

**UNIVERSIDADE REGIONAL DE BLUMENAU**  
**CENTRO DE CIÊNCIAS EXATAS E NATURAIS**  
**CURSO DE CIÊNCIA DA COMPUTAÇÃO – BACHARELADO**

**PROTÓTIPO DE UMA FERRAMENTA PARA IDENTIFICAR  
PADRÕES NOS TRATAMENTOS CLÍNICOS**

**ALAN MICHEL LEICHT**

**BLUMENAU**  
**2016**

**ALAN MICHEL LEICHT**

**PROTÓTIPO DE UMA FERRAMENTA PARA IDENTIFICAR  
PADRÕES NOS TRATAMENTOS CLÍNICOS**

Trabalho de Conclusão de Curso apresentado ao curso de graduação em Ciência da Computação do Centro de Ciências Exatas e Naturais da Universidade Regional de Blumenau como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

Prof(a). Francisco Adell Péricas, Mestre - Orientador

**BLUMENAU  
2016**

# **PROTÓTIPO DE UMA FERRAMENTA PARA IDENTIFICAR PADRÕES NOS TRATAMENTOS CLÍNICOS**

Por

**ALAN MICHEL LEICHT**

Trabalho de Conclusão de Curso aprovado para obtenção dos créditos na disciplina de Trabalho de Conclusão de Curso II pela banca examinadora formada por:

Presidente: \_\_\_\_\_  
Prof(a). Francisco Adell Péricas, Mestre – Orientador, FURB

Membro: \_\_\_\_\_  
Prof(a). Mauro Marcelo Mattos, Doutor – FURB

Membro: \_\_\_\_\_  
Prof(a). Roberto Heinzle, Doutor – FURB

Blumenau, 06 de dezembro de 2016

## **AGRADECIMENTOS**

Agradeço aos meus pais, Iris Leicht e Wolfgang Leicht, que sempre me apoiaram em todas as fases da minha vida, e em especial durante a graduação onde a jornada se tornou mais árdua.

Ao meu orientador Francisco Adell Péricas pela dedicação na orientação deste trabalho.

Sonhos determinam o que você quer. Ação determina o que você conquista.

Aldo Novak

## RESUMO

Este trabalho apresenta o desenvolvimento de um protótipo de uma ferramenta para mineração de dados sobre uma base com registros hospitalares, para detectar padrões nos tratamentos clínicos. De forma mais detalhada, objetiva-se implementar o algoritmo C4.5 para realizar a mineração de dados e demonstrar a consulta de predição sobre árvore de decisão através de uma interface gráfica, com recursos visuais e interativos para análise clínica. O processo de mineração de dados foi realizado através da técnica de árvore de decisão implementada com o algoritmo C4.5. A árvore de decisão gerada é apresentada em dois formatos, para que o médico possa ter uma visão geral de todos os fatores que podem estar relacionados ao diagnóstico de determinada doença e outra visão apenas com os fatores respondidos na avaliação clínica, representado apenas a realidade de um paciente.

Palavras-chave: Mineração de dados. Árvore de decisão. Registros clínicos.

## **ABSTRACT**

This work presents the development of a prototype of a tool for data mining on a database with hospital records to detect patterns in clinical treatments. In a more detailed way, the objective is to implement the C4.5 algorithm to perform the data mining and to demonstrate the decision tree prediction through a graphical interface with visual and interactive resources for clinical analysis. The data mining process was performed using the decision tree technique implemented with the algorithm C4.5. The decision tree generated appear in two formats, so that the doctor can have an overview of all factors that may be related to the diagnosis of a particular disease, and another view only with the factors answered in the clinical evaluation, represented only reality of a patient.

