brasileiras foram bastante significativos, estimados em 24,6 bilhões de reais; custos estes devidos principalmente às despesas com cuidados em saúde, à perda de produção, relativo à morte das pessoas ou interrupção de suas atividades, seguido dos custos associados aos veículos. Além dos custos diretos, há vários outros custos indiretos, muitas vezes invisíveis, de impactos imensuráveis, que acabam promovendo uma desestruturação familiar e pessoal ¹⁵⁰.

Por fim, apesar de poucas regras geradas e consideradas, têm-se elementos suficientes para propor medidas de prevenção desses acidentes por colisão envolvendo motoristas do sexo masculino. Salienta-se que a conclusão e obtenção destas regras foram baseadas em dados reais, que mostram efetivamente o cenário e podem sustentar programas dedicados de prevenção à violência no trânsito.

7 CONCLUSÕES

Este estudo foi realizado a partir de dados relacionados pelo método probabilístico e aplicação da tecnologia de mineração de dados, por meio do algoritmo *APriori*, que gerou regras associativas sobre as vítimas de acidentes de transporte terrestre ocorridos em Cuiabá / MT, no ano de 2006. Desta forma, conclui-se que:

- A tecnologia de mineração de dados pode ser utilizada para analisar os dados de um banco relacionado de dados de saúde e da segurança pública, apesar das limitações dos dados secundários;
- Por meio da tecnologia de mineração de dados, foram identificadas as cinquenta regras geradas pelo algoritmo *APriori*, segundo a confiança e suporte selecionados para o banco de dados de vítimas de acidentes de transporte terrestre em Cuiabá, Mato Grosso, no ano de 2006;
- As seis melhores regras, consideradas, segundo os parâmetros estipulados e pelo especialista, indicaram um conhecimento compreensível e útil à caracterização das vítimas de acidentes em Cuiabá.

8 CONSIDERAÇÕES FINAIS

Diante dos resultados e conclusões observadas, pode-se considerar a tecnologia de mineração de dados como uma poderosa ferramenta na análise de dados secundários e também como ferramenta para subsidiar o processo de tomada de decisões, com a extração de conhecimento útil de bases de dados provenientes dos sistemas de informações da saúde e da segurança pública.

Esta análise permitirá, também, estabelecer estratégia de participação intersetorial e interinstitucional para tomada decisão – sem deixar de respeitar as inúmeras diferenças na gênese e na vulnerabilidade desse agravo, frente à maioria dos problemas de saúde – além de influenciar a elaboração de políticas públicas com a finalidade de prevenção de acidentes.

Finalmente, espera-se que os resultados identificados neste estudo possam contribuir para uma reflexão sobre os acidentes de trânsito em Cuiabá e que possibilitem a apresentação de mais uma tecnologia para obtenção de conhecimento a partir de uma base de dados.

E nos trabalhos futuros sejam incorporados outras fontes de dados para ampliação dos pares verdadeiros, e também utilizando outras tarefas, técnicas e bem como outras medidas avaliação das regras objetivas e subjetivas.

9 REFERÊNCIAS BIBLOGRÁFICAS

1. Cardoso ONP, Machado RTM. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. *Rev Adm Pública*. 2008; 42(3):495-528.
2. Goldschmidt R, Passos E. *Data Mining: Guia Prático. Conceitos, técnicas, ferramentas, orientações e aplicações*. Rio de Janeiro: Elsevier; 2005.
3. Marcano AYJ, Talavera PR. Minería de datos como soporte a la toma de decisiones empresariales. *Opcion*. 2007; 23(52):104-118.
4. Araujo Junior RH, Tarapanoff K. Precisão no processo de busca e recuperação da informação: uso da mineração de textos. *Ci. Inf*. 2006; 35(3):236-247.
5. Steiner MTA, Soma NY, Shimizu T, Nievola JC, Neto-Steiner PJ. Abordagem de um problema médico por meio do processo de KDD com ênfase à análise exploratória dos dados. *Gest & Prod*. 2006; 13(2):325-337.
6. Costa LF. Bioinformatics: perspectives for the future *Genet Mol Res*. 2004; 3(4):564-574.
7. Quoniam L, Tarapanoff K, Araujo Junior RH, Alvares L. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. *Ci Inf*. 2001; 30(2):20-28.
8. Marin HF, Whitaker IY, Souza MJC, Saporoli ECL. Informática em saúde: uma nova proposta aos profissionais de enfermagem. *Rev Paul Enf*. 1990; 3(4):160-160.
9. Moraes HIS. *Política, tecnologia e informação em saúde*. Salvador: Casa da Qualidade Editora; 2002.
10. Halamka JD. Health Information Technology: Shall We Wait for the Evidence? *Ann Intern Med*. 2006; 144(10):775-776.
11. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of

REFERÊNCIAS BIBLIOGRÁFICAS

Medical Care. Ann Intern Med. 2006; 144(10):742-752.

12. Moresi EAD. Delineando o valor do sistema de informação de uma organização. Ci Inf. 2000; 29(1):14-24.

13. Lobo LCG. Sistemas de informação e planejamento de saúde. Brasília: UNB; 1993.

14. Minayo MCS. Violência: um problema para a saúde dos brasileiros. In: Impacto da violência na saúde dos brasileiros. Brasília: Ministério da Saúde. Secretaria de vigilância em saúde; 2005. p. 9-33.

15. Secretaria da Saúde de Mato Grosso. Análise de situação das doenças e agravos não transmissíveis (DANT) 2004 e 2005. Cuiabá: KCM, 2008. 128p.

16. Souza ER, Minayo MCS, Malaquias JV. Violência no trânsito: expressão da violência social. In: Impacto da violência na saúde dos brasileiros. Brasília: [MS] Ministério da Saúde. Secretaria de Vigilância em Saúde; 2005. p. 279-312.

17. [MS] Ministério da Saúde. Mortalidade por armas de fogo no Brasil: 1991-2000. Brasília: Ministério da Saúde; 2004.

18. Oliveira LR, Mello Jorge MHP. Análise epidemiológica das causas externas em unidades de urgência e emergência em Cuiabá/Mato Grosso. Rev Bras Epidemiol. 2008;11(3):420-30.

19. [OMS] Organização Mundial de Saúde. Centro Colaborador da OMS para a Classificação de Doenças em Português. Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID-10). 8^a ed. São Paulo: Editora da Universidade de São Paulo (EDUSP); 2000.

20. [MS] Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise de Situação de Saúde Política Nacional de Redução da Morbimortalidade por Acidentes e Violências: Portaria GM 737, de 16.5.2001, publicada no Diário Oficial da União nº 96, Brasília, 18 mai. 2001. Seção 1E. 2ed Brasília: Editora do Ministério da Saúde, 2005.

21. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de

REFERÊNCIAS BIBLIOGRÁFICAS

Análise de Situação em Saúde. Mortalidade por acidentes de transporte terrestre no Brasil / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise de Situação em Saúde. – Brasília : Ministério da Saúde, 2007. 80 p

22. Duarte EC, Duarte E, Sousa MC, Tauil PL, Monteiro RA. Mortalidade por acidente de transporte terrestre e homicídios em homens jovens das capitais das Regiões Norte e Centro-Oeste do Brasil, 1980-2005. Epidemiol Serv Saúde. 2008;17(1):7-20.

23. [OMS] Organização Mundial de Saúde. Relatório mundial sobre violência e saúde. Brasília: OMS/OPS, 2002.

24. OMS Organización Mundial de la Salud. Informe mundial sobre prevención de los traumatismos causados por el tránsito: resumen. Washington DC: Organización Mundial de la Salud; 2004.

24. OMS Organización Mundial de la Salud. Informe mundial sobre prevención de los traumatismos causados por el tránsito: resumen. Washington DC: Organización Mundial de la Salud; 2004.

25. [MS] Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde (DATASUS) [homepage na Internet] Banco de dados dos Sistemas de Informações sobre Mortalidade (SIM). Brasil; 2009. acesso em 29 jul 2009. disponível em <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obtm.def>

26. IPEA (Instituto de Pesquisa Econômica Aplicada). Impactos sociais e econômicos dos acidentes de trânsito nas aglomerações urbanas brasileiras: Relatório Executivo / IPEA. 2003. 45 p. acesso em 15 jun 09. Disponível em < <http://www.ipea.gov.br/default.jsp> >

27. Brasil. Lei nº 9.503, de 23 de setembro de 1997. Dispõe sobre o Novo Código de Trânsito Brasileiro. -em vigor em 22 de janeiro de 1998.

28. Ministério da Saúde. Saúde Brasil, 2005. Brasília: MS; 2005.

REFERÊNCIAS BIBLIOGRÁFICAS

29. Bastos YGL, Andrade SM, Cordoni JL. Acidentes de trânsito e o novo código de trânsito brasileiro em cidade da região sul do Brasil. *Inf Epidemiol SUS*. 1999;(2):37-45.
30. Liberatti CLB, Andrade SM, Soares DA. The new Brazilian traffic code and some characteristics of victims in southern Brazil. *Inj Prev*. 2001;7(3):190-3.
31. [MS] Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde (DATASUS) [homepage na Internet] Banco de dados dos Sistemas de Informação de Morbidade Hospitalar. Brasil; 2008. [acesso em 10 abr 2008]. Disponível em: <http://w3.datasus.gov.br/datasus> .
32. Gawryszewski VP, Koizumi MS, Mello Jorge MHP. As causas externas no Brasil no ano 2000: comparando a mortalidade e a morbidade. *Cad Saúde Pública*. 2004;20:109-18.
33. Deslandes SF, Silva CMFP. Análise da morbidade hospitalar por acidentes de trânsito em hospitais públicos do Rio de Janeiro, RJ, Brasil. *Rev Saúde Pública*. 2000; 34(4):367-372.
34. Marin L, Queiroz MS. A atualidade dos acidentes de trânsito na era da velocidade: uma visão geral. *Cad Saúde Pública*. 2000;16(1):7-21.
35. Souza ERS, Njaine K, Minayo MCS. Qualidade da informação sobre violência: um caminho para construção da cidadania. *Cad Prog pós-grad em Ciênc Inf*. 1996;2(1):104- 112.
36. Brasil. Ministério da saúde. Portaria nº 142 de 13 de novembro de 1997 regulamenta.O preenchimento dos campos caráter e internação, diagnóstico principal e diagnóstico secundário das autorizações de internação hospitalar – AIH a serem processadas pelo SUS e dá outras providências. *Diário Oficial da União*. 17 de nov. 1997; seção 1.
37. Carvalho DM. Grandes Sistemas Nacionais de Informação em Saúde. Revisão e discussão da situação atual. *Inf Epidemiol do SUS*. 1997;6(4):7-46.

REFERÊNCIAS BIBLIOGRÁFICAS

38. Camargo Junior KR, Coeli CM. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. *Cad Saúde Pública*. 2000;16(2):439-447.
39. Blakely T, Salmond C. Probabilistic Record Linkage and a method to calculate the positive predictive value. *International Epidemiological Association*. 2002;31:1246-1252.
40. Winkler WE. *Automatically Estimating Record Linkage False Match Rates*. Washington, DC: Statistical Research Division U.S. Census Bureau; 2007.
41. Camargo Júnior KR, Coeli CM. *Tutorial do Reclink III*. Rio de Janeiro:2007. 137p.
42. Coutinho RGM, Coeli CM, Faerstein E, Chor D. Sensibilidade do linkage probabilístico na identificação de nascimentos informados: estudo Pró-Saúde. *Rev Saúde Pública [serial on the Internet]*. 2008;42(6):1097-1100.
43. Silva CF, Leite AJM, Almeida NMGS, Gondim RC. Fatores de risco para mortalidade infantil em município do Nordeste do Brasil: linkage entre bancos de dados de nascidos vivos e óbitos infantis – 2000 a 2002. *Rev bras epidemiol*. 2006;9(1):69-80.
44. Sousa MH, Cecatti JG, Hardy EE, Serruya SJ. Morte materna declarada e o relacionamento de sistemas de informações em saúde. *Rev Saúde*. 2007; 41(2):181-189.
45. Machado CJ. A literature review of record linkage procedures focusing on infant health outcomes. *Cad Saúde Pública* 2004;20:362-71.
46. Waldvogel BC. Pesquisa pioneira recupera casos de Aids no Estado de São Paulo: integração das bases do Sinan-Aids e da Fundação Seade. *Rev bras estud popul*. 2006; 23(1):187-190.
47. Lucena FFA, Fonseca MGP, Sousa AIA, Coeli CM. O relacionamento de bancos de dados na implementação da vigilância da AIDS. *Relacionamento de dados e vigilância da AIDS. Cad saúde colet*. 2006 Abr-Jun;14(2):305-312.

REFERÊNCIAS BIBLIOGRÁFICAS

48. Teixeira CLS, Klein CH, Bloch KV et al. Reclassificação dos grupos de causas prováveis dos óbitos de causa mal definida, com base nas Autorizações de Internação Hospitalar no Sistema Único de Saúde, Estado do Rio de Janeiro, Brasil . *Cad Saúde Pública*. 2006;22(6):1315-1324.
49. Coeli CM, Veras RP, Coutinho ESF. Metodologia de captura-recaptura: uma opção para a vigilância das doenças não transmissíveis na população idosa. *Cad Saúde Pública*. 2000;16(4):1071-1082.
50. Coeli CM, Coutinho ESF, Veras RP. The challenge of applying capture-recapture methodology to diabetes mellitus surveillance in the elderly: lessons from a Brazilian experience. *Cad Saúde Pública*. 2004;20(6):1709-1720.
51. Leles AD. Técnicas de record linkage e inteligência artificial para a construção de um data warehouse aplicado à área de Saúde. *Revista Científica do IMAPES*. 2004;2(2):41-46.
52. Moraes IHS. Informação em Saúde: da prática fragmentada ao exercício da cidadania. São Paulo: Hucitec; Rio de Janeiro: Abrasco; 1994.
53. Laurenti R, Mello Jorge MHP, Lebrão ML, Gotlieb SLD. Estatísticas de Saúde. 2ª ed. São Paulo: EPU; 2005.
54. McGarry KL. Da documentação à informação: conceito em evolução. Lisboa: Editorial Presença; 1984. p. 196
55. Mota E, Carvalho DMT. Sistemas de Informação em Saúde. In: Rouquarol MZ, Almeida NF. *Epidemiologia & Saúde*. 6ª ed. Rio de Janeiro: Medsi; 2003. p. 605-628.
56. Angeloni MT. Elementos intervenientes na tomada de decisão. *Ci Inf*. 2003 Jan-Apr; 32(1):17-22.
57. Silveira DT, Marin HF. Conjunto de Dados Mínimos de Enfermagem: construindo um modelo em saúde ocupacional. *Acta paul enferm*. 2006;19(2):218-227.
58. Sousa PAF. Sistema de partilha de informação de enfermagem entre contextos de cuidados de saúde – um modelo explicativo. Porto: Editora Formação e Saúde Ltda; 2006. p. 34.

REFERÊNCIAS BIBLIOGRÁFICAS

59. Fayyad UM, Shapiro GP, Smyth P, Uthurusamy R. *Advances in Knowledge Discovery and Data Mining*. Califórnia: American Association for Artificial Intelligence; 1996. 611p.
60. Safran C, Perreault LE. Management of information in integrated delivery networks. In: Shortliffe EH, Perrault LE, Wiederhold G, Fagan LM. (eds). *Medical Informatics: computer applications in health care and biomedicine*. 2^a ed. New York: Springer-Verlag; 2001. p. 359-396.
61. Pelletier, D; Diers, D, *Developing data for practice and management : Na Australian*
62. Sabbatini RME. História da Informática em Saúde no Brasil, *Informática Médica*. 1998 Set/Out:1(5) [acesso em 2006 Maio 20]. Disponível em <http://www.informaticamedica.org.br/informaticamedica/n0105/sabbatini.htm>
63. Nakayama T. Evidence-based Healthcare and Health Informatics: Derivations and Extension of Epidemiology. *Journal of Epidemiology*. 2006;16:93-100.
64. Wainer J, Campos CJR, Ramos MP, Lopes P, Salomão P, Sigulem D. O que é pesquisa em informática em saúde? *Revista de Informática Teórica e Aplicada (RITA)*. 2005;1(1):2-14.
65. Greenes RA, Shortliffe EH. Medical informatics: an emerging academic discipline and institucional priority. *Journal of American Medical Association*. 1990;236(8):1114-1120.
66. Coiera EW. Inteligência Artificial na medicina. *Revista de informática médica*. 1998;1(4). Disponível em <http://www.informaticamedica.org.br/informaticamedica/n0104/index.html>
67. Degoulet P, Fieschi M. *Introduction to Clinical Informatics*. New York: Springer-Verlage; 1997.
68. Shortliffe EH. What is medical informatics? 2003 [acesso em 2006 Ago 26]. Disponível em: <http://smi.stanford.edu> .

