

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE**

MARCELO ROSANO DALLAGASSA

**CONCEPÇÃO DE UMA METODOLOGIA PARA IDENTIFICAÇÃO DE
BENEFICIÁRIOS COM INDICATIVOS DE DIABETES MELLITUS TIPO 2**

CURITIBA

2009

MARCELO ROSANO DALLAGASSA

**CONCEPÇÃO DE UMA METODOLOGIA PARA IDENTIFICAÇÃO DE
BENEFICIÁRIOS COM INDICATIVOS DE DIABETES MELLITUS TIPO 2.**

Dissertação apresentada ao Programa de Pós-graduação em Tecnologia em Saúde da PUCPR como requisito parcial para obtenção do título de Mestre em Tecnologia em Saúde.

Linha de Pesquisa: Sistemas de Informação e Apoio à Decisão em Saúde

Orientadora: Dr^a. Sandra Honorato da Silva

Co-orientadora: Dr^a. Deborah Ribeiro Carvalho

CURITIBA

2009

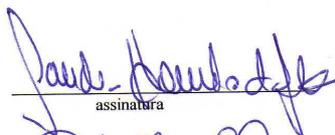
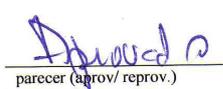
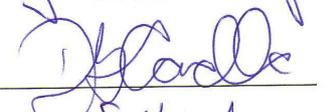
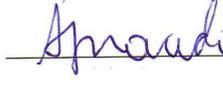
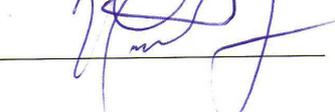
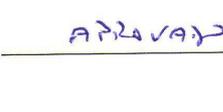


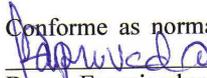
Pontifícia Universidade Católica do Paraná
 Centro de Ciências Biológicas e da Saúde
 Programa de Pós-Graduação em Tecnologia em Saúde

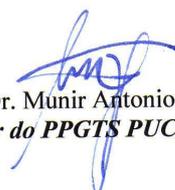
ATA DA SESSÃO PÚBLICA DE DEFESA DE DISSERTAÇÃO DE MESTRADO
 DO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE
 DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

DEFESA DE DISSERTAÇÃO Nº 106

Aos 31 de agosto de 2009 realizou-se a sessão pública de defesa da dissertação “**Concepção de uma Metodologia para Identificação de Beneficiários com Indicativos de Diabetes Mellitus Tipo 2**”, apresentada por **Marcelo Rosano Dallagassa** como requisito parcial para a obtenção do título de Mestre em Tecnologia em Saúde, – Área de Concentração – Informática em Saúde perante uma Banca Examinadora composta pelos seguintes membros:

Profª. Drª. Sandra Honorato da Silva, PUCPR (Orientadora)	 _____	 parecer (aprov/ reprov.)
Profª. Drª. Deborah Ribeiro Carvalho, UTP (Co-orientadora)	 _____	 _____
Prof. Dr. Edson Emílio Scalabrin, PUCPR/PPG1a	 _____	 _____
Prof. Dr. Edson Luiz Almeida Tizzot, UTFPR	 _____	 _____

Conforme as normas regimentais do PPGTS e da PUCPR, o trabalho apresentado foi considerado  (aprovado/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do Programa.


 Prof. Dr. Munir Antonio Gariba,
 Diretor do PPGTS PUCPR



DEDICATÓRIA

Em memória aos meus pais adotivos, Pedro Ronconi e Silta Dallagassa Ronconi,
pela formação moral e educacional para que eu pudesse chegar até esse momento;

Aos meus filhos Daniel e Matheus;

AGRADECIMENTOS

Os sinceros agradecimentos, à professora orientadora Dra. Sandra Honorato da Silva, pela sua paciência, orientações, troca de experiências e dedicação. A professora Dra. Deborah Ribeiro Carvalho pela sua sempre atenção e disponibilidade para a troca rica de discussões e experiências na área tecnológica.

A Diretoria da Unimed Federação do Estado do Paraná que subsidiou financeiramente a minha participação neste programa e em especial ao Dr. Faustino Garcia Alferez pelo apoio e incentivo, a minha gerência Sra. Cristina Leal de Castro, Rogério Goulart e Dr. Marlus Volney de Moraes pelas contribuições, incentivo e o aval técnico da minha participação neste programa e aos demais colegas da Unimed Federação do PR que sempre me apoiaram e auxiliaram nesse trabalho.

A coordenação do Programa PPGTS / PUC-PR, professores e demais funcionários, pela dedicação e paciência.

RESUMO

As empresas de saúde de uma maneira em geral, possuem uma grande quantidade de informações transacionadas e processadas em relação aos atendimentos realizados e com isto, uma grande dificuldade na extração de novos conhecimentos dessas bases de dados. A complexidade de interpretar e avaliar estas informações, demanda a utilização de ferramentas e técnicas que permitam uma análise mais automática e inteligente para a geração de novos conhecimentos. Este trabalho apresenta uma metodologia baseada no processo de descoberta de conhecimento em base de dados (KDD), proposto por Fayyat *et al* (1996), para a classificação de beneficiários da carteira de uma operadora de plano de saúde, por meio das informações dos procedimentos realizados, em relação ao indicativo de diabetes mellitus tipo 2. Trata-se de um estudo quantitativo, retrospectivo, de cunho descritivo. O estudo contemplou como etapas *Identificação das variáveis – Análise inicial - Seleção de dados para o Pré-processamento - Mineração de Dados - Validação e Interpretação dos Resultados (Validação Especialistas)*. Para a seleção das variáveis relevantes, realizou-se a exploração retrospectivo da base de dados, com 59 casos de internamento por diabetes mellitus tipo 2, e outros 59 casos com internamento clínico excluído o diabetes mellitus tipo 2. Comparou-se os atendimentos realizados no período de 5 anos e em conjunto com a literatura sobre o assunto e o apoio de um especialista obteve-se o quadro de variáveis. Para a etapa de pré processamento a amostra foi constituída por 60000 beneficiários que realizaram atendimentos num período de 6 anos na operadora . Para a etapa de mineração de dados, optou-se pelo algoritmo C4.5 proposto por Quinlan (1993) para a classificação dos registros e geração das regras. Selecionou-se 18 regras das 1106 encontradas, para a avaliação de 4 especialistas que por meio do IVC (índice de validade de conteúdo) obteve-se o nível de concordância proposto por Waltz *et al* (1991) para a validação do estudo. Desta maneira apresenta-se uma metodologia dirigida para subsidiar ações relacionadas a promoção e prevenção a saúde.

Palavras-chave: Diabetes Mellitus, Mineração de Dados, Descoberta do Conhecimento em Base de Dados

ABSTRACT

In a manner, healthcare companies generally have a great amount of traded and processed information in relation to accomplished services and with this, a great difficulty to extract new knowledge of those databases. The complexity of interpreting and evaluating these information, demand the use of tools and techniques that allow a more automatic and intelligent analysis to generate new knowledge. This paper presents a methodology based on the process knowledge discovery in databases (KDD), proposed by Fayyat et al (1996) for the classification of a health plan operator portfolio's beneficiaries, through the information of the procedures made in relation to the indicative of diabetes mellitus type 2. This is a quantitative and retrospective study, a descriptive. The study contemplated as stages: Identification of variables - Initial analysis - Selection of data for the Pre-processing - Data Mining - Validation and Interpretation of Results (Validation Experts). For the selection of relevant variables, there was the retrospective operation of the database, with 59 cases of hospitalization for diabetes mellitus type 2, and 59 other cases related to clinical excluding type 2 diabetes mellitus. It was compared the care provided in the period of 5 years and together with the literature on the subject and a specialist's support, the table of variables was obtained. For the stage of pre-processing the sample consisted of 60,000 beneficiaries who accomplished services in a period of 6 years in the operator. For the stage of data mining, it was chosen a C4.5 algorithm proposed by Quinlan (1993) for the classification of records and generation of rules. It was selected 18 of the 1106 rules found for the four experts' evaluation which through the IVC index (content validity) showed the level of agreement proposed by Waltz et al (1991) to validate the study. In this way it is presented a methodology directed to support programs related to promotion and preventive health.

Keywords: Diabetes Mellitus, Data Mining, KDD - Knowledge Discovery in Databases

LISTA DE FIGURAS

Figura 1 - Etapas do processo de descoberta de conhecimento	36
Figura 2 - Modelo dimensional	40
Figura 3 - Um típico modelo dimensional	41
Figura 4 – Arquitetura OLAM	42
Figura 5 – Representação de um processo lógico	44
Figura 6 – Estrutura de aprendizagem de máquina	45
Figura 7 – Exemplo árvore de decisão	47
Figura 8 – Matriz de confusão	48
Figura 9 – Avaliação retrospectiva dos atendimentos grupo DM2	53
Figura 10 – Avaliação retrospectiva dos atendimentos grupo controle	54
Figura 11 – Modelo dimensional atendimentos	55
Figura 12 – Estrutura árvore de decisão	69
Figura 13 – Metodologia proposta	87

LISTA DE TABELAS

Tabela 1 – Resultado matriz de confusão

69

LISTA DE GRÁFICOS

Gráfico 1 - Beneficiários do Grupo 1 com internamento por diabetes segundo idade	61
Gráfico 2 - Beneficiários do Grupo 2 com internamento clínico, segundo idade	61
Gráfico 3 - Análise comparativa das especialidades entre Grupo 1 e Grupo 2.	63
Gráfico 4 - Análise comparativa dos exames solicitados entre Grupo 1 e Grupo 2.	65

LISTA DE QUADROS

Quadro 1 - Mudanças estruturais do setor de saúde suplementar	23
Quadro 2 - Principais tarefas de mineração de dados	38
Quadro 3 – Consultas por especialidades do grupo 1 e 2	62
Quadro 4 - Exames solicitados a beneficiários do grupo 1 e 2	64
Quadro 5– Atributos selecionados para a mineração de dados	66
Quadro 6 – Estrutura do registro para a mineração de dados	68
Quadro 7 - Estrutura da tabela para ranqueamento das regras.	70
Quadro 8 – Resultado do ranqueamento e seleção das regras	71
Quadro 9 – Regras selecionadas para o processo de avaliação	72
Quadro 10 – Regras selecionadas classe sem indicativo	73
Quadro 11 – Regras selecionadas classe com indicativo	75
Quadro 12 – Regras selecionadas classe com forte indicativo	76
Quadro 13 - Regra da classe sem indicativo	79
Quadro 14 - Regra da classe com indicativo	82
Quadro 15 - Regra da classe com forte indicativo	84
Quadro 16 – Demonstrativo geral da validação das regras por especialistas	85

LISTA DE ABREVIATURAS E SIGLAS

ANS	Agência Nacional de Saúde Suplementar
CID10	Classificação internacional de doenças e problemas relacionados à saúde, décima revisão.
CEP	Comitê de Ética em Pesquisa
CONASS	Conselho Nacional de Secretários de Saúde
DCNT	Doenças Crônicas Não Transmissíveis
DM	Data Mining – Mineração de Dados
DM1	Diabetes Mellitus Tipo 1
DM2	Diabetes Mellitus Tipo 2
DW	<i>Data Warehouse</i>
IMC	Índice de Massa Corporal
KDD	<i>Knowledge Discovery in Data Bases</i> – Descoberta de Conhecimento em Base de Dados
OLAP	Processamento Analítico On-line
OLAM	Processamento Analítico On-line e Mineração de Dados
OLTP	Processamento Transações On-line
OMS	Organização Mundial da Saúde
PUCPR	Pontifícia Universidade Católica do Paraná
SBD	Sociedade Brasileira de Diabetes
SUS	Sistema Único de Saúde
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

SUMÁRIO

1 INTRODUÇÃO	16
1.1 OBJETIVOS	19
1.1.1 Objetivo geral	19
1.1.2 Objetivos específicos	20
1.2 ORGANIZAÇÃO DO TRABALHO	20
2 REVISÃO DA LITERATURA	22
2.1 SAÚDE E SAÚDE SUPLEMENTAR	22
2.1.1 Modelo assistencial: prevenção e promoção à saúde	25
2.1.2 Doença crônica não transmissíveis (DCNT)	27
2.2 BASE DE DADOS EM SAÚDE E INTELIGÊNCIA ARTIFICIAL	33
2.2.1 Descoberta de conhecimentos em base de dados	35
2.2.2 Representação do conhecimento	42
2.2.3 Aprendizagem de máquina	45
2.2.4 Algoritmo de árvore de decisão	46
2.2.5 Raciocínio baseado em regras	49
3 METODOLOGIA	51
3.1 ETAPAS DO ESTUDO	51
3.1.1 Identificação das variáveis – análise inicial	52
3.1.2 Pré-processamento	54
3.1.3 Mineração de dados	56
3.1.4 Avaliação e interpretação dos resultados	57
3.1.5 Validação das regras	58
3.2 ASPECTOS ÉTICOS	59
4 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS	60
4.1 ANÁLISE INICIAL – IDENTIFICAÇÃO DAS VARIÁVEIS	60
4.2 PRÉ-PROCESSAMENTO	67
4.3 MINERAÇÃO DE DADOS	68

4.4 VALIDAÇÃO DAS REGRAS	70
4.4.1 Seleção das regras	70
4.4.2 Avaliação dos especialistas	78
4.4.3 Índice de validade de conteúdo (IVC)	85
4.5 METODOLOGIA PROPOSTA	86
5 CONSIDERAÇÕES FINAIS	90
5.1 TRABALHOS FUTUROS	93
REFERÊNCIAS	94
ANEXOS	101
APÊNDICES	103

1 INTRODUÇÃO

O sistema de saúde no Brasil vive atualmente uma crise sem precedentes, tanto no setor público, Sistema Único de Saúde (SUS), quanto no setor privado constituído predominantemente pela Saúde Suplementar.

No setor privado, os segmentos da saúde suplementar estão sob tensão. Os beneficiários dos planos de saúde estão cada vez mais exigentes e buscando, cada vez mais, empresas que ofereçam os menores preços. A rede de atendimento faz críticas à baixa remuneração e as operadoras de planos de saúde, com raras exceções, vivem uma situação financeira abaixo do padrão de exigências para a segurança de suas atividades (MIRANDA, 2003).

Para Miranda (2003) como fatores que contribuem para este cenário, estão a retração do mercado, o excesso de controles burocráticos exigidos pela Agência Nacional de Saúde Suplementar (ANS) e a falta de ferramentas eficazes para o gerenciamento e controle de custos. Outros fatores como o aumento real do custo assistencial, o envelhecimento populacional, a incorporação de novas tecnologias na execução de procedimentos e a constatação de aumento na utilização dos serviços por parte dos beneficiários, concorrem para o agravamento da situação.

Neste contexto, as operadoras de planos de saúde se deparam com uma nova realidade econômica, em que o conhecimento tornou-se o insumo vital para a sua sobrevivência e a adoção de ferramentas eficazes de gerenciamento e controle para o apoio à tomada de decisões, tornam-se imprescindíveis para a sua sustentabilidade econômica (BRASIL, 2007).

Miranda (2003) enfatiza que o grande desafio encontra-se em desenvolver novos modelos de gestão para a saúde e que, as informações referentes aos atendimentos dos beneficiários, possam ser acompanhadas e prontamente analisadas pelos gestores, permitindo desta maneira, um acompanhamento mais eficaz em relação à gestão dos recursos.

Desenvolver um processo assistencial que contemple ações de promoção e prevenção e programas de gerenciamento de doenças e de casos permite, além da redução de custos, uma melhoria da qualidade de vida dos pacientes, influenciando

na redução das necessidades de atenção a saúde que demandem ações de maior complexidade assistencial.

Assim, as ações gerenciais devem prever e possibilitar a identificação dos pacientes para inserção em programas de promoção a saúde, prevenção de doenças, influenciando conseqüentemente na melhoria das suas condições de saúde e qualidade de vida e na redução dos gastos com assistência médica de alto custo (CZERESNIA,2003).

Kobus (2006) ressalta que os prestadores de serviços de saúde devem prover um melhor atendimento, por meio da expansão de dados de identificação, admissão ou de exame físico, coletados para a obtenção de um perfil mais detalhado do seu cliente e, conseqüentemente, o conhecimento das suas necessidades assistenciais.

Observa-se que as operadoras de plano de saúde, não possuem, via de regra, uma rede própria de atendimento, fato que obstaculiza a coleta das informações clínicas, coletando somente, os dados, relativos ao histórico dos atendimentos realizados pelo beneficiário, como: consultas realizadas em determinada especialidade, exames solicitados e realizados, atendimentos realizados por uma causa específica, além das informações cadastrais dos beneficiários como; sexo, idade, profissão e demais informações. O conjunto de dados gerado é armazenado nas bases de dados é grande, o que dificulta, muitas vezes seu tratamento, interpretação e transformação para a geração de novos conhecimentos.

Embora Porter; Teisberg (2007) ressaltem a importância em determinar o valor da assistência à saúde considerando a condição de saúde do paciente durante todo o ciclo de seu atendimento, e não medi-lo tomando como base um procedimento, serviço, consulta ou um exame isoladamente, esta ainda, não configura a realidade nas operadoras de saúde no Brasil, que conseqüentemente não cumprem um de seus papéis, que consiste em apoiar ativamente os pacientes nas suas escolhas, ajudá-los a percorrer pelo ciclo do atendimento e assisti-los no gerenciamento do seu histórico médico.

Assim, transita-se por um cenário de custos crescentes devido a tecnologias emergentes e alto índice de utilização dos beneficiários das operadoras de plano de saúde, que geram uma grande quantidade de dados com ausência, no entanto, de informações essenciais, o que torna imperiosa a necessidade de aplicação de

técnicas e métodos para descoberta de conhecimento em base de dados, para a identificação de beneficiários, com propensão a determinadas doenças, e favorecer a adoção e consolidação de um novo paradigma assistencial.

Considerando o potencial de obtenção de informações a partir das bases de dados no setor de saúde, é indicada a utilização de técnicas inteligentes para a descoberta de novos conhecimentos. Desta maneira, justifica-se a utilização de ferramentas de descoberta de conhecimentos em base de dados (KDD), que permitam detectar precocemente pacientes com propensão a determinadas doenças.

A informação é o elemento básico para a geração de novos conhecimentos que expressem resultados o mais próximo da sua origem e não se pautem apenas por indicadores financeiros, que muitas vezes expressam conhecimentos defasados em 30 a 60 dias, após o evento gerador, e que não representam os vetores de desempenho futuro da organização (KAPLAN, NORTON, 2001).

Vianna (2007), salienta que devido à grande quantidade de informações existentes nos serviços de saúde, a tecnologia da informação vem demonstrando sua utilidade com a nova geração de teorias e ferramentas computacionais.

Para Fayyad *et al* (1996), os sistemas de informação e as técnicas de gerenciamento de informações, como o KDD, caracterizam uma ferramenta de suporte para a análise de dados que permite a geração de novos conhecimentos em grande base de dados, utilizando a busca de padrões e extração de informações, condição adequada para uma base de saúde.

Neste contexto surge a necessidade da utilização da tecnologia da informação, como ferramenta essencial para o controle e monitoramento da utilização dos beneficiários, possibilitando o gerenciamento de ações, para a redução de custos e melhoria no atendimento dos seus clientes, sobremaneira no tangente às doenças crônicas não transmissíveis (DCNT).

Segundo a Organização Mundial de Saúde, estima-se que nos próximos dez anos os óbitos devido as DCNTs como as cardiovasculares, o câncer, o diabetes, as pulmonares obstrutivas crônicas e os transtornos mentais, devem aumentar em 17% e estima-se que haverá 64 milhões de óbitos em 2015, sendo 41 milhões em decorrência de uma DCNT (OMS,2005).

Nas últimas décadas no Brasil, as Doenças Crônicas Não Transmissíveis passaram a determinar a maioria das causas de óbito e incapacidade prematura, ultrapassando as taxas de mortalidade por doenças infecciosas e parasitárias e a representar uma grande parcela das despesas com assistência hospitalar no SUS e no Setor Suplementar. O impacto das DCNTs cresce continuamente e aumenta as dificuldades socioeconômicas, criando um ciclo vicioso com a pobreza, uma vez que impactam negativamente no desenvolvimento macroeconômico do país (BRASIL,2008).

Entre as DCNTs, o Diabetes Mellitus configura-se como uma doença em forte expansão mundial, tendo-se no envelhecimento da população, na urbanização crescente e na adoção de modos de vida pouco saudáveis, como o sedentarismo, a dieta inadequada e a obesidade, como fatores responsáveis pelo aumento da incidência e prevalência do diabetes em todo o mundo (BRASIL,2006).

Assim, este estudo propõe como questão norteadora: que metodologia possibilitaria a descoberta de novos conhecimentos em base de dados para a classificação de beneficiários em relação ao Diabetes Mellitus Tipo 2, na carteira de uma operadora de plano de saúde?

1.1 OBJETIVOS

1.1.1 Objetivo geral

Propor uma metodologia para a identificação de beneficiários com indicativo de Diabetes Mellitus Tipo 2, para o ingresso em programas de promoção da saúde, prevenção de doenças e gerenciamento de caso.

1.1.2 Objetivos específicos

- a) identificar as variáveis relevantes para o processo de descoberta de conhecimento, referente ao diabetes mellitus tipo 2;
- b) definir uma metodologia para o processo de classificação de beneficiários com indicativos a uma doença crônica não transmissível.
- c) validar as regras de produção selecionadas, com especialistas.

1.2 ORGANIZAÇÃO DO TRABALHO

Este trabalho apresenta no Capítulo 2 o referencial teórico abordando aspectos relevantes do segmento da saúde suplementar, a importância do modelo de promoção e prevenção à saúde e gerenciamento de caso, o panorama de doenças crônicas não transmissíveis, mais especificamente o Diabetes Mellitus. Neste mesmo capítulo, detalham-se os conceitos sobre as tecnologias disponíveis para o processo de apoio a tomada de decisões, o processo de descoberta de conhecimentos em base de dados KDD e o processo de mineração de dados, mais especificamente os algoritmos de árvore de decisões.

O Capítulo 3 apresenta a metodologia utilizada, enfocando o processo de classificação de beneficiário de uma operadora de plano de saúde em; sem indicativo, com indicativo e com forte indicativo para diabetes mellitus. Descreve cada um das etapas do processo de descoberta de conhecimento, apresentando toda a tarefa de exploração de dados para o elenco das variáveis importantes para o processo de mineração de dados.

No Capítulo 4 estão apresentados e discutidos os resultados relacionados ao desenvolvimento da metodologia e dos experimentos realizados para a definição das variáveis relevantes, bem como a avaliação das regras com especialistas e o cálculo do índice de concordância sobre o conhecimento gerado pelas regras do processo de mineração de dados.

O Capítulo 5 apresenta as considerações finais relacionadas à metodologia proposta e os trabalhos futuros.

2 REVISÃO DA LITERATURA

Este capítulo apresenta um panorama da saúde suplementar e as dificuldades do setor decorrentes do aumento dos custos assistenciais, as doenças crônico-degenerativas e o diabetes mellitus. Aborda ainda, questões relacionadas à utilização das ferramentas computacionais baseadas em inteligência artificial, para a extração de novos conhecimentos em base de dados, como ferramenta para identificação de clientes para ingresso em programas de promoção da saúde, prevenção de doenças e gerenciamento de caso.

2.1 SAÚDE E SAÚDE SUPLEMENTAR

O mundo moderno trouxe fortes mudanças nos hábitos de vida da população mundial, acarretando uma mudança significativa para a vida em sociedade.

Na última década percebe-se uma preocupação mais acentuada sobre o processo de cuidar em saúde, de modo a reduzir a vulnerabilidade ao adoecer e as chances de que o indivíduo seja produtor de incapacidades, de sofrimento crônico e de morte prematura (BRASIL, ANS, 2006).