**Key-words:** Data mining. Decision tree. Clinical records.

## LISTA DE FIGURAS

Figura 1 - Funções da Informática clínica .....	14
Figura 2 - Sistema de informação hospitalar integrado .....	15
Figura 3 - Registro médico eletrônico .....	16
Figura 4 - Modelos de decisão médica .....	17
Figura 5 - Etapas da mineração de dados .....	18
Figura 6 - Técnicas de classificação de dados .....	19
Figura 7 - Grafo de Árvore de decisão .....	20
Figura 8 - Exemplo de árvore de decisão .....	21
Figura 9 - Núcleo do algoritmo C4.5.....	21
Figura 10 - Arquitetura do InteliMED.....	22
Figura 11 - Telas do InteliMED .....	23
Figura 12 - Árvore de decisão gerada pelo InteliMED .....	24
Figura 13 - Camadas da aplicação .....	25
Figura 14 - Diagrama de casos de uso .....	25
Figura 15 - Principal tela do SADM-DST Desktop .....	26
Figura 16 - Diagrama de caso de uso .....	28
Figura 17 - Diagrama de atividades.....	29
Figura 18 - Diagrama de pacotes.....	30
Figura 19 - Diagrama de classes do pacote view.....	31
Figura 20 - Diagrama de classes do pacote dao .....	32
Figura 21 - Diagrama de classes do pacote inducaoArvore.....	33
Figura 22 - Menu principal.....	38
Figura 23 - Cadastro de usuários .....	38
Figura 24 - Tela de configuração da base de dados.....	39
Figura 25 - Configuração da árvore.....	39
Figura 26 - Configuração dos atributos da árvore .....	40
Figura 27 - Visualização em modo detalhe do cadastro de atributos .....	40
Figura 28 - Execução da mineração .....	41
Figura 29 - Grid com as árvores geradas .....	42
Figura 30 - Avaliação clínica .....	42
Figura 31 - Árvore resultante com o diagnóstico geral de uma doença .....	43



## LISTA DE QUADROS

Quadro 1 - Requisitos funcionais .....	27
Quadro 2 - Requisitos não funcionais .....	27
Quadro 3 - Mineração de todas as árvores .....	34
Quadro 4 - Preparação dos atributos para mineração .....	35
Quadro 5 - Função que inicia a indução .....	36
Quadro 6 - Indução da árvore .....	37
Quadro 7 - Criação da árvore para apresentação .....	37
Quadro 8 - Comparativo com os trabalhos correlatos .....	44
Quadro 9 - Caso de uso: Permitir visualizar árvore com o diagnóstico da avaliação clínica...	50
Quadro 10 - Caso de uso: Permitir executar avaliação clínica .....	50
Quadro 11 - Caso de uso: Permitir visualizar árvore com diagnóstico geral .....	50

## **LISTA DE ABREVIATURAS E SIGLAS**

DST – Doenças Sexualmente Transmissíveis

SQL – Structured Query Language

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>11</b>
1.1 OBJETIVOS.....	12
1.2 ESTRUTURA.....	12
<b>2 FUNDAMENTAÇÃO TEÓRICA .....</b>	<b>13</b>
2.1 INFORMÁTICA NA MEDICINA.....	13
2.2 SISTEMAS DE APOIO À DECISÃO EM MEDICINA.....	14
2.2.1 Registros clínicos .....	15
2.2.2 Decisão clínica .....	16
2.3 MINERAÇÃO DE DADOS.....	17
2.3.1 Etapas da mineração de dados.....	18
2.3.2 Árvore de decisão.....	20
2.4 TRABALHOS CORRELATOS .....	22
2.4.1 InteliMED - Sistema Móvel de Apoio a Decisão Médica Aplicado ao Diagnóstico de Asma .....	22
2.4.2 SADM-DST Desktop .....	24
<b>3 DESENVOLVIMENTO .....</b>	<b>27</b>
3.1 REQUISITOS.....	27
3.2 ESPECIFICAÇÃO .....	27
3.2.1 Diagrama de Casos de uso .....	27
3.2.2 Diagrama de atividades .....	28
3.2.3 Diagrama de pacotes .....	30
3.3 IMPLEMENTAÇÃO .....	33
3.3.1 Técnicas e ferramentas utilizadas.....	33
3.3.2 Operacionalidade da implementação .....	38
3.4 ANÁLISE DOS RESULTADOS .....	43
<b>4 CONCLUSÕES.....</b>	<b>46</b>
4.1 EXTENSÕES .....	47
<b>REFERÊNCIAS .....</b>	<b>48</b>
<b>APÊNDICE A – DETALHAMENTO DOS CASOS DE USO .....</b>	<b>50</b>

## 1 INTRODUÇÃO

Atualmente os serviços de saúde no Brasil estão sobrecarregados, e diversos fatores contribuem para o aumento da procura por atendimento médico. Durante o processo de diagnóstico e definição do tratamento do paciente, um grande volume de informações é manipulado e a agilidade no atendimento e do processamento dessas informações é um fator importante e pode representar um risco à saúde do paciente.