REFERÊNCIAS BIBLIOGRÁFICAS

69. Massad E, Ortega NRS, Silveira PSP, Menezes RX. Métodos quantitativos em medicina. São Paulo: Manole; 2004.
70. Hannah KJ, Ball MJ, Edwards MJA. Introduction to nursing informatics. 2^a ed. New York: Springer-Verlag; 1999.
71. Eysenbach G. Recent advances: Consumer health informatics. BMJ. 2000;320:1713-1716.
72. [SBIS] Sociedade Brasileira de Informática em Saúde. [acesso 2007 Maio 19] Disponível em <http://www.sbis.org.br/indexframe.html> .
73. Sigulem D. Um Novo Paradigma de Aprendizado na Prática Médica da UNIFESP/EPM [Tese Livre-Docência]. São Paulo (SP): Universidade Federal de São Paulo – Escola Paulista de Medicina; 1997.
74. Abbott PA, Lee SM. Data mining and Knowledge Discovery. In: Saba VK, McCormick KA. Essentials of Nursing Informatics. 4^a ed. Nova Iorque: Mcgraw Hill – Medical Publishing Division; 2006.
75. Bayam E, Liebowitz J, Agresti W. Older drivers and accidents: A meta analysis and data mining application on traffic accident data. Expert Systems with Applications. 2005; 29:598-629.
76. Navega S. Princípios essenciais do Data Mining. Anais do Infoimagem, Cenadem., 2002 Nov.
77. Matos G, Chalmeta R, Coltell O. Metodología para la Extracción del Conocimiento Empresarial a partir de los Datos. Inf tecnol. 2006;17(2):81-88.
78. Febles RJP, González PA. Aplicación de la minería de datos en la bioinformática. ACIMED. 2002;10(2):69-76.
79. Chen SY, Liu X. The contribution of data mining to information science. Journal of Information Science. 2004;30(6):550-558.
80. Chen L, Lu L, Feng K, Li W, Song J, Zheng L et al. Multiple classifier integration

REFERÊNCIAS BIBLIOGRÁFICAS

for the prediction of protein structural classes. *J Comput Chem*. 2009 (Epub ahead of print).

81. Kaur H, Wasan SK. Empirical study on applications of Data Mining techniques in healthcare. *Journal of Computer Science*. 2006;2(2):194-200.

82. Weiss SM, Indurkha N. *Predictive Data Mining: A Practical Guide*. San Francisco (CA): Morgan Kaufmann Publishers Inc.; 1998.

83. Hand DJ. Data Mining: New Challenges for Statisticians. *Social Science Computer Review*. 2000;18(4):442-449.

84. Agrawal R, Srikant R, Vu Q. Mining association rules with item constraints. *Future Generations Computer System, Elsevier, Netherlands*. 1997;13(2-3):161-80.

85. Berry MIA, Linoff G. *Data mining techniques for marketing, sales, and customer support*. New York : John Wiley & Sons; 1997. 454 p

86. Brusso MJ. *Access Miner: Uma Proposta para Extração de Regras de Associação Aplicada à Mineração do Uso da Web*. [dissertação de Mestrado em Ciência da Computação]. Porto Alegre: Instituto de Informática, Universidade Federal do Rio Grande do Sul; 2000.

87. Witten IH, Frank E. *Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations*. Newark: Morgan Kaufmann Publishers; 2000.

88. Penna MLF. Rede neural artificial para detecção de sobremortalidade atribuível à cólera no Ceará. *Rev Saúde Pública*. 2004;38(3):351-357.

89. Zanetti Sidney S. et al . Estimacão da evapotranspiracão de referênciac no estado do Rio de Janeiro usando redes neurais artificiais. *Rev bras eng agríc ambient*. 2008;12(2).

90. Kovács ZL. *Redes neurais artificiais: fundamentos e aplicacões*. 3ª ed. São Paulo: Livraria da Física; 2002.

REFERÊNCIAS BIBLIOGRÁFICAS

91. Sudheer KP, Gosain AK, Ramasastry KS. Estimating actual evapotranspiration from limited climatic data using neural computing technique. *Journal of Irrigation and Drainage Engineering*. 2003;129(3):214-218.
92. Baesens B, Setiono R, Mues C, Vanthienen J. Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation. *Management Science*, Evanston, IL, INFORMS. 2003; 49(3):312-329.
93. Lin A, Hoffman D, Gaasterland DE, Caprioli J. Neural networks to identify glaucomatous visual field progression. *Am J Ophthalmol*. 2003;135(1):49-54.
94. Goulart DA, Tacla MA, Marback PMF, Solé D, Paranhos Junior A, Perez HB et al. . Redes neurais artificiais aplicadas no estudo de questionário de varredura para conjuntivite alérgica em escolares. *Arq Bras Oftalmol*. 2006;69(5):707-713.
95. Han J, Kamber M. *Data mining: Concepts and Techniques*. 2^a ed. San Francisco (CA): Morgan Kaufmann; 2006.
96. Meira CAA, Rodrigues LHA, Moraes SA. Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. *Trop plant pathol*. 2008;33(2):114-124.
97. Apte C, Weiss S. Data mining with decision trees and decision rules. *Future Generation Computer Systems*. 1997;13:197-210.
98. Faria MHG, Patrocínio RMS, Vidal E, Barem SH. Estratégias auxiliares para graduação dos tumores astrocíticos segundo os critérios histopatológicos estabelecidos pela OMS. *J Bras Patol Med Lab*. 2006;42(5):401-410.
99. Cardoso CS, Caiaffa WT, Bandeira M et al. Depressão na esquizofrenia: prevalência e relação com a qualidade de vida. *Cad Saúde Pública*. 2007;23(9):2035-2048.
100. Leite PT; Carneiro AAFM, Carvalho ACPLF. Aplicação de algoritmos genéticos na determinação da operação ótima de sistemas hidrotérmicos de potência. *Sba Controle & Automação*. 2006;17(1):81-88.
101. Liannoni AP, Morabito R. Modelo hipercubo integrado a um algoritmo genético para análise de sistemas médicos emergenciais em rodovias. *Gestão & Produção*.

REFERÊNCIAS BIBLIOGRÁFICAS

2006a; 13(1):93-104.

102. Liannoni AP, Morabito R. A discrete simulation analysis of a logistics supply system. *Transportation Research*. 2006b;42(3):191-210.

103. Liannoni AP, Morabito R. Modelo de fila hipercubo com múltiplo despacho e backup parcial para análise de sistemas de atendimento médico emergenciais em rodovias. *Pesquisa Operacional*. 2006c;26(3):493-519.

104. Liannoni AP, Morabito R. Otimização da localização das bases de ambulâncias e do dimensionamento das suas regiões de cobertura em rodovias. *Prod*. 2008;18(1):47-63.

105. Sousa CA, Duarte PS, Pereira JCR. Lógica fuzzy e regressão logística na decisão para prática de cintilografia das paratiróides. *Rev Saúde Pública*. 2006;40(5):898-906.

106. Pereira JCR, Tonelli PA, Barros LC et al. Clinical signs of pneumonia in children: association with and prediction of diagnosis by fuzzy sets theory. *Braz J Med Biol Res*. 2004;37(5):701-709.

107. Nascimento LFC, Ortega NRS. Fuzzy linguistic model for evaluating the risk of neonatal death. *Rev Saúde Pública*. 2002;36(6):686-692.

108. Reis MAM, Ortega NRS, Silveira PSP. Fuzzy expert system in the prediction of neonatal resuscitation. *Braz J Med Biol Res*. 2004;37(5):755-764.

109. Marques IR; Barbosa SF; Basile ALO, Marin HF. Guia de apoio à decisão em enfermagem obstétrica: aplicação da técnica da lógica Fuzzy. *Rev bras enferm*. 2005; 58(3):349-354.

110. Duarte PS, Mastrocolla LE, Farsky PS, Sampaio CREPS, Tonelli PA, Barros LC et al. Selection of patients for myocardial perfusion scintigraphy based on fuzzy sets theory applied to clinical-epidemiological data and treadmill test results. *Braz J Med Biol Res*. 2006;39(1):9-18.

111. Carvalho LMF, Nassar SM, Azevedo FM, Carvalho HJT, Monteiro LL, Rech CMZ. A neuro-fuzzy system to support in the diagnostic of epileptic events and non-

REFERÊNCIAS BIBLIOGRÁFICAS

epileptic events using different fuzzy arithmetical operations. *Arq Neuro-Psiquiatr.* 2008; 66(2a):179-183.

112. Roazzi A, O'brien DP, Dias MGBB. Sobre o debate frequentista versus probabilista: "sorte de tolo" torna-se uma explicação plausível. *Psicol Reflex Crit.* 2003;16(2):389-402.

113. Assunção RM, Barreto SM, Guerra HL, Sakurai E. Mapas de taxas epidemiológicas: uma abordagem Bayesiana. *Cad Saúde Pública.* 1998;14(4):713-723.

114. Souza WV, Barcellos CC, Brito AM, Carvalho MS, Cruz OG, Albuquerque MFM et al. Aplicação de modelo bayesiano empírico na análise espacial da ocorrência de hanseníase. *Rev Saúde Pública.* 2001;35(5):474-480.

115. Cassou MF, Gonçalves PCZ, Engelhorn CA. Probabilidade de refluxo nas veias safenas de mulheres com diferentes graus de insuficiência venosa crônica. *J vasc Bras.* 2007;6(3):238-245.

116. Morisson P, Neves DD. Avaliação da adenosina desaminase no diagnóstico da tuberculose pleural: uma metanálise brasileira. *J bras pneumol.* 2008;34(4):217-224.

117. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. *Bioinformatics.* 2004;20(15):2479-2481.

118. Rakotomalala R. "TANAGRA: un logiciel gratuit pour l'enseignement et la recherche", in *Actes de EGC'2005. RNTI-E-3.* 2005;2:697-702.

119. Reis VC, Torres FA, Poças-Fonseca MJ, De-Souza MT, Souza DP et al. Cell cycle, DNA replication, repair, and recombination in the dimorphic human pathogenic fungus *Paracoccidioides brasiliensis*. *Genet mol res.* 2005;4(2):232-250.

120. Jupiter DC, VanBuren V. A visual data mining tool that facilitates reconstruction of transcription regulatory networks. *PLoS ONE.* 2008;(3):e1717.

REFERÊNCIAS BIBLIOGRÁFICAS

121. Madigan EA, Curet OL, Zrinyi M. Workforce analysis using data mining and linear regression to understand HIV/AIDS prevalence patterns. *Human Resources for Health*. 2008;6:2.
122. Pfaff M. et al . Prediction of Cardiovascular Risk in Hemodialysis Patients by Data Mining. *Methods of Information in Medicine*. 2004;43(1):106-113.
123. Mattos MC, Pelegrin DC, Casagrande DP, Simões PWTA, Charnovscki R, Bettiol J. Análise de Alergia e Rinite em Escolares por meio da Shell Orion Data Mining Engine. *Anais do Congresso da SBIS; 2006; Florianópolis / SC*. Disponível em <http://www.sbis.org.br/cbis/arquivos/858.pdf>
124. Pelegrin DC, Casagrande DP, Mattos MC, Simões PWTA, Charnovscki R, Bettiol J. As Tarefas de Associação e de Classificação na Shell de Data Mining Orion II. *Anais do II Congresso Sul-Catarinense de Computação; 2006; Criciúma / SC*. Disponível em <http://www.dcc.unesc.net/sulcomp/06/artigos.htm>
125. IBGE – Instituto Brasileiro de Geografia e Estatística. Censo demográfico: [homepage na internet]. Acesso em 25 de maio de 2007. Disponível em <http://www.ibge.gov.br/home/estatistica/populacao/censo2001>.
126. IBGE – Instituto Brasileiro de Geografia e Estatística; 2007 [acesso em 14 de maio de 2009]. Disponível em <http://www.sidra.ibge.gov.br/bda/tabela/protabl.asp?c=793&z=cd&o=17&i=P>
127. Anuario Estatístico de Mato Grosso. Mato Grosso: Secretaria de Estado de Planejamento. 2007: 29:540-545. [acesso em 10 jun. 2009]. Disponível em: http://www.anuario.seplan.mt.gov.br/2007/pdf/cap._15.pdf.
128. Gewehr JE, Szugat M, Zimmer R. BioWeka – extending the Weka framework for bioinformatics. *Bioinformatics*. 2007;1;23(5):651-3.
129. Torkamani A, Schork NJ. Predicting functional regulatory polymorphisms. *Bioinformatics*. 2008;15;24(16):1787-92.

REFERÊNCIAS BIBLIOGRÁFICAS

130. Base cartográfica de Cuiabá. Cuiabá: Instituto de Pesquisa e Desenvolvimento Urbano (IPDU). 2009. [acesso em 10 jun. 2009]. Disponível em: <http://www.cuiaba.mt.gov.br/orgaos/ipdu/mapas.jsp>
131. Marchese VS, Scatena JHG, Ignoti E. Caracterização das vítimas de acidentes e violências atendidas em serviço de emergência: município de Alta Floresta, MT (Brasil). *Revista Brasileira de Epidemiologia* . 2008;11(4):648-659.
132. Soares BAC, Scatena JHG, Galvão ND. Evolução e Características da Morbidade por Acidentes e Violências na Grande Cuiabá – Mato Grosso. *Revista Espaço para a Saúde*. 2008;9(2):26-38.
133. Firouzi F, Rashidi M, Hashemi S, Kangavari M, Bahari A, Daryani NE et al. A decision tree-based approach for determining low bone mineral density in inflammatory bowel disease using WEKA software. *Eur J Gastroenterol Hepatol*. 2007;19(12):1075-81.
134. Chaichian S, Shoaee S, Saremi A, Pedar S, Firouzi F. Factors influencing success rate of leukocyte immunization and anti-paternal antibodies in spontaneous recurrent miscarriage. *Am J Reprod Immunol*; 2007;57(3):169-76.
135. Steiner MTA, Nievola JC, Soma NY, Shimizu T, Steiner Neto PJ. Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. *Pesqui Oper*. 2007;27(3):407-426.
136. Sigurdardottir AK, Jonsdottir H, Benediktsson R. Outcomes of educational interventions in type 2 diabetes: WEKA data-mining analysis. *Patient Educ Couns*. 2007 Jul;67(1-2):21-31.
137. Batinic D, Milosevic D, Konjevoda P, Nizic L, Vrljicak K, Matkovic M, et al. The value of urine citrate/calcium ratio in the estimation of risk of urolithiasis. *Clin Nephrol*. 2004;61(6):387-91.
138. Holmes G, Hall MA. A development environment for predictive modelling in foods. *Int J Food Microbiol*. 2002;73(2-3):351-62.

REFERÊNCIAS BIBLIOGRÁFICAS

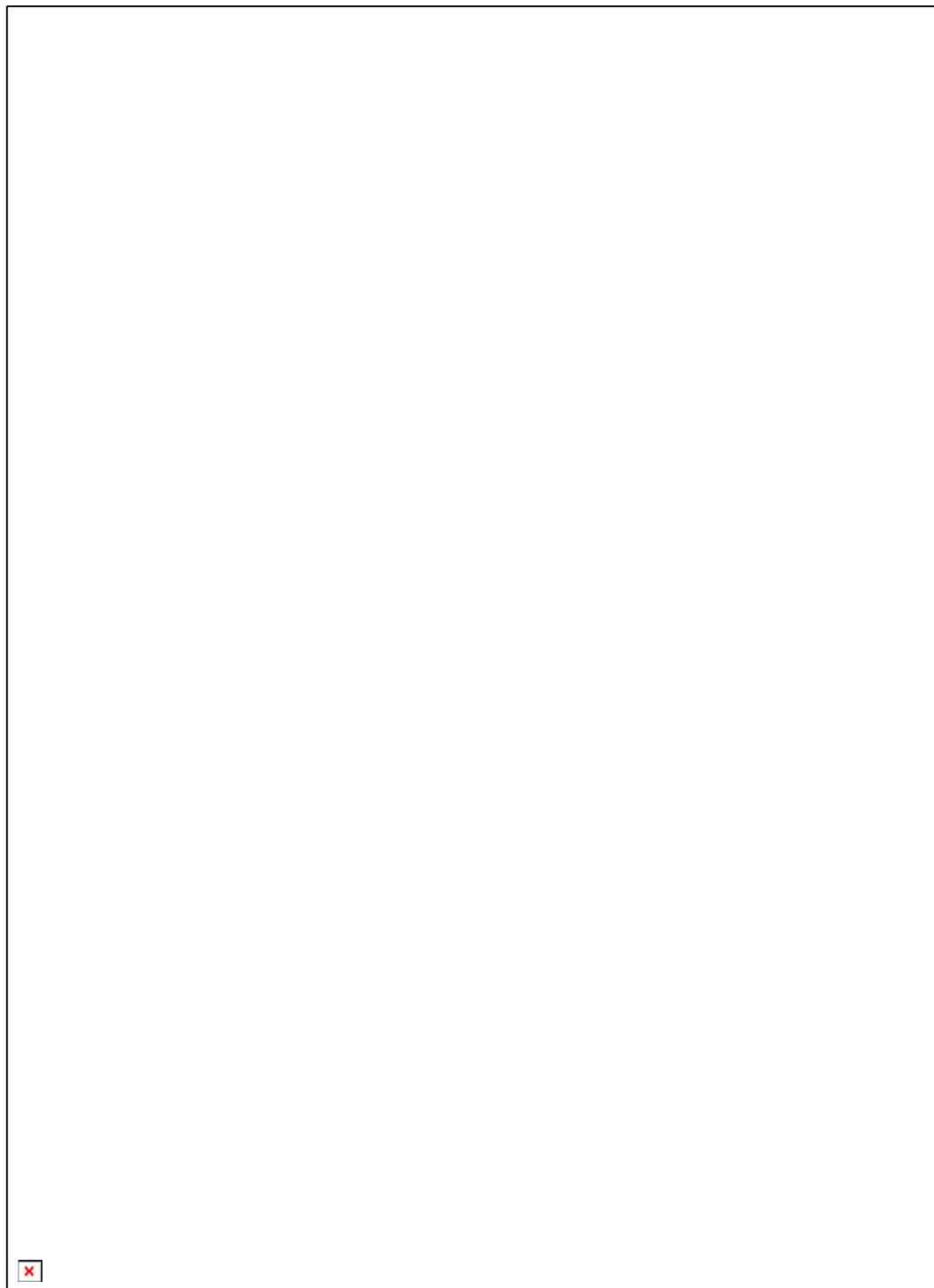
139. Soares BAC, Scatena JHG, Galvão ND. Acidentes e Violência na Grande Cuiabá: o que retrata a demanda dos serviços de emergência. *Epidemiol Serv Saúde*. 2009; 18(3):265-276.
140. Andrade SM, Mello Jorge MHP. Acidentes de transporte terrestre em município da Região Sul do Brasil. *Revista de Saúde Pública* 2001;35:318-320.
141. Mato Grosso. Anuario Estatístico de Mato Grosso, 2000. Secretaria de Estado de Planejamento. 2000. Disponível em: <
<http://www.anuario.seplan.mt.gov.br/2000/>> Acesso em: 10 jun. 2009
142. Andrade SM, Mello Jorge MHP. Características das vítimas por acidentes de transporte terrestre em município da Região Sul do Brasil. *Rev Saúde Pública* 2000;34:149-56.
143. Modelli MES, Pratesi R, Tauil PL. Alcoolemia em vítimas fatais de acidentes de trânsito no Distrito Federal, Brasil. *Rev Saúde Pública*. 2008;42(2):350-352.
144. Souza MFM, Malta DC, Conceicao GMS, Silva MMA, Gazal-Carvalho C, Moraes Neto OL. Análise descritiva e de tendência de acidentes de transporte terrestre para políticas sociais no Brasil. *Epidemiol Serv Saúde*. 2007;16(1):33-44.
145. Malvestio MAA, Sousa RMC. Suporte avançado à vida: atendimento a vítimas de acidentes de trânsito. *Rev Saúde Pública*. 2002;36(5):584-9.
146. Bahten LC, Alcantara EM, Pimenta APP, Dallagnol JC, Yoshizumi KO, Dresch MF. O impacto econômico do trauma em um hospital universitário. *Rev Col Bras Cir*. 2003; 30(3):224-229.
147. Souza ER, Minayo MCS, Njaine K, Cruz AJM, Silva HB, Santos FCL, Moreira CO. Tendências da produção científica brasileira sobre violência e acidentes na década de 90. In: Minayo MCS, Souza ER, organizadores. *Violência sob o olhar da saúde: a infrapolítica da contemporaneidade brasileira*. Rio de Janeiro: Editora FIOCRUZ; 2003a. p. 49-81.