No Brasil, conforme o Conselho Nacional de Secretários de Saúde (CONASS), no segmento das operadoras de planos de saúde existem: **as empresas de Medicina de Grupo** onde a gestão é realizada por uma empresa privada, **as empresas de Seguro de Saúde** onde o objeto da apólice é a de garantia de assistência médica, e **as Cooperativas Médicas** que são organizações formadas por profissionais vinculados, denominados cooperados. Estas cooperativas, inicialmente, atuavam apenas no atendimento a saúde e com o passar do tempo, passaram a administrar planos de saúde, da mesma forma que as empresas de medicina de grupo. Finalmente, **as empresas de autogestão** que configuram serviços voltados aos próprios funcionários, e muitas vezes seus familiares, podendo ser organizados pelos próprios funcionários ou pela própria empresa. (BRASIL, 2007).

Com a criação da ANS em novembro de 1999, pela medida provisória nº 1928, convertida na lei 9.961, como órgão de regulação, normatização, controle e fiscalização das atividades que garantem a assistência suplementar à saúde ocorreram diversas mudanças na estrutura do setor de saúde suplementar no Brasil, conforme CONASS, apresentado no quadro 1,(BRASIL, 2007).

	ANTES DA REGULAMENTAÇÃO Livre Atuação	DEPOIS DA REGULAMENTAÇÃO Atuação Controlada
Operadoras (empresas)	<ul style="list-style-type: none"> • Legislação do tipo societário • Controle deficiente 	<ul style="list-style-type: none"> • Autorização de funcionamento • Regras de operação, sujeitas a intervenção e liquidação • Exigência de garantias financeiras • Profissionalização da Gestão • Qualificação da atenção integral à saúde • Proibição da seleção de risco • Proibição da rescisão unilateral dos contratos
Assistência à saúde e acesso (produto)	<ul style="list-style-type: none"> • Livre definição da cobertura assistencial • Seleção de risco • Exclusão de usuários • Livre definição de carências • Livre definição de reajustes • <u>Modelo centrado na doença</u> • <u>Ausência de sistemas de informações</u> • Contratos nebulosos 	<ul style="list-style-type: none"> • Definição e limitação das carências • Reajustes controlados • Internação sem limites • <u>Modelo de atenção com ênfase nas ações de promoção à saúde e prevenção de doenças</u> • <u>Sistemas de informações como insumo estratégico</u> • Contratos mais transparentes

Quadro 1 - Mudanças estruturais do setor de saúde suplementar. Fonte: ANS,2007.

Neste cenário de mudanças ocorridas no setor da saúde suplementar percebe-se a importância do modelo focado nas ações de promoção à saúde e prevenção de doenças, e a necessidade das empresas possuírem sistemas de informações gerenciais, como insumo estratégico para auxílio e apoio à tomada de decisões.

O impacto da era da informação revolucionou mais o setor de serviços do que o segmento industrial. Muitas empresas do setor de serviços, em especial as empresas do setor de assistência médica, conviveram por muitas décadas com um

confortável ambiente não-competitivo. Nas duas últimas décadas ocorreram iniciativas de privatização e desregulamentação, à medida que a tecnologia da informação gerou as “sementes da destruição”, tornando-se um elemento vital, para a sobrevivência das empresas de serviços regulamentadas (KAPLAN; NORTON,1997).

Este ambiente da era da informação exige novas capacidades para assegurar o sucesso competitivo. A capacidade de mobilização e exploração dos ativos intangíveis ou invisíveis tornou-se muito mais decisiva, do que investir e gerenciar ativos físicos tangíveis. Os ativos intangíveis segundo Kaplan e Norton (1997), permitem que uma empresa:

- a) desenvolva relacionamentos que fortaleçam a fidelização do cliente e que se expandam novos segmentos de clientes;
- b) promova o lançamento de novos produtos e serviços inovadores que atendam e superem a expectativa dos clientes;
- c) desenvolva a customização de produtos e serviços de alta qualidade e preços baixos e com ciclos de produção mais curtos;
- d) mobilize as habilidades e motivação dos funcionários para a melhoria contínua do processo, qualidade e os tempos de resposta;
- e) utilize a tecnologia da informação.

Miranda (2003) comenta sobre esse modelo de assistência à saúde, focado na doença e não na saúde e enfatiza que segundo a Organização Mundial da Saúde, dentre os diversos fatores que fazem com que um indivíduo ultrapasse os 65 anos apenas 10% estão ligados à assistência médica. O estilo de vida é responsável pelos outros 53%, o meio ambiente por 20% e a herança genética pelos demais 17%. Tal achado evidencia o baixo alcance do modelo vigente para a melhoria da qualidade de vida e de saúde da população e impulsiona a adoção de um novo paradigma assistencial.

2.1.1 Modelo assistencial: prevenção e promoção à saúde

Na visão das perspectivas da regulação na saúde suplementar diante dos modelos assistenciais, os modelos atuais devem ser focados no compromisso ético com a vida e com a promoção e a recuperação da saúde e com assistência abordada de forma integrada, visualizando-se todos os passos na produção do cuidado e no restabelecimento da saúde, sendo cada beneficiário acompanhado, segundo um determinado projeto terapêutico instituído, comandado por um processo de trabalho cuidador, e não por uma lógica indutora de consumo (MALTA *et al*, 2004).

Desenvolver medidas e reunir informações de resultados sobre tratamentos é um dos focos mencionados por Porter e Teisberg (2007), justificando a importância para os planos de saúde, de conhecer os ciclos de atendimento: diagnóstico, gerenciamento e prevenção de doenças, como estratégia para melhorar a qualidade dos serviços e reduzir os custos (PORTER; TEISBERG, 2007).

Ressaltam ainda, a importância das operadoras conhecerem e dominarem as abordagens de gerenciamento de doenças e as medidas de eficácia deste gerenciamento afirmando que:

[...] os planos de saúde têm que se tornarem especialistas, em ajudar os clientes a compreender os fatores que afetam sua saúde, as melhores abordagens à prevenção de doenças e as taxas de êxito dos prestadores de serviços de prevenção de doenças [...], concluindo, que os planos de saúde devem transformar-se em conselheiros e defensores da saúde (PORTER; TEISBERG,2007).

Combinar a visão de indicadores focado na estratégia das operadoras, com ênfase nas ações de promoção à saúde e prevenção, reforça este conceito de transformar um indicador intangível em um resultado direto para o beneficiário e, conseqüentemente, para a operadora, no tangente a redução de custos.

Para obtenção de indicadores de qualidade como instrumento de gestão voltada à estratégia das cooperativas médicas, torna-se necessário o

desenvolvimento de estudos que ancorem seu estabelecimento, considerando que estes indicadores, além de auxiliar no processo de gestão, deverão alavancar a adoção de ações preventivas e negociais junto aos seus contratantes.

Para tanto, são fundamentais os conhecimentos sobre Medicina Preventiva, definido por Leavell e Clark (1976), como: “a arte de evitar doenças, prolongar a vida e desenvolver a saúde física e mental e a eficiência”.

Os indicadores deverão servir de elementos para a execução de ações de prevenção primária, conceituada como:

[...] prevenção realizada no período de pré-patogênese e que tal prevenção exige uma ação antecipada, baseada no conhecimento, a fim de tornar improvável o progresso posterior da doença.(LEAVELL; CLARK, 1976).

Praticam medicina preventiva aqueles que utilizam o conhecimento moderno, na medida de sua capacidade, para desenvolver a saúde, evitar a doença e a invalidez e prolongar a vida com qualidade.

Araújo (2004) apresenta conceitos modernamente utilizados de promoção à saúde que compreendem três níveis de prevenção:

- a) prevenção primária: compreende ações que permitem a redução da ocorrência de doenças. Inclui, não somente campanhas de vacinação como também, investimentos em saneamento básico, sugestões sobre os hábitos de vida e alimentares, campanhas antitabagismo, entre outras que apontem para ganhos em qualidade de vida;
- b) prevenção secundária: envolve ações que objetivam a redução ou a eliminação de conseqüências para a saúde, decorrente de doenças crônicas como; câncer, diabetes, doenças cardiovasculares entre outras. Recebe também a denominação de gerenciamento de doenças;
- c) prevenção terciária: envolve ações que permitem minimizar o sofrimento causado pelas limitações impostas, às pessoas já

acometidas por doenças crônicas. Recebe também a denominação de gerenciamento de caso.

Os programas de promoção à saúde são dirigidos aos clientes das operadoras de planos de saúde e considerados eficientes e estratégicos para o seu equilíbrio financeiro (CZERESINA, 2003).

Miranda (2005) afirma que o gerenciamento de doença está baseado no entendimento de que é possível atuar na rede causal relacionada com determinadas doenças que ocorrem com significativa magnitude em termos de morbimortalidade, propiciando a intervenção num momento mais precoce da sua história natural, de forma a reduzir a ocorrência de suas manifestações e complicações tendo, por consequência, uma melhor qualidade de vida para os beneficiários com menor custo para o sistema.

O gerenciamento de caso pressupõe a identificação de casos, em geral de natureza crônica, com potencial para geração de demanda por serviços e a introdução de atividade gerencial com o objetivo de adequar a aplicação de recursos para a otimização da assistência no sentido de obter uma relação entre o custo e qualidade mais favorável. Deve ainda ter por finalidade diagnosticar as reais necessidades dos pacientes e seus familiares e propor, na medida do possível, as melhores práticas para atendê-las (MIRANDA,2003).

Araújo (2004) posiciona-se sugerindo que as operadoras de plano de saúde executem ações que permitam práticas de promoção à saúde com a adoção de uma postura voltada a prevenção de doenças crônicas.

2.1.2 Doenças crônicas não transmissíveis

As Doenças Crônicas Não Transmissíveis (DCNT) se caracterizam por apresentar, de forma geral, um longo período de latência, que muitas vezes não resultam em morte súbita, levando os indivíduos a se tornarem progressivamente enfermos, especialmente se não tiverem um tratamento adequado. Este fato reforça

a preocupação das instituições com os custos da assistência a saúde, fortalecendo a motivação pela prevenção e o controle das DCNT (BRASIL,MS, 2006).

Verifica-se uma preocupação muito grande dos órgãos e instituições de saúde na prevenção das DCNT's, como as cardiovasculares, o câncer, o diabetes mellitus, a cirrose hepática, as pulmonares obstrutivas crônicas e os transtornos mentais. Segundo estimativas da OMS (2005), a cada ano morrem, pelos menos: 4,9 milhões de pessoas em decorrência do consumo do tabaco; 2,6 milhões de pessoas como consequência de estarem acima do peso ou serem obesas; 4,4 milhões de pessoas em decorrência de níveis de colesterol elevados; 7,1 milhões de pessoas em decorrência de pressão sanguínea elevada.

2.1.2.1 Diabetes Mellitus

Conforme a definição do Ministério da Saúde:

[...] O Diabetes é um grupo de doenças metabólicas caracterizadas por hiperglicemia e associadas a complicações, disfunções e insuficiência de vários órgãos, especialmente olhos, rins, nervos, cérebro, coração e vasos sanguíneos. Pode resultar de defeitos de secreção e/ou ação da insulina envolvendo processos patogênicos específicos, por exemplo, destruição das células beta do pâncreas (produtoras de insulina), resistência à ação da insulina, distúrbios da secreção da insulina, entre outros [...]
(BRASIL,MS, 2006).

A Sociedade Brasileira de Diabetes (SBD) descreve a classificação etiológica do Diabetes Mellitus em:

- a) Diabetes Mellitus Tipo 1 (DM1) representa de 5% a 10% dos casos. É resultado da destruição das células beta pancreáticas e como consequência a deficiência de insulina.

- b) Diabetes Mellitus Tipo 2 (DM2) é a forma presente em 90% a 95% dos casos e caracteriza-se por defeitos na ação e na secreção da insulina, geralmente diagnosticado após os 40 anos. A maioria dos pacientes com este tipo apresenta sobrepeso ou obesidade.
- c) Outros Tipos de Diabetes Mellitus referem-se as formas menos comuns do DM, cujos defeitos ou processos causadores podem ser identificados. Incluem-se os: defeitos genéticos das células beta; defeitos genéticos na ação da insulina; doenças do pâncreas exócrino, endocrinopatias induzidas por medicamentos ou agentes químicos e infecções.
- d) Diabetes Mellitus Gestacional, qualquer intolerância a glicose, de magnitude variável, com início ou diagnóstico durante a gestação. Não está excluída a possibilidade de existir esta condição antes da gravidez. Semelhante ao DM2, o Diabetes Mellitus gestacional é associado tanto a resistência a insulina, quanto à diminuição da função das células beta.

Dados da Organização Mundial da Saúde mostram uma incidência crescente do diabetes. Estatísticas apontam 5% das causas de morte por diabetes a cada ano e que estas mortes crescem em 50% a cada 10 anos. (OMS, 2009).

Pesquisa recentemente publicada na revista *Panamericana de Salud Publica* e realizada entre os anos de 1999 e 2003 comprova o aumento do diabetes em adultos brasileiros, e mostra que 9% das mortes na região sul e sudeste do Brasil estão relacionadas ao diabetes (BARRETO *et al*, 2007).

O número de diabéticos é crescente, devido ao envelhecimento populacional, à maior urbanização, à crescente prevalência de obesidade e sedentarismo, bem como a maior sobrevida do paciente diabético (SBD, 2007).

A natureza crônica, a gravidade de suas complicações e os meios necessários para controlá-lo, tornam o Diabetes Mellitus uma doença muito onerosa, não apenas para os indivíduos e suas famílias, mas também, para todo o sistema de saúde. Devem ser considerados ainda, os custos intangíveis como dor, ansiedade,

inconveniência e perda de qualidade de vida por apresentarem um grande impacto na vida destes diabéticos e suas famílias, porém de difícil mensuração (SBD, 2007).

A prevenção primária poderá ser adotada em diversas áreas como: nutrição em medicina preventiva, doenças crônicas, efeitos do envelhecimento sobre a saúde, prevenção do câncer e a prevenção de lesões e morte por acidentes, prevenção da doença oral, genética, saúde materno–infantil, saúde mental e saúde ocupacional (LEAVELL; CLARK, 1976).

Nesta perspectiva, a prevenção do Diabetes Mellitus pode ser realizada de duas maneiras; a prevenção da manifestação do Diabetes Mellitus, prevenção primária e a relacionada às complicações agudas ou crônicas a prevenção secundária. No caso da prevenção primária, a iniciativa objetiva proteger indivíduos suscetíveis de desenvolverem a doença, de maneira a reduzir ou retardar a necessidade de atenção a saúde e ainda de tratar as suas complicações (BRASIL, MS, 2006).

A prevenção secundária ou gerenciamento da doença é realizada quando o paciente possui a patologia já instalada, podendo atingir níveis de complexidade elevados e, agravado por comorbidades que causam danos irreversíveis a qualidade de vida do paciente e geram um alto custo assistencial (ARAUJO,2004).

No caso do Diabetes Mellitus tipo 1, não existe uma base científica que possa ser aplicada ao seu controle. As propostas melhor aceitas estão baseadas em programas de estímulo ao aleitamento materno e em evitar a introdução do leite de vaca, nos primeiros três meses de vida. Quanto ao Diabetes Mellitus Tipo 2, o de maior incidência, existem evidências de que as alterações no modo de vida (estilo de vida), com ênfase na alimentação inadequada e na redução de atividade física, estão associadas ao acréscimo da prevalência do Diabetes Mellitus Tipo 2 (SBD,2007).

Conforme o Caderno de Atenção Básica sobre o Diabetes Mellitus do Ministério da Saúde, as mudanças do estilo de vida reduziram 58% da incidência de diabetes em 3 anos. Essas mudanças visavam discreta redução de peso (5-10% do peso), manutenção do peso perdido, aumento da ingestão de fibras, restrição energética moderada, restrição de gorduras, especialmente as saturadas e aumento de atividade física regular, além de intervenções farmacológicas. Alguns medicamentos utilizados no tratamento do diabetes, como a metformina, também

foram eficazes, reduzindo em 31% a incidência de diabetes em 3 anos. Esse efeito foi mais acentuado em pacientes com IMC > 35 kg/m² (BRASIL, MS, 2006).

Desta maneira, comprova-se a importância e a efetividade de programas de prevenção, que se tornam mais efetivos se aplicados a indivíduos com indicativos para a doença (pré-diabéticos). O termo pré-diabetes é utilizado para designar o estágio intermediário entre a homeostase normal da glicose e o Diabetes Mellitus.

Para o diagnóstico do Diabetes Mellitus tipo 2 o critério foi modificado, em 1997, pela *American Diabetes Association* (ADA), posteriormente aceito pela OMS e pela Sociedade Brasileira de Diabetes (SBD) sendo atualmente aceitos três critérios para o diagnóstico do Diabetes Mellitus:

- a) sintomas clássicos do diabetes: poliúria, polidipsia, polifagia e perda involuntária de peso, acrescidos de glicemia casual acima de 200mg/dL. Compreende-se por glicemia casual aquela realizada a qualquer hora do dia, independentemente do horário das refeições;
- b) Glicemia de jejum entre 70 a 99mg/dL e inferior a 140 mg/dL 2 horas após a sobre carga de glicose, considera-se como normal. Considera-se como Intolerância à glicose ou glicose alterada quando valores de glicemia de jejum estiverem entre 100 a 126 mg/dL. Caracteriza-se como Diabético valores de glicemia de jejum igual ou superior a 126 mg/dL ou a se glicemia aleatória (realizada a qualquer momento) estiver igual ou superior a 200mg/dL;
- c) Hemoglobina glicosilada uma forma de avaliação do controle glicêmico, que deve ser medida rotineiramente nos pacientes com diabetes mellitus para documentar o grau de controle glicêmico. O nível sobre o controle da hemoglobina glicada para adultos deve ser inferior a 7% (SBD, 2007).

Outros sintomas que levantam a suspeita clínica são: fadiga, fraqueza, letargia, prurido cutâneo e vulvar, balanopostite (inflamação do prepúcio e da glândula), e infecções de repetição. Algumas vezes o diagnóstico é realizado a partir

de complicações crônicas como neuropatia, retinopatia ou doença cardiovascular aterosclerótica (BRASIL, MS 2006).

Verifica-se ainda, que 10 a 20% dos doentes têm evidências de retinopatia e nefropatia quando é realizado o diagnóstico clínico da doença, demonstrando que as lesões microvasculares podem ocorrer antes que a glicemia atinja o limite de 140mg/dL em jejum (PAIVA, 2001).

Uma grande parte da população desconhece o fato de ser portador da doença, sendo muitas vezes necessária, a realização de testes de rastreamento em indivíduos assintomáticos que apresentem maior risco da doença, tais como:

- a) idade superior a 45 anos;
 - b) sobrepeso (IMC > 25);
 - c) obesidade central (cintura abdominal > 102 cm para homens e > 88 cm para mulheres, medida na altura das cristas ilíacas);
 - d) antecedente familiar (mãe ou pai) de diabetes;
 - e) hipertensão arterial (> 140/90 mmHg);
 - f) colesterol HDL 35 mg/dL e/ou triglicerídeos 150mg/dL;
 - g) história de macrossomia ou diabetes gestacional;
 - h) diagnóstico prévio de síndrome de ovário policísticos;
 - i) doença cardiovascular, cerebrovascular ou vascular periférica definida.
- (BRASIL, 2006)

Considerando que as informações clínicas dos beneficiários, que permitam a identificação dos indivíduos com risco para determinadas doenças, muitas vezes, não estão contidas nas bases de dados dos sistemas das operadoras de plano de saúde, existe uma grande dificuldade para identificação dos beneficiários para ingresso em programas de promoção, prevenção e gerenciamento de doenças e de caso.

Conforme Porter e Teisberg (2007) existem grandes lacunas de informações que limitam a capacidade dos planos de saúde de conhecerem o ciclo completo de atendimento do paciente para a mensuração de resultados e dos custos gerais,

porém afirmam que, não se deve esperar a obtenção das informações completas sobre os registros médicos, para se iniciar um controle, gerenciamento e mensuração de resultados.

Neste contexto, a utilização de técnicas sofisticadas de inteligência artificial, para melhor aproveitamento das bases de dados, justifica-se, tanto pelo grande volume de informações existentes nos serviços de saúde, quanto pela ausência de informações clínicas. Embora os dados armazenados sejam um bem valioso, muitos se deparam com o problema de “muitos dados, mas pouco conhecimento” sobre eles (LU *et al*, 1995).

Observa-se que apesar da ausência de informações clínicas dos pacientes, já existem experiências de identificação e reconhecimento de usuários com propensão a determinadas doenças, por meio de dados financeiros e de utilização do plano.

[...] A localização do paciente, identificação de seu endereço e realização do contato, é uma fase trabalhosa do processo. Embora achar os pacientes possa ser difícil, uma operadora americana descobriu que conseguia identificar mais de dois terços de seus diabéticos, através das consultas agendadas, e quase metade, via dados de assistência farmacêutica e dados de pagamento e utilização [...] (ZITTER, 1997).

Desta maneira, percebe-se a viabilidade do emprego de técnicas avançadas para a descoberta em base de dados, para a identificação e seleção de beneficiários elegíveis para ações programadas nos diferentes níveis de atenção, melhorando a qualidade de vida e de saúde dos beneficiários, bem como contribuindo para a redução de custos nas operadoras de saúde.

2.2 BASE DE DADOS EM SAÚDE E INTELIGÊNCIA ARTIFICIAL

Devido ao grande volume de informações nas bases de dados geradas pelas solicitações de atendimentos de uma operadora e a complexidade de interpretar e avaliar as informações para gerar conhecimentos, torna-se necessária a utilização

de ferramentas e técnicas que permitam uma análise mais automática e inteligente como DW – *Data Warehouse*, técnicas oriundas da Inteligência Artificial, como por exemplo; KDD – *Knowledge Discovery in Databases – Data Mining* e Sistemas Especialistas.

Para um melhor entendimento sobre Inteligência Artificial, são apresentadas algumas definições:

“Automação das atividades associadas ao pensamento humano, atividade de tomada de decisão, solução de problemas e aprendizado” (BELLMAM, 1978)

“Um dispositivo capaz de colecionar, selecionar entre, de compreender, de distinguir e de saber” (FEIGENBAUM; MCCORDUCK, 1983)

“Capacidade de adquirir e aplicar conhecimentos implementados por humanos” (ARNOLD; BOWIE, 1986)

“A Inteligência Artificial é o estudo de conceitos que permitem aos computadores serem inteligentes” (WINSTON, 1987)

“Resultado da aplicação de técnicas e recursos, especialmente de natureza não numérica, viabilizando a solução de problemas que exigiriam do humano, certo grau de raciocínio e de perícia” (RABUSKE, 1995).

Uma relação com seu objeto de estudo semelhante à psicologia, mas com uma importante diferença: os modelos e teorias da IA são implementadas em um computador, o que os torna de certa forma, autônomos. Assim, a validade de um modelo ou de uma teoria de IA não precisa ser provada através da comparação de seus resultados previstos com o comportamento psíquico humano, como no caso da psicologia, mas pode ser implementada em um computador e demonstrada diretamente através da ação inteligente do programa no mundo (BITTENCOURT, 1998).

“Inteligência Artificial é uma subdivisão da ciência da computação, voltada à criação de software e hardware que objetiva a produção de conhecimentos, tal como os produzidos pelos seres humanos”. (TURBAN, 1995).

“A Inteligência Artificial sistematiza e automatiza tarefas intelectuais e, portanto, é potencialmente relevante para qualquer esfera da atividade intelectual humana” (RUSSEL; NORVIG, 2004).