Para determinadas doenças, quando a equipe clínica não tem tempo hábil para buscar embasamento em análises e pesquisas, fica evidente a necessidade de sistemas que analisem de uma forma rápida as informações existentes, buscando relações desconhecidas e ainda não testadas entre os dados já conhecidos (MOURA; ANDRADE; KIKUCHI, 2005). Essas relações possibilitaram a geração de novas informações, ou apontam quais medidas devem ser tomadas para que o paciente envolvido tenha maiores possibilidades de cura.

Os sistemas que fazem a análise das informações clínicas e fornecem apoio aos médicos na tomada de decisão podem utilizar o conceito de mineração de dados, que segundo Tang, Steinbach e Kumar (2009, p. 3), “é o processo de descoberta automática de informações úteis em grandes depósitos de dados.” De acordo com Côtés, Porcaro e Lifschitz (2002, p. 1) “o conceito de mineração de dados está se tornando cada vez mais popular como uma ferramenta de descoberta de informações, que podem revelar estrutura de conhecimento, que possam guiar decisões em condições de certeza limitada.”

Conforme Tang, Steinbach e Kumar (2009, p. 3), “as técnicas de mineração de dados são organizadas para agir sobre grandes bancos de dados com o intuito de descobrir padrões úteis e recentes que poderiam, de outra forma, permanecer ignorados.” A árvore de decisão é umas das técnicas utilizadas na tarefa de classificação, e segundo Hosokawa (2011, p. 24) “uma árvore de decisão é uma estrutura que pode ser utilizada para, por meio de uma simples regra de decisão, dividir sucessivamente uma grande coleção de registros em conjuntos menores.”

Nos sistemas convencionais o usuário informa a entrada, que pode indicar a hipótese de uma doença, com base em alguns sintomas do paciente e o sistema verifica se esta é verdadeira ou falsa, portanto é o usuário quem deve ter o “*insight*” para verificar as normas ou regras, assim a descoberta das informações escondidas fica dependente da racionalidade limitada do usuário. Na mineração de dados, ao contrário, o sistema retorna todas as regras e a pessoa faz uso da forma que achar mais conveniente (GONÇALVES, 1999, p. 5).

Com o avanço da tecnologia, é possível manter grandes bases de dados, com informações sobre os sintomas, resultados de exames, diagnósticos, tratamentos e curso das doenças dos pacientes. Com a aplicação da mineração de dados na medicina, é possível gerar um conhecimento, como a relação entre determinadas doenças e é possível desenhar um perfil sociocultural relacionado com os hábitos pessoais e local de moradia. Com essas relações, é possível estabelecer um melhor entendimento em relação às doenças e seus respectivos tratamentos (CARVALHO, 2001, p. 10).

Diante deste contexto, percebe-se a importância de se criar aplicações voltadas para a detecção de padrões nos tratamentos clínicos. Sendo assim, propõe-se neste trabalho desenvolver um protótipo de uma ferramenta que irá aplicar essas técnicas, possibilitando identificar com mais rapidez a doença do paciente, baseando-se nos sintomas por ele sentidos.

## 1.1 OBJETIVOS

O objetivo deste trabalho é desenvolver uma ferramenta para detectar padrões em uma base de dados de tratamentos clínicos.

Os objetivos específicos são:

- a) implementar o algoritmo C4.5 para realizar a mineração de dados;
- b) disponibilizar uma interface gráfica com recursos visuais e interativos para análise clínica.