REFERÊNCIAS BIBLIOGRÁFICAS

148. Souza ER, Ximenes LF, Alves F, Magalhães C, Bilate D, Szuchmacher AM, Malaquias J. Avanços do conhecimento sobre as causas externas no Brasil e no mundo: enfoque quantitativo e qualitativo. In: Minayo MCS, Souza ER, organizadores. Violência sob o olhar da saúde: A infrapolítica da contemporaneidade brasileira. Rio de Janeiro: Editora FIOCRUZ; 2003b. p. 131-60.
149. Dunham MH. Data Mining Introductory and Advanced Topics. USA: Prentice Hall; 2002.
150. IPEA (Instituto de Pesquisa Econômica Aplicada). Impactos sociais e econômicos dos acidentes de trânsito nas rodovias brasileiras. Relatório Executivo: IPEA/DENATRAN/ANTP; 2006 [80 p.]. [acesso em 2009 Jun 15] Disponível em <http://www.ipea.gov.br/default.jsp>.

Anexo 1

Mapa em que divide a cidade de Cuiabá-MT em zonas e bairros



Anexo 2
Autorizações dos Gestores de SMS de Cuiabá e da SEJUSP



AUTORIZAÇÃO

Eu, **Guilherme Antonio Maluf** na qualidade de Secretário Municipal de Saúde de Cuiabá - Mato Grosso, portador (a) do CPF nº. 314450471-87 do RG nº. 008054SSP/MT, autorizo a utilização dos dados secundários do Sistema de Informações sobre Mortalidade (SIM), do Sistema de Informações Hospitalares do SUS (SIH-SUS), para a Tese de Doutorado de **Noemi Dreyer Galvão**, que possui objetivo de "Analisar um conjunto de dados de acidentes a partir dos bancos Sistemas de Informação de Saúde Pública de Cuiabá - Mato Grosso, utilizando a técnica de data mining".

Meu parecer é favorável que os resultados da pesquisa sejam analisados, divulgados e publicados em periódicos para conhecimento e também para elaboração de políticas públicas.

Guilherme Antonio Maluf
Secretário Municipal de Saúde de Cuiabá.

Guilherme Maluf
Secretário Municipal de Saúde

Cuiabá, 25 de setembro de 2007.

Continuação do Anexo 2



ESTADO DE MATO GROSSO
CORPO DE BOMBEIROS MILITAR
COMANDO DE BOMBEIROS MILITAR METROPOLITANO
COMPANHIA DE URGÊNCIA E RESGATE DE ACIDENTADOS
C.U.R.A.



AUTORIZAÇÃO

Eu, Tarciana Nogueira Ramos, na qualidade de Comandante da Companhia de Urgência e Resgate de Acidentados, portadora da matrícula 946270015 e do RG nº 000.883 CBM-MT, autorizo a utilização do banco de dados e/ou as fichas de atendimento de ocorrência de vítimas pelo Serviço de Atendimento de Urgência (SAMU) do município de Cuiabá, do ano de 2006, para a tese de Doutorado da Servidora *Noemi Dreyer Galvão*, que possui objetivo de Analisar um conjunto de dados de acidentes a partir dos bancos sistemas de informação de saúde pública de Cuiabá - Mato Grosso, utilizando a técnica de data mining.

Meu parecer é favorável que os resultados da pesquisa sejam analisados, divulgados e publicados em periódicos para conhecimento e também para elaboração de políticas públicas.

TARCIANA NOGUEIRA RAMOS - 1º TEN BM

Tarciana Nogueira Ramos - Ten BM
RG BM 000883

Cuiabá, 24 de agosto de 2007.

Anexo 3

Protocolo e parecer do Comitê de Ética



Universidade Federal de São Paulo
Escola Paulista de Medicina

Comitê de Ética em Pesquisa
Hospital São Paulo

São Paulo, 30 de novembro de 2007.
CEP 1595/07

Ilmo(a). Sr(a).
Pesquisador(a) NOEMI DREYER GALVÃO
Co-Investigadores: Heimar de Fátima Marin (orientadora)
Disciplina/Departamento: Disc. De Enfermagem Obstétrica do Departamento de Enfermagem da Universidade Federal de São Paulo/Hospital São Paulo
Patrocinador: Recursos Próprios.

PARECER DO COMITÊ DE ÉTICA INSTITUCIONAL

Ref: Projeto de pesquisa intitulado: “**Mineração de dados nos bancos dos sistemas de informação da saúde pública para acidentes**”.

CARACTERÍSTICA PRINCIPAL DO ESTUDO: Pesquisa exploratória e retrospectiva de dados secundários.

RISCOS ADICIONAIS PARA O PACIENTE: Sem risco, sem contato com paciente.

OBJETIVOS: Analisar um conjunto de dados de acidentes a partir dos bancos sistemas de informação de saúde pública de Cuiabá, Mato Grosso, utilizando a técnica de data mining..

RESUMO: Estudo de dados secundários de acidentes, de morbidade e mortalidade, de análise exploratória, retrospectiva através do data minig-mineração de dados dos bancos dos sistemas de informação da saúde pública, no município de Cuiabá-MT. A população do estudo compreende as pessoas que residiam no município de Cuiabá durante o ano de 2006 que estavam expostas aos acidentes. A amostra será as vítimas de acidentes durante o período de 01 de janeiro a 31 de dezembro de 2006. Os dados serão coletados dos sistemas de informação que contém registro de acidentes de qualquer faixa etária e sexo. O setor saúde possui sistemas que registram as internações no SUS- SIH/SUS, os óbitos (SIM) e o relatório de ocorrência -APH dos atendimentos realizados pelo Serviço de Atendimento Móvel de Urgência (SAMU) e Corpo de Bombeiros. A mineração de dados será desenvolvida em várias fases: identificação das variáveis nos diversos bancos de dados selecionados, a limpeza dos bancos de dados; reformatação dos dados para o relacionamento bancos de dados; aplicação de algoritmos através de softwares; mineração dos dados automatizados; formação de um conjunto de dados essenciais, a partir da mineração; apresentação do conjunto de dados minerados..

FUNDAMENTOS E RACIONAL: Este estudo pretende responder as seguintes questões: qual é o conjunto de dados essenciais para acidentes de transporte? Quais são os fatores que determinam a morbidade e mortalidade por acidentes de transporte em Cuiabá-MT no ano de 2006.

MATERIAL E MÉTODO: Descreve os procedimentos, e apresenta cartas de concordância dos locais envolvidos com a pesquisa.

TCLE: Não se aplica.

ANEXOS

Continuação do Anexo 3



Universidade Federal de São Paulo
Escola Paulista de Medicina

Comitê de Ética em Pesquisa
Hospital São Paulo

DETALHAMENTO FINANCEIRO: Sem financiamento externo - R\$ 630,00.

CRONOGRAMA: 24 meses.

OBJETIVO ACADÊMICO: Doutorado.

ENTREGA DE RELATÓRIOS PARCIAIS AO CEP PREVISTOS PARA: 29/11/2008 e 29/11/2009.

O Comitê de Ética em Pesquisa da Universidade Federal de São Paulo/Hospital São Paulo **ANALISOU** e **APROVOU** o projeto de pesquisa referenciado.

1. Comunicar toda e qualquer alteração do projeto e termo de consentimento livre e esclarecido. Nestas circunstâncias a inclusão de pacientes deve ser temporariamente interrompida até a resposta do Comitê, após análise das mudanças propostas.
2. Comunicar imediatamente ao Comitê qualquer evento adverso ocorrido durante o desenvolvimento do estudo.
3. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos para possível auditoria dos órgãos competentes.

Atenciosamente,

Prof. Dr. José Osmar Medina Pestana
Coordenador do Comitê de Ética em Pesquisa da
Universidade Federal de São Paulo/ Hospital São Paulo

CEP 1595/07

Anexo 4

Relação dos bairros segundo as zonas das regiões do município de Cuiabá-MT

• **Região Norte:** Jardim Florianópolis, Paiaguas, Jardim Vitória, Novo Paraíso, Centro Político Administrativo, Morada do Ouro, Morada da Serra, Três Barras, Nova Conquista, 1º de Março, Jardim Umuarama, Três Barras.

• **Região Oeste:** Parque Residencia, Nações Indígenas, Jardim Ubirajara, Ribeirão do Lipa, Novo Colorado, Despraiado, Jardim Santa Marta, Alvorada, Jardim Mariana, Ribeirão da Ponte, Santa Rosa, Duque de Caxias, Quilombo, Araés, Alvorada, Jardim Cuiabá, Popular, Centro Norte, Centro Sul, Barra do Pari, Jardim Santa Isabel, Cidade Verde, Novo Terceiro, Cidade Alta, Porto, Goiabeiras, Coophamil.

• **Região Leste:** Jardim Aclimação, Terra Nova, Bela Vista, São Roque, Novo Mato Grosso, Novo Horizonte, Residencial São Carlos, Planalto, Sol Nascente, Residencial Itamarati, Jardim Eldorado, Carumbé, Dom Bosco, Canjica, Campo Verde, Bosque da Saúde, Pedregal, Jardim Itália, Morada dos Nobres, Lixeira, Baú, Areão, Jardim Leblon, Jardim das Américas, UFMT, Santa Cruz, Recanto dos Pássaros, Jardim Imperial, Bandeirantes, Poção, Jardim das Américas, Pico do Amor, Dom Aquino, Jardim Tropical, Campo Velho, Jardim Petrópolis, Boa Esperança, Jardim Universitário, Terceiro, Jardim Europa, Grande Terceiro, Jardim Califórnia, Jardim Shangri-lá, Praeiro, Bela Marina, Cachoeira das Garças.

• **Região Sul:** Coxipó, Jordão, Parque Ohara, Vista Alegre, Jardim Gramado, CoopHEMA, São Gonçalo Beira Rio, Nossa Senhora Aparecida, Jardim Comodoro, São José, Altos do Coxipó, São Francisco, Jardim dos Ipês, Tijucal, Lagoa Azul, Jardim Passaredo, Jardim Fortaleza, São João Del Rei, Santa Laura, São Sebastião, Pascoal Ramos, Zona de Expansão, Urbana do Manduri, Nova Esperança, Pedra 90, Osmar Cabral, Parque Georgea, COHAB São Gonçalo, Jardim Presidente, Jardim Industriário II, Jardim Mossoró, Residencial Coxipó, Parque Cuiabá, Parque Itália, Parque Nova Esperança (III Etapa), Distrito Industrial.

APÊNDICE 1

Regras geradas pelo algoritmo APriori

=== Run information ===

Scheme: weka.associations.Apriori -N 1000 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
 Relation: ACIDENTECUIABA-weka.filters.unsupervised.attribute.Remove-R1,3,13
 Instances: 139
 Attributes: 12
 SEXO
 FX_ET
 MES_OC
 ZONA_RES
 ZONA_OC
 TP_ACID
 TP_VIT
 TRANSP_VIT
 ASSIST_MED
 LC_ASSIST
 EVOL
 CAT_DIAS_INTER

=== Associator model (full training set) ===

Primeiro Modelo

Apriori

=====

Minimum support: 0.1 (14 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 43

Size of set of large itemsets L(2): 162

Size of set of large itemsets L(3): 208

Size of set of large itemsets L(4): 107

Size of set of large itemsets L(5): 21

Best rules found:

1. EVOL=ALTA 110 ==> ASSIST_MED=SIM 110 conf:(1)
2. SEXO=M EVOL=ALTA 89 ==> ASSIST_MED=SIM 89 conf:(1)
3. TP_ACID=COLISÃO EVOL=ALTA 57 ==> ASSIST_MED=SIM 57 conf:(1)
4. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM 51 conf:(1)
5. SEXO=M TP_ACID=COLISÃO EVOL=ALTA 50 ==> ASSIST_MED=SIM 50 conf:(1)