Russel e Norvig (2004) sintetizam das diversas definições sobre inteligência artificial, quatro categorias: Sistemas que pensam como seres humanos, sistemas

que pensam racionalmente, sistemas que atuam como seres humanos e sistemas que atuam racionalmente.

Desta maneira, percebe-se pelos conceitos apresentados, que a inteligência artificial baseia-se em estruturas de equipamentos e sistemas, dirigidas para o auxílio dos processos de formulação de hipóteses, criação de modelos e na capacidade de síntese e aquisição de novos conhecimentos.

2.2.1 Descoberta de conhecimento em base de dados

O processo de descoberta de Conhecimentos em Base de Dados (KDD - *Knowledge Discovery in Databases*) foi definido por Fayyad et al (1996) como:

um processo não trivial para identificar padrões que sejam válidos e previamente desconhecidos, potencialmente úteis e compreensíveis, visando melhorar o entendimento do problema, possibilitando o auxílio no processo de tomada de decisões (FAYYAD et al., 1996).

Este processo pode ser aplicado em diversas áreas entre elas: mercado, análise de crédito, análise de tendências, marketing, astronomia e medicina. Na medicina são encontrados diversos trabalhos aplicados a análise do histórico de pacientes, para aperfeiçoar o tratamento e auxiliar na obtenção de conhecimento implícito, para sugerir ações na área da saúde (VIANNA, 2007 KOBUS, 2006).

O processo de Descoberta de Conhecimento a partir de base de dados, apresentado por Fayyad *et al.* (1996), envolve as etapas mostradas na Figura 1.

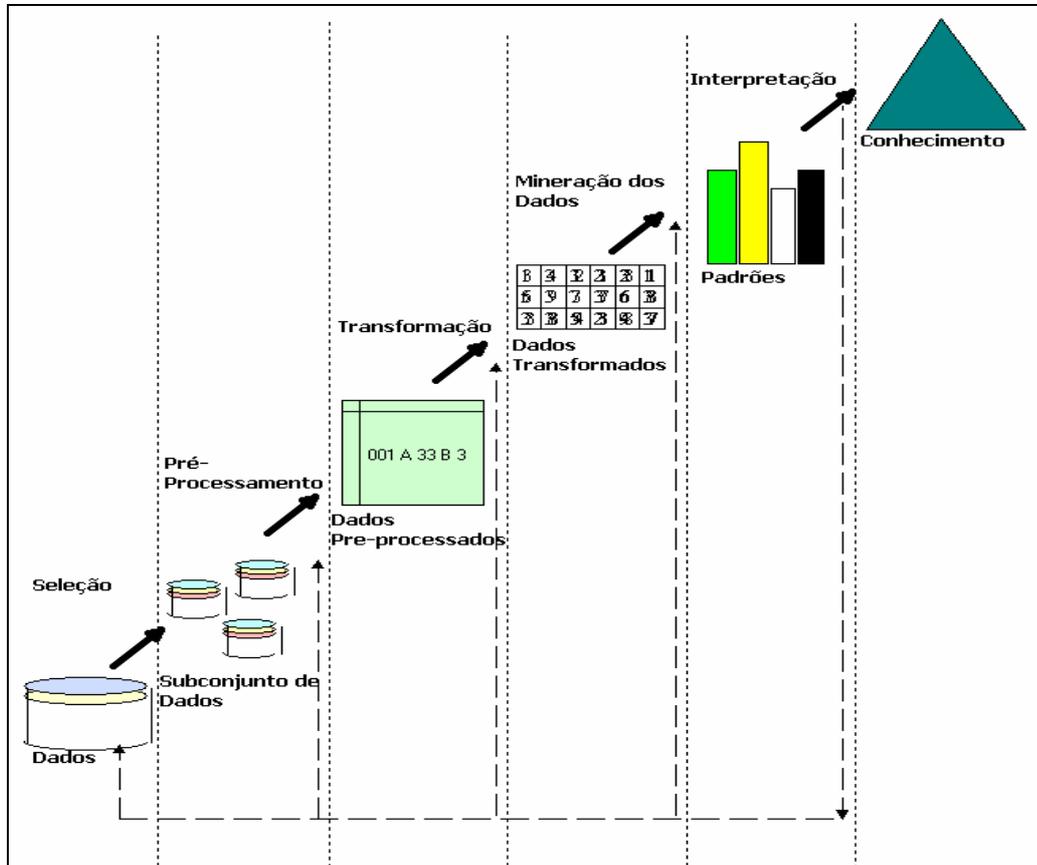


Figura 1 - Etapas do processo de descoberta de conhecimentos. Fonte: Fayyad et al,1996
(Adaptação do autor)

Este processo inclui as etapas:

- Análise inicial** e domínio sobre a aplicação: engloba o tipo de conhecimento que se deseja descobrir, além de uma compreensão e domínio sobre a aplicação e o tipo de decisão que poderá contribuir para o processo.
- Seleção:** criação de um conjunto de dados (atributos) visando a realização da tarefa da descoberta de conhecimentos. Geralmente esta etapa exige consultas à base de dados de sistemas transacionais das organizações.

- c) **Limpeza dos dados e pré-processamento:** envolve operações como a remoção de ruídos, a coleta de informações necessária para modelagem, a escolha de estratégias para a manipulação de dados ausente, a formatação de dados para adequá-los a mineração.
- d) **Redução de dados e projeção - Transformação:** Identificação de características para a identificação dos dados em função do objetivo da tarefa, visando a redução do número de variáveis.
- e) **Mineração de Dados:** seleção dos métodos a serem utilizados para a tarefa de mineração de dados. Esta etapa requer a utilização do algoritmo mais apropriado para a geração de padrões e a descoberta de novos conhecimentos. Os algoritmos são específicos para cada tipo de tarefas e finalidade, como apresentado resumidamente por LÂVOR (2003), no quadro 2:
- f) **Interpretação e avaliação dos padrões selecionados:** são avaliados e adotados os padrões extraídos e seus modelos.
- g) **Geração do conhecimento:** consiste na própria utilização do conhecimento, documentação e publicação para as partes interessadas. Nesta etapa inclui-se a verificação e potenciais conflitos entre outros conhecimentos gerados previamente.

Grupo de Tarefas	Tarefa	Descrição	Tipo de Tarefa	Métodos
Análise de Dependências	Associação	Descoberta de regras que descrevem dependências significativas entre os itens de uma base de dados que ocorrem na mesma transação.	Tarefa Descritiva	Estatística, teoria de conjuntos, modelos probabilísticos de gráficos de dependência.
	Detecção de Sequências Temporais	Descoberta de regras que descrevem dependências entre itens que ocorrem ao longo do tempo.	Tarefa Descritiva	Estatística, teoria de conjuntos.
Identificação de Classes	Regressão	Identificação de um padrão de comportamento dos itens da base de dados, descobrindo uma função que mapeie os novos itens.	Tarefa Preditiva	Regressão linear, redes neurais, ajuste de curvas, árvores de decisão
	Classificação	Identificação de um padrão de comportamento dos itens da base de dados, mapeando os novos itens em classes pré-existentes	Tarefa Preditiva	Árvore de decisão, redes neurais e métodos baseados em exemplo
	Clusterização/Agrupamento	Geração e identificação de grupos, clusters, por meio de agrupamento natural de itens da base de dados, em função da similaridade entre eles.	Tarefa Descritiva	Árvore de decisão, redes neurais e métodos baseados em exemplo
Descrição de Classes	Sumarização	Descoberta de regras que descrevam sucintamente uma classe	Tarefa Descritiva	Árvores de decisão, redes neurais, algoritmos genéticos
Detecção de Desvios	Detecção de Desvios	Detecção de desvios e anomalias nos itens de uma base de dados	Tarefa Descritiva	Teoria de conjuntos, estatística.

Quadro 2 - Principais tarefas de mineração de dados. Fonte: Lavor, 2003

Para auxiliar as etapas de análise inicial e seleção, torna-se necessária a utilização de técnicas de exploração de dados, que compreende o pré-processamento do KDD. Assim, apresenta-se o conceito de *Data Warehouse* (DW), que permitirá uma análise mais eficiente dos dados.

Para Harjinder e Prakash (1996), em função de uma diversidade de base de dados em ambientes transacionais, faz-se necessário um processo para concentrá-las, incluindo dados históricos e dados externos, usados para atender a necessidade de consultas estruturadas, relatórios analíticos e de suporte à decisão, o que justifica a utilização de Data Warehouse na etapa de pré-processamento do KDD.

Segundo Inmon (1997), “Data Warehouse é uma coleção de dados integrados, orientados por assunto, não-volátil e variáveis em relação ao tempo, utilizados para o apoio às decisões gerenciais”.

“Dados Integrados: num ambiente DW é usual a construção de um conjunto de dados de fontes heterogêneas, de bases de dados relacionais, arquivos isolados e dados de transações on-line”.

“Dados Orientados por Assuntos: o DW é organizado sobre um assunto, ou seja, os dados são concentrados sobre a operação e processamento das transações de um determinado objetivo, excluindo as informações da estrutura desnecessárias para o processo de tomada de decisões”.

“Dados variáveis em relação ao tempo: os dados são armazenados para prover informações de perspectivas históricas. Em ambientes de DW sempre se mantêm na estrutura dos modelos, implicitamente ou explicitamente, o elemento de tempo”.

“Dados não voláteis: no DW o armazenamento das informações é fisicamente independente dos ambientes transacionais. Esta separação faz-se necessária para a manutenção dos dados, sem a possibilidade de sua alteração. Em um DW são permitidas apenas duas operações; carga inicial e acesso aos dados”.

KIMBAL (1996), ao conceituar Data Warehouse apresenta uma diferenciação entre o ambiente OLTP (*Processamento Transacional Online*) como o ambiente que armazena as informações diárias de uma organização e utilizadas para executar operações pré-definidas e as informações que sofrem constantes mudanças (transações). No Ambiente OLAP (*Processamento Analítico Online*) os dados armazenados são analíticos, destinados as necessidades de apoio ao processo de

tomada de decisões, com grande quantidade de informações, pois são armazenados diversos períodos, mantendo as informações históricas sem sofrer mudanças.

Data Marts são subconjuntos de dados da empresa armazenados fisicamente em mais de um local, geralmente divididos por assuntos (departamentais). Os Data Marts diferenciam-se dos DW pelos seguintes fatores: os dados são personalizados e atendem as necessidades específicas de um departamento; possuem um volume menor de dados e com isto, um histórico mais limitado e geralmente não mantêm os dados no mesmo nível de detalhe que no Data Warehouse (INMON, 1997).

A representação por meio do modelo multidimensional refere-se a dados que podem ser visualizados como um “cubo”, com três, quatro, cinco ou mais dimensões. (KIMBALL, 1996). O modelo dimensional do negócio é representado a cada ponto do cubo, como contendo a combinação de suas dimensões. Na figura 2, cada ponto do cubo, representa a combinação entre o Usuários, Serviços e o Tempo.

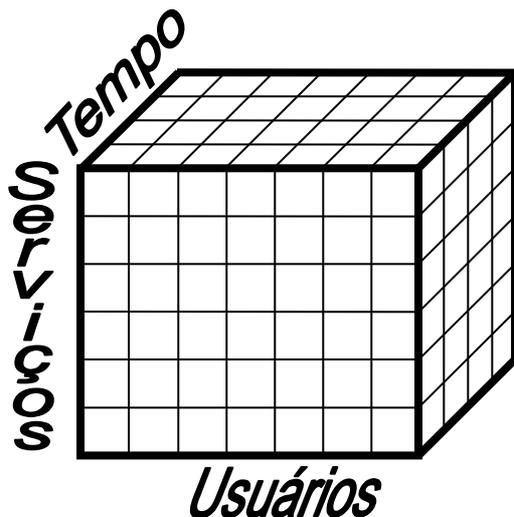


Figura 2 – Modelo dimensional – Fonte: Kimbal, 1996 (Adaptação do autor).

Kimball (1996) comenta ainda, que o modelo multidimensional é também chamado de modelo estrela, porque o diagrama assemelha-se a uma estrela, onde no centro, encontra-se a tabela principal e ao seu redor, as tabelas com volumes menores, também chamadas de dimensões. Cita ainda, que o modelo diferentemente de um modelo entidade-relacionamento, é assimétrico com uma tabela ao centro e com diversos relacionamentos conectando-se as demais tabelas. A tabela central, também pode ser chamada de tabela fato.

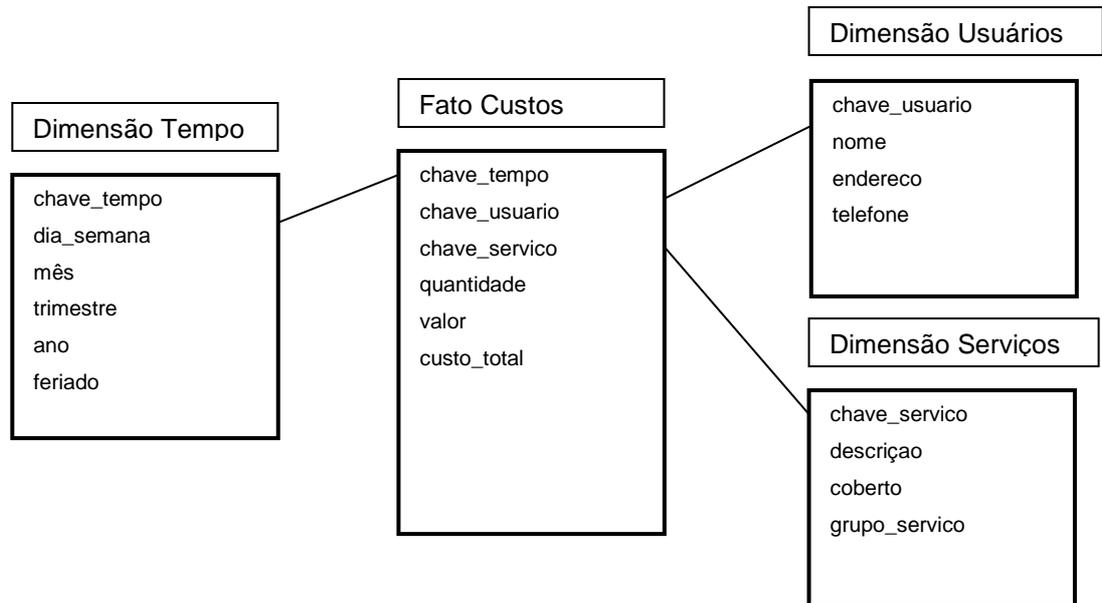


Figura 3 – Um típico modelo dimensional. Fonte: Kimbal, 1996 (adaptação do autor).

Han e Kamber (2001), justificando a importância da integração entre ambientes Data Warehouse e Tecnologias OLAP (Processamento Analítico Online) com o de Mineração de Dados, mencionam que na construção de um Data Warehouse são necessárias as etapas de limpeza, organização e agregação dos dados e a integração de múltiplas fontes de dados, permitindo que se obtenha uma base de dados estruturada, padronizada e consistente, sendo estas, as características fundamentais para o início do processo de mineração de dados.

As funcionalidades de tecnologias OLAP e mineração de dados podem ser visualizadas como distintas. As tecnologias OLAP são ferramentas de agregação e sumarização dos dados que auxiliam na análise dos dados, enquanto a Mineração de Dados possibilita a descoberta automática de padrões implícitos e de conhecimentos escondidos em grande quantidade de dados (HAN; KAMBER, 2001).

Han e Kamber (2001), propõem o ambiente OLAM (Processamento Analítico *Online* e Mineração de Dados), ou seja, o ambiente DW e tecnologias OLAP relacionando-se automaticamente com o ambiente de Mineração de Dados.

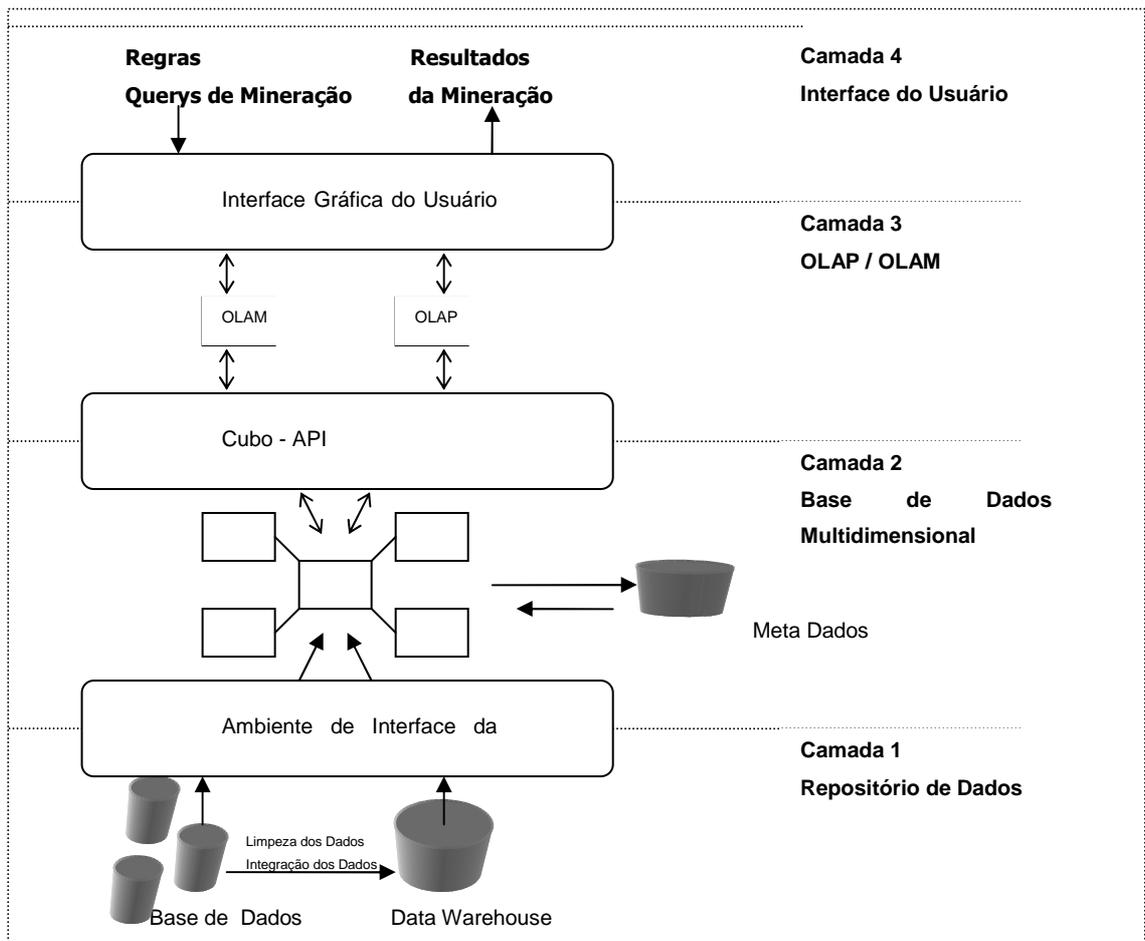


Figura 4 - Arquitetura OLAM . Fonte: Han e Kamber, 2001 (Adaptação do autor).

Por meio dos conceitos apresentados, verifica-se além da potencialidade do DW e do ambiente OLAP, no processo de extração de informações para a etapa de pré-processamento do KDD, a possibilidade de integração entre um ambiente transacional (sistema de gestão), com um processo de mineração de dados, permitindo com isto a integração das ferramentas e a possibilidade de geração de novos conhecimentos de forma direta e on-line.

2.2.2 Representação do conhecimento

Na utilização de ferramentas que geram novos conhecimentos, faz-se necessária à introdução de alguns conceitos, como: dado, informação e

conhecimento. Dado é um indicador bruto, quantificável, sem a capacidade de levar à compreensão de uma situação ou realidade. Já a informação é o dado de forma contextualizado e o conhecimento pode ser entendido como sendo o conjunto de padrões que podem ser obtidos por meio do relacionamento entre dados e informações (GOLDCHMIDT; PASSOS, 2005).

Rezende (2005) descreve que o conhecimento pode ser representado por meio de técnicas, que devem ser suficientemente expressivas a respeito do domínio escolhido de maneira completa e eficiente.

A representação do conhecimento é uma forma sistemática de estruturar e codificar o que se sabe sobre uma determinada aplicação e deve observar as seguintes características:

- a) ser compreensível para poder avaliar e interpretá-la;
- b) abstrair-se dos detalhes do funcionamento do processador de conhecimento;
- c) ser robusta, permitindo sua utilização, mesmo que não aborde todas as situações possíveis;
- d) possibilitar a generalização, necessitando de pontos de vistas diferentes do mesmo conhecimento, permitindo diversas situações e interpretações (REZENDE,2005).

Os sistemas especialistas são construídos baseados em duas partes: a base do conhecimento e o mecanismo de inferência. A base do conhecimento contém fatos sobre objetos dentro dos domínios selecionados e seus relacionamentos. Pode conter também, conceitos, teorias, programas e outras associações. Esta base de conhecimento forma a fonte dos sistemas especialistas e são baseados no mecanismo de inferência para a definição das conclusões (TURBAN, 1995).



Figura 5 - Representação de um processo lógico. Fonte: Turban, 1995 (Adaptação do autor).

O mecanismo de inferência é um conjunto de programas utilizados para examinar a base do conhecimento dentro de uma classificação das respostas referentes as suas questões e para resolver problemas ou criar decisões sobre um determinado domínio (TURBAN, 1995).

Turban (1995), comenta ainda, que os processos lógicos são as maneiras mais simples de discutir o conteúdo e organizar o conhecimento. A organização dos objetos em categorias é uma parte vital para a representação de conhecimento e são subdivididos em lógica descritiva (proposições ou instrução) e lógica preditiva (por analogia ou por indução).

O método de aprendizagem por indução é a maneira mais utilizada pela mente para produzir um conhecimento novo. Por meio de exemplos com objetos pertencentes a uma classe pré-definida, e outros com objetos a ela não pertencentes, pode-se verificar os conjuntos de objetos e reconhecer similaridades, a fim de encontrar uma descrição da classe que englobe os exemplos pertencentes (positivos) e não pertencentes (negativos). No método de aprendizagem dedutiva, as descobertas são realizadas por meio de uma premissa maior para uma menor, e também por meio de generalizações e especializações, até que se alcance o critério previamente exigido (CARVALHO, 1999).

2.2.3 Aprendizado de máquina

O Aprendizado de máquina pode ser: supervisionado ou não supervisionado, porém neste estudo será utilizado o aprendizado de máquina supervisionado.

O aprendizado supervisionado utiliza um conjunto de exemplos de treinamento com um rótulo da classe associado e conhecido, em conjunto com um algoritmo de aprendizagem, ou indutor. O objetivo do algoritmo é construir uma função ou modelo que permita determinar corretamente a classe de novos exemplos, ainda não rotulados (REZENDE, 2005).