## 1.2 ESTRUTURA

Esta monografia está organizada em capítulos. No primeiro capítulo é apresentada a introdução ao assunto da pesquisa e descritos os objetivos do trabalho. O segundo capítulo descreve a fundamentação teórica, na qual apresenta a informática na medicina juntamente com os sistemas de apoio a decisão, e também conceitos sobre mineração de dados. O terceiro capítulo descreve o desenvolvimento da solução proposta, que inclui os requisitos propostos, as ferramentas e técnicas utilizadas, assim como a operacionalidade da aplicação. Por fim, o quarto capítulo descreve as conclusões sobre a pesquisa e apresentada alternativas para o desenvolvimento de trabalhos futuros relacionados a este.

## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são abordados os aspectos teóricos relacionados ao trabalho. Inicialmente conceitua-se Informática na medicina, depois apresenta-se uma fundamentação sobre Sistemas de apoio a decisão clínica e seus componentes. Em seguida, é apresentada a mineração de dados e as estruturas que a compõem. Ao final são apresentados os trabalhos correlatos.

### 2.1 INFORMÁTICA NA MEDICINA

O avanço da informática na medicina não ocorreu somente com o avanço da computação e das telecomunicações. A quantidade de informações resultantes do atendimento médico, sejam elas clínicas ou administrativas, é tão grande que métodos que tenham o papel como substrato não atendem as necessidades para fazer um gerenciamento eficiente dessas informações. Além disso, ocorreu, por parte do meio médico, o reconhecimento de que não é somente o processo de tomada de decisão que importa, mas devem ser levados em consideração os fatos em que este processo se apoia (LEITE; PAES, 2003).

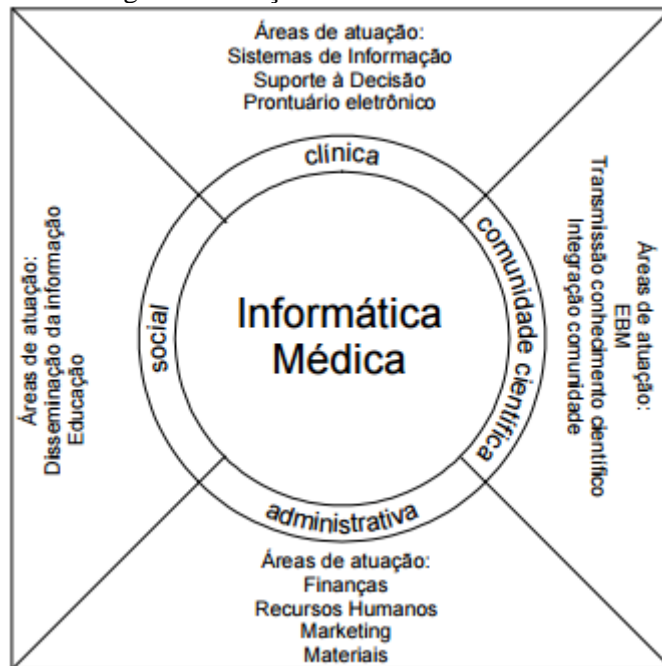
De acordo com Leite e Paes (2003, p. 5), “a preocupação com a organização dos dados que envolvem as atividades médicas resultou no desenvolvimento e aprimoramento de ferramentas que pudessem, apoiadas nas tecnologias emergentes, auxiliar na atuação dos profissionais de saúde.”

Conforme Leite e Paes (2003):

Na atividade clínica, a informática é utilizada em diversas fases do processo de interação entre o paciente e o médico: na coleta de dados (através uso de instrumentos como aparelhos de ressonância magnética), no monitoramento (utilizado principalmente em pacientes internados ou em processo cirúrgico) e no tratamento (no controle de dosagem de medicamentos). Também pode ser utilizada no fornecimento de informações antigas sobre o paciente ao médico através de ferramentas como o Prontuário Eletrônico (LEITE; PAES, 2003, p. 6).