APÊNDICE

6. FX_ET=3 EVOL=ALTA 47 ==> ASSIST_MED=SIM 47 conf:(1)
7. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM 45 conf:(1)
8. LC_ASSIST=HPSMCBA EVOL=ALTA 44 ==> ASSIST_MED=SIM 44 conf:(1)
9. EVOL=ALTA CAT_DIAS_INTER=1 44 ==> ASSIST_MED=SIM 44 conf:(1)
10. SEXO=M LC_ASSIST=HPSMCBA 41 ==> ASSIST_MED=SIM 41 conf:(1)
11. SEXO=M FX_ET=3 EVOL=ALTA 41 ==> ASSIST_MED=SIM 41 conf:(1)
12. TRANSP_VIT=MOTOCICLETA EVOL=ALTA 40 ==> ASSIST_MED=SIM 40 conf:(1)
13. CAT_DIAS_INTER=2 39 ==> ASSIST_MED=SIM 39 conf:(1)
14. EVOL=ALTA CAT_DIAS_INTER=2 38 ==> ASSIST_MED=SIM 38 conf:(1)
15. TP_VIT=CONDUTOR EVOL=ALTA 37 ==> ASSIST_MED=SIM 37 conf:(1)
16. TP_VIT=IGNORADO EVOL=ALTA 35 ==> ASSIST_MED=SIM 35 conf:(1)
17. SEXO=M TRANSP_VIT=MOTOCICLETA EVOL=ALTA 35 ==> ASSIST_MED=SIM 35 conf:(1)
18. SEXO=M CAT_DIAS_INTER=1 34 ==> ASSIST_MED=SIM 34 conf:(1)
19. SEXO=M CAT_DIAS_INTER=2 34 ==> ASSIST_MED=SIM 34 conf:(1)
20. SEXO=M LC_ASSIST=HPSMCBA EVOL=ALTA 34 ==> ASSIST_MED=SIM 34 conf:(1)
21. SEXO=M EVOL=ALTA CAT_DIAS_INTER=1 33 ==> ASSIST_MED=SIM 33 conf:(1)
22. SEXO=M EVOL=ALTA CAT_DIAS_INTER=2 33 ==> ASSIST_MED=SIM 33 conf:(1)
23. SEXO=M TP_VIT=CONDUTOR EVOL=ALTA 31 ==> ASSIST_MED=SIM 31 conf:(1)
24. CAT_DIAS_INTER=3 29 ==> ASSIST_MED=SIM 29 conf:(1)
25. TRANSP_VIT=VEICULO EVOL=ALTA 29 ==> ASSIST_MED=SIM 29 conf:(1)
26. ZONA_OC=OESTE EVOL=ALTA 28 ==> ASSIST_MED=SIM 28 conf:(1)
27. EVOL=ALTA CAT_DIAS_INTER=3 28 ==> ASSIST_MED=SIM 28 conf:(1)
28. SEXO=M TP_VIT=IGNORADO EVOL=ALTA 28 ==> ASSIST_MED=SIM 28 conf:(1)
29. ZONA_OC=SUL EVOL=ALTA 27 ==> ASSIST_MED=SIM 27 conf:(1)
30. ZONA_OC=SINF EVOL=ALTA 27 ==> ASSIST_MED=SIM 27 conf:(1)
31. CAT_DIAS_INTER=0 26 ==> EVOL=OBITO 26 conf:(1)
32. TP_ACID=SINF EVOL=ALTA 26 ==> ASSIST_MED=SIM 26 conf:(1)
33. TP_VIT=PASSAGEIRO EVOL=ALTA 26 ==> ASSIST_MED=SIM 26 conf:(1)
34. FX_ET=3 TP_ACID=COLISÃO EVOL=ALTA 26 ==> ASSIST_MED=SIM 26 conf:(1)
35. ZONA_RES=SINF EVOL=ALTA 25 ==> ASSIST_MED=SIM 25 conf:(1)
36. ZONA_RES=SINF ASSIST_MED=SIM 25 ==> EVOL=ALTA 25 conf:(1)
37. SEXO=M TP_VIT=PASSAGEIRO EVOL=ALTA 25 ==> ASSIST_MED=SIM 25 conf:(1)
38. SEXO=M TRANSP_VIT=VEICULO EVOL=ALTA 25 ==> ASSIST_MED=SIM 25 conf:(1)
39. TRANSP_VIT=SINF 24 ==> TP_VIT=IGNORADO 24 conf:(1)
40. LC_ASSIST=HMBOMJESUS 24 ==> ASSIST_MED=SIM 24 conf:(1)
41. LC_ASSIST=HMBOMJESUS 24 ==> EVOL=ALTA 24 conf:(1)
42. SEXO=M CAT_DIAS_INTER=3 24 ==> ASSIST_MED=SIM 24 conf:(1)
43. TP_ACID=COLISÃO CAT_DIAS_INTER=2 24 ==> ASSIST_MED=SIM 24 conf:(1)
44. LC_ASSIST=HMBOMJESUS EVOL=ALTA 24 ==> ASSIST_MED=SIM 24 conf:(1)
45. ASSIST_MED=SIM LC_ASSIST=HMBOMJESUS 24 ==> EVOL=ALTA 24 conf:(1)
46. LC_ASSIST=HMBOMJESUS 24 ==> ASSIST_MED=SIM EVOL=ALTA 24 conf:(1)
47. SEXO=M ZONA_OC=SUL EVOL=ALTA 24 ==> ASSIST_MED=SIM 24 conf:(1)
48. FX_ET=3 LC_ASSIST=HPSMCBA 23 ==> ASSIST_MED=SIM 23 conf:(1)
49. ZONA_RES=SUL EVOL=ALTA 23 ==> ASSIST_MED=SIM 23 conf:(1)
50. ZONA_RES=OESTE EVOL=ALTA 23 ==> ASSIST_MED=SIM 23 conf:(1)
51. TP_ACID=COLISÃO CAT_DIAS_INTER=1 23 ==> ASSIST_MED=SIM 23 conf:(1)
52. TP_ACID=COLISÃO CAT_DIAS_INTER=1 23 ==> EVOL=ALTA 23 conf:(1)
53. TRANSP_VIT=MOTOCICLETA LC_ASSIST=HPSMCBA 23 ==> ASSIST_MED=SIM 23 conf:(1)
54. SEXO=M EVOL=ALTA CAT_DIAS_INTER=3 23 ==> ASSIST_MED=SIM 23 conf:(1)
55. TP_ACID=COLISÃO TRANSP_VIT=MOTOCICLETA EVOL=ALTA 23 ==> ASSIST_MED=SIM 23 conf:(1)
56. TP_ACID=COLISÃO EVOL=ALTA CAT_DIAS_INTER=1 23 ==> ASSIST_MED=SIM 23 conf:(1)
57. TP_ACID=COLISÃO ASSIST_MED=SIM CAT_DIAS_INTER=1 23 ==> EVOL=ALTA 23 conf:(1)

APÊNDICE

58. TP_ACID=COLISÃO CAT_DIAS_INTER=1 23 ==> ASSIST_MED=SIM EVOL=ALTA 23
conf:(1)
59. TP_ACID=COLISÃO EVOL=ALTA CAT_DIAS_INTER=2 23 ==> ASSIST_MED=SIM 23
conf:(1)
60. SEXO=M LC_ASSIST=HMBOMJESUS 22 ==> ASSIST_MED=SIM 22 conf:(1)
61. SEXO=M LC_ASSIST=HMBOMJESUS 22 ==> EVOL=ALTA 22 conf:(1)
62. ZONA_RES=LESTE EVOL=ALTA 22 ==> ASSIST_MED=SIM 22 conf:(1)
63. TP_ACID=COLISÃO LC_ASSIST=HPSMCBA 22 ==> ASSIST_MED=SIM 22 conf:(1)
64. SEXO=M FX_ET=3 LC_ASSIST=HPSMCBA 22 ==> ASSIST_MED=SIM 22 conf:(1)
65. SEXO=M ZONA_OC=OESTE EVOL=ALTA 22 ==> ASSIST_MED=SIM 22 conf:(1)
66. SEXO=M TP_ACID=COLISÃO CAT_DIAS_INTER=2 22 ==> ASSIST_MED=SIM 22 conf:(1)
67. SEXO=M LC_ASSIST=HMBOMJESUS EVOL=ALTA 22 ==> ASSIST_MED=SIM 22 conf:(1)
68. SEXO=M ASSIST_MED=SIM LC_ASSIST=HMBOMJESUS 22 ==> EVOL=ALTA 22 conf:(1)
69. SEXO=M LC_ASSIST=HMBOMJESUS 22 ==> ASSIST_MED=SIM EVOL=ALTA 22 conf:(1)
70. SEXO=F EVOL=ALTA 21 ==> ASSIST_MED=SIM 21 conf:(1)
71. FX_ET=3 CAT_DIAS_INTER=1 21 ==> ASSIST_MED=SIM 21 conf:(1)
72. FX_ET=4 EVOL=ALTA 21 ==> ASSIST_MED=SIM 21 conf:(1)
73. TRANSP_VIT=SINF ASSIST_MED=SIM 21 ==> TP_VIT=IGNORADO 21 conf:(1)
74. SEXO=M ZONA_RES=SUL EVOL=ALTA 21 ==> ASSIST_MED=SIM 21 conf:(1)
75. SEXO=M TP_ACID=SINF EVOL=ALTA 21 ==> ASSIST_MED=SIM 21 conf:(1)
76. TP_ACID=COLISÃO TP_VIT=IGNORADO EVOL=ALTA 21 ==> ASSIST_MED=SIM 21 conf:(1)
77. SEXO=M FX_ET=3 TP_ACID=COLISÃO EVOL=ALTA 21 ==> ASSIST_MED=SIM 21 conf:(1)
78. SEXO=M TP_ACID=COLISÃO EVOL=ALTA CAT_DIAS_INTER=2 21 ==> ASSIST_MED=SIM 21
conf:(1)
79. SEXO=M CAT_DIAS_INTER=0 20 ==> EVOL=OBITO 20 conf:(1)
80. TRANSP_VIT=SINF EVOL=ALTA 20 ==> TP_VIT=IGNORADO 20 conf:(1)
81. TRANSP_VIT=SINF EVOL=ALTA 20 ==> ASSIST_MED=SIM 20 conf:(1)
82. LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 20 ==> ASSIST_MED=SIM 20 conf:(1)
83. LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 20 ==> EVOL=ALTA 20 conf:(1)
84. SEXO=M ZONA_RES=OESTE EVOL=ALTA 20 ==> ASSIST_MED=SIM 20 conf:(1)
85. SEXO=M TP_ACID=COLISÃO LC_ASSIST=HPSMCBA 20 ==> ASSIST_MED=SIM 20 conf:(1)
86. SEXO=M TRANSP_VIT=MOTOCICLETA LC_ASSIST=HPSMCBA 20 ==> ASSIST_MED=SIM
20 conf:(1)
87. FX_ET=3 TRANSP_VIT=MOTOCICLETA EVOL=ALTA 20 ==> ASSIST_MED=SIM 20 conf:(1)
88. FX_ET=3 EVOL=ALTA CAT_DIAS_INTER=1 20 ==> ASSIST_MED=SIM 20 conf:(1)
89. TP_ACID=COLISÃO TP_VIT=CONDUTOR EVOL=ALTA 20 ==> ASSIST_MED=SIM 20
conf:(1)
90. TRANSP_VIT=SINF ASSIST_MED=SIM EVOL=ALTA 20 ==> TP_VIT=IGNORADO 20 conf:(1)
91. TP_VIT=IGNORADO TRANSP_VIT=SINF EVOL=ALTA 20 ==> ASSIST_MED=SIM 20 conf:(1)
92. TRANSP_VIT=SINF EVOL=ALTA 20 ==> TP_VIT=IGNORADO ASSIST_MED=SIM 20 conf:(1)
93. TRANSP_VIT=MOTOCICLETA LC_ASSIST=HPSMCBA EVOL=ALTA 20 ==>
ASSIST_MED=SIM 20 conf:(1)
94. LC_ASSIST=HPSMCBA EVOL=ALTA CAT_DIAS_INTER=3 20 ==> ASSIST_MED=SIM 20
conf:(1)
95. ASSIST_MED=SIM LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 20 ==> EVOL=ALTA 20
conf:(1)
96. LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 20 ==> ASSIST_MED=SIM EVOL=ALTA 20
conf:(1)
97. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=MOTOCICLETA EVOL=ALTA 20 ==>
ASSIST_MED=SIM 20 conf:(1)
98. SEXO=M TRANSP_VIT=SINF 19 ==> TP_VIT=IGNORADO 19 conf:(1)
99. TP_VIT=IGNORADO CAT_DIAS_INTER=1 19 ==> ASSIST_MED=SIM 19 conf:(1)
100. TP_VIT=IGNORADO CAT_DIAS_INTER=1 19 ==> EVOL=ALTA 19 conf:(1)
101. SEXO=M ZONA_RES=SINF EVOL=ALTA 19 ==> ASSIST_MED=SIM 19 conf:(1)
102. SEXO=M ZONA_RES=SINF ASSIST_MED=SIM 19 ==> EVOL=ALTA 19 conf:(1)
103. FX_ET=3 LC_ASSIST=HPSMCBA EVOL=ALTA 19 ==> ASSIST_MED=SIM 19 conf:(1)

APÊNDICE

104. TP_VIT=IGNORADO EVOL=ALTA CAT_DIAS_INTER=1 19 ==> ASSIST_MED=SIM 19
conf:(1)
105. TP_VIT=IGNORADO ASSIST_MED=SIM CAT_DIAS_INTER=1 19 ==> EVOL=ALTA 19
conf:(1)
106. TP_VIT=IGNORADO CAT_DIAS_INTER=1 19 ==> ASSIST_MED=SIM EVOL=ALTA 19
conf:(1)
107. SEXO=M FX_ET=3 TRANSP_VIT=MOTOCICLETA EVOL=ALTA 19 ==> ASSIST_MED=SIM 19
conf:(1)
108. LC_ASSIST=NENHUM 18 ==> ASSIST_MED=NÃO 18 conf:(1)
109. ASSIST_MED=NÃO 18 ==> LC_ASSIST=NENHUM 18 conf:(1)
110. ASSIST_MED=NÃO 18 ==> EVOL=OBITO 18 conf:(1)
111. ASSIST_MED=NÃO 18 ==> CAT_DIAS_INTER=0 18 conf:(1)
112. LC_ASSIST=NENHUM 18 ==> EVOL=OBITO 18 conf:(1)
113. LC_ASSIST=NENHUM 18 ==> CAT_DIAS_INTER=0 18 conf:(1)
114. ZONA_OC=OESTE CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM 18 conf:(1)
115. ZONA_OC=OESTE CAT_DIAS_INTER=1 18 ==> EVOL=ALTA 18 conf:(1)
116. TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM 18 conf:(1)
117. TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=1 18 ==> EVOL=ALTA 18 conf:(1)
118. LC_ASSIST=NENHUM EVOL=OBITO 18 ==> ASSIST_MED=NÃO 18 conf:(1)
119. ASSIST_MED=NÃO EVOL=OBITO 18 ==> LC_ASSIST=NENHUM 18 conf:(1)
120. ASSIST_MED=NÃO LC_ASSIST=NENHUM 18 ==> EVOL=OBITO 18 conf:(1)
121. LC_ASSIST=NENHUM 18 ==> ASSIST_MED=NÃO EVOL=OBITO 18 conf:(1)
122. ASSIST_MED=NÃO 18 ==> LC_ASSIST=NENHUM EVOL=OBITO 18 conf:(1)
123. LC_ASSIST=NENHUM CAT_DIAS_INTER=0 18 ==> ASSIST_MED=NÃO 18 conf:(1)
124. ASSIST_MED=NÃO CAT_DIAS_INTER=0 18 ==> LC_ASSIST=NENHUM 18 conf:(1)
125. ASSIST_MED=NÃO LC_ASSIST=NENHUM 18 ==> CAT_DIAS_INTER=0 18 conf:(1)
126. LC_ASSIST=NENHUM 18 ==> ASSIST_MED=NÃO CAT_DIAS_INTER=0 18 conf:(1)
127. ASSIST_MED=NÃO 18 ==> LC_ASSIST=NENHUM CAT_DIAS_INTER=0 18 conf:(1)
128. ASSIST_MED=NÃO CAT_DIAS_INTER=0 18 ==> EVOL=OBITO 18 conf:(1)
129. ASSIST_MED=NÃO EVOL=OBITO 18 ==> CAT_DIAS_INTER=0 18 conf:(1)
130. ASSIST_MED=NÃO 18 ==> EVOL=OBITO CAT_DIAS_INTER=0 18 conf:(1)
131. LC_ASSIST=NENHUM CAT_DIAS_INTER=0 18 ==> EVOL=OBITO 18 conf:(1)
132. LC_ASSIST=NENHUM EVOL=OBITO 18 ==> CAT_DIAS_INTER=0 18 conf:(1)
133. LC_ASSIST=NENHUM 18 ==> EVOL=OBITO CAT_DIAS_INTER=0 18 conf:(1)
134. SEXO=M TP_ACID=COLISÃO CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM 18 conf:(1)
135. SEXO=M TP_ACID=COLISÃO CAT_DIAS_INTER=1 18 ==> EVOL=ALTA 18 conf:(1)
136. ZONA_OC=OESTE EVOL=ALTA CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM 18 conf:(1)
137. ZONA_OC=OESTE ASSIST_MED=SIM CAT_DIAS_INTER=1 18 ==> EVOL=ALTA 18 conf:(1)
138. ZONA_OC=OESTE CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM EVOL=ALTA 18 conf:(1)
139. TP_ACID=COLISÃO TRANSP_VIT=VEICULO EVOL=ALTA 18 ==> ASSIST_MED=SIM 18
conf:(1)
140. TP_ACID=COLISÃO TRANSP_VIT=VEICULO ASSIST_MED=SIM 18 ==> EVOL=ALTA 18
conf:(1)
141. TP_ACID=COLISÃO LC_ASSIST=HPSMCBA EVOL=ALTA 18 ==> ASSIST_MED=SIM 18
conf:(1)
142. TRANSP_VIT=MOTOCICLETA EVOL=ALTA CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM
18 conf:(1)
143. TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM CAT_DIAS_INTER=1 18 ==> EVOL=ALTA
18 conf:(1)
144. TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM EVOL=ALTA
18 conf:(1)
145. LC_ASSIST=NENHUM EVOL=OBITO CAT_DIAS_INTER=0 18 ==> ASSIST_MED=NÃO 18
conf:(1)
146. ASSIST_MED=NÃO EVOL=OBITO CAT_DIAS_INTER=0 18 ==> LC_ASSIST=NENHUM 18
conf:(1)