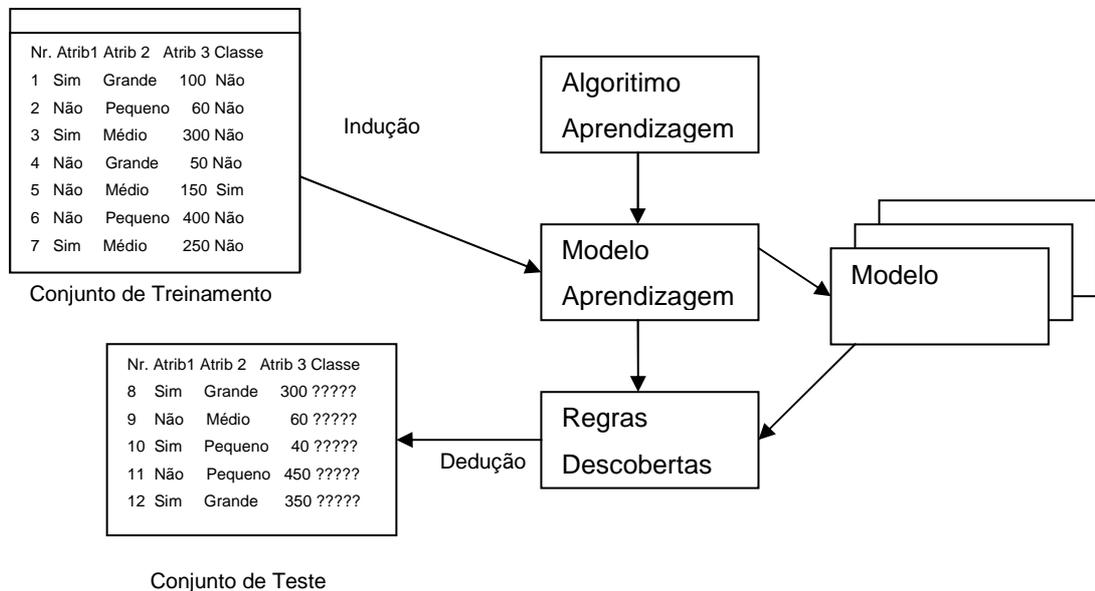


Figura 6 – Estrutura de aprendizagem de máquina. Fonte: TAN *et al*, 2005 (Adaptação do autor).

A tarefa de mineração de dados é uma das etapas do KDD, que é a técnica que permite a aplicação de algoritmos que possibilitem a extração de padrões em base de dados, ou seja, a explicitação de relações entre elementos de uma base de dados. A importância destas técnicas está diretamente relacionada à sua capacidade de extrair conhecimento útil e previamente desconhecido e que está implícito nas bases de dados (FAYYAD *et al*, 1996).

Para Goldschmidt e Passos (2005), a tarefa de classificação é um método que consiste em “descobrir uma função que mapeia um conjunto de rótulos

categóricos predefinidos chamados “classe”. Uma vez descoberta, pode ser aplicada a novos registros de forma a prever a classe em que tais registros se enquadram”.

2.2.4. Algoritmo de árvore de decisão

A divisão de cada nó interno é definida utilizando um conjunto de dados de treinamento que possui uma classe de caracterização do registro, e por meio de procedimentos estatísticos os nós são divididos, sempre utilizando a estratégia chamada *dividir-para-conquistar*, onde um problema complexo é decomposto em subproblemas mais simples; recursivamente a mesma estratégia é aplicada a cada subproblema. Esta técnica é chamada de algoritmo de aprendizado ou *Learning Algorithms* (WITTEN E FRANK, 2000).

Uma das formas de representação do resultado da tarefa de mineração, objeto deste estudo é a de árvore de decisão. A estrutura de uma árvore de decisão indutiva é composta de um nó raiz, o topo da árvore “*root node*” e representado como sendo a divisão mais relevante, um conjunto de nós internos (“*internal node*”) que representam uma avaliação sobre um determinado atributo e de nós terminais chamados de folhas “*leaves*” que representam os identificadores das classes (HAN, 2003).

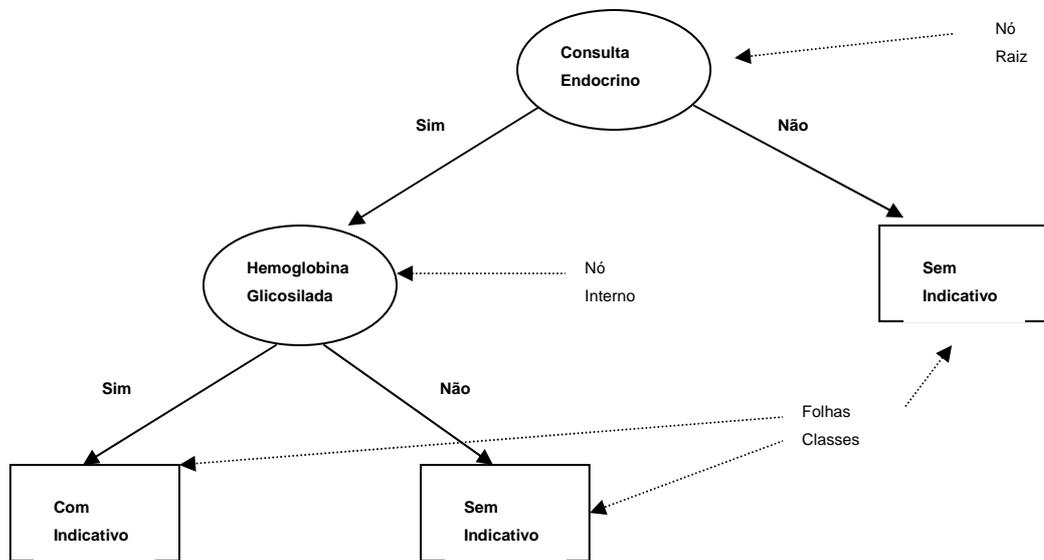


Figura 7 – Exemplo árvore de decisão. Fonte: TAN *et al*, 2005 (Adaptação do autor).

Como vantagens na utilização de árvores de decisão têm-se: facilidade de construção; boa performance para a classificação de registros desconhecidos; boa compreensibilidade para árvores pequenas e precisão semelhante às de outras técnicas de classificação (TAN *et al*, 2005).

Também podem ser citadas como vantagens; a derivação de regras representadas como conjuntos do tipo “Se – Então”, que podem ser aplicados diretamente em uma determinada função, tornando-o um modelo claro e demonstrando os atributos que estão discriminando os padrões (STEINER *et al*, 2006).

Como exemplo, apresenta-se a extração das regras de classificação de beneficiários com indicativo de diabetes mellitus 2, demonstrando a clareza e facilidade de interpretação dos padrões encontrados.

Se *Consulta Endócrino* = “Não” Então “*Sem Indicativo*”
 Se *Consulta Endócrino* = “Sim”
 E
 Se *Hemoglobina Glicolisada* = “Não” Então “*Sem Indicativo*”

Se *Consulta Endócrino* = “Sim”
 E
 Se *Hemoglobina Glicolisada* = “Sim” Então “*Com Indicativo*”

Outras vantagens da utilização dos algoritmos de árvore de decisão podem ser destacadas: não assumem nenhuma distribuição particular para os dados; as características ou atributos podem ser categóricos (qualitativos) ou numéricos (quantitativos); permitem construir modelos para qualquer função, desde que o número de exemplos de treinamento, seja suficiente e de elevado grau de compreensão (STEINER *et al*, 2006).

Para a análise e a avaliação de algoritmos de classificação, de acordo com os modelos de indução são utilizadas algumas métricas, como a **acurácia**, que é a quantidade de registros classificados corretamente e a **taxa de erro** que são os registros classificados incorretamente, que indicam o quão o modelo é confiável.

Para a representação destas métricas, utiliza-se uma matriz, chamada de matriz de confusão ou matriz de contingência.

	Classe Prevista	
Classe Real	Classe =1	Classe =0
Classe = 1	f_{11}	f_{10}
Classe = 0	f_{01}	f_{00}

Figura 8 – Matriz de confusão ou de contingência. Fonte: O próprio autor, 2008.

Onde: f_{11} – Verdadeiro Positivo, f_{10} – Verdadeiro Negativo, f_{01} – Falso Negativo e f_{00} – Falso Positivo

$$\text{Acurácia ou Precisão} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}}$$

$$\text{Taxa de Erro} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}}$$

Existem vários algoritmos de árvores de decisão, entre eles o pioneiro CART (*Classification and Regression Trees*), sugerido por Breiman et al (1984). Sua principal característica é ser binário, contendo apenas duas divisões para cada nó de decisão e suas divisões são baseadas na medida de impureza, por meio do índice GINI. O ID3 e o C4.5, proposto por Quinlan (1993), utilizam o conceito de “ganho da informação” como critério de ramificação, para a construção da árvore de decisão. O ganho da informação é baseado no índice de entropia para medição da homogeneidade de cada nó (LAROSE, 2005).

$$\text{Ganho da Informação} = \text{Entropia}(t) = - \sum p(j/t) \log_2 p(j/t)$$

Onde; j/t é a frequência relativa da classe j no nó t .

Para uma representação que facilite a apresentação dos resultados gerados pelo algoritmo de árvore de decisões, pode-se citar o raciocínio baseado em regras.

2.2.5 Raciocínio baseado em regras

As abordagens baseadas em regras esperam usar como fundamento o sucesso dos sistemas baseados em regras lógicas, acrescidas de um “fator de improvisação” a cada regra, para acomodar a incerteza. Esses métodos foram desenvolvidos na década de 1970 e formaram a base para um grande número de sistemas especialistas em saúde e em outras áreas (RUSSELL e NORVIG, 2004).

A forma de representação, a facilidade de explanação e a modelagem do conhecimento, de uma forma muito próxima ao processo cognitivo humano tornam a regra muito vantajosa.

Uma regra consiste de uma parte SE, o lado esquerdo da regra, e de outra parte ENTÃO, que representa, a ação a ser executada ou a conclusão a ser deduzida, caso todas as condições da parte SE tenham sido satisfeitas.

As regras podem ser representadas da seguinte forma:

Se Cond1, Cond2, ... Cond _n	Então
Conclu1, ... , Concl _n	
Ou	
Se Cond1, Cond2, ... Cond _n	Então
Ação 1, ..., Ação _n	

Todos os conceitos apresentados propõem o processo de descoberta de conhecimento em base de dados, como uma técnica para o apoio na seleção de beneficiários elegíveis para ações programáticas de saúde, por meio de informações de atendimentos construídos pelo beneficiário da operadora e os conceitos de *data warehouse*, que atuam como ferramentas facilitadoras para a extração das informações e processamento destas informações para todo o processo.

3 METODOLOGIA

Este trabalho tem como objetivo conceber uma metodologia para a identificação de beneficiários com indicativos para o Diabetes Mellitus Tipo 2, para o ingresso em programas de promoção da saúde e prevenção de doenças.

Trata-se de um estudo quantitativo, retrospectivo, de cunho descritivo. A pesquisa quantitativa tem como objetivo a análise e compreensão de informações e fatos traduzidos em números, a fim de quantificá-los, utilizando-se de recursos e técnicas da estatística (SILVA; MENEZES, 2001).

Um estudo retrospectivo caracteriza-se pela análise retroativa de casos, com a análise das conseqüências de fatos passados. Um aspecto importante deste tipo de estudo é que como os fatos analisados não foram observados diretamente, sua quantificação depende das informações disponíveis serem suficientes e confiáveis. Muitas vezes, os dados encontrados são incompletos ou inconsistentes, tal como nas bases de dados da saúde, em que diversos profissionais têm acesso à documentação, porém trabalham com diferentes padrões ou critérios de registro (SILVA, 1999).

Já estudos descritivos têm como característica observar e descrever aspectos de uma situação (POLIT; HUNGLER, 1995).

3.1 ETAPAS DO ESTUDO

O estudo foi desenvolvido utilizando-se as técnicas de KDD, sobre a base de dados do Data Warehouse da Unimed Federação do Estado do Paraná, estrutura que possibilita gerar as informações de atendimentos de beneficiários ativos, do Estado do Paraná, conforme autorização da instituição para utilização dos dados.

Para o processo de descoberta de novos conhecimentos, aplicou-se o processo de KDD, segundo os conceitos aplicados por FAYYAD *et al.* (1996).

Etapas: Levantamento Inicial, Seleção dos Dados, Pré-Processamento, Mineração de Dados e Análise e Avaliação dos Conhecimentos Descobertos.

Assim, o estudo contemplou como etapas:

- a) identificação das variáveis – Análise inicial;
- b) seleção de dados para o Pré-processamento;
- c) mineração de Dados;
- d) avaliação e Interpretação dos Resultados (Validação Especialistas).

Para as etapas de análise inicial e seleção de dados para o pré-processamento, utilizou-se o modelo de dados multidimensional – ambiente Data Warehouse da Federação das Unimed do Paraná, para a exploração e análise das informações.

No processo de mineração de dados, aplicou-se o algoritmo C4.5 (QUINLAN,1993) por meio do software livre WEKA (Waikato Environment for Knowledge Analysis).

Para o processo de avaliação e interpretação dos resultados, utilizou-se ferramentas de planilha eletrônica e de processador de texto para a seleção da regras, elaboração do formulário e avaliação dos resultados.

3.1.1 Identificação das variáveis - análise inicial

Para esta etapa aplicou-se um desenho individuado – observacional – longitudinal-retrospectivo, chamado também, de estudo de caso-controle, concebido especialmente para investigar associações etiológicas em doenças de baixa

incidência e/ou condições com período de latência prolongado. Neste tipo de estudo, após a identificação dos elementos diagnosticadores, é realizada uma retrospectiva na sua historia, para investigação dos possíveis fatores, que possam ser considerados importantes para a identificação e classificação de novos elementos (ALMEIDA; ROUQUAYROL, 2006).

Para esta investigação histórica, foram selecionados na base de dados 2 grupos distintos.

O Grupo 1 foi constituído por 59 beneficiários com internamento pelo Código Internacional de Doenças, 10ª revisão (CID-10) de Diabetes Mellitus tipo 2 (E11), no terceiro trimestre de 2007, com idade igual ou superior a 25 anos, sendo avaliados os seus atendimentos específicos, consultas por especialidade médica e solicitação de exames específicos, de forma retrospectiva nos 5 anos anteriores, ou seja o período compreendido entre 01 de janeiro de 2003 à 31 de dezembro de 2007.

O número de beneficiários selecionados neste grupo, constitui a totalidade dos indivíduos que tiveram como causa do internamento o CID do Diabetes Mellitus Tipo 2, no terceiro trimestre de 2007 na base de dados da Unimed PR.

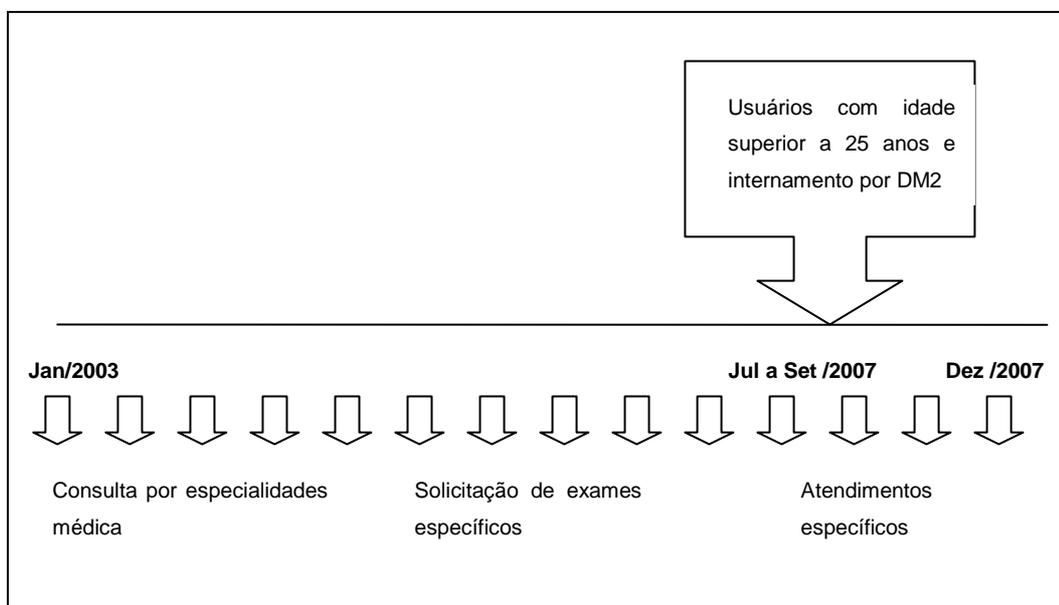


Figura 9 – Avaliação retrospectiva dos atendimentos grupo DM2. Fonte: O próprio autor, 2009.

Foi criado ainda, um grupo controle, denominado Grupo 2, constituído também por 59 beneficiários, que tiveram internamento clínicos por outras causas que não Diabetes Mellitus tipo 2, no terceiro trimestre de 2007.

Considerando o número de internamentos clínicos no período adotou-se como critério a seleção dos primeiros 59 internamentos, sendo analisados os eventos, consultas em especialidades, solicitação de exames e atendimentos específicos, de forma retrospectiva nos 5 anos anteriores, objetivando uma comparação com os dados de internamento por Diabetes Mellitus tipo 2, para definição e confirmação das variáveis do estudo.

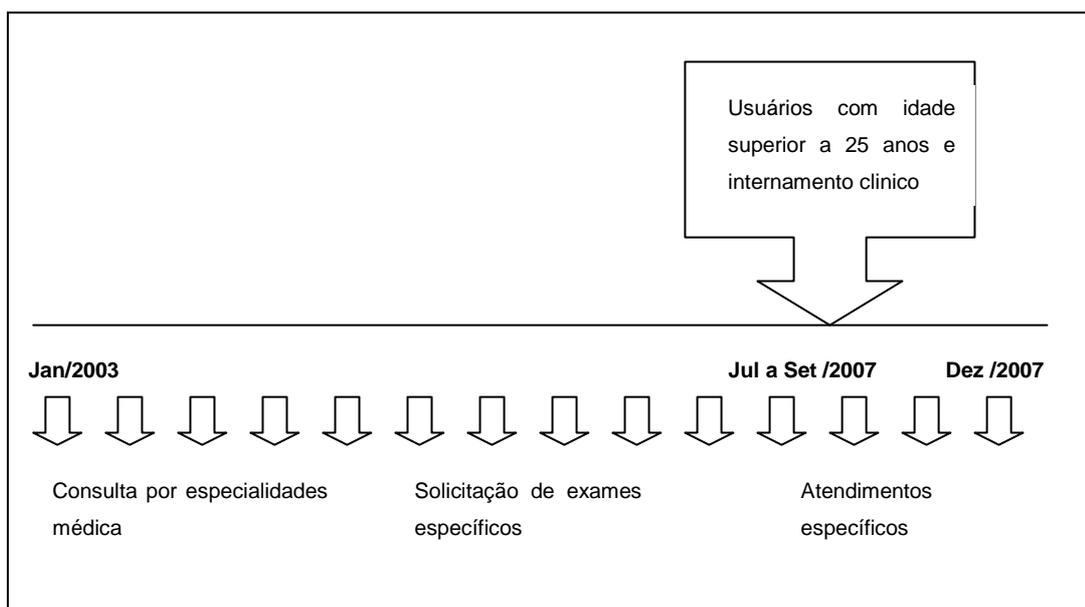


Figura 10 – Avaliação retrospectiva dos atendimentos grupo controle. Fonte: O próprio autor, 2009.

Desta maneira foram levantadas informações, comparadas e analisadas o que possibilitaram a identificação das variáveis relevantes para o estudo.

3.1.2 Pré-processamento

Para esta etapa, foi utilizada a base de dados de “Atendimentos” da Federação das Unimeds do Estado do Paraná, estrutura que possibilita gerar as informações de atendimentos de 1.105.301 beneficiários do Estado do Paraná (UNIMEDPR,2008).

O modelo de dados de Atendimentos, utilizado para esta fase do estudo configura um modelo estrela apresentado na Figura 11.

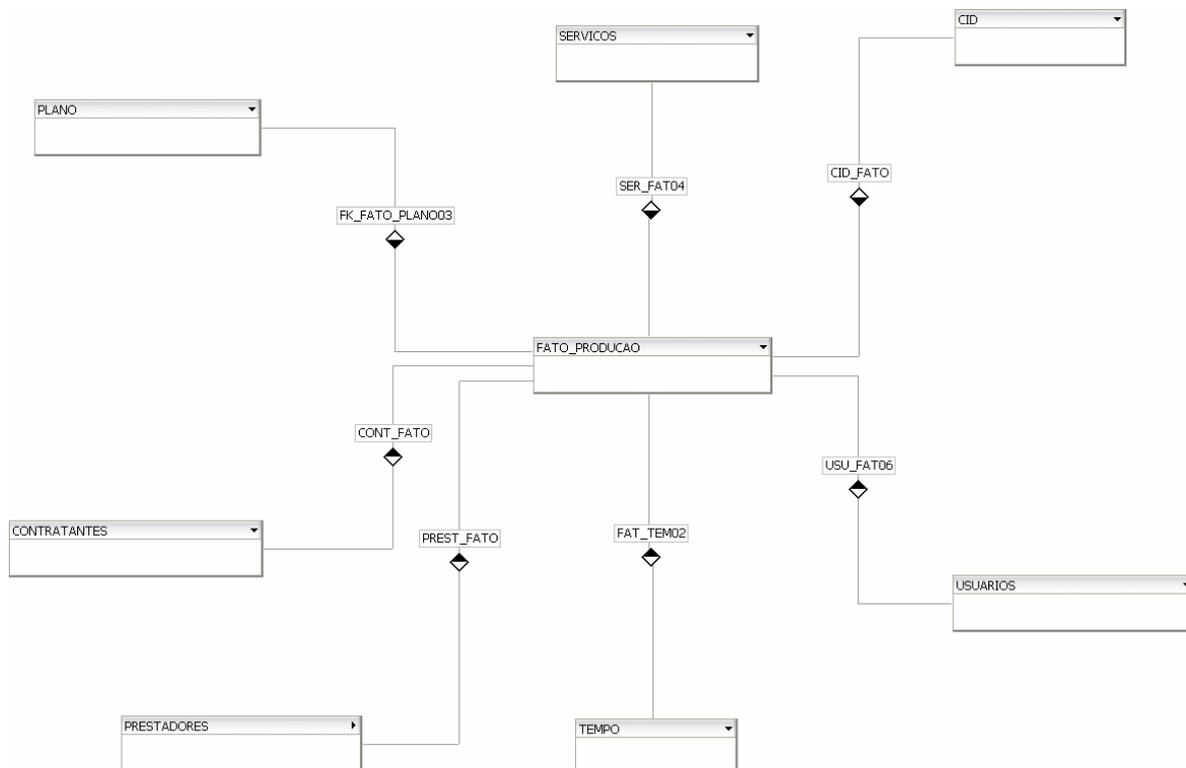


Figura 11 - Modelo multidimensional de atendimentos. Fonte: UNIMEDPR,2008.

Na estrutura multidimensional apresenta-se a tabela objetivo no centro do modelo, com o nome Fato_Producao, que se refere às informações sobre os atendimentos.

Nas extremidades do modelo multidimensional, apresentam-se as tabelas dimensões, como: Prestadores e suas especialidades, a tabela CID, beneficiários, contratantes, serviços e tempo (ano e mês de atendimento)

Utilizou-se para esta etapa, a mesma base do ambiente Data Warehouse da Federação das Unimeds do Estado do Paraná, no período histórico de 6 anos, de 01 de janeiro de 2002 a 31 de dezembro de 2007, constituída por 60.000 beneficiários ativos na operadora durante todo esse período e que corresponde quase a totalidade dos beneficiários da Operadora Unimed do Estado do Paraná, para treinamento e validação.

Os dados foram agrupados e sumarizados por beneficiário, omitindo-se sua identificação. Para a fase de treinamento e validação, determinou-se a formação do atributo rótulo, com a seguinte regra:

- a) se Quantidade de atendimento com CID de diabetes e número de solicitações de hemoglobina glicosilada ≤ 2 então ***sem indicativo***;
- b) se Quantidade de atendimento com CID de diabetes e número de solicitações de hemoglobina glicosilada ≤ 6 então ***com indicativo***;
- c) se Quantidade de atendimento com CID de diabetes e número de solicitações de hemoglobina glicosilada > 6 então ***com forte indicativo***.