Está crescendo a quantidade de profissionais da saúde que estão dispostos a utilizar, como vantagem na execução das suas atividades profissionais, os benefícios dos sistemas e tecnologias da informação (ROCHA; VASCONCELOS; MOREIRA, 2005). A informática na área da saúde possui diversas aplicações desde o gerenciamento de estoques, até métodos de diagnóstico. A Figura 1 demonstra o agrupamento das funções da informática médica que são classificadas em: administrativa, social, clínica e científica (LEITE; PAES, 2003).

Figura 1 - Funções da Informática clínica



Fonte: Leite e Paes (2003).

## 2.2 SISTEMAS DE APOIO À DECISÃO EM MEDICINA

Os sistemas de apoio à decisão em medicina buscam reproduzir a perícia de um profissional humano numa área específica do conhecimento. Assim sendo, podem ser utilizados como ferramenta de apoio pelos profissionais da saúde durante o processo de tomada de decisão clínica.

A qualidade das informações coletadas do paciente pelo médico é parte fundamental do processo de diagnóstico de uma doença. Segundo Sigulem et al. (2010, p.2), “no que diz respeito à coleta de dados é necessário que estes sejam em número suficiente e que contenham os tópicos de maior relevância para a identificação do diagnóstico e o subsequente planejamento terapêutico.”

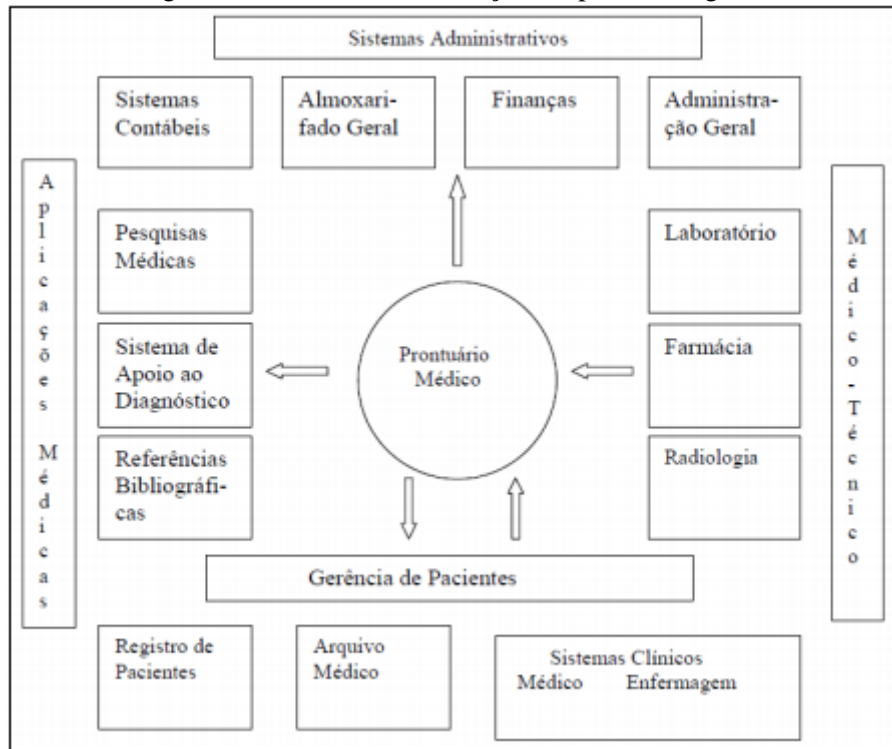
Como o processo de diagnóstico de uma doença pode ser complexo, o profissional envolvido pode raciocinar com incerteza, tendo em vista que:

[...] na medicina um médico habitualmente raciocina com incerteza, pois, em geral, os sintomas (por exemplo, febre, falta de apetite, etc.) não determinam uma única doença já que pode ser comuns a várias (por exemplo, gripe, amidalite e diversas outras infecções). Assim, os médicos lidam todo o tempo com incerteza, decidindo o tratamento das doenças de seu paciente, baseando-se em evidências obtidas através da análise, exames clínicos e laboratoriais (ANDRADE, 2005, p.1).