APÊNDICE

147. ASSIST_MED=NÃO LC_ASSIST=NENHUM CAT_DIAS_INTER=0 18 ==> EVOL=OBITO 18
conf:(1)
148. ASSIST_MED=NÃO LC_ASSIST=NENHUM EVOL=OBITO 18 ==> CAT_DIAS_INTER=0 18
conf:(1)
149. LC_ASSIST=NENHUM CAT_DIAS_INTER=0 18 ==> ASSIST_MED=NÃO EVOL=OBITO 18
conf:(1)
150. LC_ASSIST=NENHUM EVOL=OBITO 18 ==> ASSIST_MED=NÃO CAT_DIAS_INTER=0 18
conf:(1)
151. ASSIST_MED=NÃO CAT_DIAS_INTER=0 18 ==> LC_ASSIST=NENHUM EVOL=OBITO 18
conf:(1)
152. ASSIST_MED=NÃO EVOL=OBITO 18 ==> LC_ASSIST=NENHUM CAT_DIAS_INTER=0 18
conf:(1)
153. ASSIST_MED=NÃO LC_ASSIST=NENHUM 18 ==> EVOL=OBITO CAT_DIAS_INTER=0 18
conf:(1)
154. LC_ASSIST=NENHUM 18 ==> ASSIST_MED=NÃO EVOL=OBITO CAT_DIAS_INTER=0 18
conf:(1)
155. ASSIST_MED=NÃO 18 ==> LC_ASSIST=NENHUM EVOL=OBITO CAT_DIAS_INTER=0 18
conf:(1)
156. SEXO=M FX_ET=3 LC_ASSIST=HPSMCBA EVOL=ALTA 18 ==> ASSIST_MED=SIM 18
conf:(1)
157. SEXO=M TP_ACID=COLISÃO TP_VIT=IGNORADO EVOL=ALTA 18 ==> ASSIST_MED=SIM
18 conf:(1)
158. SEXO=M TP_ACID=COLISÃO EVOL=ALTA CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM
18 conf:(1)
159. SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM CAT_DIAS_INTER=1 18 ==> EVOL=ALTA
18 conf:(1)
160. SEXO=M TP_ACID=COLISÃO CAT_DIAS_INTER=1 18 ==> ASSIST_MED=SIM EVOL=ALTA
18 conf:(1)
161. SEXO=M ZONA_RES=NORTE 17 ==> ASSIST_MED=SIM 17 conf:(1)
162. FX_ET=5 EVOL=ALTA 17 ==> ASSIST_MED=SIM 17 conf:(1)
163. ZONA_RES=NORTE EVOL=ALTA 17 ==> ASSIST_MED=SIM 17 conf:(1)
164. ZONA_OC=OESTE LC_ASSIST=HPSMCBA 17 ==> ASSIST_MED=SIM 17 conf:(1)
165. TP_ACID=ATROPELAMENTO EVOL=ALTA 17 ==> ASSIST_MED=SIM 17 conf:(1)
166. SEXO=M FX_ET=4 EVOL=ALTA 17 ==> ASSIST_MED=SIM 17 conf:(1)
167. ZONA_RES=SUL ZONA_OC=SUL ASSIST_MED=SIM 17 ==> SEXO=M 17 conf:(1)
168. SEXO=M TRANSP_VIT=SINF ASSIST_MED=SIM 17 ==> TP_VIT=IGNORADO 17 conf:(1)
169. FX_ET=3 TP_VIT=CONDUTOR EVOL=ALTA 17 ==> ASSIST_MED=SIM 17 conf:(1)
170. SEXO=M TP_ACID=COLISÃO TP_VIT=CONDUTOR EVOL=ALTA 17 ==> ASSIST_MED=SIM
17 conf:(1)
171. SEXO=M TRANSP_VIT=MOTOCICLETA LC_ASSIST=HPSMCBA EVOL=ALTA 17 ==>
ASSIST_MED=SIM 17 conf:(1)
172. MES_OC=SET 16 ==> ASSIST_MED=SIM 16 conf:(1)
173. ZONA_RES=SUL LC_ASSIST=HPSMCBA 16 ==> ASSIST_MED=SIM 16 conf:(1)
174. ZONA_OC=LESTE EVOL=ALTA 16 ==> ASSIST_MED=SIM 16 conf:(1)
175. TP_VIT=CONDUTOR LC_ASSIST=HPSMCBA 16 ==> ASSIST_MED=SIM 16 conf:(1)
176. TP_VIT=IGNORADO LC_ASSIST=HPSMCBA 16 ==> ASSIST_MED=SIM 16 conf:(1)
177. SEXO=M FX_ET=3 CAT_DIAS_INTER=1 16 ==> ASSIST_MED=SIM 16 conf:(1)
178. SEXO=M ZONA_OC=SINF EVOL=ALTA 16 ==> ASSIST_MED=SIM 16 conf:(1)
179. SEXO=M TRANSP_VIT=SINF EVOL=ALTA 16 ==> TP_VIT=IGNORADO 16 conf:(1)
180. SEXO=M TRANSP_VIT=SINF EVOL=ALTA 16 ==> ASSIST_MED=SIM 16 conf:(1)
181. SEXO=M LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 16 ==> ASSIST_MED=SIM 16
conf:(1)
182. SEXO=M LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 16 ==> EVOL=ALTA 16 conf:(1)
183. TP_ACID=COLISÃO TP_VIT=PASSAGEIRO EVOL=ALTA 16 ==> ASSIST_MED=SIM 16
conf:(1)

APÊNDICE

184. TP_VIT=CONDUTOR TRANSP_VIT=VEICULO EVOL=ALTA 16 ==> ASSIST_MED=SIM 16 conf:(1)
185. SEXO=M TP_ACID=COLISÃO LC_ASSIST=HPSMCBA EVOL=ALTA 16 ==> ASSIST_MED=SIM 16 conf:(1)
186. SEXO=M TRANSP_VIT=SINF ASSIST_MED=SIM EVOL=ALTA 16 ==> TP_VIT=IGNORADO 16 conf:(1)
187. SEXO=M TP_VIT=IGNORADO TRANSP_VIT=SINF EVOL=ALTA 16 ==> ASSIST_MED=SIM 16 conf:(1)
188. SEXO=M TRANSP_VIT=SINF EVOL=ALTA 16 ==> TP_VIT=IGNORADO ASSIST_MED=SIM 16 conf:(1)
189. SEXO=M LC_ASSIST=HPSMCBA EVOL=ALTA CAT_DIAS_INTER=3 16 ==> ASSIST_MED=SIM 16 conf:(1)
190. SEXO=M ASSIST_MED=SIM LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 16 ==> EVOL=ALTA 16 conf:(1)
191. SEXO=M LC_ASSIST=HPSMCBA CAT_DIAS_INTER=3 16 ==> ASSIST_MED=SIM EVOL=ALTA 16 conf:(1)
192. TP_VIT=PEDESTRE 15 ==> TP_ACID=ATROPELAMENTO 15 conf:(1)
193. TRANSP_VIT=APÉ 15 ==> TP_ACID=ATROPELAMENTO 15 conf:(1)
194. TRANSP_VIT=APÉ 15 ==> TP_VIT=PEDESTRE 15 conf:(1)
195. TP_VIT=PEDESTRE 15 ==> TRANSP_VIT=APÉ 15 conf:(1)
196. LC_ASSIST=STACASA 15 ==> ASSIST_MED=SIM 15 conf:(1)
197. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> SEXO=M 15 conf:(1)
198. FX_ET=3 TP_VIT=IGNORADO 15 ==> ASSIST_MED=SIM 15 conf:(1)
199. FX_ET=3 TP_VIT=IGNORADO 15 ==> EVOL=ALTA 15 conf:(1)
200. MES_OC=JUN EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
201. MES_OC=SET EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
202. MES_OC=NOV EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
203. ZONA_OC=OESTE TRANSP_VIT=MOTOCICLETA 15 ==> ASSIST_MED=SIM 15 conf:(1)
204. TP_VIT=PEDESTRE TRANSP_VIT=APÉ 15 ==> TP_ACID=ATROPELAMENTO 15 conf:(1)
205. TP_ACID=ATROPELAMENTO TRANSP_VIT=APÉ 15 ==> TP_VIT=PEDESTRE 15 conf:(1)
206. TP_ACID=ATROPELAMENTO TP_VIT=PEDESTRE 15 ==> TRANSP_VIT=APÉ 15 conf:(1)
207. TRANSP_VIT=APÉ 15 ==> TP_ACID=ATROPELAMENTO TP_VIT=PEDESTRE 15 conf:(1)
208. TP_VIT=PEDESTRE 15 ==> TP_ACID=ATROPELAMENTO TRANSP_VIT=APÉ 15 conf:(1)
209. TP_ACID=COLISÃO TRANSP_VIT=SINF 15 ==> TP_VIT=IGNORADO 15 conf:(1)
210. TP_VIT=CONDUTOR CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM 15 conf:(1)
211. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> ASSIST_MED=SIM 15 conf:(1)
212. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> EVOL=ALTA 15 conf:(1)
213. LC_ASSIST=HPSMCBA CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM 15 conf:(1)
214. ZONA_RES=SUL ZONA_OC=SUL EVOL=ALTA 15 ==> SEXO=M 15 conf:(1)
215. SEXO=M ZONA_RES=LESTE EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
216. SEXO=M ZONA_OC=LESTE EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
217. TP_VIT=CONDUTOR ASSIST_MED=SIM CAT_DIAS_INTER=2 15 ==> SEXO=M 15 conf:(1)
218. SEXO=M TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> ASSIST_MED=SIM 15 conf:(1)
219. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> SEXO=M ASSIST_MED=SIM 15 conf:(1)
220. TP_VIT=CONDUTOR EVOL=ALTA CAT_DIAS_INTER=2 15 ==> SEXO=M 15 conf:(1)
221. SEXO=M TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> EVOL=ALTA 15 conf:(1)
222. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> SEXO=M EVOL=ALTA 15 conf:(1)
223. SEXO=M TP_VIT=IGNORADO CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM 15 conf:(1)
224. SEXO=M TP_VIT=IGNORADO CAT_DIAS_INTER=1 15 ==> EVOL=ALTA 15 conf:(1)
225. FX_ET=3 ZONA_OC=SUL EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
226. FX_ET=3 TP_VIT=IGNORADO EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
227. FX_ET=3 TP_VIT=IGNORADO ASSIST_MED=SIM 15 ==> EVOL=ALTA 15 conf:(1)
228. FX_ET=3 TP_VIT=IGNORADO 15 ==> ASSIST_MED=SIM EVOL=ALTA 15 conf:(1)
229. ZONA_RES=SUL ZONA_OC=SUL EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
230. ZONA_OC=OESTE TP_VIT=IGNORADO EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)

APÊNDICE

231. TP_VIT=CONDUTOR TRANSP_VIT=MOTOCICLETA EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
232. TP_VIT=CONDUTOR EVOL=ALTA CAT_DIAS_INTER=2 15 ==> ASSIST_MED=SIM 15 conf:(1)
233. TP_VIT=CONDUTOR ASSIST_MED=SIM CAT_DIAS_INTER=2 15 ==> EVOL=ALTA 15 conf:(1)
234. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> ASSIST_MED=SIM EVOL=ALTA 15 conf:(1)
235. TP_VIT=IGNORADO LC_ASSIST=HPSMCBA EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
236. SEXO=M FX_ET=3 TP_VIT=CONDUTOR EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
237. SEXO=M FX_ET=3 EVOL=ALTA CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM 15 conf:(1)
238. ZONA_RES=SUL ZONA_OC=SUL ASSIST_MED=SIM EVOL=ALTA 15 ==> SEXO=M 15 conf:(1)
239. SEXO=M ZONA_RES=SUL ZONA_OC=SUL EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
240. ZONA_RES=SUL ZONA_OC=SUL EVOL=ALTA 15 ==> SEXO=M ASSIST_MED=SIM 15 conf:(1)
241. SEXO=M TP_ACID=COLISÃO TP_VIT=PASSAGEIRO EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
242. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=VEICULO EVOL=ALTA 15 ==> ASSIST_MED=SIM 15 conf:(1)
243. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=VEICULO ASSIST_MED=SIM 15 ==> EVOL=ALTA 15 conf:(1)
244. TP_VIT=CONDUTOR ASSIST_MED=SIM EVOL=ALTA CAT_DIAS_INTER=2 15 ==> SEXO=M 15 conf:(1)
245. SEXO=M TP_VIT=CONDUTOR EVOL=ALTA CAT_DIAS_INTER=2 15 ==> ASSIST_MED=SIM 15 conf:(1)
246. SEXO=M TP_VIT=CONDUTOR ASSIST_MED=SIM CAT_DIAS_INTER=2 15 ==> EVOL=ALTA 15 conf:(1)
247. TP_VIT=CONDUTOR EVOL=ALTA CAT_DIAS_INTER=2 15 ==> SEXO=M ASSIST_MED=SIM 15 conf:(1)
248. TP_VIT=CONDUTOR ASSIST_MED=SIM CAT_DIAS_INTER=2 15 ==> SEXO=M EVOL=ALTA 15 conf:(1)
249. SEXO=M TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> ASSIST_MED=SIM EVOL=ALTA 15 conf:(1)
250. TP_VIT=CONDUTOR CAT_DIAS_INTER=2 15 ==> SEXO=M ASSIST_MED=SIM EVOL=ALTA 15 conf:(1)
251. SEXO=M TP_VIT=IGNORADO EVOL=ALTA CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM 15 conf:(1)
252. SEXO=M TP_VIT=IGNORADO ASSIST_MED=SIM CAT_DIAS_INTER=1 15 ==> EVOL=ALTA 15 conf:(1)
253. SEXO=M TP_VIT=IGNORADO CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM EVOL=ALTA 15 conf:(1)
254. FX_ET=3 CAT_DIAS_INTER=3 14 ==> SEXO=M 14 conf:(1)
255. TP_VIT=PASSAGEIRO TRANSP_VIT=MOTOCICLETA 14 ==> SEXO=M 14 conf:(1)
256. FX_ET=3 CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM 14 conf:(1)
257. FX_ET=3 CAT_DIAS_INTER=3 14 ==> ASSIST_MED=SIM 14 conf:(1)
258. FX_ET=3 CAT_DIAS_INTER=2 14 ==> EVOL=ALTA 14 conf:(1)
259. TP_ACID=COLISÃO CAT_DIAS_INTER=0 14 ==> EVOL=OBITO 14 conf:(1)
260. TP_VIT=PASSAGEIRO TRANSP_VIT=MOTOCICLETA 14 ==> ASSIST_MED=SIM 14 conf:(1)
261. TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM 14 conf:(1)
262. LC_ASSIST=HMBOMJESUS CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM 14 conf:(1)
263. LC_ASSIST=STACASA EVOL=ALTA 14 ==> ASSIST_MED=SIM 14 conf:(1)
264. LC_ASSIST=HMBOMJESUS CAT_DIAS_INTER=2 14 ==> EVOL=ALTA 14 conf:(1)
265. FX_ET=3 ASSIST_MED=SIM CAT_DIAS_INTER=3 14 ==> SEXO=M 14 conf:(1)

APÊNDICE

266. SEXO=M FX_ET=3 CAT_DIAS_INTER=3 14 ==> ASSIST_MED=SIM 14 conf:(1)
 267. FX_ET=3 CAT_DIAS_INTER=3 14 ==> SEXO=M ASSIST_MED=SIM 14 conf:(1)
 268. SEXO=M FX_ET=5 EVOL=ALTA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 269. SEXO=M ZONA_RES=NORTE EVOL=ALTA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 270. SEXO=M ZONA_RES=SUL LC_ASSIST=HPSMCBA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 271. TP_VIT=PASSAGEIRO TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 14 ==> SEXO=M 14 conf:(1)
 272. SEXO=M TP_VIT=PASSAGEIRO TRANSP_VIT=MOTOCICLETA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 273. TP_VIT=PASSAGEIRO TRANSP_VIT=MOTOCICLETA 14 ==> SEXO=M ASSIST_MED=SIM 14 conf:(1)
 274. SEXO=M TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=1 14 ==> ASSIST_MED=SIM 14 conf:(1)
 275. SEXO=M TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=1 14 ==> EVOL=ALTA 14 conf:(1)
 276. FX_ET=3 EVOL=ALTA CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM 14 conf:(1)
 277. FX_ET=3 ASSIST_MED=SIM CAT_DIAS_INTER=2 14 ==> EVOL=ALTA 14 conf:(1)
 278. FX_ET=3 CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM EVOL=ALTA 14 conf:(1)
 279. ZONA_OC=OESTE TP_ACID=COLISÃO EVOL=ALTA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 280. ZONA_OC=SINF TP_ACID=COLISÃO EVOL=ALTA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 281. ZONA_OC=SINF TP_ACID=COLISÃO ASSIST_MED=SIM 14 ==> EVOL=ALTA 14 conf:(1)
 282. TP_ACID=COLISÃO TRANSP_VIT=SINF ASSIST_MED=SIM 14 ==> TP_VIT=IGNORADO 14 conf:(1)
 283. TP_VIT=CONDUTOR EVOL=ALTA CAT_DIAS_INTER=1 14 ==> ASSIST_MED=SIM 14 conf:(1)
 284. LC_ASSIST=HMBOMJESUS EVOL=ALTA CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM 14 conf:(1)
 285. ASSIST_MED=SIM LC_ASSIST=HMBOMJESUS CAT_DIAS_INTER=2 14 ==> EVOL=ALTA 14 conf:(1)
 286. LC_ASSIST=HMBOMJESUS CAT_DIAS_INTER=2 14 ==> ASSIST_MED=SIM EVOL=ALTA 14 conf:(1)
 287. LC_ASSIST=HPSMCBA EVOL=ALTA CAT_DIAS_INTER=1 14 ==> ASSIST_MED=SIM 14 conf:(1)
 288. SEXO=M TP_VIT=CONDUTOR TRANSP_VIT=VEICULO EVOL=ALTA 14 ==> ASSIST_MED=SIM 14 conf:(1)
 289. SEXO=M TRANSP_VIT=MOTOCICLETA EVOL=ALTA CAT_DIAS_INTER=1 14 ==> ASSIST_MED=SIM 14 conf:(1)
 290. SEXO=M TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM CAT_DIAS_INTER=1 14 ==> EVOL=ALTA 14 conf:(1)
 291. SEXO=M TRANSP_VIT=MOTOCICLETA CAT_DIAS_INTER=1 14 ==> ASSIST_MED=SIM EVOL=ALTA 14 conf:(1)
 292. CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44 conf:(0.98)
 293. ASSIST_MED=SIM CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44 conf:(0.98)
 294. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM EVOL=ALTA 44 conf:(0.98)
 295. CAT_DIAS_INTER=2 39 ==> EVOL=ALTA 38 conf:(0.97)
 296. ASSIST_MED=SIM CAT_DIAS_INTER=2 39 ==> EVOL=ALTA 38 conf:(0.97)
 297. CAT_DIAS_INTER=2 39 ==> ASSIST_MED=SIM EVOL=ALTA 38 conf:(0.97)
 298. SEXO=M CAT_DIAS_INTER=1 34 ==> EVOL=ALTA 33 conf:(0.97)
 299. SEXO=M CAT_DIAS_INTER=2 34 ==> EVOL=ALTA 33 conf:(0.97)
 300. SEXO=M ASSIST_MED=SIM CAT_DIAS_INTER=1 34 ==> EVOL=ALTA 33 conf:(0.97)
 301. SEXO=M CAT_DIAS_INTER=1 34 ==> ASSIST_MED=SIM EVOL=ALTA 33 conf:(0.97)
 302. SEXO=M ASSIST_MED=SIM CAT_DIAS_INTER=2 34 ==> EVOL=ALTA 33 conf:(0.97)
 303. SEXO=M CAT_DIAS_INTER=2 34 ==> ASSIST_MED=SIM EVOL=ALTA 33 conf:(0.97)
 304. CAT_DIAS_INTER=3 29 ==> EVOL=ALTA 28 conf:(0.97)
 305. TP_VIT=PASSAGEIRO ASSIST_MED=SIM 29 ==> SEXO=M 28 conf:(0.97)
 306. SEXO=M TRANSP_VIT=VEICULO 29 ==> ASSIST_MED=SIM 28 conf:(0.97)