As regras exibem valores relacionados à quantidade de atendimento no CID de diabetes e número de exames de hemoglobina glicosilada referente ao período de 6 anos, consonante com as Diretrizes da Sociedade Brasileira do Diabetes – SBD, que recomenda como frequência que o exame de hemoglobina glicosilada, seja realizado pelo menos duas vezes ao ano, para todos os diabéticos e quatro vezes ao ano (a cada três meses) para pacientes submetidos a alterações no esquema terapêutico (SBD, 2007).

Além das informações da literatura expostas acima, contou-se ainda para a construção das regras, com o apoio de um especialista que atestou a validade das regras para a formação do atributo rótulo.

3.1.3 Mineração de Dados

Definiu-se pelo algoritmo com aprendizagem supervisionada e a representação do conhecimento, utilizando o método de árvore de decisão, considerando que estes métodos são preditivos, pois desempenham inferência nos

dados com o intuito de fornecer previsões ou tendências. Utilizou-se como ferramenta o software livre WEKA (Waikato Environment for Knowledge Analysis, disponível no site www.cs.waikato.ac.nz/ml/weka), com a utilização do algoritmo de classificação J4.8, uma versão do algoritmo C4.5 proposto por Quinlan (QUINLAN,1993).

Para a técnica de teste aplicou-se o método *holdout* estratificado (WITTEN; FRANK, 2000) com 10 repetições (10 simulações), para a matriz M^1 e 10 para M^2 e, para a execução das 10 simulações, foram divididos os subconjuntos aleatoriamente, sendo um dos conjuntos chamado de “Conjunto para Treinamento”, e o outro, “Conjunto de Teste”

3.1.4 Avaliação e interpretação dos resultados

Os resultados do processo de mineração de dados estão analisados utilizando-se da declaração das regras e padrões descobertos, bem como, a matriz de confusão, derivando a acurácia e taxa de erro de toda a amostra e de cada regra.

As categorias adotadas para classificação dos beneficiários foram:

- a) **sem indicativo** representando beneficiários que não possuem atendimentos característicos em relação do diabetes mellitus tipo 2;
- b) **com indicativo** essa classe caracteriza por haver indícios nos atendimentos referentes à doença;
- c) **com forte indicativo** que se caracteriza por haver fortes indícios ao Diabetes Mellitus tipo 2, pelos atendimentos realizados na operadora.

3.1.5 Validação das regras

A seleção das regras para validação pelos especialistas, utilizou o proposto por Carvalho (2005), que sugere o ranqueamento das regras em relação a taxa de acerto e de cobertura. Assim, foram selecionadas as 3 (três) regras com melhor taxa de acerto e cobertura em cada uma das categorias. e as 3 (três) regras da posição mediana do ranqueamento em relação a cobertura e taxa de acerto para cada categoria totalizando 18 regras.

O objetivo da validação das regras foi verificar o quanto as mesmas agregavam valor ao que os especialistas já conheciam sobre o problema ou assunto. Desta maneira pode-se identificar se as regras geraram novos conhecimentos em relação aos conhecimentos prévios dos especialistas.

A população para esta etapa do estudo foi constituída por 4 especialistas da área médica, com atuação também, na área gerencial.

Para esta validação foi elaborado um formulário (Apêndice A) apresentado aos especialistas que tiveram a orientação de preenchimento e deram anuência para participação no estudo. O prazo para preenchimento e devolução do formulário foi de 2 semanas.

Para a validação das 9 (nove) regras com melhores coberturas e taxa de acerto, os dados foram tabulados e avaliados pelo índice de validade de conteúdo (IVC), proposto por Waltz *et al.* (1991), que tem como proposta estimar a validade de um item de mensuração, com base no julgamento dos especialistas.

Para a determinação do índice foi definido um escore para cada uma das três opções de resposta:

- a) **C** - Concordo “a regra confirma o conhecimento do especialista” escore = 2;
- b) **CP** - Concordo parcialmente “a regra não confirma o conhecimento do especialista, mas não apresenta erro ou equívoco nos antecedentes da regra”, escore = 1;
- c) **D** - Discordo “a regra não confirma o conhecimento do especialista e apresenta erro ou equívoco, nos antecedentes da regra” escore = 0.

O processo de geração do índice é obtido considerando as seguintes etapas:

1º em função dos escores e das respostas dos especialistas, somam-se os escores de cada regra.

2º soma-se os totais das regras, resultando o total geral.

3º o total máximo possível é obtido considerando concordância de todos os especialistas em todas as regras.

4º o resultado obtido (IVC) e o valor resultante da tabulação, pelo total máximo possível, multiplicando-se por 100.

$$\text{IVC} = \frac{\text{Total obtido}}{\text{Total máximo possível}} * 100$$

3.2 ASPECTOS ÉTICOS

É importante ressaltar que todos os dados relacionados à identificação dos usuários foram omitidos. Após as primeiras buscas na base de dados, os números de identificação dos usuários para o sistema foram substituídos por números em ordem crescente, a fim de manter a privacidade dos indivíduos.

O projeto foi submetido e autorizado pelo Comitê de Ética em Pesquisa da PUC-PR, nº 2383, anexo II.

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Neste capítulo estão apresentados os resultados observando as etapas propostas para o estudo. Para o processo de análise inicial e levantamento de informações, são apresentadas as informações exploradas pelo ambiente DW para a seleção das variáveis relevantes para o estudo.

No processo de mineração de dados, estão detalhados os resultados alcançados, bem como a avaliação do processo, por meio da matriz de confusão.

No processo de avaliação das regras por especialistas estão apresentados os resultados observando os critérios propostos para o estudo. Ao final deste capítulo, apresenta-se a metodologia proposta para a identificação de beneficiários em relação ao diabetes mellitus tipo 2.

4.1 ANÁLISE INICIAL – IDENTIFICAÇÃO DAS VARIÁVEIS

As variáveis identificadas nesta etapa relacionam-se a idade dos beneficiários, número de consultas por especialidade e exames.

No conjunto selecionado para o estudo retrospectivo do Grupo 1 foram identificadas 59 instâncias¹ e 20.525 eventos realizados no período de 5 anos. No Grupo 2 foram identificadas 59 instâncias¹ e 25.942 eventos realizados. Os internamentos clínicos ocorreram no terceiro trimestre de 2007 e foram avaliados os seus atendimentos de forma retrospectiva nos 5 anos anteriores, ou seja, no período compreendido entre 1 de janeiro de 2003 a 31 de dezembro de 2007.

Para este conjunto de dados foi realizada uma avaliação da idade dos beneficiários, apresentada nos Gráfico 1 e 2.

¹ Instância: A coleção de informações armazenadas no banco de dados (KORTH, 1993);

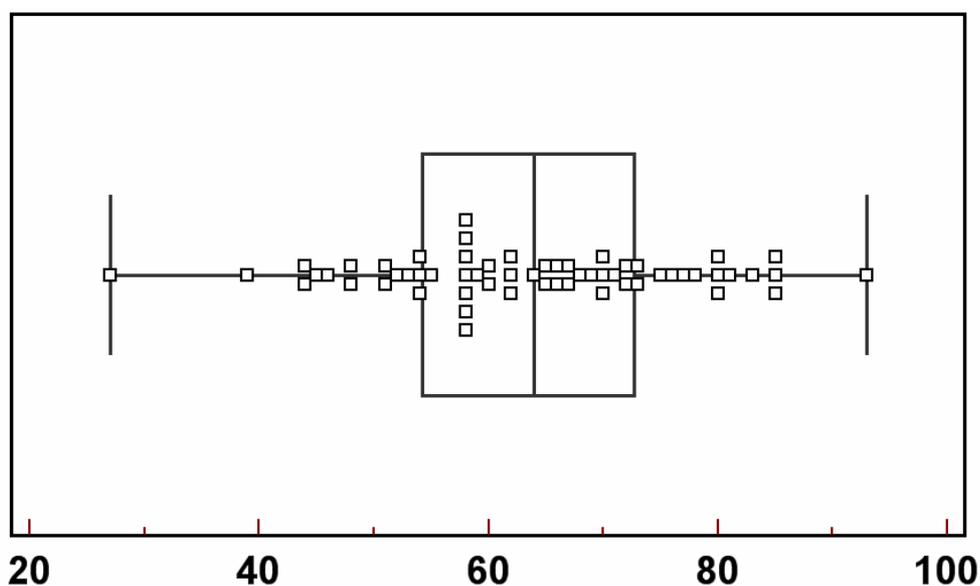


Gráfico 1 – Beneficiários do Grupo 1 com internamento por diabetes segundo idade. Fonte: Base de dados da UNIMEDPR. Jan/2003 a dez/2007

Observa-se como idade mínima, 27 anos, o primeiro quartil 54 anos, a média de 66 anos, o terceiro quartil 75 anos e o valor máximo 93 anos.

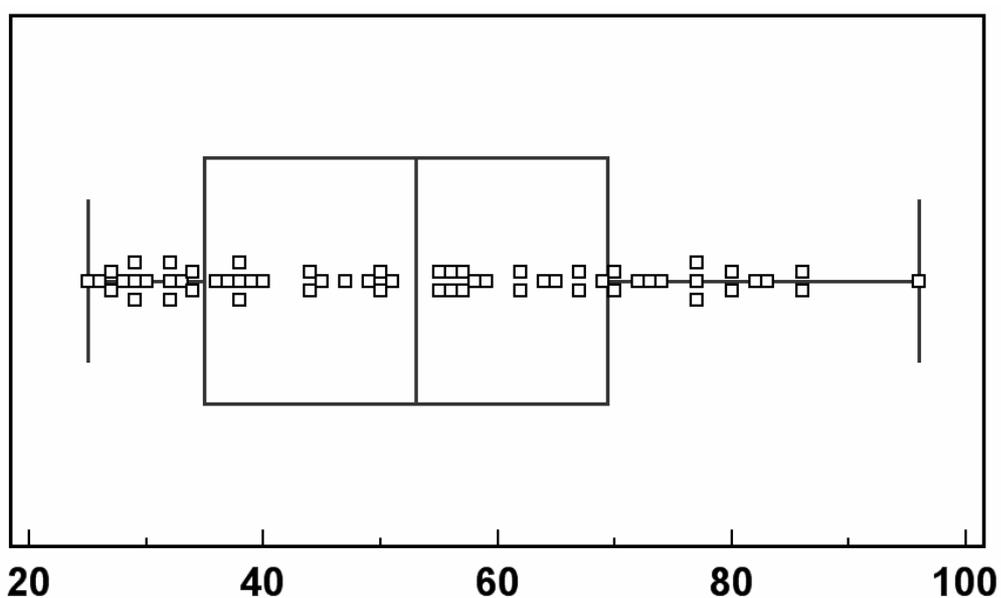


Gráfico 2 – Beneficiários do Grupo 2 com internamento clínico excluídos aqueles por diabetes, segundo idade. Fonte: Base de dados da UNIMEDPR. Jan/2003 a dez/2007

A idade mínima foi de 25 anos (sugestão do estudo), sendo o primeiro quartil de 35 anos, a média de 53 anos, o terceiro quartil 69 anos e o valor máximo de 96 anos.

Desta maneira, verifica-se a diferença encontrada no Grupo do DM2, que possui uma concentração a partir de 54 anos (1º quartil), e o grupo 2 com uma idade menor, a partir de 35 anos.

A informação encontrada é corroborada com o exposto no Caderno de Atenção Básica do Ministério da Saúde sobre o DM2, que referencia como fatores indicativos de maior risco ao DM2, a idade superior a 45 anos (BRASIL,MS 2006) e com as diretrizes da Sociedade Brasileira de Diabetes que referenciam que em geral o DM2 é diagnosticado após os 40 anos (SBD, 2007).

Para a avaliação dos eventos que caracterizaram o atendimento, de cada um dos 59 beneficiários neste período foram realizadas pesquisas considerando as consultas por especialidades e exames solicitados, para o levantamento dos elementos importantes para o estudo.

Inicialmente desenvolveu-se a pesquisa relacionando as consultas, para o levantamento do quantitativo por especialidade médica, obtendo-se as informações do Grupo 1 – caracterizado pelo internamento pela causa do DM2 e o Grupo 2 – internamento clínica, apresentadas no Quadro 3.

Especialidades	Grupo 1	Grupo 2
Clinica Médica	432	365
Cardiologia	236	153
Endocrinologia	177	94
Oftalmologia	166	189
Ginecologia	84	191
Nefrologia	75	1
Psiquiatria	66	19
Gastroenterologia	54	23
Geriatria	54	5
Ortopedia	54	151
Neurologia	46	68

Quadro 3 - Consultas por especialidade realizadas por beneficiários do Grupo 1 e Grupo 2. Fonte:

Base de dados da UNIMEDPR. Jan/2003 a dez/2007.

Ficaram evidenciadas as especialidades cardiologia, endocrinologia e oftalmologia pela freqüência de consultas apresentada, 236, 177 e 166

respectivamente. Foi desprezada a clinica médica com 431 casos, por não favorecer uma associação direta ao diabetes mellitus. Desta maneira selecionaram-se as especialidades cardiologia, endocrinologia e oftalmologia, como variáveis importantes para o estudo. Foi realizada ainda, a inclusão da especialidade Nefrologia por delinear-se como uma especialidade associada diretamente às complicações do Diabetes Mellitus.

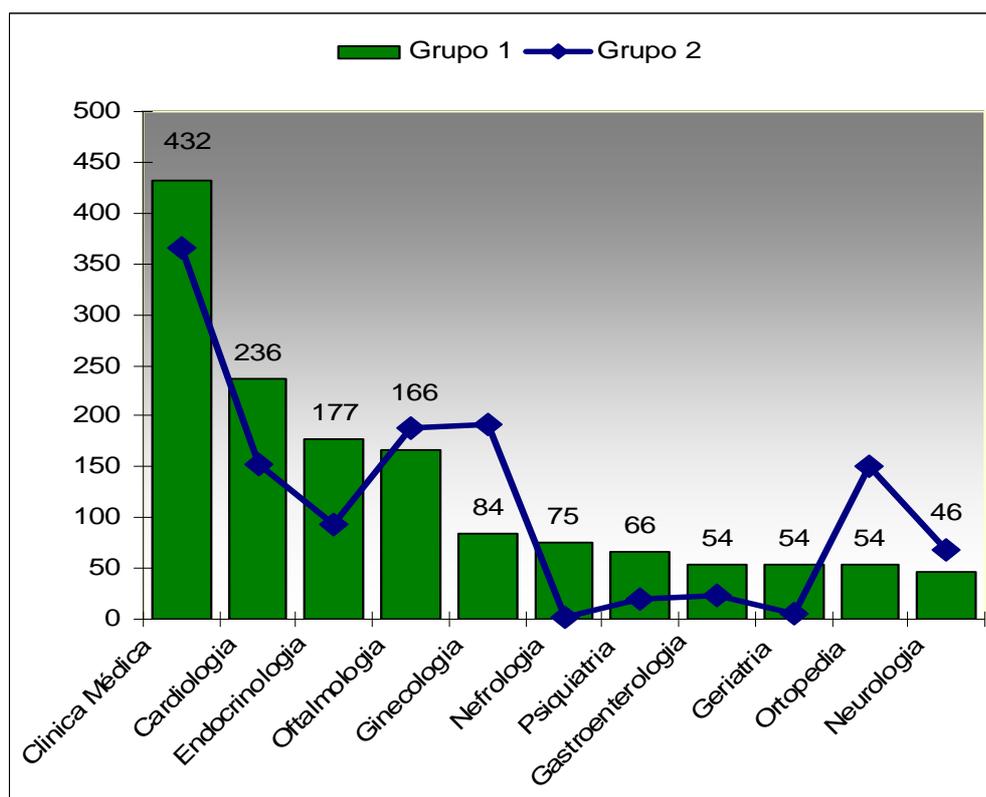


Gráfico 3 – Análise comparativa das especialidades entre Grupo 1 e Grupo 2. Fonte: Base de dados da UNIMEDPR. Jan/2003 a dez/2007.

Na comparação dos resultados do Grupo 1 e do Grupo 2, verificou-se a evidência da endocrinologia, como a especialidade característica do tratamento do DM2 e a cardiologia pela identificação de problemas cardiovasculares e cerebrovasculares, como complicações do DM2 (SBD,2007). A especialidade oftalmologia foi mantida no estudo, considerando que a literatura refere que algumas vezes, o diagnóstico do DM2 é realizado por meio das suas complicações crônicas como a neuropatia e a retinopatia (BRASIL,MS,2006).

Como outra variável realizou-se a pesquisa relacionada a quantidade de exames solicitados aos beneficiários, classificando-os em ordem decrescente de quantidade e selecionando os mais solicitados, sendo os resultados dos Grupos 1 e 2, apresentados no Quadro 4.

Exame Solicitado	Grupo 1	Grupo 2
Glicose	1512	704
Creatinina	887	659
Hemograma Completo	840	838
Potássio	613	430
Sódio	496	378
Uréia	459	415
Colesterol Total	447	343
Hemoglobina Glicosilada	444	57
Rotina de Urina	398	314
Triglicerídios	379	328

Quadro 4 - Exames solicitados por beneficiários do Grupo 1 e Grupo 2. Fonte: Base de dados da UNIMEDPR. Jan/2003 a dez/2007.

Em relação aos exames realizados pelos beneficiários que apresentaram internamento clínico, observa-se também uma diferenciação se comparada pelo grupo de beneficiários internados por DM2, conforme o gráfico a seguir.

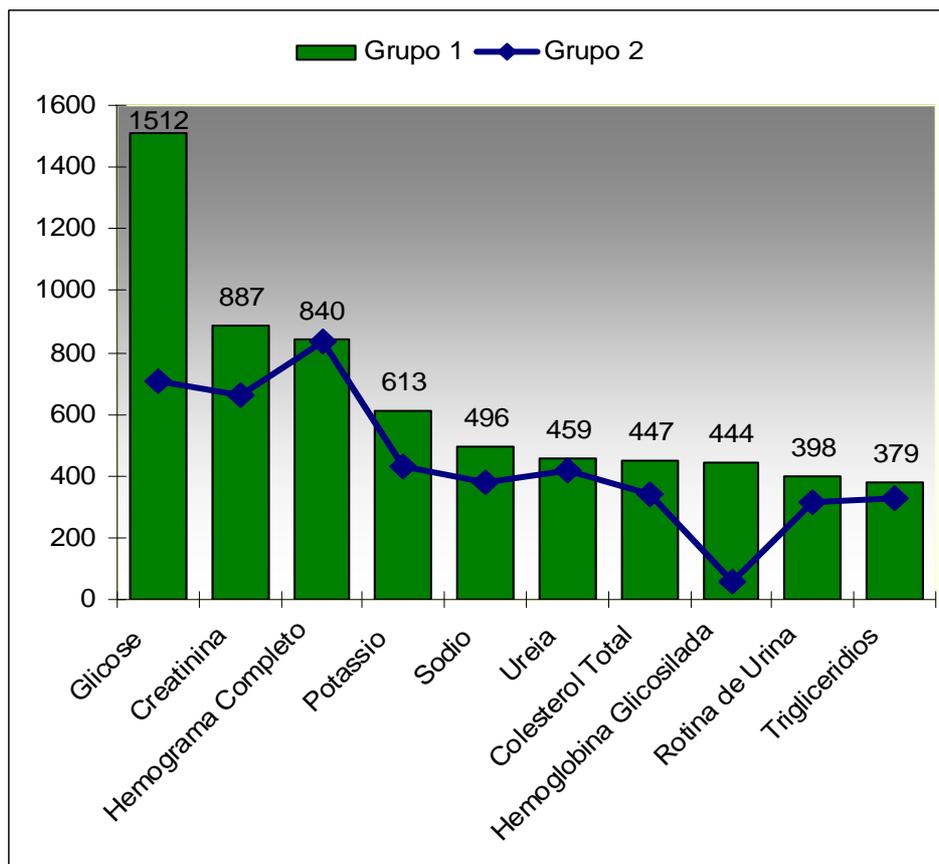


Gráfico 4 – Comparativo dos exames solicitados entre Grupo 1 e Grupo 2. Fonte: Base de dados da UNIMEDPR. Jan/2003 a dez/2007.

Evidencia-se na comparação dos Grupos 1 e 2, a importância e a caracterização do exame Hemoglobina Glicada ou Hemoglobina Glicosilada, como um exame de avaliação inicial para o diagnóstico do diabetes (BRASIL,2006), variável utilizada como identificadora do atributo classe do estudo e conforme o preconizado a avaliação clínica inicial de pacientes com diabetes. O Caderno de Atenção Básica do Ministério da Saúde preconiza que o exame físico para o DM2, deve contemplar o exame de fundo de olho (mapeamento de retina) e como avaliação laboratorial os exames de Glicose (glicemia de jejum), hemoglobina glicosilada (A1C), colesterol total, HDL-C e triglicerídeos, creatinina e microalbuminúria (BRASIL,MS,2006). Kobus (2006), em seu estudo, também referencia a microalbuminúria, hemoglobina glicosilada e o mapeamento de retina como procedimentos específicos e indicativos ao DM2.

Considerando os resultados apresentados, foram selecionados os exames; glicose, creatinina e hemoglobina glicosilada, acrescidos dos exames:

microalbuminúria, curva glicêmica e mapeamento de retina, como exames adicionais e relevantes para o estudo, conforme a literatura e evidências obtidas nos resultados deste estudo.

O Quadro 5, apresenta os atributos relevantes e identificados na análise inicial e selecionados para o estudo.

Exames Laboratoriais	Exames Especiais	Consultas Especialidades	Outras
Glicose Creatinina Microalbuminúria Colesterol Total	Curva Glicêmica Mapeamento de Retina	Oftalmologia Endocrinologia Nefrologia Cardiologia	Idade CID Obesidade

Quadro 5 - Atributos selecionados para o processo de mineração de dados. Fonte: Base de dados da UNIMEDPR. Jan/2003 a dez/2007.

Para cada atributo observou-se:

- a) **exames laboratoriais:** quantidade de exames realizados no período de 5 anos;
- b) **exames especiais:** quantidade dos exames realizados/solicitados no período de 5 anos;
- c) **consultas especialidades:** quantidade de consultas realizadas na especialidade médica dentro do período de 5 anos;
- d) **outras variáveis:** idade, descrita como um atributo numérico identificando a idade do indivíduo. CID de obesidade, ou seja, a quantidade de atendimentos realizados no CID sendo ele o código 'E66', que, refere-se a atendimento de obesidade.

Desta maneira, nesta fase de análise inicial, foram validadas as variáveis relevantes e a regra de formação do atributo classe para o estudo, consonante ao referencial teórico que embasou o estudo.

4.2 PRÉ-PROCESSAMENTO

Utilizou-se para esta etapa, a base de dados de atendimentos da Federação das Unimed's do Estado do Paraná, no período histórico de 6 anos, de 01 de janeiro de 2002 a 31 de dezembro de 2007, sendo selecionados aleatoriamente 60.000 beneficiários, para treinamento e validação de um conjunto total de aproximadamente 1.000.000 de beneficiários ativos do Estado do Paraná (UNIMEDPR,2008).

Os dados foram agrupados e sumarizados por beneficiário, considerando a seguinte estrutura:

	Nome	Domínio	Descrição
1	sexo	Nominal	Sexo do beneficiário; (M) masculino ou (F) feminino.
2	idade	Contínuo	Idade do beneficiário, atributo numérico representando a idade do beneficiário.
3	Qt_cons_nefro	Contínuo	Quantidade total de consultas, realizadas no período de 6 anos, na especialidade médica nefrologia.
4	Qt_cons_oftal	Contínuo	Quantidade total de consultas, realizadas no período de 6 anos, na especialidade médica oftalmologia.
5	Qt_cons_endo	Contínuo	Quantidade total de consultas, realizadas no período de 6 anos, na especialidade médica endocrinologia.
6	Qt_cons_cardio	Contínuo	Quantidade total de consultas realizadas, no período de 6 anos, na especialidade médica cardiologia.
7	Qt_exa_creatinina	Contínuo	Quantidade total de exames de creatinina solicitados no período de 6 anos.
8	Qt_exa_glicose	Contínuo	Quantidade total de exames de glicose (glicemia) solicitados no período de 6 anos.
9	Qt_exa_cur_gli	Contínuo	Quantidade total de exames de curva glicêmica solicitados no período de 6 anos.