De acordo com Moraes et al. (2012), o uso de sistemas de apoio à decisão em medicina podem facilitar o processo de diagnóstico, e também diminuir a necessidade de consulta da literatura técnica, sem, no entanto, reduzir a qualidade do diagnóstico definitivo. Portanto pode implicar em melhora da qualidade do diagnóstico médico, redução do tempo necessário

para realização do diagnóstico definitivo e redução de custos. A Figura 2 mostra como exemplo a organização de um sistema hospitalar integrado.

Figura 2 - Sistema de informação hospitalar integrado



Fonte: Leite e Paes (2003).

### 2.2.1 Registros clínicos

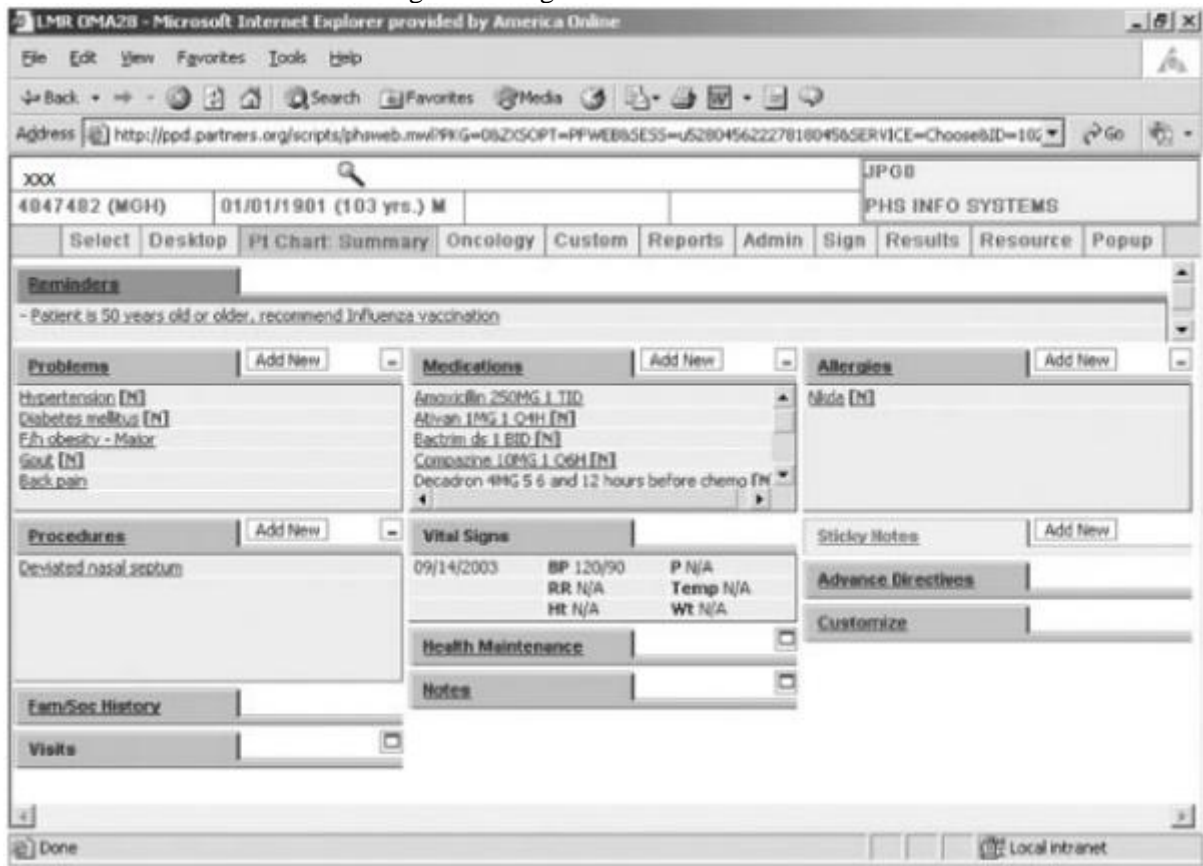
Os registros de pacientes são amplamente utilizados como base no processo de diagnóstico e tratamento do paciente, e este pode ser considerado um dos principais motivos para se manter registros dos pacientes. Os registros são um importante meio pelo qual uma equipe de profissionais podem se comunicar uns com os outros sobre as necessidades do paciente (CHIMIESKI, 2013).