APÊNDICE

307. ASSIST_MED=SIM CAT_DIAS_INTER=3 29 ==> EVOL=ALTA 28 conf:(0.97)
308. CAT_DIAS_INTER=3 29 ==> ASSIST_MED=SIM EVOL=ALTA 28 conf:(0.97)
309. ZONA_OC=SINF ASSIST_MED=SIM 28 ==> EVOL=ALTA 27 conf:(0.96)
310. TP_ACID=SINF ASSIST_MED=SIM 27 ==> EVOL=ALTA 26 conf:(0.96)
311. TP_VIT=PASSAGEIRO EVOL=ALTA 26 ==> SEXO=M 25 conf:(0.96)
312. TP_VIT=PASSAGEIRO ASSIST_MED=SIM EVOL=ALTA 26 ==> SEXO=M 25 conf:(0.96)
313. TP_VIT=PASSAGEIRO EVOL=ALTA 26 ==> SEXO=M ASSIST_MED=SIM 25 conf:(0.96)
314. SEXO=M CAT_DIAS_INTER=3 24 ==> EVOL=ALTA 23 conf:(0.96)
315. ZONA_RES=OESTE ASSIST_MED=SIM 24 ==> EVOL=ALTA 23 conf:(0.96)
316. TP_ACID=COLISÃO CAT_DIAS_INTER=2 24 ==> EVOL=ALTA 23 conf:(0.96)
317. SEXO=M ASSIST_MED=SIM CAT_DIAS_INTER=3 24 ==> EVOL=ALTA 23 conf:(0.96)
318. SEXO=M CAT_DIAS_INTER=3 24 ==> ASSIST_MED=SIM EVOL=ALTA 23 conf:(0.96)
319. TP_ACID=COLISÃO ASSIST_MED=SIM CAT_DIAS_INTER=2 24 ==> EVOL=ALTA 23
conf:(0.96)
320. TP_ACID=COLISÃO CAT_DIAS_INTER=2 24 ==> ASSIST_MED=SIM EVOL=ALTA 23
conf:(0.96)
321. ZONA_OC=LESTE 23 ==> SEXO=M 22 conf:(0.96)
322. FX_ET=3 LC_ASSIST=HPSMCBA 23 ==> SEXO=M 22 conf:(0.96)
323. FX_ET=3 ASSIST_MED=SIM LC_ASSIST=HPSMCBA 23 ==> SEXO=M 22 conf:(0.96)
324. FX_ET=3 LC_ASSIST=HPSMCBA 23 ==> SEXO=M ASSIST_MED=SIM 22 conf:(0.96)
325. SEXO=M ZONA_RES=OESTE 22 ==> ASSIST_MED=SIM 21 conf:(0.95)
326. SEXO=F ASSIST_MED=SIM 22 ==> EVOL=ALTA 21 conf:(0.95)
327. FX_ET=4 ASSIST_MED=SIM 22 ==> EVOL=ALTA 21 conf:(0.95)
328. FX_ET=3 TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 22 ==> SEXO=M 21 conf:(0.95)
329. SEXO=M FX_ET=3 TRANSP_VIT=MOTOCICLETA 22 ==> ASSIST_MED=SIM 21 conf:(0.95)
330. SEXO=M TP_ACID=COLISÃO CAT_DIAS_INTER=2 22 ==> EVOL=ALTA 21 conf:(0.95)
331. SEXO=M TP_ACID=SINF ASSIST_MED=SIM 22 ==> EVOL=ALTA 21 conf:(0.95)
332. TP_ACID=COLISÃO TP_VIT=IGNORADO ASSIST_MED=SIM 22 ==> EVOL=ALTA 21
conf:(0.95)
333. SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM CAT_DIAS_INTER=2 22 ==> EVOL=ALTA
21 conf:(0.95)
334. SEXO=M TP_ACID=COLISÃO CAT_DIAS_INTER=2 22 ==> ASSIST_MED=SIM EVOL=ALTA
21 conf:(0.95)
335. ZONA_RES=NORTE 21 ==> ASSIST_MED=SIM 20 conf:(0.95)
336. FX_ET=3 CAT_DIAS_INTER=1 21 ==> EVOL=ALTA 20 conf:(0.95)
337. TRANSP_VIT=SINF ASSIST_MED=SIM 21 ==> EVOL=ALTA 20 conf:(0.95)
338. SEXO=M ZONA_RES=OESTE ASSIST_MED=SIM 21 ==> EVOL=ALTA 20 conf:(0.95)
339. FX_ET=3 ASSIST_MED=SIM CAT_DIAS_INTER=1 21 ==> EVOL=ALTA 20 conf:(0.95)
340. FX_ET=3 CAT_DIAS_INTER=1 21 ==> ASSIST_MED=SIM EVOL=ALTA 20 conf:(0.95)
341. TP_VIT=IGNORADO TRANSP_VIT=SINF ASSIST_MED=SIM 21 ==> EVOL=ALTA 20
conf:(0.95)
342. TRANSP_VIT=SINF ASSIST_MED=SIM 21 ==> TP_VIT=IGNORADO EVOL=ALTA 20
conf:(0.95)
343. SEXO=M TRANSP_VIT=MOTOCICLETA 40 ==> ASSIST_MED=SIM 38 conf:(0.95)
344. FX_ET=3 TRANSP_VIT=MOTOCICLETA EVOL=ALTA 20 ==> SEXO=M 19 conf:(0.95)
345. SEXO=M TP_ACID=COLISÃO TP_VIT=IGNORADO 20 ==> ASSIST_MED=SIM 19 conf:(0.95)
346. FX_ET=3 TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM EVOL=ALTA 20 ==> SEXO=M 19
conf:(0.95)
347. FX_ET=3 TRANSP_VIT=MOTOCICLETA EVOL=ALTA 20 ==> SEXO=M ASSIST_MED=SIM 19
conf:(0.95)
348. ZONA_RES=SUL ZONA_OC=SUL 19 ==> SEXO=M 18 conf:(0.95)
349. FX_ET=3 LC_ASSIST=HPSMCBA EVOL=ALTA 19 ==> SEXO=M 18 conf:(0.95)
350. FX_ET=3 ASSIST_MED=SIM LC_ASSIST=HPSMCBA EVOL=ALTA 19 ==> SEXO=M 18
conf:(0.95)
351. FX_ET=3 LC_ASSIST=HPSMCBA EVOL=ALTA 19 ==> SEXO=M ASSIST_MED=SIM 18
conf:(0.95)

APÊNDICE

352. SEXO=M TP_ACID=COLISÃO TP_VIT=IGNORADO ASSIST_MED=SIM 19 ==> EVOL=ALTA 18 conf:(0.95)
353. TP_VIT=IGNORADO ASSIST_MED=SIM 37 ==> EVOL=ALTA 35 conf:(0.95)
354. SEXO=M FX_ET=5 18 ==> ASSIST_MED=SIM 17 conf:(0.94)
355. SEXO=M ZONA_OC=SINF 18 ==> ASSIST_MED=SIM 17 conf:(0.94)
356. FX_ET=3 ZONA_OC=SUL 18 ==> ASSIST_MED=SIM 17 conf:(0.94)
357. ZONA_OC=OESTE TP_VIT=IGNORADO 18 ==> ASSIST_MED=SIM 17 conf:(0.94)
358. TP_ACID=ATROPELAMENTO ASSIST_MED=SIM 18 ==> EVOL=ALTA 17 conf:(0.94)
359. TP_VIT=CONDUTOR TRANSP_VIT=VEICULO 18 ==> ASSIST_MED=SIM 17 conf:(0.94)
360. SEXO=M FX_ET=4 ASSIST_MED=SIM 18 ==> EVOL=ALTA 17 conf:(0.94)
361. SEXO=M ZONA_RES=SUL ZONA_OC=SUL 18 ==> ASSIST_MED=SIM 17 conf:(0.94)
362. TP_ACID=COLISÃO TP_VIT=PASSAGEIRO ASSIST_MED=SIM 18 ==> SEXO=M 17 conf:(0.94)
363. TRANSP_VIT=VEICULO 34 ==> ASSIST_MED=SIM 32 conf:(0.94)
364. ZONA_OC=LESTE ASSIST_MED=SIM 17 ==> SEXO=M 16 conf:(0.94)
365. ZONA_OC=LESTE ASSIST_MED=SIM 17 ==> EVOL=ALTA 16 conf:(0.94)
366. SEXO=M ZONA_OC=SINF ASSIST_MED=SIM 17 ==> EVOL=ALTA 16 conf:(0.94)
367. SEXO=M TRANSP_VIT=SINF ASSIST_MED=SIM 17 ==> EVOL=ALTA 16 conf:(0.94)
368. TP_VIT=CONDUTOR TRANSP_VIT=VEICULO ASSIST_MED=SIM 17 ==> EVOL=ALTA 16 conf:(0.94)
369. SEXO=M TP_VIT=IGNORADO TRANSP_VIT=SINF ASSIST_MED=SIM 17 ==> EVOL=ALTA 16 conf:(0.94)
370. SEXO=M TRANSP_VIT=SINF ASSIST_MED=SIM 17 ==> TP_VIT=IGNORADO EVOL=ALTA 16 conf:(0.94)
371. MES_OC=SET 16 ==> EVOL=ALTA 15 conf:(0.94)
372. FX_ET=3 ZONA_RES=SUL 16 ==> SEXO=M 15 conf:(0.94)
373. FX_ET=3 TP_VIT=PASSAGEIRO 16 ==> SEXO=M 15 conf:(0.94)
374. FX_ET=4 TP_ACID=COLISÃO 16 ==> SEXO=M 15 conf:(0.94)
375. ZONA_OC=LESTE EVOL=ALTA 16 ==> SEXO=M 15 conf:(0.94)
376. FX_ET=3 TRANSP_VIT=VEICULO 16 ==> ASSIST_MED=SIM 15 conf:(0.94)
377. MES_OC=JUN ASSIST_MED=SIM 16 ==> EVOL=ALTA 15 conf:(0.94)
378. MES_OC=SET ASSIST_MED=SIM 16 ==> EVOL=ALTA 15 conf:(0.94)
379. MES_OC=SET 16 ==> ASSIST_MED=SIM EVOL=ALTA 15 conf:(0.94)
380. MES_OC=NOV ASSIST_MED=SIM 16 ==> EVOL=ALTA 15 conf:(0.94)
381. TP_VIT=IGNORADO LC_ASSIST=HPSMCBA 16 ==> EVOL=ALTA 15 conf:(0.94)
382. SEXO=M FX_ET=3 ZONA_OC=SUL 16 ==> ASSIST_MED=SIM 15 conf:(0.94)
383. SEXO=M FX_ET=3 CAT_DIAS_INTER=1 16 ==> EVOL=ALTA 15 conf:(0.94)
384. ZONA_OC=LESTE ASSIST_MED=SIM EVOL=ALTA 16 ==> SEXO=M 15 conf:(0.94)
385. SEXO=M ZONA_OC=LESTE ASSIST_MED=SIM 16 ==> EVOL=ALTA 15 conf:(0.94)
386. ZONA_OC=LESTE EVOL=ALTA 16 ==> SEXO=M ASSIST_MED=SIM 15 conf:(0.94)
387. SEXO=M ZONA_OC=OESTE TP_ACID=COLISÃO 16 ==> ASSIST_MED=SIM 15 conf:(0.94)
388. TP_ACID=COLISÃO TP_VIT=PASSAGEIRO EVOL=ALTA 16 ==> SEXO=M 15 conf:(0.94)
389. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=VEICULO 16 ==> ASSIST_MED=SIM 15 conf:(0.94)
390. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=VEICULO 16 ==> EVOL=ALTA 15 conf:(0.94)
391. SEXO=M TP_VIT=CONDUTOR TRANSP_VIT=VEICULO 16 ==> ASSIST_MED=SIM 15 conf:(0.94)
392. TP_VIT=CONDUTOR TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 16 ==> EVOL=ALTA 15 conf:(0.94)
393. TP_VIT=IGNORADO ASSIST_MED=SIM LC_ASSIST=HPSMCBA 16 ==> EVOL=ALTA 15 conf:(0.94)
394. TP_VIT=IGNORADO LC_ASSIST=HPSMCBA 16 ==> ASSIST_MED=SIM EVOL=ALTA 15 conf:(0.94)
395. SEXO=M FX_ET=3 ASSIST_MED=SIM CAT_DIAS_INTER=1 16 ==> EVOL=ALTA 15 conf:(0.94)

APÊNDICE

396. SEXO=M FX_ET=3 CAT_DIAS_INTER=1 16 ==> ASSIST_MED=SIM EVOL=ALTA 15
conf:(0.94)
397. TP_ACID=COLISÃO TP_VIT=PASSAGEIRO ASSIST_MED=SIM EVOL=ALTA 16 ==> SEXO=M
15 conf:(0.94)
398. TP_ACID=COLISÃO TP_VIT=PASSAGEIRO EVOL=ALTA 16 ==> SEXO=M ASSIST_MED=SIM
15 conf:(0.94)
399. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=VEICULO 16 ==> ASSIST_MED=SIM
EVOL=ALTA 15 conf:(0.94)
400. TRANSP_VIT=MOTOCICLETA 46 ==> ASSIST_MED=SIM 43 conf:(0.93)
401. ZONA_OC=SINF 30 ==> ASSIST_MED=SIM 28 conf:(0.93)
402. SEXO=M ZONA_OC=OESTE 30 ==> ASSIST_MED=SIM 28 conf:(0.93)
403. SEXO=M TP_VIT=IGNORADO ASSIST_MED=SIM 30 ==> EVOL=ALTA 28 conf:(0.93)
404. MES_OC=ABR 15 ==> SEXO=M 14 conf:(0.93)
405. LC_ASSIST=STACASA 15 ==> EVOL=ALTA 14 conf:(0.93)
406. SEXO=M MES_OC=NOV 15 ==> ASSIST_MED=SIM 14 conf:(0.93)
407. ZONA_RES=SUL TP_ACID=COLISÃO 15 ==> SEXO=M 14 conf:(0.93)
408. ZONA_OC=SUL TP_VIT=CONDUTOR 15 ==> SEXO=M 14 conf:(0.93)
409. ZONA_OC=SUL TP_VIT=CONDUTOR 15 ==> ASSIST_MED=SIM 14 conf:(0.93)
410. ZONA_OC=SINF TP_ACID=COLISÃO 15 ==> ASSIST_MED=SIM 14 conf:(0.93)
411. ZONA_OC=SINF TP_ACID=COLISÃO 15 ==> EVOL=ALTA 14 conf:(0.93)
412. TP_ACID=COLISÃO TRANSP_VIT=SINF 15 ==> ASSIST_MED=SIM 14 conf:(0.93)
413. TP_ACID=COLISÃO EVOL=OBITO 15 ==> CAT_DIAS_INTER=0 14 conf:(0.93)
414. TP_VIT=CONDUTOR CAT_DIAS_INTER=1 15 ==> EVOL=ALTA 14 conf:(0.93)
415. ASSIST_MED=SIM LC_ASSIST=STACASA 15 ==> EVOL=ALTA 14 conf:(0.93)
416. LC_ASSIST=STACASA 15 ==> ASSIST_MED=SIM EVOL=ALTA 14 conf:(0.93)
417. LC_ASSIST=HPSMCBA CAT_DIAS_INTER=1 15 ==> EVOL=ALTA 14 conf:(0.93)
418. ZONA_OC=SINF TP_ACID=COLISÃO 15 ==> ASSIST_MED=SIM EVOL=ALTA 14 conf:(0.93)
419. TP_ACID=COLISÃO TP_VIT=IGNORADO TRANSP_VIT=SINF 15 ==> ASSIST_MED=SIM 14
conf:(0.93)
420. TP_ACID=COLISÃO TRANSP_VIT=SINF 15 ==> TP_VIT=IGNORADO ASSIST_MED=SIM 14
conf:(0.93)
421. TP_VIT=CONDUTOR ASSIST_MED=SIM CAT_DIAS_INTER=1 15 ==> EVOL=ALTA 14
conf:(0.93)
422. TP_VIT=CONDUTOR CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM EVOL=ALTA 14
conf:(0.93)
423. ASSIST_MED=SIM LC_ASSIST=HPSMCBA CAT_DIAS_INTER=1 15 ==> EVOL=ALTA 14
conf:(0.93)
424. LC_ASSIST=HPSMCBA CAT_DIAS_INTER=1 15 ==> ASSIST_MED=SIM EVOL=ALTA 14
conf:(0.93)
425. SEXO=M TP_VIT=CONDUTOR TRANSP_VIT=VEICULO ASSIST_MED=SIM 15 ==>
EVOL=ALTA 14 conf:(0.93)
426. TP_ACID=SINF 29 ==> ASSIST_MED=SIM 27 conf:(0.93)
427. ZONA_OC=SUL ASSIST_MED=SIM 29 ==> EVOL=ALTA 27 conf:(0.93)
428. TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 43 ==> EVOL=ALTA 40 conf:(0.93)
429. ZONA_RES=OESTE 26 ==> ASSIST_MED=SIM 24 conf:(0.92)
430. SEXO=M ZONA_OC=SUL ASSIST_MED=SIM 26 ==> EVOL=ALTA 24 conf:(0.92)
431. SEXO=M TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 38 ==> EVOL=ALTA 35
conf:(0.92)
432. SEXO=M TP_ACID=COLISÃO TRANSP_VIT=MOTOCICLETA 25 ==> ASSIST_MED=SIM 23
conf:(0.92)
433. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> EVOL=ALTA 57 conf:(0.92)
434. LC_ASSIST=HMBOMJESUS 24 ==> SEXO=M 22 conf:(0.92)
435. FX_ET=3 TRANSP_VIT=MOTOCICLETA 24 ==> SEXO=M 22 conf:(0.92)
436. TP_ACID=COLISÃO CAT_DIAS_INTER=2 24 ==> SEXO=M 22 conf:(0.92)
437. SEXO=M TP_ACID=SINF 24 ==> ASSIST_MED=SIM 22 conf:(0.92)
438. ASSIST_MED=SIM LC_ASSIST=HMBOMJESUS 24 ==> SEXO=M 22 conf:(0.92)