10	Qt_exa_microal	Contínuo	Quantidade total de exames de microalbuminúria solicitados no período de 6 anos.
11	St_cid_obesidade	Contínuo	Quantidade total de atendimento realizados no período de 6 anos com o código CID de obesidade.
12	Qt_exa_map_ret	Contínuo	Quantidade total de exames de mapeamento de retina solicitados no período de 6 anos.
13	classe	Rótulo (Categórico)	Atributo meta (classe), podendo ser: sem indicativo, com indicativo e com forte indicativo

Quadro 6 - **Estrutura do registro para mineração de dados.** Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

4.3 MINERAÇÃO DE DADOS

Dentre os algoritmos disponíveis no software WEKA, software de livre utilização, produzido pela Universidade de Waikato – Nova Zelândia, optou-se pelos métodos de classificação C4.5 (QUINLAN,1993). A metodologia para o teste consistiu em aplicar o método de referência cruzada, repetido 10 vezes (10 conjuntos).

Os resultados foram obtidos com a aplicação do algoritmo de árvore de decisão C4.5 (QUINLAN,1993) ou J48 como é denominado na ferramenta, e avaliados com a utilização da matriz de confusão com o cálculo da acurácia e da taxa de erro.

Após a aplicação do método de, em referência cruzada, obteve-se um resultado com taxa de acurácia de aproximadamente 88%, ou seja, dos 60.000 registros utilizados para a etapa de treinamento e validação, 52.817 foram classificados corretamente.

Tabela 1 – **Resultado matriz de confusão.** Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

PREVISTO			REAL
Sem indicativo	Com indicativo	Com forte Indicativo	
50418	1010	362	Sem Indicativo
2469	120	398	Com Indicativo
2692	252	2279	Com forte Indicativo

Na classe sem indicativo obteve-se uma cobertura de 51.790, com 50.418 registros classificados corretamente, o que corresponde a 97,35% de acurácia.

Observa-se ainda, a principal característica das técnicas de algoritmos de árvore de decisão, relacionada à compreensibilidade do resultado final, que pode ser visualizado por meio da estrutura (SE - ENTAO), evidenciando em cada resultado da classificação (folhas), um indicativo dos registros classificados corretamente em relação aos classificados incorretamente.

A Figura 12, apresenta as ramificações da árvore de decisão, extraída pelo algoritmo.

```

QT_EXA_MICROAL <= 0
|  QT_EXA_GLICOSE <= 4: SEM_INDICATIVO (40429.0/1670.0)
|  QT_EXA_GLICOSE > 4
|  |  QT_CONS_ENDO <= 2
|  |  |  QT_EXA_GLICOSE <= 7: SEM_INDICATIVO (7597.0/1090.0)
|  |  |  QT_EXA_GLICOSE > 7
|  |  |  |  idade <= 48
|  |  |  |  |  QT_EXA_GLICOSE <= 15: SEM_INDICATIVO (841.0/101.0)
|  |  |  |  |  QT_EXA_GLICOSE > 15
|  |  |  |  |  |  QT_CONS_ENDO <= 1
|  |  |  |  |  |  |  QT_EXA_CREATININA <= 4
|  |  |  |  |  |  |  |  QT_EXA_MAP_RET <= 1
|  |  |  |  |  |  |  |  |  ST_CID_OBESIDADE<= 7 COM_IND (3.0)

```

Figura 12 – Árvore de decisão, obtida a partir do C4.5 (J48) na base de treinamento. Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

4.4 VALIDAÇÃO DAS REGRAS

4.4.1 Seleção das regras

Para a avaliação das regras por especialistas, foram selecionadas algumas regras adotando os critérios propostos por Carvalho (2005): a) ranqueamento de todas as regras subdivididas em suas classes; b) para cada classe, as regras foram ranqueadas em função da taxa de acerto e cobertura, observando a estrutura do registro conforme Quadro 7:

Nr Regra	Rotulo	Cobertura	Erros	Acertos	Taxa	Ranque
X	Classe	C	E	A	T	N

Quadro 7 - **Estrutura da tabela para ranqueamento das regras.** Fonte: autor da pesquisa.

Onde:

X: numero da regra

Classe: classe identificadora da regra

C: número de elementos classificados pela regra ou cobertura.

E: número de elementos classificados com erro

A: número de elementos classificados com acerto

T: taxa de acerto que corresponde a seguinte fórmula: $A/C * 100$

N: número da posição no ranqueamento.

O ranqueamento gerado foi obtido em função de T e C e nesta ordem respectivamente para cada classe. As regras selecionadas foram identificadas em cada classe, como as melhores e as regras da posição média do ranque em relação

a taxa de acerto e a cobertura. A seguir apresenta-se no Quadro 10 o número total de regras selecionadas e a identificação das regras (melhores e de posição média).

Classe	número Regras	Melhores	Medias	Posição Média no Ranque
Com_indicativo	214	1,2,3	106,107,108	107
Forte_indicativo	411	1,2,3	204,205,206	205,5
Sem_indicativo	481	1,2,3	239,240,241	240,5

Quadro 8 – **Resultado do ranqueamento e seleção das regras.** Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

[...] A seleção das regras com os menores, medianos e maiores ranqueamentos produz grupos distintos de regras, os quais idealmente também deveriam receber graus de interesse distintos pelos usuários. Desta forma, o valor de correlação entre medidas *data-driven* de interesse e o verdadeiro interesse do usuário, obtido a partir destes distintos grupos de regras, se torna mais confiável do que se o valor de correlação tivesse sido calculado selecionando-se apenas regras com ranqueamentos similares das medidas de interesse [...] (CARVALHO, 2005)

Carvalho (2005) comenta ainda, que a opção pela posição mediana do ranque, não sofre influência de valores extremos (*outliers*). Desta maneira foram selecionadas as regras, conforme Quadro 9:

Melhores Regras						
Ordem	Nr. Regra	Classe	Cobertura	Erros	Acertos	Taxa
1	1	sem indicativo	40429	1670	38759	95,87
2	2	sem indicativo	7597	1090	6507	85,65
3	3	sem indicativo	841	101	740	87,99
1	831	com indicativo	8	0	8	100,00
2	835	com indicativo	9	1	8	88,89
3	174	com indicativo	12	3	9	75,00
1	1106	forte indicativo	690	35	655	94,93
2	1094	forte indicativo	269	58	211	78,44
3	1102	forte indicativo	165	10	155	93,94
		Total	50020			
Regras Posição Média						
Ordem	Nr Regra	Classe	Total	Erros	Acertos	Taxa
239	669	sem indicativo	5	1	4	80,00
240	718	sem indicativo	5	1	4	80,00
241	789	sem indicativo	5	1	4	80,00
106	164	com indicativo	2		2	100,00
107	167	com indicativo	2		2	100,00
108	171	com indicativo	2		2	100,00
204	615	forte indicativo	3		3	100,00
205	651	forte indicativo	3		3	100,00
206	747	forte indicativo	3		3	100,00
		Total	30			

Quadro 9 – **Regras selecionadas para o processo de avaliação.** Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

Foram selecionadas 9 (nove) regras de melhor taxas de acerto e cobertura utilizadas para validação de conteúdo pelos especialistas, que representam da população total do estudo, 60.000 registros, 83,37% dos registros classificados pelo algoritmo de árvore de decisão, ou 50.020 registros. Selecionou-se ainda, as 9 (nove) regras de posição mediana no ranqueamento, para análise dos especialistas.

O Quadro 10 apresenta o descritivo de regras com as colunas referentes ao número de acertos e número de erros, com as taxas de acerto e erro, relacionadas à categoria “sem indicativo para o diabetes mellitus”.

Nº da regra	Regra	Nº acertos	Nº erros	% Acerto	% Erro
1	Exame Microalbuminúria = 0 e Exame Glicose <= 4 Então "sem indicativo"	38759	1670	95,87	4,13
2	Exame Microalbuminúria = 0 e Consulta Endocrinologista <= 2 e Exame Glicose >4 e <= 7 Então "sem indicativo"	6507	1090	85,65	14,35
3	Exame Microalbuminúria = 0 e Idade <= 48 e Consulta Endocrinologista <= 2 e Exame Glicose > 7 e <= 15 Então "sem indicativo"	740	101	87,99	12,01
669	Exame Microalbuminúria = 0 e Consulta Endocrinologista > 2 e Sexo Masculino e Exame Creatinina > 6 e Exame Curva Glicêmica = 0 e Consulta Oftalmologista > 0 e Exame Glicose = 7 e Idade > 51 e <= 74 e Atendimento Obesidade <=1 e Consulta Nefrologista = 0 Então "sem indicativo"	4	1	80,00	20,00
718	Exame Microalbuminúria = 0 e Exame Glicose > 8 e Exame Curva Glicêmica <= 0 e Consulta Endocrinologista > 2 e <= 12 e Sexo Feminino e Exame Mapeamento de Retina > 1 e Idade > 37 e <= 55 e Consulta Nefrologista <= 5 e Exame Creatinina > 9 e Consulta Cardiologista <= 7 Então "sem indicativo"	4	1	80,00	20,00
789	Exame Microalbuminúria = 0 e Consulta Endocrinologista >2 e <= 7 e Exame Curva Glicêmica = 0 e Exame Glicose = 9 e Sexo Feminino e Idade > 65 e Exame Mapeamento de Retina > 3 e Atendimento Obesidade = 0 e Exame Creatinina > 5 Então "sem indicativo"	4	1	80,00	20,00

Quadro 10 - Regras selecionadas da classe sem indicativo. Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

O Quadro 11 apresenta o descritivo de regras com as colunas referentes ao número de acertos e número de erros, com as taxas de acerto e erro, relacionadas à categoria “com indicativo para o diabetes mellitus”.

Nº da regra	Regra	Nº acerto	Nº erros	% Acerto	% Erro
831	Exame Microalbuminúria = 0 e Idade > 55 e Atendimento Obesidade <= 27 e Consulta Endocrinologista > 2 e <= 7 e Exame Curva Glicêmica <= 0 e Exame Glicose > 9 e <= 13 e Sexo Feminino e Consulta Nefrologista > 4 e <= 17 e Exame Creatinina > 3 e <= 17 e Consulta Oftalmologista > 2 Então “com indicativo”	8	0	100	0
835	Exame Microalbuminúria = 0 e Idade > 55 e Atendimento Obesidade <= 27 e Consulta Endocrinologista > 2 e <= 5 e Exame Curva Glicêmica > 0 e Exame Glicose > 8 e <= 12 e Sexo Feminino e Consulta Nefrologista <= 8 Então “com indicativo”	8	1	88,89	11,11
174	Exame Microalbuminúria = 0 e Idade > 48 e Exame Glicose > 7 e <= 9 e Consulta Nefrologista <= 1 e Exame Curva Glicêmica = 1 e Consulta Endocrinologista > 0 e <= 1 e Exame Mapeamento de Retina <= 2 Então “com indicativo”	9	3	75,00	25,00
164	Exame Microalbuminúria = 0 e Idade > 48 e Exame Glicose > 4 e <= 11 e Exame Curva Glicêmica = 0 e Exame Creatinina > 6 e Consulta Endocrinologista = 1 e Consulta Nefrologista > 3 e Sexo Masculino e Consulta Cardiologista > 10 Então “com indicativo”	2	0	100,00	0
167	Exame Microalbuminúria = 0 e Exame Curva Glicêmica > 0 e Exame Glicose = 9 e Consulta Nefrologista <= 1 e Consulta Endocrinologista = 0 e	2	0	100,00	0

	Exame Mapeamento de Retina ≤ 7 e Idade > 48 e ≤ 58 e Exame Creatinina = 3 Então "com indicativo"				
171	Exame Microalbuminúria = 0 e Exame Glicose > 7 e ≤ 11 e Exame Curva Glicêmica > 0 e Consulta Nefrologista ≤ 1 e Exame Mapeamento de Retina ≤ 7 e Idade > 58 e ≤ 73 e Consulta Endocrinologista > 0 Então "com indicativo"	2	0	100,00	0

Quadro 11 - Regras selecionadas da classe com indicativo. Fonte: Base de dados da UNIMEDPR.
Jan/2002 a Dez/2007.

O Quadro 12 apresenta o descritivo de regras com as colunas referentes ao número de acertos e número de erros, com as taxas de acerto e erro, relacionadas à classificação com forte indicativo para o diabetes mellitus tipo 2.

Nº da regra	Regra	Nº acerto	Nº erros	% Acerto	% Erro
1106	Exame Microalbuminúria > 0 e Exame Glicose > 4 e Consulta Endocrinologista > 8 Então "com forte indicativo"	655	35	94,93	5,07
1094	Exame Microalbuminúria > 0 e Consulta Endocrinologista > 1 e ≤ 8 e Atendimento Obesidade ≤ 35 e Exame Glicose > 4 e ≤ 13 e Idade > 42 e Exame Creatinina > 2 e ≤ 9 e Consulta Nefrologista ≤ 8 Então "com forte indicativo"	211	58	78,44	21,56
1102	Exame Microalbuminúria > 0 e Exame Glicose > 13 e Consulta Endocrinologista > 1 e ≤ 8 e Atendimento Obesidade ≤ 35 Então "com forte indicativo"	155	10	93,94	6,06
615	Exame Microalbuminúria = 0 e Sexo Feminino e Consulta Endocrinologista > 6 e Atendimento Obesidade ≤ 18 e Idade > 47 e Exame Curva Glicêmica > 0 e Consulta Cardiologista > 11 e Exame Creatinina > 3 e Exame Glicose > 5 e ≤ 7 e	3	0	100,00	0

	Consulta Oftalmologista <= 4 Então "com forte indicativo"				
651	Exame Microalbuminúria = 0 e Sexo Masculino e Atendimento Obesidade <= 9 e Exame Mapeamento de Retina <= 3 e Consulta Oftalmologista <= 10 e Consulta Endocrinologista > 3 e <= 8 e Exame Creatinina > 1 e <= 5 e Exame Curva Glicêmica <= 0 e Exame Glicose = 7 Idade > 75 e Exame Mapeamento de Retina > 1 Então "com forte indicativo"	3	0	100,00	0
747	Exame Microalbuminúria = 0 e Idade > 33 e <= 55 Exame Curva Glicêmica <= 0 e Consulta Endocrinologista > 12 e Sexo Feminino e Atendimento Obesidade > 6 e Exame Glicose = 9 e Consulta Nefrologista <= 1 e Consulta Oftalmologista <= 4 Então "com forte indicativo"	3	0	100,00	0

Quadro 12 - Regras selecionadas da classe com forte indicativo. Fonte: Base de dados da UNIMEDPR. Jan/2002 a Dez/2007.

Em relação às regras que evidenciaram as variáveis que se destacaram como relevantes para o processo de classificação, alguns aspectos demandam análise.

O "Exame Microalbuminúria" apresenta-se na regra 1, a de maior cobertura, com 40.429 registros dos 60.000 contemplados no estudo e com uma taxa de acerto de aproximadamente 96% evidenciando a relevância da variável. Representa a quantidade de exames solicitados de microalbuminúria e configurou-se como o elemento divisor entre indivíduo "sem indicativo" ou "com indicativo ao DM2" em relação à aqueles "com forte indicativo ao DM2". No primeiro caso, são inexistentes as solicitações de microalbuminúria, diferentemente da classe "com forte indicativo", pois das 6 regras, apenas as duas de menor cobertura e taxa de acerto, apresentaram valor inexistente de solicitação do exame.

Almeida (2001), sugere a solicitação do exame de microalbuminúria anualmente, como parte da avaliação e acompanhamento de pacientes com risco de hipertensão ou diabetes.

O “**Exame de glicose**” mostrou-se decisório na classificação. Na regra 1, que refere até 4 solicitações deste exame, ou seja, menos de um ao ano, os registros foram classificados como “sem indicativo ao DM2”. Já a regra 1106, a melhor regra referente a cobertura e taxa de acerto da classe, “forte indicativo ao DM2”, apresenta essa variável, como superior a 4 solicitações em 6 anos.

A glicemia de jejum, ou exame de glicose é considerado pela SBD, como o exame característico para o diagnóstico de pacientes referente ao Diabetes Mellitus (SBD,2007).

Para a “**Consulta na especialidade endocrinologia**” percebe-se na regra 2, “sem indicativo ao DM2” que apresenta uma boa cobertura e taxa de acerto, que o valor não deve ser superior a 2 consultas, diferentemente das regras do rótulo “com indicativo ao DM2”, que se mostram com valores mais expressivos.

Com relação à “**idade**” verifica-se a consonância com a literatura, onde na classe “sem indicativo ao DM2”, a regra 3 exibe a faixa de até 48 anos e na classe “forte indicativo ao DM2”, a regra 1094, com idade superior a 42 anos. Constatou-se ainda, que na classe “com indicativo ao DM2”, todas as regras apresentam-se com valor superior a 48 anos.

Quanto a “**Solicitação de exame de curva glicêmica**” observou-se que as regras classificadas sem indicativo ao DM2 apresentaram na totalidade a inexistência de solicitação do exame em seis anos (669,718,789), diferentemente das regras classificadas como com indicativo e com forte indicativo ao DM2 (835,167,171 e 615).

Em relação ao “**Atendimento de obesidade**” os valores limites de atendimento de obesidade apresentaram-se altos na classe com indicativo e com forte indicativo (regras 831,835,1094,1102,651,747). Para a classe sem indicativo a regra 669 apresentou o valor de até 1 (um) atendimento e a regra 789 apresentou a negativa do atendimento de obesidade.

Para as “consultas na especialidade Nefrologia” observa-se que a regra 1094 da classe com forte indicativo, apresentou um limite de até 8 atendimentos, sendo que na classe com indicativo ao DM2 nas duas regras de melhores coberturas e taxa de acerto, esses valores apresentaram-se de forma semelhantes. Na classe sem indicativo as regras que demonstraram a condição, 669 e 718,

apresentam-se ou com a inexistência do atendimento ou com valor inferior a 5 atendimentos em 6 anos respectivamente.

As demais variáveis como: sexo, consultas na especialidade oftalmologia e cardiologia, exame de creatinina e mapeamento de retina, não apresentaram uma associação representativa entre as regras. Desta maneira acredita-se que não foram determinantes na identificação e diferenciação das classes.

4.4.2 Avaliação dos especialistas

Nesta etapa solicitou-se que cada especialista, por meio do instrumento utilizado realizasse sua avaliação, assinalando como opções, concordo (C) - concordo parcialmente (CP) e discordo (D):

- a) C = A regra confirma o conhecimento;
- b) CP = A regra contraria o conhecimento, mas não apresenta uma ou mais condições no antecedente da regra que representa equívoco ou erro;
- c) NC = A regra contraria o conhecimento, mas apresenta uma ou mais condições no antecedente da regra, que representa equívoco ou erro.

Nº	REGRAS	ESPECIALISTAS			
		A.	B.	C.	D.
1	Se Exame Microalbuminuria = 0 e Exame Glicose <= 4 então "sem indicativo"	C	C	C	C
2.	Exame Microalbuminuria = 0 e Consulta Endocrinologista <= 2 e Exame Glicose >4 e <= 7 então "sem indicativo"	C	C	C	C
3	Exame Microalbuminuria = 0 e Idade <= 48 e Consulta Endocrinologista <= 2 e Exame Glicose > 7 e <= 15	D	C	D	D

	Então "sem indicativo"				
669	Se Exame Microalbuminúria = 0 e Consulta Endocrinologista > 2 e Sexo Masculino e Exame Creatinina > 6 e Exame Curva Glicêmica = 0 e Consulta Oftalmologista > 0 e Exame Glicose = 7 e Idade > 51 e <= 74 e Atendimento Obesidade <=1 e Consulta Nefrologista = 0 Então "sem indicativo"	C	C	C	CP
718	Se Exame Microalbuminúria = 0 e Exame Glicose > 8 e Exame Curva Glicêmica <= 0 e Consulta Endocrinologista > 2 e <= 12 e Sexo Feminino e Exame Mapeamento de Retina > 1 e Idade > 37 e <= 55 e Consulta Nefrologista <= 5 e Exame Creatinina > 9 e Consulta Cardiologista <= 7 Então "sem indicativo"	D	CP	CP	D
789	Se Exame Microalbuminúria = 0 e Consulta Endocrinologista >2 e <= 7 e Exame Curva Glicêmica = 0 e Exame Glicose = 9 e Sexo Feminino e Idade > 65 e Exame Mapeamento de Retina > 3 e Atendimento Obesidade = 0 e Exame Creatinina > 5 Então "sem indicativo"	D	CP	CP	CP

Quadro 13 - Regras da classe sem indicativo. Fonte: Base de dados da UNIMEDPR.

Jan/2002 a Dez/2007.

Observa-se no Quadro 13 que a regra número 1 da categoria "sem indicativo ao diabetes", registrou um consenso dos 4 especialistas, relativo a regra que indica a ausência do exame de microalbuminúria associado com até 5 exames de glicose no período de 6 anos, como uma regra que confirma o conhecimento dos especialistas.

A regra 2 da categoria "sem indicativo ao diabetes", igualmente a anterior, apresentou um consenso referente a confirmação do conhecimento. Ressalta-se que nesta regra, apesar do aumento do número de exames de glicose, até 7 em seis

anos, vinculou-se a variável da quantidade de consultas realizadas na especialidade Endocrinologia, que foi limitada a até duas, no período de 6 anos.

Referente a regra 3 observa-se que 3 dos 4 especialistas consultados, manifestaram que a regra não confirma o conhecimento (D) e que apresenta um erro ou equívoco. O especialista A, considerou um número elevado de exames de glicemia em adultos jovens (idade inferior a 48 anos) e o especialista C, igualmente ao A, afirmou que a regra contraria o seu conhecimento, pois possui um número elevado de solicitação de exames de glicemia de jejum, o que caracteriza a indicação de diabetes.

Para a regra 669, 3 dos especialistas, que afirmaram que a regra confirma o conhecimento e não apresenta nenhum erro ou equívoco, e apenas o especialista D indicou que a regra, apesar de não confirmar o seu conhecimento, não apresentou erros ou equívocos em seus antecessores.

Com relação a regra 718 observa-se que o Especialista A comenta sobre a limitação de 2 até 12 consultas na especialidade Endocrinologia, associada à realização de exame de mapeamento de retina, considerando que estas variáveis associadas à quantidade de exames de glicemia superior a 8, representam um equívoco ou erro. Os especialistas B e C consideraram que a regra não confirma os seus conhecimentos, porém as regras não representam um erro ou equívoco, pois associadas tornam-se mais indicativas de obesidade e hipertensão, do que de Diabetes.

Na regra 789 o especialista A relata que a regra contraria o seu conhecimento, e que os antecedentes da regra referentes a idade superior a 65, mapeamento de retina superior a 3, creatinina superior a 5 e número elevado de exames de glicemia, representa equívoco ou erro, para o conseqüente (classe) “sem indicativo”. Já os especialistas B, C e D, afirmaram que a regra contraria o conhecimento, mas não apresenta erro ou equívoco.