Os registros clínicos armazenam um histórico do paciente, são mantidos ao longo do tempo, e podem ser utilizados para relacionar doenças e tratamentos que o paciente tenha experimentado. Todos os envolvidos no cuidado do paciente podem contribuir com novos registros, mantendo uma perspectiva histórica dos episódios do paciente (CHIMIESKI, 2013).

A Figura 3 mostra um exemplo de um registro médico eletrônico, com informações de um paciente.



Figura 3 - Registro médico eletrônico



Fonte: Chimieski (2013).

De acordo com Chimieski (2003):

A importância de manter registros dos pacientes completos e precisos não pode ser subestimada. Eles não servem apenas como base para o planejamento de assistência ao paciente, mas também como o registro legal, documentando o cuidado que foi fornecido aos pacientes por parte da organização. Os prontuários de pacientes fornecem grande parte dos dados de origem de informações de saúde que são gerados dentro e entre as organizações de cuidados de saúde. Os dados capturados em um registro médico do paciente pode se tornar um registro permanente de diagnósticos de paciente, de tratamentos e de resposta aos tratamentos (CHIMIESKI, 2013, p. 33).

### 2.2.2 Decisão clínica

O processo diagnóstico acaba se tornando um processo espontâneo e inconsciente porque é algo realizado diariamente pelos médicos. A experiência dos profissionais de saúde torna tão fácil a realização de tarefas ordinárias, da mesma forma como é simples reconhecer os rostos que são familiares. Mas, mesmo sendo de extrema importância o processo de diagnóstico, ele pode falhar quando o problema apresenta um alto grau de complexidade ou quando se trata de algo não habitual (RÉA NETO, 1998).

De acordo com Réa Neto (1998):

Diagnósticos corretos são baseados em raciocínios adequados e em informações válidas. O médico que descarta a lógica pode assumir ingenuamente que provou um

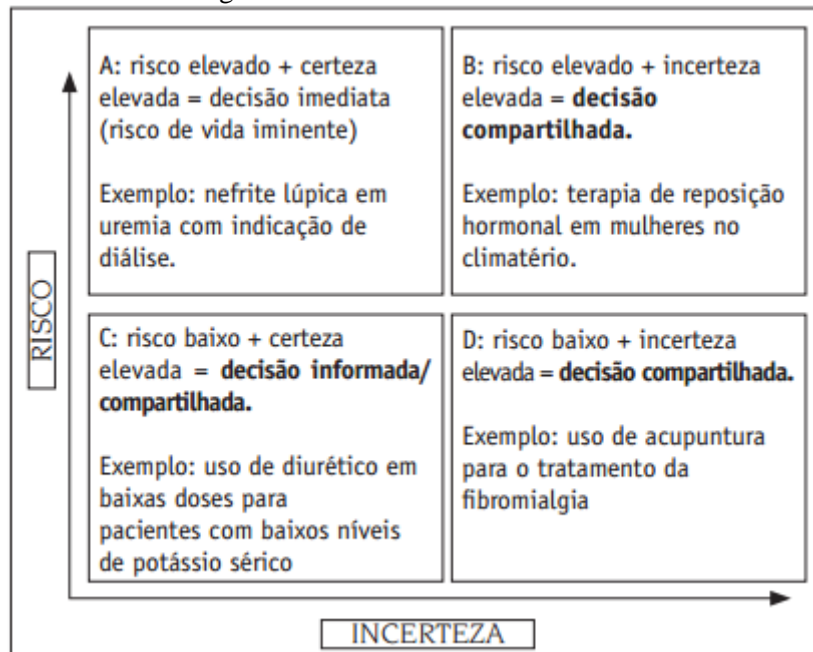
diagnóstico, quando apenas estabeleceu um diagnóstico provável ou possível. O conhecimento da base lógica da prova ou da refutação pode não somente dar maior precisão ao diagnóstico individual como também fornece uma base racional para avaliar as decisões diagnósticas (RÉA NETO, 1998, p. 307).