APÊNDICE

439. LC_ASSIST=HMBOMJESUS 24 ==> SEXO=M ASSIST_MED=SIM 22 conf:(0.92)
 440. LC_ASSIST=HMBOMJESUS EVOL=ALTA 24 ==> SEXO=M 22 conf:(0.92)
 441. LC_ASSIST=HMBOMJESUS 24 ==> SEXO=M EVOL=ALTA 22 conf:(0.92)
 442. FX_ET=3 TRANSP_VIT=MOTOCICLETA 24 ==> ASSIST_MED=SIM 22 conf:(0.92)
 443. ZONA_RES=LESTE ASSIST_MED=SIM 24 ==> EVOL=ALTA 22 conf:(0.92)
 444. TP_ACID=COLISÃO TP_VIT=IGNORADO 24 ==> ASSIST_MED=SIM 22 conf:(0.92)
 445. TP_ACID=COLISÃO ASSIST_MED=SIM CAT_DIAS_INTER=2 24 ==> SEXO=M 22 conf:(0.92)
 446. TP_ACID=COLISÃO CAT_DIAS_INTER=2 24 ==> SEXO=M ASSIST_MED=SIM 22 conf:(0.92)
 447. ASSIST_MED=SIM LC_ASSIST=HMBOMJESUS EVOL=ALTA 24 ==> SEXO=M 22 conf:(0.92)
 448. LC_ASSIST=HMBOMJESUS EVOL=ALTA 24 ==> SEXO=M ASSIST_MED=SIM 22 conf:(0.92)
 449. ASSIST_MED=SIM LC_ASSIST=HMBOMJESUS 24 ==> SEXO=M EVOL=ALTA 22 conf:(0.92)
 450. LC_ASSIST=HMBOMJESUS 24 ==> SEXO=M ASSIST_MED=SIM EVOL=ALTA 22 conf:(0.92)
 451. ZONA_RES=SUL EVOL=ALTA 23 ==> SEXO=M 21 conf:(0.91)
 452. ZONA_RES=SUL ASSIST_MED=SIM EVOL=ALTA 23 ==> SEXO=M 21 conf:(0.91)
 453. ZONA_RES=SUL EVOL=ALTA 23 ==> SEXO=M ASSIST_MED=SIM 21 conf:(0.91)
 454. TP_ACID=COLISÃO EVOL=ALTA CAT_DIAS_INTER=2 23 ==> SEXO=M 21 conf:(0.91)
 455. TP_ACID=COLISÃO ASSIST_MED=SIM EVOL=ALTA CAT_DIAS_INTER=2 23 ==> SEXO=M 21 conf:(0.91)
 456. TP_ACID=COLISÃO EVOL=ALTA CAT_DIAS_INTER=2 23 ==> SEXO=M ASSIST_MED=SIM 21 conf:(0.91)
 457. TP_VIT=PASSAGEIRO 34 ==> SEXO=M 31 conf:(0.91)
 458. ASSIST_MED=SIM 121 ==> EVOL=ALTA 110 conf:(0.91)
 459. SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM 55 ==> EVOL=ALTA 50 conf:(0.91)
 460. SEXO=M TP_VIT=IGNORADO 33 ==> ASSIST_MED=SIM 30 conf:(0.91)
 461. FX_ET=5 22 ==> ASSIST_MED=SIM 20 conf:(0.91)
 462. SEXO=M ZONA_RES=OESTE 22 ==> EVOL=ALTA 20 conf:(0.91)
 463. TP_ACID=COLISÃO LC_ASSIST=HPSMCBA 22 ==> SEXO=M 20 conf:(0.91)
 464. SEXO=M ZONA_RES=OESTE 22 ==> ASSIST_MED=SIM EVOL=ALTA 20 conf:(0.91)
 465. TP_ACID=COLISÃO ASSIST_MED=SIM LC_ASSIST=HPSMCBA 22 ==> SEXO=M 20 conf:(0.91)
 466. TP_ACID=COLISÃO LC_ASSIST=HPSMCBA 22 ==> SEXO=M ASSIST_MED=SIM 20 conf:(0.91)
 467. FX_ET=3 TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 22 ==> EVOL=ALTA 20 conf:(0.91)
 468. TP_ACID=COLISÃO TP_VIT=CONDUTOR ASSIST_MED=SIM 22 ==> EVOL=ALTA 20 conf:(0.91)
 469. TRANSP_VIT=VEICULO ASSIST_MED=SIM 32 ==> EVOL=ALTA 29 conf:(0.91)
 470. SEXO=M ZONA_RES=SINF 21 ==> ASSIST_MED=SIM 19 conf:(0.9)
 471. SEXO=M ZONA_RES=SINF 21 ==> EVOL=ALTA 19 conf:(0.9)
 472. TP_ACID=COLISÃO TP_VIT=PASSAGEIRO 21 ==> SEXO=M 19 conf:(0.9)
 473. SEXO=M ZONA_RES=SINF 21 ==> ASSIST_MED=SIM EVOL=ALTA 19 conf:(0.9)
 474. SEXO=M FX_ET=3 TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 21 ==> EVOL=ALTA 19 conf:(0.9)
 475. FX_ET=3 ASSIST_MED=SIM 52 ==> EVOL=ALTA 47 conf:(0.9)
 476. SEXO=M TP_VIT=PASSAGEIRO 31 ==> ASSIST_MED=SIM 28 conf:(0.9)
 477. TP_VIT=IGNORADO 41 ==> ASSIST_MED=SIM 37 conf:(0.9)
 478. SEXO=M FX_ET=3 51 ==> ASSIST_MED=SIM 46 conf:(0.9)
 479. SEXO=M TP_VIT=CONDUTOR 40 ==> ASSIST_MED=SIM 36 conf:(0.9)
 480. ZONA_OC=SINF 30 ==> EVOL=ALTA 27 conf:(0.9)
 481. ZONA_OC=SINF 30 ==> ASSIST_MED=SIM EVOL=ALTA 27 conf:(0.9)

APÊNDICE

482. SEXO=M FX_ET=4 20 ==> ASSIST_MED=SIM 18 conf:(0.9)
 483. ZONA_OC=OESTE TP_ACID=COLISÃO 20 ==> ASSIST_MED=SIM 18 conf:(0.9)
 484. TP_ACID=COLISÃO TRANSP_VIT=VEICULO 20 ==> ASSIST_MED=SIM 18 conf:(0.9)
 485. TP_ACID=COLISÃO TRANSP_VIT=VEICULO 20 ==> EVOL=ALTA 18 conf:(0.9)
 486. FX_ET=3 TP_VIT=CONDUTOR ASSIST_MED=SIM 20 ==> SEXO=M 18 conf:(0.9)
 487. SEXO=M FX_ET=3 TP_VIT=CONDUTOR 20 ==> ASSIST_MED=SIM 18 conf:(0.9)
 488. SEXO=M TP_ACID=COLISÃO TP_VIT=IGNORADO 20 ==> EVOL=ALTA 18 conf:(0.9)
 489. TP_ACID=COLISÃO TRANSP_VIT=VEICULO 20 ==> ASSIST_MED=SIM EVOL=ALTA 18
 conf:(0.9)
 490. SEXO=M TP_ACID=COLISÃO TP_VIT=IGNORADO 20 ==> ASSIST_MED=SIM EVOL=ALTA
 18 conf:(0.9)

Segundo Modelo

Apriori

=====

Minimum support: 0.2 (28 instances)

Minimum metric <confidence>: 0.8

Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 22

Size of set of large itemsets L(2): 44

Size of set of large itemsets L(3): 32

Size of set of large itemsets L(4): 8

Best rules found:

1. EVOL=ALTA 110 ==> ASSIST_MED=SIM 110 conf:(1)
2. SEXO=M EVOL=ALTA 89 ==> ASSIST_MED=SIM 89 conf:(1)
3. TP_ACID=COLISÃO EVOL=ALTA 57 ==> ASSIST_MED=SIM 57 conf:(1)
4. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM 51 conf:(1)
5. SEXO=M TP_ACID=COLISÃO EVOL=ALTA 50 ==> ASSIST_MED=SIM 50 conf:(1)
6. FX_ET=3 EVOL=ALTA 47 ==> ASSIST_MED=SIM 47 conf:(1)
7. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM 45 conf:(1)
8. LC_ASSIST=HPSMCBA EVOL=ALTA 44 ==> ASSIST_MED=SIM 44 conf:(1)
9. EVOL=ALTA CAT_DIAS_INTER=1 44 ==> ASSIST_MED=SIM 44 conf:(1)
10. SEXO=M LC_ASSIST=HPSMCBA 41 ==> ASSIST_MED=SIM 41 conf:(1)
11. SEXO=M FX_ET=3 EVOL=ALTA 41 ==> ASSIST_MED=SIM 41 conf:(1)
12. TRANSP_VIT=MOTOCICLETA EVOL=ALTA 40 ==> ASSIST_MED=SIM 40 conf:(1)
13. CAT_DIAS_INTER=2 39 ==> ASSIST_MED=SIM 39 conf:(1)
14. EVOL=ALTA CAT_DIAS_INTER=2 38 ==> ASSIST_MED=SIM 38 conf:(1)
15. TP_VIT=CONDUTOR EVOL=ALTA 37 ==> ASSIST_MED=SIM 37 conf:(1)
16. TP_VIT=IGNORADO EVOL=ALTA 35 ==> ASSIST_MED=SIM 35 conf:(1)
17. SEXO=M TRANSP_VIT=MOTOCICLETA EVOL=ALTA 35 ==> ASSIST_MED=SIM 35 conf:(1)
18. SEXO=M CAT_DIAS_INTER=1 34 ==> ASSIST_MED=SIM 34 conf:(1)
19. SEXO=M CAT_DIAS_INTER=2 34 ==> ASSIST_MED=SIM 34 conf:(1)
20. SEXO=M LC_ASSIST=HPSMCBA EVOL=ALTA 34 ==> ASSIST_MED=SIM 34 conf:(1)
21. SEXO=M EVOL=ALTA CAT_DIAS_INTER=1 33 ==> ASSIST_MED=SIM 33 conf:(1)
22. SEXO=M EVOL=ALTA CAT_DIAS_INTER=2 33 ==> ASSIST_MED=SIM 33 conf:(1)
23. SEXO=M TP_VIT=CONDUTOR EVOL=ALTA 31 ==> ASSIST_MED=SIM 31 conf:(1)
24. CAT_DIAS_INTER=3 29 ==> ASSIST_MED=SIM 29 conf:(1)

APÊNDICE

25. TRANSP_VIT=VEICULO EVOL=ALTA 29 ==> ASSIST_MED=SIM 29 conf:(1)
 26. ZONA_OC=OESTE EVOL=ALTA 28 ==> ASSIST_MED=SIM 28 conf:(1)
 27. EVOL=ALTA CAT_DIAS_INTER=3 28 ==> ASSIST_MED=SIM 28 conf:(1)
 28. SEXO=M TP_VIT=IGNORADO EVOL=ALTA 28 ==> ASSIST_MED=SIM 28 conf:(1)
 29. CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44 conf:(0.98)
 30. ASSIST_MED=SIM CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44 conf:(0.98)
 31. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM EVOL=ALTA 44 conf:(0.98)
 32. CAT_DIAS_INTER=2 39 ==> EVOL=ALTA 38 conf:(0.97)
 33. ASSIST_MED=SIM CAT_DIAS_INTER=2 39 ==> EVOL=ALTA 38 conf:(0.97)
 34. CAT_DIAS_INTER=2 39 ==> ASSIST_MED=SIM EVOL=ALTA 38 conf:(0.97)
 35. SEXO=M CAT_DIAS_INTER=1 34 ==> EVOL=ALTA 33 conf:(0.97)
 36. SEXO=M CAT_DIAS_INTER=2 34 ==> EVOL=ALTA 33 conf:(0.97)
 37. SEXO=M ASSIST_MED=SIM CAT_DIAS_INTER=1 34 ==> EVOL=ALTA 33 conf:(0.97)
 38. SEXO=M CAT_DIAS_INTER=1 34 ==> ASSIST_MED=SIM EVOL=ALTA 33 conf:(0.97)
 39. SEXO=M ASSIST_MED=SIM CAT_DIAS_INTER=2 34 ==> EVOL=ALTA 33 conf:(0.97)
 40. SEXO=M CAT_DIAS_INTER=2 34 ==> ASSIST_MED=SIM EVOL=ALTA 33 conf:(0.97)
 41. CAT_DIAS_INTER=3 29 ==> EVOL=ALTA 28 conf:(0.97)
 42. TP_VIT=PASSAGEIRO ASSIST_MED=SIM 29 ==> SEXO=M 28 conf:(0.97)
 43. SEXO=M TRANSP_VIT=VEICULO 29 ==> ASSIST_MED=SIM 28 conf:(0.97)
 44. ASSIST_MED=SIM CAT_DIAS_INTER=3 29 ==> EVOL=ALTA 28 conf:(0.97)
 45. CAT_DIAS_INTER=3 29 ==> ASSIST_MED=SIM EVOL=ALTA 28 conf:(0.97)
 46. SEXO=M TRANSP_VIT=MOTOCICLETA 40 ==> ASSIST_MED=SIM 38 conf:(0.95)
 47. TP_VIT=IGNORADO ASSIST_MED=SIM 37 ==> EVOL=ALTA 35 conf:(0.95)
 48. TRANSP_VIT=VEICULO 34 ==> ASSIST_MED=SIM 32 conf:(0.94)
 49. TRANSP_VIT=MOTOCICLETA 46 ==> ASSIST_MED=SIM 43 conf:(0.93)
 50. ZONA_OC=SINF 30 ==> ASSIST_MED=SIM 28 conf:(0.93)
 51. SEXO=M ZONA_OC=OESTE 30 ==> ASSIST_MED=SIM 28 conf:(0.93)
 52. SEXO=M TP_VIT=IGNORADO ASSIST_MED=SIM 30 ==> EVOL=ALTA 28 conf:(0.93)
 53. TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 43 ==> EVOL=ALTA 40 conf:(0.93)
 54. SEXO=M TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 38 ==> EVOL=ALTA 35
 conf:(0.92)
 55. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> EVOL=ALTA 57 conf:(0.92)
 56. TP_VIT=PASSAGEIRO 34 ==> SEXO=M 31 conf:(0.91)
 57. ASSIST_MED=SIM 121 ==> EVOL=ALTA 110 conf:(0.91)
 58. SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM 55 ==> EVOL=ALTA 50 conf:(0.91)
 59. SEXO=M TP_VIT=IGNORADO 33 ==> ASSIST_MED=SIM 30 conf:(0.91)
 60. TRANSP_VIT=VEICULO ASSIST_MED=SIM 32 ==> EVOL=ALTA 29 conf:(0.91)
 61. FX_ET=3 ASSIST_MED=SIM 52 ==> EVOL=ALTA 47 conf:(0.9)
 62. SEXO=M TP_VIT=PASSAGEIRO 31 ==> ASSIST_MED=SIM 28 conf:(0.9)
 63. TP_VIT=IGNORADO 41 ==> ASSIST_MED=SIM 37 conf:(0.9)
 64. SEXO=M FX_ET=3 51 ==> ASSIST_MED=SIM 46 conf:(0.9)
 65. SEXO=M TP_VIT=CONDUTOR 40 ==> ASSIST_MED=SIM 36 conf:(0.9)
 66. SEXO=M ASSIST_MED=SIM 99 ==> EVOL=ALTA 89 conf:(0.9)
 67. ZONA_OC=OESTE 39 ==> ASSIST_MED=SIM 35 conf:(0.9)
 68. FX_ET=3 58 ==> ASSIST_MED=SIM 52 conf:(0.9)
 69. SEXO=M FX_ET=3 ASSIST_MED=SIM 46 ==> EVOL=ALTA 41 conf:(0.89)
 70. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> SEXO=M 55 conf:(0.89)
 71. SEXO=M TP_ACID=COLISÃO 62 ==> ASSIST_MED=SIM 55 conf:(0.89)
 72. FX_ET=3 ASSIST_MED=SIM 52 ==> SEXO=M 46 conf:(0.88)
 73. SEXO=M 112 ==> ASSIST_MED=SIM 99 conf:(0.88)
 74. TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 43 ==> SEXO=M 38 conf:(0.88)
 75. ZONA_RES=SUL 34 ==> SEXO=M 30 conf:(0.88)
 76. FX_ET=3 58 ==> SEXO=M 51 conf:(0.88)
 77. ZONA_OC=SUL 33 ==> SEXO=M 29 conf:(0.88)
 78. ZONA_OC=SUL 33 ==> ASSIST_MED=SIM 29 conf:(0.88)
 79. FX_ET=3 TP_ACID=COLISÃO 33 ==> ASSIST_MED=SIM 29 conf:(0.88)