Nº	REGRAS	ESPECIALISTAS			
		A.	B.	C.	D.
831	Se Exame Microalbuminuria = 0 e Idade > 55 e Atendimento Obesidade <= 27 e Consulta Endocrinologista > 2 e <= 7 e Exame Curva Glicemica <= 0 e Exame Glicose > 9 e <= 13 e Sexo Feminino e Consulta Nefrologista > 4 e <= 17 e Exame Creatinina > 3 e <= 17 e Consulta Oftalmologista > 2 Então "com indicativo"	C	C	C	D
835	Se Exame Microalbuminuria = 0 e Idade > 55 e Atendimento Obesidade <= 27 e Consulta Endocrinologista > 2 e <= 5 e Exame Curva Glicemica > 0 e Exame Glicose > 8 e <= 12 e Sexo Feminino e Consulta Nefrologista <= 8 então "com indicativo"	C	C	C	D
174	Se Exame Microalbuminuria = 0 e Idade > 48 e Exame Glicose > 7 e <= 9 e Consulta Nefrologista <= 1 e Exame Curva Glicemica = 1 e Consulta Endocrinologista > 0 e <= 1 e Exame Mapeamento de Retina <= 2 Então "com indicativo"	CP	C	C	CP
164	Se Exame Microalbuminuria = 0 e Idade > 48 e Exame Glicose > 4 e <= 11 e Exame Curva Glicemica = 0 e então "com indicativo" Exame Creatinina > 6 e Consulta Endocrinologista = 1 e Consulta Nefrologista > 3 e Sexo Masculino e Consulta Cardiologista > 10 Então "com indicativo"	CP	CP	CP	D
167	Se Exame Microalbuminuria = 0 e Exame Curva Glicemica > 0 e Exame Glicose = 9 e Consulta Nefrologista <= 1 e Consulta Endocrinologista = 0 e Exame Mapeamento de Retina <= 7 e Idade > 48 e <= 58 e Exame Creatinina = 3 Então "com indicativo"	CP	CP	C	CP
171	Se Exame Microalbuminuria = 0 e	CP	C	C	C

Exame Glicose > 7 e <= 11 e				
Exame Curva Glicêmica > 0 e				
Consulta Nefrologista <= 1 e				
Exame Mapeamento de Retina <= 7 e				
Idade > 58 e <= 73 e				
Consulta Endocrinologista > 0				
Então "com indicativo"				

Quadro 14 - Regras da classe com indicativo. Fonte: Base de dados da UNIMEDPR.

Jan/2002 a Dez/2007.

No Quadro 14, a regra 831 mostra a concordância dos especialistas A, B e C de que a regra confirma o conhecimento, considerando a utilização dos beneficiários em 6 anos. O especialista D afirma que a regra não representa o seu conhecimento e que apresenta erro ou equívoco em seus antecessores.

Para a regra 835 observa-se a concordância dos especialistas A, B e C que a consideraram consistente em relação às regras e desta forma confirmam seus conhecimentos, e o especialista D discorda totalmente da regra.

Para a regra 174, constatou-se que os especialistas A e D afirmaram que a regra não confirma o seu conhecimento, considerando que as variáveis são pouco específicas para a classe "com indicativo", porém consideraram também, que as variáveis não apresentam erro ou equívoco. Os especialistas B e C, afirmaram concordarem com a regra e que a mesma reflete os seus conhecimentos.

Referente a regra 164, o especialista A salienta que a mesma não confirma o seu conhecimento, pois as variáveis são pouco específicas para a classe com indicativo, porém considerou, que as variáveis não apresentam erro ou equívoco. Os especialistas B e C concordaram com A, porém afirmaram a associação das variáveis a pacientes com indicativo de hipertensão arterial e a problemas cardíacos. Já o especialista D indicou discordância total das informações.

Referente a regra 167 os especialistas A, B e D afirmaram que a regra não confirma o conhecimento, porém não representa erro ou equívoco. O especialista C manifestou concordar com o conhecimento gerado pelo algoritmo.

Na regra 171, o especialista A afirmou não confirmar o seu conhecimento, pois as variáveis são pouco específicas para a consideração da classe "com indicativo", porém considerou que os antecessores da regra, não apresentam erro

ou equívoco. Os especialistas B, C e D, afirmaram concordância com a regra e que a mesma reflete os seus conhecimentos.

Nº	REGRAS	ESPECIALISTAS			
		A.	B.	C.	D.
1106	Exame Microalbuminúria > 0 e Exame Glicose > 4 e Consulta Endocrinologista > 8 Então "com forte indicativo"	C	C	C	C
1094	Se Exame Microalbuminúria > 0 e Consulta Endocrinologista > 1 e <= 8 e Atendimento Obesidade <= 35 e Exame Glicose > 4 e <= 13 e Idade > 42 e Exame Creatinina > 2 e <= 9 e Consulta Nefrologista <= 8 Então "com forte indicativo"	C	C	C	C
1102	Se Exame Microalbuminúria > 0 e Exame Glicose > 13 e Consulta Endocrinologista > 1 e <= 8 e Atendimento Obesidade <= 35 Então "com forte indicativo"	C	C	C	C
615	Se Exame Microalbuminúria = 0 e Sexo Feminino e Consulta Endocrinologista > 6 e Atendimento Obesidade <= 18 e Idade > 47 e Exame Curva Glicêmica > 0 e Consulta Cardiologista > 11 e Exame Creatinina > 3 e Exame Glicose > 5 e <= 7 e Consulta Oftalmologista <= 4 Então "com forte indicativo"	CP	CP	CP	D
651	Se Exame Microalbuminúria = 0 e Sexo Masculino e Atendimento Obesidade <= 9 e Exame Mapeamento de Retina <= 3 e Consulta Oftalmologista <= 10 e Consulta Endocrinologista > 3 e <= 8 e Exame Creatinina > 1 e <= 5 e Exame Curva Glicêmica <= 0 e Exame Glicose = 7 Idade > 75 e Exame Mapeamento de Retina > 1 Então "com forte indicativo"	D	CP	CP	CP
747	Se Exame Microalbuminúria = 0 e Idade > 33 e <= 55	D	C	D	C

Exame Curva Glicêmica <= 0 e Consulta Endocrinologista > 12 e Sexo Feminino e Atendimento Obesidade > 6 e Exame Glicose = 9 e Consulta Nefrologista <= 1 e Consulta Oftalmologista <= 4 Então "com forte indicativo"				
---	--	--	--	--

Quadro 15 - Regras da classe com forte indicativo Fonte: Base de dados da UNIMEDPR .

Jan/2002 a Dez/2007.

A primeira regra 1106 da categoria "com forte indicativo" do Quadro 15, registrou a concordância de todos os especialistas A, B, C e D referentes a confirmar o conhecimento gerado pelo algoritmo. O especialista A, citou a presença da solicitação do exame de microalbuminúria, como uma variável importante para a identificação da classe "com forte indicativo".

As regras 1094 e 1102 obtiveram concordância dos especialistas A, B, C e D.

A regra 615 apresentou concordância entre os especialistas que afirmaram que a regra não confirma o conhecimento, porém representam erro ou equívoco. O especialista C comenta estar surpreso sobre a baixa solicitação de exames de glicose e a inexistência da solicitação do exame de microalbuminúria, porém não a considera como um erro ou equívoco. O especialista D afirmou discordar totalmente da regra.

Na regra 651, o especialista A afirmou que a regra não confirma o seu conhecimento e que apresenta equívoco por apresentar uma quantidade muito pequena de solicitação do exame de glicemia, associada a inexistência da solicitação do exame de curva glicêmica e microalbuminúria. Os especialistas B, C e D afirmaram que a regra não confirma o conhecimento, porém não apresenta equívoco ou erro nos antecessores da regra.

Para a regra 747, o especialista B e D afirmaram que a mesma valida os seus conhecimentos, porém os especialistas A e C discordam que a regra confirma o conhecimento e que os antecedentes da regra apresentam erros ou equívocos, considerando a idade, baixa solicitação de exame de microalbuminúria e curva glicêmica. O especialista C descreveu que essa regra representa melhor a classe "com indicativo" de que a "com forte indicativo".

O Quadro 16 apresenta o resultado geral da validação do conteúdo das regras.

Nº	A.	B.	C.	D.	Total
1	2	2	2	2	8
2	2	2	2	2	8
3	1	2	1	0	4
831	2	2	2	0	6
835	2	2	2	0	6
174	1	2	2	1	6
1106	2	2	2	2	8
1094	2	2	2	2	8
1102	2	2	2	2	8

Quadro 16 - Demonstrativo geral da validação (pontuação) das regras por especialistas.

Curitiba. Maio / 2009. Fonte: autor da pesquisa

Observa-se no quadro 16, referente a avaliação das regras com melhores taxas de acerto e cobertura, ter havido concordância entre a avaliação dos especialistas com o resultado obtido pela matriz de contingência, onde as melhores taxas de acertos ocorreram nas regras das classes sem indicativo e com forte indicativo (1,2,3 e 1106,1094,1102).

4.4.3 Índice de validade de conteúdo (IVC)

Nesta etapa os dados foram tratados para a verificação do grau de validade do conteúdo das regras obtidas, por meio do índice de validade de conteúdo proposto por Waltz *et al.* (1991). Este índice possibilita estimar a validade de um item de mensuração, com base nas respostas obtidas dos especialistas, quanto à relação entre o objeto utilizado para mensuração e os objetivos da medida.

Para a determinação do índice, determinou-se o seguinte escore para cada opção da resposta, como segue: Para a opção, concordo “a regra confirma o

conhecimento do especialista”, escore = 2, para a opção concordo parcialmente, “a regra não confirma o conhecimento do especialista, mas não apresenta erro ou equívoco nos antecedentes da regra”, escore = 1 e para a opção discordo, “a regra não confirma o conhecimento do especialista e apresenta erro ou equívoco, nos antecedentes da regra”, escore = 0.

O processo de geração do índice foi obtido considerando:

1º Em função dos escores e das respostas dos especialistas, foram somados os escores de cada regra.

2º Soma-se os totais das regras, obtendo-se o total geral.

3º O total máximo foi obtido considerando concordância de todos os especialistas em todas as regras, ou seja, $4 \times 2 \times 9 = 72$

4º O resultado obtido e o valor resultante da tabulação, pelo total máximo possível multiplicando-se por 100.

Total máximo possível = 72

Total obtido = 63

IVC = $62/72 \times 100 = 86,11\%$

O resultado deste estudo permite concluir que as 9 (nove) regras com melhor taxa de acerto e cobertura encontradas pelo processo de mineração de dados, após a análise por profissionais da área, obteve-se 86.11% de concordância, desta maneira verificou-se a pertinência do conteúdo do item no domínio específico e foi considerada adequada, ou seja, concordância superior a 80%. (WALTZ *et al*, 1991).

4.5 METODOLOGIA PROPOSTA

Apresenta-se a proposta metodológica desenvolvida neste estudo, com etapas e acessos às informações necessárias para o processo de descoberta de novos conhecimentos.

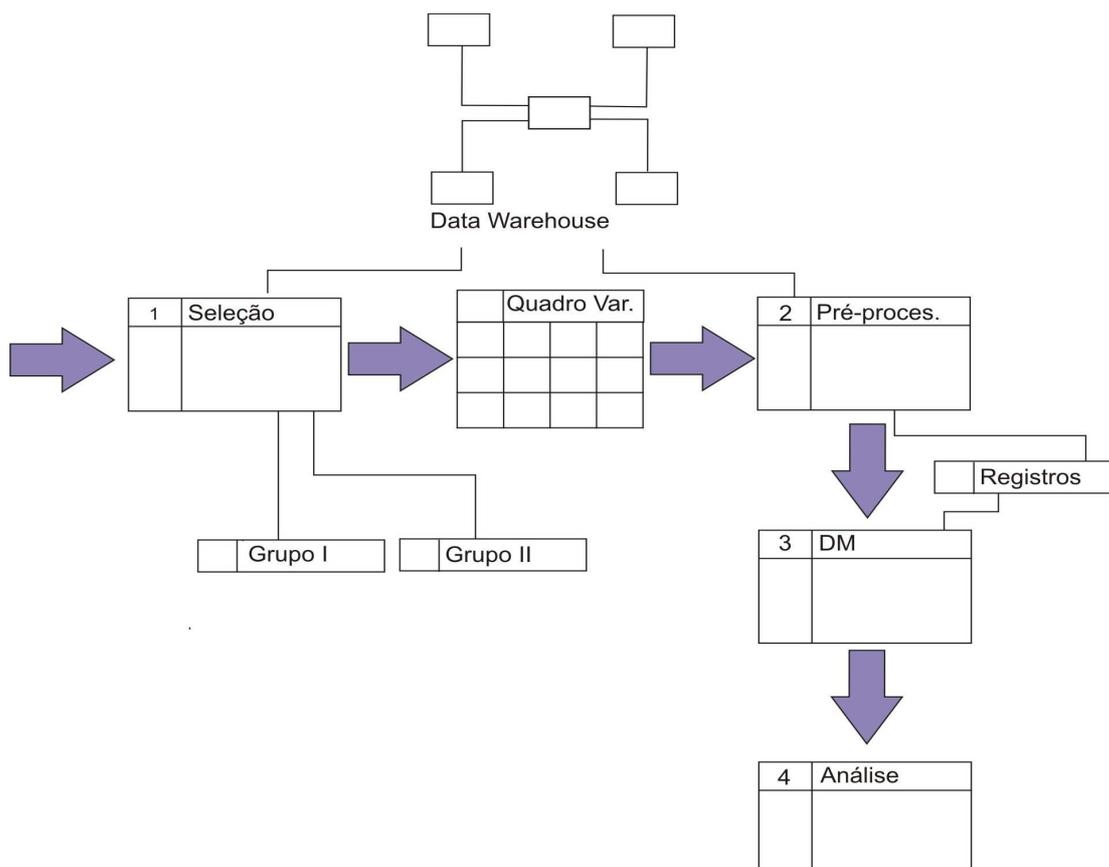


Figura 13 – Representação da Metodologia Proposta. Fonte: O próprio autor.

A primeira etapa, seleção e análise das variáveis para o estudo, se subdivide em: **análise retrospectiva**, com criação de dois grupos – Grupo I, de pacientes internados pelo motivo da doença do estudo e Grupo II, de pacientes internados por motivo clínico excetuando-se o CID da doença em estudo.

Os dados obtidos são submetidos a **análise comparativa**, onde se avaliam os grupos e se comparam os eventos realizados pelos beneficiários no período estabelecido, de forma retrospectiva tomando como ponto de partida o internamento.

O levantamento e **análise da literatura** são realizados para identificação de informações específicas, relacionadas às variáveis analisadas. As informações e formação dos grupos para análise são obtidas por meio de exploração de dados, o que facilita a exploração e seleção das informações. O resultado desta etapa é a seleção das variáveis, que servirão de subsídio para a estruturação do arquivo.

Na segunda etapa, pré-processamento com as informações do quadro de variáveis obtido na etapa anterior e por meio da extração de informações do

ambiente *data warehouse* é gerado o arquivo com a estrutura baseada nas variáveis selecionadas, por beneficiário e com os valores cumulativos no período de 6 anos. Nesta etapa se determina também a formação do atributo rótulo, que identifica a classe de cada registro do arquivo, de acordo com as categorias estabelecidas: “sem indicativo”, “com indicativo” e “com forte indicativo” a doença.

Na terceira etapa, a mineração de dados utiliza-se um algoritmo para a tarefa de classificação o C4.5 (QUINLAN,1993) para a descoberta das regras. Neste trabalho utilizou-se o algoritmo de árvore de decisão, devido a sua principal característica que é a compreensibilidade na apresentação das regras.

Na quarta etapa são analisados e validados os resultados obtidos, por especialistas, para verificar a concordância em relação aos seus conhecimentos em relação a uma amostra das regras encontradas pelo algoritmo. Para a seleção, utilizou-se o método de amostragem aleatória estratificada, onde de cada estrato obteve-se as 3 primeiras regras relacionadas a melhor cobertura e menor taxa de acerto e as 3 medianas também, referente a cobertura e taxa de acerto. Desta maneira selecionou 18 (dezoito) regras de 1.106 para a avaliação dos especialistas.

Esta metodologia foi aplicada com informações de um ambiente real de uma operadora de plano de saúde, extraíndo regras pelo processo de mineração de dados, que puderam ser avaliadas por especialistas da área médica, para a classificação de beneficiários em relação ao Diabetes Mellitus Tipo 2.

Portanto, apresenta-se uma metodologia que contempla o processo de descoberta de conhecimento de base de dados e na etapa de mineração de dados utiliza o algoritmo de árvores de decisão C45 (QUINLAN,1993), para a identificação e seleção de beneficiários com indicativos ao diabetes, por meio de seu ciclo de atendimento na operadora, para de forma a favorecer uma atenção gerenciada à saúde nos diferentes níveis de atenção.

Vislumbra-se a partir desta metodologia, a possibilidade de estabelecimento de indicadores que monitorem todo o processo de prestação de serviços, considerando que combinar a visão de indicadores focado na estratégia das operadoras, com ênfase nas ações de promoção à saúde e prevenção, reforça o conceito de transformar um indicador intangível em um resultado direto para o beneficiário e, conseqüentemente, reduzir os custos das operadoras de planos de saúde (PORTER e TEISBERG, 2007).

Desta forma, os indicadores deverão servir de elementos para a execução de ações de promoção, prevenção, gerenciamento de doença e gerenciamento de caso. Praticam um novo paradigma assistencial aqueles que utilizam o conhecimento moderno, na medida de sua capacidade, para desenvolver e manter a saúde, evitar a doença e suas conseqüências e prolongar a vida com qualidade.

5 CONSIDERAÇÕES FINAIS

O usuário da saúde se diferencia do consumidor comum, por não deter os conhecimentos necessários para a sua tomada de decisão sobre o que irá dos serviços e desta maneira não cabem as leis comuns do mercado, como a livre escolha e a concorrência.

Muitas vezes o consumo é imposto ao usuário por situações de emergência e desta maneira os gestores da saúde devem trabalhar com saberes que resultem em produção do cuidado em saúde. O diferencial de ação na produção da saúde relaciona-se às tecnologias voltadas para a produção do cuidado apostando em novas relações entre os prestadores e pacientes, tentando gerir um serviço de saúde centrado nos clientes e suas necessidades (MALTA *et al*, 2004).

Ceclilio (2009) salienta a importância das empresas e profissionais de saúde de serem capazes ou de estarem abertos para propor novas formas de cuidado, adequados a cada situação singular do cliente, associadas a construção de práticas gerenciais que valorizem estas necessidades.

As empresas de saúde de uma maneira em geral, possuem uma grande quantidade de informações transacionadas e processadas em relação aos atendimentos realizados e com isto, uma grande dificuldade na extração de novos conhecimentos dessas bases de dados. Além desta dificuldade, as operadoras de planos de saúde apresentam ainda como agravante a ausência de informações clínicas dos pacientes, pois muitas vezes, não possuem uma rede própria de atendimento, o que dificulta o acesso e a concentração dessas informações.

Associando-se a esta questão a posição recente da classe médica e do judiciário brasileiro em relação à privacidade das informações dos usuários de plano de saúde, que confere a não obrigatoriedade da informação do CID realizadas nos atendimentos, o que dificulta ainda mais a realização dos programas de promoção e prevenção de doenças considerando as atuais bases de dados disponibilizadas nas operadoras de saúde (CREMESP,2007)

Neste contexto, surgiu a motivação pelo desenvolvimento de uma metodologia que, mesmo sem as informações clínicas dos pacientes, possa explorar as informações dos atendimentos realizados pelos beneficiários no ciclo de vida da

operadora, para a identificação e classificação em relação a doenças crônicas não transmissíveis.

Ressalta-se que deve sempre existir o compromisso de se manter a privacidade e a confidencialidade das informações e conhecimentos gerados. Freire (2004) comenta que no estágio atual de desenvolvimento tecnológico, o potencial para a invasão da privacidade e uso inescrupuloso das informações não deve ser desprezado e medidas devem ser tomadas para disciplinar a disponibilidade e o acesso e garantir a segurança, privacidade e confidencialidade das informações.

Em um processo de exploração de dados e geração de novos conhecimentos da área de saúde, a troca de informação deve ser usada de forma correta, sem ameaça a privacidade desta ou daquela pessoa, em especial observando as disposições trazidas pela Constituição da República, pelo Código Civil, pelo Código de Defesa do Consumidor e pelas normas do Conselho Federal de Medicina e outros Conselhos de Classe (CAMPOS *et al*,2008).

O conhecimento gerado por essa metodologia deve ser garantido no tangente à segurança, privacidade e confidencialidade das informações e os resultados direcionados principalmente para a melhoria da qualidade de vida e de saúde de seus beneficiários, e não apenas para a otimização dos custos das operadoras.

O desenvolvimento deste estudo possibilitou a implementação e validação de um processo de descoberta de conhecimento em base de dados de uma operadora de plano de saúde, para o auxílio na seleção de beneficiários elegíveis para programas de promoção, prevenção e gerenciamento de caso direcionados ao Diabetes Mellitus tipo 2, utilizando dados e informações dos atendimentos realizados por estes beneficiários.

A metodologia contemplou a fase de concepção do modelo inicial, propondo a manipulação e a exploração das informações para a seleção das variáveis relevantes para o processo de mineração de dados, onde além da exploração das informações, por meio da base data warehouse, contou-se com o apoio de um especialista para o auxílio em sua definição. Ainda, neste modelo observou-se a etapa que avaliou o processo de mineração de dados realizado pelo algoritmo de árvore de decisão C4.5 (QUINLAN,1993), onde especialistas avaliaram uma amostra das regras encontradas, validando-as.

Esta metodologia mostrou-se um processo que pode ser adotado para a seleção de beneficiários para programas de promoção e prevenção e gerenciamento de doenças crônicas não transmissíveis, programas estes, que se apresentam mais eficientes na redução de custos para a operadora e na melhoria da qualidade de vida para os beneficiários, quando utilizados no conceito de prevenção primária, ou seja, antes da manifestação da doença.

A seleção e inserção desses beneficiários, por meio desta metodologia, em programas de promoção e prevenção da saúde, permitirá a identificação das necessidades de saúde dos clientes, bem como a organização e o auxílio no dimensionamento dos serviços para supri-las.

Esta metodologia possibilitará as operadoras de plano de saúde, abordar os diversos graus de necessidades de saúde, contemplando a gestão de fatores ou comportamentos de risco, a fim de evitar ou minimizar os agravamentos das condições de saúde dos clientes, promovendo desta maneira, a detecção precoce e um tratamento adequado das doenças e suas complicações.

Identifica-se ainda, a possibilidade do gerenciamento dos pacientes elegíveis por esta metodologia, a um controle por meio de sistemas específicos a esse fim de forma integrada, como exemplo o GERHAS, que direciona o paciente a uma melhor condição clínica e estabilização de risco do paciente e conseqüentemente pelo ganho na qualidade de vida. Salienta a importância da utilização de ferramentas que permitam levantar o perfil de pacientes, e a descoberta de novos conhecimentos integradas com a sua ferramenta. (AQUINO,2006).

5.1 TRABALHOS FUTUROS

Pretende-se em trabalhos futuros, estender esta metodologia para outras doenças crônicas não transmissíveis, numa perspectiva de adoção de um novo paradigma assistencial e melhoria de qualidade de vida dos beneficiários.

Para trabalhos futuros vislumbra-se a utilização de agentes inteligentes de forma integrada a base de dados dos de sistema de processamento analítico - OLAP e integrado com sistema de mineração de dados - OLAM, de maneira que possam selecionar novos casos de elegíveis, de maneira automática e de forma integrada, para o processo de prevenção e promoção a saúde.

Um agente é uma entidade capaz de perceber seu ambiente por meio de sensores e de agir sobre o ambiente por meio de atuadores. A função do agente relativa a um agente especifica a ação executada pelo agente em resposta a qualquer seqüência de percepções e a medida de desempenho avalia o comportamento do agente no ambiente. Um agente racional age para realizar uma melhor medida de desempenho (RUSSEL e NORVIG, 2004).

Uma especificação de ambiente da tarefa inclui a medida de desempenho, o ambiente externo, os atuadores e os sensores e eles variam ao longo de diversas dimensões. O programa de um agente implementa a função do agente e pode melhorar seu desempenho por meio do aprendizado, o que permite afirmar que agentes inteligentes são mecanismos integrados as bases de dados para gerar, de forma automatizada, novos conhecimentos.

Desta maneira a incorporação dessa metodologia aplicando a integração de ambientes OLAP e OLAM em conjunto com agentes inteligentes, permitirá uma automação do processo, possibilitando a identificação ágil e automática dos beneficiários elegíveis aos programas de promoção e prevenção da saúde, com base em informações de seus atendimentos realizados na operadora.

REFERÊNCIAS

ALMEIDA, Fernando A. Microalbuminúria como marcador precoce de comprometimento da função renal. **Revista Brasileira de Hipertensão**. São Paulo, v. 8, n.3, p 347-348, jul. 2001.

ALMEIDA, Naomar F.; ROUQUAYROL, Maria Z. **Introdução a Epidemiologia**. 4. ed Rio de Janeiro: Guanabara Koogan, 2006. 296 p.

ARNOLD, William R; BOWIE John S. **Artificial Intelligence: a personal, commonsense journey**. New Jersey: Prentice– Hall, Inc, 1986. 219 p.

ARAÚJO, Mário Luiz Cardoso. **Gerencia de Assistência à Saúde no Setor de Saúde Suplementar: Uma Experiência**. Dissertação de Mestrado. Disponível em:<http://www.ans.gov.br/portal/upload/biblioteca/DIS_Gerencia_Assistencia_Saud_e_Setor_Saude_Suplementar_Mario_Araujo.pdf>> Publicado em jun. 2004. Acesso em 8 ago 2008.

BARRETO et al. The Increase of diabetes mortality burden among Brazilian adults. **Revista Panamericana de Salud Publica**. Washington, v.22. n.4, p. 239-245 out. 2007.

BELLMAN, Richard E. **An Introduction to Artificial Intelligence: Can Computers Think?** San Francisco: Boyd & Fraser Publishing Company, 1978. 146 p.

BITTENCOURT, Guilherme. **Inteligência Artificial: Ferramentas e Teorias**. Florianópolis: Editora da UFSC, 1998. 362 p.

BRASIL. Agência Nacional de Saúde Suplementar. **Manual técnico de promoção da saúde e prevenção de riscos e doenças**. Rio de Janeiro: ANS, 2006. 65 p.

BRASIL. Conselho Nacional de Secretários de Saúde. **Saúde Suplementar**. Brasília: CONASS, 2007. 234 p.

BRASIL. Ministério da Saúde. **Cadernos de Atenção Básica - n. 16, Serie A. Normas e Manuais Técnicos**. Brasília: Secretaria de Atenção à Saúde, Ministério da Saúde, 2006. 64 p.

BRASIL. Ministério da Saúde. **Diretrizes e Recomendações para o Cuidado Integral de Doenças Crônicas Não-Transmissíveis: Promoção da Saúde**. Brasília: Ministério da Saúde: 2008.

CAMPOS et al. **Modelo Cuidador da Federação das Unimeds de Minas Gerais**. Belo Horizonte, MG: Federação das Unimeds de Minas Gerais, 2008. 429 p.

CARVALHO, Deborah R. **Mineração de Dados através de indução de regras e algoritmos genéticos**. Dissertação de Mestrado. Programa de Pós-Graduação em Informática Aplicada. Pontifícia Universidade Católica do Paraná. Curitiba, 1999.

CARVALHO, Deborah R. **Algoritmo Genético para Tratar o Problema de Pequenos Disjuntos em Classificação de Dados**. Dissertação de Doutorado. COPPE, Ed. UFRJ, 2005.

CECILIO, Luiz C. O. A morte de Ivan Litch, de Leon Tostói: Elementos Para Se Pensar as Múltiplas Dimensões da Gestão do Cuidado. **Revista Interface Comunicação Saúde Educação**. V.13, supl. I, p.545-555, 2009.

São Paulo, Conselho Regional de Medicina do Estado de São Paulo, Educação Continuada, Notícias de 13 de julho 2007. Acessado em 21 jun 2009, disponível em: <http://www.cremesp.org.br/?siteAcao=Noticias&id=1377>>

CZERESNIA, Dina. Ações de Promoção à Saúde e Prevenção de Doenças: O Papel da ANS. Trabalho Técnico. Rio de Janeiro. Agencia Nacional de Saúde Suplementar, jul.2003, p. 1-35. Disponível em:

<http://www.ans.gov.br/portal/upload/biblioteca/TT_AS_02_DCzeresnia_AcoesPromocaoSaude.pdf> Acesso em 22 fev 2009.

FAYYAD, U. PIATESKY-SHAPIO, G., SMYTH, P. **Advances in knowledge discovery and data mining**. Boston: MIT Press, 1996. p 83-113.

FEIGENBAUM, Edward. A.; MCCORDUCK, Pamela. **The Fifth Generation: artificial Intelligence na Japan's Computer Challenge to the World** – Londres: Addison-Wesley, 1983. 275 p.

FREIRE, Miranda Sergio, **Sigilo das Informações**. Disponível em: <http://www.ans.gov.br/portal/upload/biblioteca/TT_AS_19_SMirandaFreire_SigiloInformacoes.pdf>. Acesso em 28 mar. 2007.

GOLDSCHMIDT, R.; PASSOS, E. **Mineração de Dados: um guia prático**. Rio de Janeiro: Elsevier, 2005. 261 p.

HAN, Jiawei; KAMBER Micheline. **Data Mining: concepts and Techniques**. Califórnia, USA: Morgan Kaufmann Publishers, 2001. 550p

HARJINDER, Gill S.; PRAKASH Rao C. **The Official Guide to Data Warehousing**, Indianapolis, USA: QUE Corporation, 1996. 382p

INMON, William H. **Como Construir o Data Warehouse**. Rio de Janeiro: Campos, 1997. 388 p.

KAPLAN, Robert S.; NORTON, David P. **Balanced Scorecard: A Estratégia em Ação**. Rio de Janeiro: Campus, 1997. 344 p.

KAPLAN, Robert S.; NORTON, David P. **Organização Orientada para a Estratégia**. Rio de Janeiro: Elsevier, 2001. 411 p.

KIMBALL, Ralph. **The Data Warehouse Toolkit**. New York, USA: John Wiley & Sons, Inc., 1996. 388 p.

KOBUS, Luciana S.G. **Aplicação Da Descoberta De Conhecimentos Em Bases De Dados Para Identificação De Usuário Com Doenças Cardiovasculares Elegíveis Para Programas De Gerenciamento De Caso**. Programa de Pós-Graduação em Tecnologia em Saúde. Pontifícia Universidade Católica do Paraná. Curitiba, 2006.

KORTH, Henry F.; SILBERSCHATZ, Abraham. **Sistema de Banco de Dados**. São Paulo: MAKRON Books, 1993. 748 p.

LÂVOR, Renata M. **Implementação de Serviços Relacionais À Mineração de Regras de Associação**. Dissertação de Mestrado em Informática, UFRJ/IM, Rio de Janeiro, 2003.

LAROSE, Daniel T. **Discovering Knowledge In Data: An Introduction To Data Mining**. New Jersey , USA: John Wiley & Sons, Inc., 2005. 222 p.

LEAVELL, Hugh; CLARK, Gurney E. **Medicina Preventiva**. Rio de Janeiro: McGraw-Hill, 1976. 744 p.

LU H.; SETIONO, R; LIU, H. NeuroRule: A Connectionist Approach to Data Mining. In: 21. VERY LARGE DATA BASE CONFERENCE, Zurich, Switzerland, Proceedings 1995. p 478-489..

MALTA, Deborah C. Saúde Suplementar e Modelos Assistenciais. Trabalho Técnico. Rio de Janeiro: Agência Nacional de Saúde Suplementar, Disponível em: < http://www.ans.gov.br/portal/upload/biblioteca/TT_AS_14_DCarvalhoMalta_Modelos_Assistenciais.pdf> Acesso em 12 dez 2008.

MALTA, Deborah C. ; CECÍLIO, Luiz C. ; MERHY, Emerson E. ; Franco, Túlio B. ; JORGE, Alzira de O. ; COSTA, Mônica A. Perspectiva da Regulação na Saúde

Suplementar Diante dos Modelos Assistenciais. **Revista Ciência & Saúde Coletiva** v.9,n.2, p.433-444, 2004.

MIRANDA, Cláudio Rocha, **Gerenciamento de Custos em Planos de Assistência à Saúde**, 2003. Trabalho Técnico. Rio de Janeiro: Agência nacional de Saúde Suplementar. Disponível em: <http://www.ans.gov.br/portal/upload/biblioteca/TT_AS_20_ClaudioMiranda_GerenciamentodeCusto.pdf> Publicado em nov. 2003. Acesso em 28 mar. 2007.

ORGANIZAÇÃO MUNDIAL DE SAÚDE (OMS). **Prevenção de Doenças Crônicas: um investimento vital**. OMS, 2005.

QUINLAN, J. Ross. **C4.5: Programs for Machine Learning**. Morgan Kaufmann Publishers. Califórnia, USA: Morgan Kaufmann, 1993. 302p.

PAIVA, Clara. Novos Critérios de Diagnóstico e Classificação do Diabetes Mellitus. **Medicina Interna**, v. 7, n. 4, 2001.

POLIT, Denise F., HUNGLER, Bernadette P. **Fundamentos de pesquisa em enfermagem**. Porto Alegre: Artes Médicas, 1995. 391 p.

PORTER, Michael E.; TEISBERG Elizabeth O. **Repensando a Saúde: Estratégias para Melhorar a Qualidade e Reduzir os Custos**. Porto Alegre: Bookman, 2007. 432 p.

RABUSKE, A. Renato. **Inteligência Artificial**. Florianópolis: Editora da UFSC, 1995. 240 p.

REZENDE, Solange Oliveira. **Sistemas Inteligentes: fundamentos e aplicações**. Barueri: Manole, 2005. 560 p.

RUSSEL, Stuart; NORVIG, Peter. **Inteligência Artificial**. Rio de Janeiro: Campus, 2004. 1040 p.

SANTOS, Alcione Miranda; SEIXAS, José Manuel; PEREIRA, Basílio de Bragança; Medronho, Roberto de Andrade. Usando Redes Neurais Artificiais e Regressão Logística na Predição da Hepatite A. **Revista Brasileira de Epidemiologia**, v.8,n.2, p.117-126, 2005.

SILVA, J. B. **O Projeto de Pesquisa: passo a passo**. Disponível em: <www.meac.ufc.br/public/projpesq/pesquisa.htm#cap3> Fortaleza, CE, 1999. Acesso em 02 maio 2007.

SILVA, E. L.; MENEZES, E. M **Metodologia da pesquisa e elaboração de dissertação**. Florianópolis: Laboratório de Ensino à Distância da UFSC, 2001.

SOARES, Pedro Paulo; NADAL, Jurandir. Aplicação de uma Rede Neural Feedforward com Algoritmo de Levenberg-Marquardt para Classificação de Alterações do Segmento ST do Eletrocardiograma. In: 4. Congresso Brasileiro de Redes Neurais. São José dos Campos: ITA, 1999, p. 384-389.

SOCIEDADE BRASILEIRA DE DIABETES (SBD). **Diretrizes da Sociedade Brasileira de Diabetes (SBD)**. Rio de Janeiro – RJ: DIAGRAPHIC, 2007.

STEINER, Maria T. A.; SOMA, Y. Nei ; SHIMIZU, Tamio; NIEVOLA, Júlio C. ; NETO, Pedro J. S. Abordagem de um Problema Médico por Meio do Processo de KDD com Ênfase à Análise Exploratória dos Dados. **Revista G&P Gestão & Produção**, v.13,n.2,p.325-337,2006

TAN Pang-Ning; STEINBACH Michael; KUMAR Vipin; **Introduction To Data Mining**. Boston, USA: Longman, 2005. 769 p.

TURBAN, Efraim. **Decision Suport and Expert Systems: management support systems**. New Jersey, USA: Prentice-Hall Inc, 1995. 887 p.

UNIMEDPR, **Relatório de Gestão 2008** – Federação de Estado do Paraná. Curitiba - PR, Unimed do Estado do Paraná, 2008.

VIANNA, Rossana Cristina Xavier Ferreira. **Identificação de Características Relacionadas à Mortalidade Infantil Utilizando a Descoberta de Conhecimento em Base de Dados de Saúde Pública**. Programa de Pós-Graduação em Tecnologia em Saúde. Pontifícia Universidade Católica do Paraná. Curitiba, 2007.

ZITTER, M. **A new paradigm in health care delivery: disease management**. Chicago, USA: American Hospital Publishing, 1997. p 207

WALTZ, Carolyn F., STRICKLAND Ora L., LENZ Elizabeth R. **Measurement in Nursing and Health Research**, Philadelphia, USA: F.A. Davis Company, 1991. 498p

WINSTON, Patrick H., **Inteligência Artificial**. Rio de Janeiro: LTC, 1987. 258 p.

WEKA, Waikato Environment for Knowledge Analysis. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka>> Acesso em 25 agosto 2007, University of Waikato, New Zealand, 2007.

WORLD HEALTH ORGANIZATION. Diabetes programme. Disponível em: <<http://www.who.int/diabetes/en>> Acesso em 24 fevereiro 2009.

WITTEN, Ian. H.; FRANK, Eibe. **Data Mining: Pratical Machine Learning Tools and Techniques with Java Implemetantions**. San Francisco, USA: Morgan Kaufmann, 2000. 525 p.

ANEXOS

Anexo B: Parecer Comitê de Ética em Pesquisa.



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
Núcleo de Bioética
Comitê de Ética em Pesquisa

PARECER CONSUBSTANCIADO DE PROTOCOLO DE PESQUISA

Parecer Nº **0001638/08**

Protocolo CEP Nº **2383**

Título do projeto **Concepção de um modelo de identificação de beneficiários com indicativo ao diabetes mellitus tipo 2**

Grupo **III**

Versão **1**

Protocolo CONEP **0162.0.084.000-08**

Pesquisador responsável **Marcelo Rosano Dallagassa**

Instituição

Objetivos

Objetivo geral

Conceber um modelo de identificação de beneficiários com indicativos ao Diabetes Mellitus Tipo 2, para programas de promoção da saúde e prevenção de doenças.

Objetivos específicos

Dentre os principais objetivos específicos destacam-se:

- Conceber uma metodologia e processos para possibilitar a criação de indicadores de Doenças Crônicas Não Transmissíveis.
- Identificar indicadores de monitoramento da saúde dos beneficiários.
- Avaliar os resultados e o impacto dos indicadores no monitoramento.

Comentários

O Protocolo é uma dissertação de mestrado e um trabalho direcionado ao setor de saúde suplementar no qual o pesquisador labuta. Trata-se do desenvolvimento de um sistema gerencial de informações, base de dados, para conhecer os ciclos de atendimentos: diagnóstico, gerenciamento e prevenção de doenças, como estratégias dos planos de saúde, para melhorar a qualidade e reduzir os custos. Esses indicadores, além de auxiliar no processo de gestão, deverão possibilitar a adoção de ações preventivas e negociais junto aos seus contratantes.

Considerações

Como não envolve nem se utiliza de sujeitos humanos de pesquisa, nem de seus prontuários, e tão somente consultas em fichas da Instituição de saúde complementar, não visualizamos problemas éticos nesse sentido, que é o objeto de nosso Comitê de Ética.

Termo de consentimento livre e esclarecido

Como não envolve pesquisa com seres humanos e sim com dados de instituição, é suficiente o Termo de Compromisso de Utilização de Dados e a Autorização da Instituição onde se realizará.



Anexo 2: Autorização utilização da Base de Dados.

SOLICITAÇÃO DE ACESSO A BASES DE DADOS PARA ESTUDO EM DISSERTAÇÃO DE MESTRADO – PUC/PR.

A/C Dr. Faustino Garcia Alferez
 Diretoria de Projetos
 Unimed do Estado do Paraná

Curitiba, 19 de Julho de 2007

Com o objetivo de desenvolver o trabalho de Dissertação de Mestrado em Tecnologia em Saúde – PUCPR, gostaria de solicitar a permissão da utilização de dados históricos da base de dados Data Warehouse, objetivando um trabalho acadêmico e novos conhecimentos e aplicações, que possam contribuir para as operações das Cooperativas Médicas.

Resumo do Trabalho e Necessidades da Utilização da Base de Dados:

Título do Trabalho:

Panel de Indicadores de Saúde para o Monitoramento da Operação de Uma Cooperativa Médica.

Resumo do Trabalho:

De modo geral, o que se pretende apresentar e discutir neste estudo é a concepção de um painel de indicadores de atenção à saúde, para auxiliar os gestores das cooperativas médicas no processo de apoio à tomada de decisões estratégicas, combinando a visão de indicadores focado na estratégia das operadoras com indicadores para o desenvolvimento de ações de promoção à saúde e prevenção.

Devido ao grande volume de informações nas bases de dados geradas pelas solicitações de atendimentos a uma operadora e a complexidade de interpretar e avaliar as informações, será necessário à utilização de ferramentas e técnicas que permitam uma análise mais automática e inteligente como *Data Mining*, Redes Neurais e Inteligência Artificial.

Na metodologia do trabalho, pretende-se gerar conhecimentos por meio do treinamento de Redes Neurais, aplicada a uma parcela de dados da base Data Warehouse, utilizando-a na geração de conhecimento sobre determinadas DCNT (Doenças Crônicas Não Transmissíveis).

Forma de Trabalho e Responsabilidade: Todos os trabalhos efetuados com os registros da base de dados, serão modificados de forma que as informações e dados não identifiquem pessoas físicas ou jurídicas como; nome da empresa, nomes de médicos, beneficiários e prestadores.

Certo de vossa compreensão,

De acordo com o combinado em reunião de 19/07/07
 Marcelo Rosano Dallagassa
 Analista de Negócios Sr. - TI

APÊNDICES

Apêndice A:

Instrumento para Avaliação de regras por Especialistas

Graduação: Especialidade:

Titulação: Instituição (Vínculo Atual):

1) Categoria “Sem Indicativo ao Diabetes”

Nr.	Regra	1	2	3	Comentários
1	Exame Microalbuminúria = 0 e Exame Glicose <= 4				
2	Exame Microalbuminúria = 0 e Consulta Endocrinologista <= 2 e Exame Glicose > 4 e <= 7				
3	Exame Microalbuminúria = 0 e Idade <= 48 e Consulta Endocrinologista <= 2 e Exame Glicose > 7 e <= 15				
669	Exame Microalbuminúria = 0 e Consulta Endocrinologista > 2 e Sexo Masculino e Exame Creatinina > 6 e Exame Curva Glicêmica = 0 e Consulta Oftalmologista > 0 e Exame Glicose = 7 e Idade > 51 e <= 74 e Atendimento Obesidade <=1 e Consulta Nefrologista = 0				
718	Exame Microalbuminúria = 0 e Exame Glicose > 8 e Exame Curva Glicêmica <= 0 e Consulta Endocrinologista > 2 e <= 12 e Sexo Feminino e Exame Mapeamento de Retina > 1 e Idade > 37 e <= 55 e Consulta Nefrologista <= 5 e Exame Creatinina > 9 e Consulta Cardiologista <= 7				
789	Exame Microalbuminúria = 0 e Consulta Endocrinologista > 2 e <= 7 e Exame Curva Glicêmica = 0 Exame Glicose = 9 e Sexo Feminino e Idade > 65 Exame Mapeamento de Retina > 3 e Atendimento Obesidade = 0 Exame Creatinina > 5				

2) Categoria “Com Indicativo ao Diabetes”

Nr.	Regra	1	2	3	Comentários
831	Exame Microalbuminúria = 0 e Idade > 55 e Sexo Feminino e Atendimento Obesidade <= 27 e Consulta Endocrinologista > 2 e <= 7 e Exame Curva Glicêmica <= 0 e Exame Glicose > 9 e <= 13 Consulta Nefrologista > 4 e <= 17 Exame Creatinina > 3 e <= 17 e Consulta Oftalmologista > 2				
835	Exame Microalbuminúria = 0 e Idade > 55 e Sexo Feminino e Atendimento Obesidade <= 27 e Consulta Endocrinologista > 2 e <= 7 e Exame Curva Glicêmica > 0 e Exame Glicose > 8 e <= 12 e Consulta Nefrologista <= 8				
174	Exame Microalbuminúria = 0 e Idade > 48 Exame Glicose > 7 e <= 9 e Consulta Nefrologista <= 1 Exame Curva Glicêmica = 1 e Consulta Endocrinologista > 0 e <= 1 e Exame Mapeamento de Retina <= 2				
164	Exame Microalbuminúria = 0 e Idade > 48 e Sexo Masculino e Exame Glicose > 4 e <= 11 e Exame Curva Glicêmica = 0 e Exame Creatinina > 6 e Consulta Endocrinologista = 1 e Consulta Nefrologista > 3 e Consulta Cardiologista > 10				
167	Exame Microalbuminúria = 0 e Exame Curva Glicêmica > 0 e Exame Glicose > 7 e <= 9 e Consulta Nefrologista <= 1 e Consulta Endocrinologista = 0 e Exame Mapeamento de Retina <= 7 e Idade > 48 e <= 58 e Exame Creatinina = 3				
171	Exame Microalbuminúria = 0 e Exame Glicose > 7 e <= 11 Exame Curva Glicêmica > 0 e Consulta Nefrologista <= 1 e Exame Mapeamento de Retina <= 7 e Idade > 58 e <= 73 e Consulta Endocrinologista > 0				

3) Categoria “Com Forte Indicativo ao Diabetes”

Nr.	Regra	1	2	3	Comentários
1106	Exame Microalbuminúria > 0 e Exame Glicose > 4 e Consulta Endocrinologista > 8				
1094	Exame Microalbuminúria > 0 e Consulta Endocrinologista > 1 e <= 8 e Atendimento Obesidade <= 35 e Exame Glicose > 4 e <= 13 Idade > 42 e Exame Creatinina > 2 e <= 9 e Consulta Nefrologista <= 8				
1102	Exame Microalbuminúria > 0 e Exame Glicose > 13 Consulta Endocrinologista > 1 e <= 8 e Atendimento Obesidade <= 35				
615	Exame Microalbuminúria = 0 e Sexo Feminino e Consulta Endocrinologista > 6 e Atendimento Obesidade <= 18 e Idade > 47 e Exame Curva Glicêmica > 0 e Consulta Cardiologista > 11 e Exame Creatinina > 3 Exame Glicose > 5 e <= 7 e Consulta Oftalmologista <= 4				
651	Exame Microalbuminúria = 0 e Sexo Masculino e Atendimento Obesidade <= 9 e Exame Mapeamento de Retina <=3 e Consulta Oftalmologista <= 10 e Consulta Endocrinologista > 3 e <= 8 e Exame Creatinina > 1 e <= 5 e Exame Curva Glicêmica <= 0 e Exame Glicose = 7 e Idade > 75 e Exame Mapeamento de Retina > 1				
747	Exame Microalbuminúria = 0 e Idade > 33 e <= 55 Exame Curva Glicêmica <= 0 e Consulta Endocrinologista > 12 Sexo Feminino e Atendimento Obesidade > 6 Exame Glicose = 9 e Consulta Nefrologista <= 1 e Consulta Oftalmologista <= 4				