APÊNDICE

80. TP_VIT=CONDUTOR 49 ==> ASSIST_MED=SIM 43 conf:(0.88)
81. TP_ACID=COLISÃO EVOL=ALTA 57 ==> SEXO=M 50 conf:(0.88)
82. TP_ACID=COLISÃO ASSIST_MED=SIM EVOL=ALTA 57 ==> SEXO=M 50 conf:(0.88)
83. TP_ACID=COLISÃO EVOL=ALTA 57 ==> SEXO=M ASSIST_MED=SIM 50 conf:(0.88)
84. TRANSP_VIT=MOTOCICLETA EVOL=ALTA 40 ==> SEXO=M 35 conf:(0.88)
85. SEXO=M TRANSP_VIT=MOTOCICLETA 40 ==> EVOL=ALTA 35 conf:(0.88)
86. TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM EVOL=ALTA 40 ==> SEXO=M 35 conf:(0.88)
87. TRANSP_VIT=MOTOCICLETA EVOL=ALTA 40 ==> SEXO=M ASSIST_MED=SIM 35 conf:(0.88)
88. SEXO=M TRANSP_VIT=MOTOCICLETA 40 ==> ASSIST_MED=SIM EVOL=ALTA 35 conf:(0.88)
89. TRANSP_VIT=VEICULO ASSIST_MED=SIM 32 ==> SEXO=M 28 conf:(0.88)
90. FX_ET=3 EVOL=ALTA 47 ==> SEXO=M 41 conf:(0.87)
91. FX_ET=3 ASSIST_MED=SIM EVOL=ALTA 47 ==> SEXO=M 41 conf:(0.87)
92. FX_ET=3 EVOL=ALTA 47 ==> SEXO=M ASSIST_MED=SIM 41 conf:(0.87)
93. CAT_DIAS_INTER=2 39 ==> SEXO=M 34 conf:(0.87)
94. ASSIST_MED=SIM CAT_DIAS_INTER=2 39 ==> SEXO=M 34 conf:(0.87)
95. CAT_DIAS_INTER=2 39 ==> SEXO=M ASSIST_MED=SIM 34 conf:(0.87)
96. TRANSP_VIT=MOTOCICLETA 46 ==> SEXO=M 40 conf:(0.87)
97. TRANSP_VIT=MOTOCICLETA 46 ==> EVOL=ALTA 40 conf:(0.87)
98. TRANSP_VIT=MOTOCICLETA 46 ==> ASSIST_MED=SIM EVOL=ALTA 40 conf:(0.87)
99. EVOL=ALTA CAT_DIAS_INTER=2 38 ==> SEXO=M 33 conf:(0.87)
100. ASSIST_MED=SIM EVOL=ALTA CAT_DIAS_INTER=2 38 ==> SEXO=M 33 conf:(0.87)
101. EVOL=ALTA CAT_DIAS_INTER=2 38 ==> SEXO=M ASSIST_MED=SIM 33 conf:(0.87)
102. LC_ASSIST=HPSMCBA 51 ==> EVOL=ALTA 44 conf:(0.86)
103. ASSIST_MED=SIM LC_ASSIST=HPSMCBA 51 ==> EVOL=ALTA 44 conf:(0.86)
104. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM EVOL=ALTA 44 conf:(0.86)
105. TP_ACID=COLISÃO 72 ==> SEXO=M 62 conf:(0.86)
106. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM 62 conf:(0.86)
107. SEXO=M TP_VIT=CONDUTOR ASSIST_MED=SIM 36 ==> EVOL=ALTA 31 conf:(0.86)
108. TP_VIT=CONDUTOR ASSIST_MED=SIM 43 ==> EVOL=ALTA 37 conf:(0.86)
109. TP_VIT=IGNORADO 41 ==> EVOL=ALTA 35 conf:(0.85)
110. TP_VIT=IGNORADO 41 ==> ASSIST_MED=SIM EVOL=ALTA 35 conf:(0.85)
111. TRANSP_VIT=VEICULO 34 ==> SEXO=M 29 conf:(0.85)
112. TP_VIT=PASSAGEIRO 34 ==> ASSIST_MED=SIM 29 conf:(0.85)
113. TRANSP_VIT=VEICULO 34 ==> EVOL=ALTA 29 conf:(0.85)
114. TRANSP_VIT=VEICULO 34 ==> ASSIST_MED=SIM EVOL=ALTA 29 conf:(0.85)
115. SEXO=M TP_VIT=IGNORADO 33 ==> EVOL=ALTA 28 conf:(0.85)
116. SEXO=M TP_VIT=IGNORADO 33 ==> ASSIST_MED=SIM EVOL=ALTA 28 conf:(0.85)
117. CAT_DIAS_INTER=2 39 ==> SEXO=M EVOL=ALTA 33 conf:(0.85)
118. ASSIST_MED=SIM CAT_DIAS_INTER=2 39 ==> SEXO=M EVOL=ALTA 33 conf:(0.85)
119. CAT_DIAS_INTER=2 39 ==> SEXO=M ASSIST_MED=SIM EVOL=ALTA 33 conf:(0.85)
120. TP_VIT=CONDUTOR EVOL=ALTA 37 ==> SEXO=M 31 conf:(0.84)
121. TP_VIT=CONDUTOR ASSIST_MED=SIM EVOL=ALTA 37 ==> SEXO=M 31 conf:(0.84)
122. TP_VIT=CONDUTOR EVOL=ALTA 37 ==> SEXO=M ASSIST_MED=SIM 31 conf:(0.84)
123. TP_VIT=CONDUTOR ASSIST_MED=SIM 43 ==> SEXO=M 36 conf:(0.84)
124. SEXO=M LC_ASSIST=HPSMCBA 41 ==> EVOL=ALTA 34 conf:(0.83)
125. SEXO=M ASSIST_MED=SIM LC_ASSIST=HPSMCBA 41 ==> EVOL=ALTA 34 conf:(0.83)
126. SEXO=M LC_ASSIST=HPSMCBA 41 ==> ASSIST_MED=SIM EVOL=ALTA 34 conf:(0.83)
127. TRANSP_VIT=MOTOCICLETA 46 ==> SEXO=M ASSIST_MED=SIM 38 conf:(0.83)
128. ZONA_RES=SUL 34 ==> ASSIST_MED=SIM 28 conf:(0.82)
129. TP_VIT=PASSAGEIRO 34 ==> SEXO=M ASSIST_MED=SIM 28 conf:(0.82)
130. TRANSP_VIT=VEICULO 34 ==> SEXO=M ASSIST_MED=SIM 28 conf:(0.82)
131. ASSIST_MED=SIM 121 ==> SEXO=M 99 conf:(0.82)
132. TP_VIT=CONDUTOR 49 ==> SEXO=M 40 conf:(0.82)

APÊNDICE

133. TRANSP_VIT=MOTOCICLETA ASSIST_MED=SIM 43 ==> SEXO=M EVOL=ALTA 35
conf:(0.81)
134. TP_VIT=IGNORADO ASSIST_MED=SIM 37 ==> SEXO=M 30 conf:(0.81)
135. FX_ET=3 58 ==> EVOL=ALTA 47 conf:(0.81)
136. FX_ET=3 58 ==> ASSIST_MED=SIM EVOL=ALTA 47 conf:(0.81)
137. EVOL=ALTA 110 ==> SEXO=M 89 conf:(0.81)
138. ASSIST_MED=SIM EVOL=ALTA 110 ==> SEXO=M 89 conf:(0.81)
139. EVOL=ALTA 110 ==> SEXO=M ASSIST_MED=SIM 89 conf:(0.81)
140. SEXO=M TP_ACID=COLISÃO 62 ==> EVOL=ALTA 50 conf:(0.81)
141. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> SEXO=M EVOL=ALTA 50 conf:(0.81)
142. SEXO=M TP_ACID=COLISÃO 62 ==> ASSIST_MED=SIM EVOL=ALTA 50 conf:(0.81)
143. TP_VIT=IGNORADO 41 ==> SEXO=M 33 conf:(0.8)
144. LC_ASSIST=HPSMCBA 51 ==> SEXO=M 41 conf:(0.8)
145. SEXO=M FX_ET=3 51 ==> EVOL=ALTA 41 conf:(0.8)
146. ASSIST_MED=SIM LC_ASSIST=HPSMCBA 51 ==> SEXO=M 41 conf:(0.8)
147. LC_ASSIST=HPSMCBA 51 ==> SEXO=M ASSIST_MED=SIM 41 conf:(0.8)
148. SEXO=M FX_ET=3 51 ==> ASSIST_MED=SIM EVOL=ALTA 41 conf:(0.8)
149. ZONA_OC=OESTE ASSIST_MED=SIM 35 ==> SEXO=M 28 conf:(0.8)
150. TP_VIT=IGNORADO EVOL=ALTA 35 ==> SEXO=M 28 conf:(0.8)
151. ZONA_OC=OESTE ASSIST_MED=SIM 35 ==> EVOL=ALTA 28 conf:(0.8)
152. TP_VIT=IGNORADO ASSIST_MED=SIM EVOL=ALTA 35 ==> SEXO=M 28 conf:(0.8)
153. TP_VIT=IGNORADO EVOL=ALTA 35 ==> SEXO=M ASSIST_MED=SIM 28 conf:(0.8)

Terceiro Modelo

Apriori

=====

Minimum support: 0.3 (42 instances)

Minimum metric <confidence>: 0.7

Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 9

Size of set of large itemsets L(2): 15

Size of set of large itemsets L(3): 8

Size of set of large itemsets L(4): 1

Best rules found:

1. EVOL=ALTA 110 ==> ASSIST_MED=SIM 110 conf:(1)
2. SEXO=M EVOL=ALTA 89 ==> ASSIST_MED=SIM 89 conf:(1)
3. TP_ACID=COLISÃO EVOL=ALTA 57 ==> ASSIST_MED=SIM 57 conf:(1)
4. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM 51 conf:(1)
5. SEXO=M TP_ACID=COLISÃO EVOL=ALTA 50 ==> ASSIST_MED=SIM 50 conf:(1)
6. FX_ET=3 EVOL=ALTA 47 ==> ASSIST_MED=SIM 47 conf:(1)
7. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM 45 conf:(1)
8. LC_ASSIST=HPSMCBA EVOL=ALTA 44 ==> ASSIST_MED=SIM 44 conf:(1)
9. EVOL=ALTA CAT_DIAS_INTER=1 44 ==> ASSIST_MED=SIM 44 conf:(1)
10. CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44 conf:(0.98)
11. ASSIST_MED=SIM CAT_DIAS_INTER=1 45 ==> EVOL=ALTA 44 conf:(0.98)
12. CAT_DIAS_INTER=1 45 ==> ASSIST_MED=SIM EVOL=ALTA 44 conf:(0.98)
13. TRANSP_VIT=MOTOCICLETA 46 ==> ASSIST_MED=SIM 43 conf:(0.93)

APÊNDICE

14. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> EVOL=ALTA 57 conf:(0.92)
15. ASSIST_MED=SIM 121 ==> EVOL=ALTA 110 conf:(0.91)
16. SEXO=M TP_ACID=COLISÃO ASSIST_MED=SIM 55 ==> EVOL=ALTA 50 conf:(0.91)
17. FX_ET=3 ASSIST_MED=SIM 52 ==> EVOL=ALTA 47 conf:(0.9)
18. SEXO=M FX_ET=3 51 ==> ASSIST_MED=SIM 46 conf:(0.9)
19. SEXO=M ASSIST_MED=SIM 99 ==> EVOL=ALTA 89 conf:(0.9)
20. FX_ET=3 58 ==> ASSIST_MED=SIM 52 conf:(0.9)
21. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> SEXO=M 55 conf:(0.89)
22. SEXO=M TP_ACID=COLISÃO 62 ==> ASSIST_MED=SIM 55 conf:(0.89)
23. FX_ET=3 ASSIST_MED=SIM 52 ==> SEXO=M 46 conf:(0.88)
24. SEXO=M 112 ==> ASSIST_MED=SIM 99 conf:(0.88)
25. FX_ET=3 58 ==> SEXO=M 51 conf:(0.88)
26. TP_VIT=CONDUTOR 49 ==> ASSIST_MED=SIM 43 conf:(0.88)
27. TP_ACID=COLISÃO EVOL=ALTA 57 ==> SEXO=M 50 conf:(0.88)
28. TP_ACID=COLISÃO ASSIST_MED=SIM EVOL=ALTA 57 ==> SEXO=M 50 conf:(0.88)
29. TP_ACID=COLISÃO EVOL=ALTA 57 ==> SEXO=M ASSIST_MED=SIM 50 conf:(0.88)
30. LC_ASSIST=HPSMCBA 51 ==> EVOL=ALTA 44 conf:(0.86)
31. ASSIST_MED=SIM LC_ASSIST=HPSMCBA 51 ==> EVOL=ALTA 44 conf:(0.86)
32. LC_ASSIST=HPSMCBA 51 ==> ASSIST_MED=SIM EVOL=ALTA 44 conf:(0.86)
33. TP_ACID=COLISÃO 72 ==> SEXO=M 62 conf:(0.86)
34. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM 62 conf:(0.86)
35. ASSIST_MED=SIM 121 ==> SEXO=M 99 conf:(0.82)
36. FX_ET=3 58 ==> EVOL=ALTA 47 conf:(0.81)
37. FX_ET=3 58 ==> ASSIST_MED=SIM EVOL=ALTA 47 conf:(0.81)
38. EVOL=ALTA 110 ==> SEXO=M 89 conf:(0.81)
39. ASSIST_MED=SIM EVOL=ALTA 110 ==> SEXO=M 89 conf:(0.81)
40. EVOL=ALTA 110 ==> SEXO=M ASSIST_MED=SIM 89 conf:(0.81)
41. SEXO=M TP_ACID=COLISÃO 62 ==> EVOL=ALTA 50 conf:(0.81)
42. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> SEXO=M EVOL=ALTA 50 conf:(0.81)
43. SEXO=M TP_ACID=COLISÃO 62 ==> ASSIST_MED=SIM EVOL=ALTA 50 conf:(0.81)
44. SEXO=M 112 ==> EVOL=ALTA 89 conf:(0.79)
45. SEXO=M 112 ==> ASSIST_MED=SIM EVOL=ALTA 89 conf:(0.79)
46. FX_ET=3 58 ==> SEXO=M ASSIST_MED=SIM 46 conf:(0.79)
47. TP_ACID=COLISÃO 72 ==> EVOL=ALTA 57 conf:(0.79)
48. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM EVOL=ALTA 57 conf:(0.79)
49. TP_ACID=COLISÃO 72 ==> SEXO=M ASSIST_MED=SIM 55 conf:(0.76)
50. ASSIST_MED=SIM 121 ==> SEXO=M EVOL=ALTA 89 conf:(0.74)

Quarto Modelo

Apriori

=====

Minimum support: 0.4 (56 instances)

Minimum metric <confidence>: 0.7

Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 2

Best rules found:

APÊNDICE

1. EVOL=ALTA 110 ==> ASSIST_MED=SIM 110 conf:(1)
2. SEXO=M EVOL=ALTA 89 ==> ASSIST_MED=SIM 89 conf:(1)
3. TP_ACID=COLISÃO EVOL=ALTA 57 ==> ASSIST_MED=SIM 57 conf:(1)
4. TP_ACID=COLISÃO ASSIST_MED=SIM 62 ==> EVOL=ALTA 57 conf:(0.92)
5. ASSIST_MED=SIM 121 ==> EVOL=ALTA 110 conf:(0.91)
6. SEXO=M ASSIST_MED=SIM 99 ==> EVOL=ALTA 89 conf:(0.9)
7. SEXO=M 112 ==> ASSIST_MED=SIM 99 conf:(0.88)
8. TP_ACID=COLISÃO 72 ==> SEXO=M 62 conf:(0.86)
9. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM 62 conf:(0.86)
10. ASSIST_MED=SIM 121 ==> SEXO=M 99 conf:(0.82)
11. EVOL=ALTA 110 ==> SEXO=M 89 conf:(0.81)
12. ASSIST_MED=SIM EVOL=ALTA 110 ==> SEXO=M 89 conf:(0.81)
13. EVOL=ALTA 110 ==> SEXO=M ASSIST_MED=SIM 89 conf:(0.81)
14. SEXO=M 112 ==> EVOL=ALTA 89 conf:(0.79)
15. SEXO=M 112 ==> ASSIST_MED=SIM EVOL=ALTA 89 conf:(0.79)
16. TP_ACID=COLISÃO 72 ==> EVOL=ALTA 57 conf:(0.79)
17. TP_ACID=COLISÃO 72 ==> ASSIST_MED=SIM EVOL=ALTA 57 conf:(0.79)
18. ASSIST_MED=SIM 121 ==> SEXO=M EVOL=ALTA 89 conf:(0.74)