Nesse contexto, Réa Neto (1998) afirma que:

[...] o uso do raciocínio lógico não é uma garantia de conclusões verdadeiras. A lógica possui regras úteis para processar as informações clínicas na busca de uma solução adequada para o problema clínico, mas não integra nenhuma segurança de que as informações clínicas e suas interpretações estão corretas. A lógica estuda somente as formas de raciocínio e não os seus conteúdos. O médico necessita obter, analisar, sintetizar e avaliar adequadamente informações clínicas precisas e acuradas para depois processá-las de forma lógica. Somente assim ele estará próximo do raciocínio correto e da decisão certa (RÉA NETO, 1998, p. 307).

A Figura 4 ilustra alguns contextos de certeza/incerteza e risco e suas aplicações nas tomadas de decisão médica.

Figura 4 - Modelos de decisão médica



Fonte: Abreu et al. (2006).

### 2.3 MINERAÇÃO DE DADOS

A mineração de dados é o processo que consiste na descoberta de informações relevantes em grandes bases de dados. As técnicas de mineração de dados são realizadas sobre grandes depósitos de dados de modo a detectar padrões úteis e recentes que dificilmente seriam encontradas a olho nu, devido às limitações e capacidade de processamento intelectual do ser humano. Além disto, fornecem a capacidade de se prever resultados de uma observação futura, como, por exemplo, a previsão de se um novo visitante de um site irá realizar uma compra no primeiro acesso (TANG; STEINBACH; KUMAR, 2009).

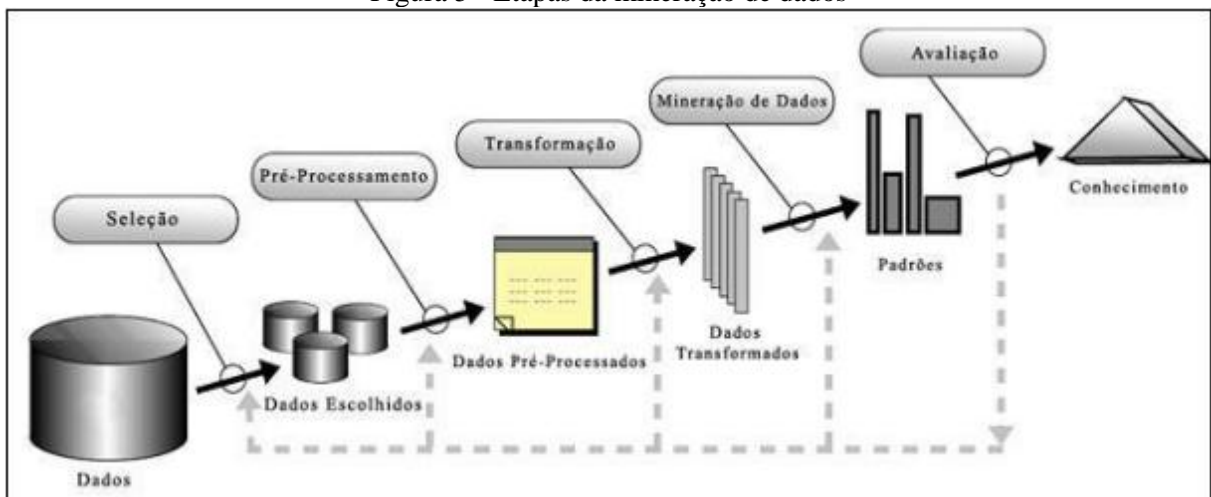
### 2.3.1 Etapas da mineração de dados

O processo de mineração de dados é formado por cinco etapas, sendo que em cada uma delas é realizada uma operação distinta, para que seja possível extrair informações úteis em uma base de dados. De acordo com Silva (2004, p.3), as etapas descritas abaixo foram o processo geral de mineração de dados:

- selecionar os dados: selecionar um conjunto de dados, ou focar num subconjunto, onde a descoberta deve ser realizada;
- pré-processamento: operações básicas tais como remoção de ruídos quando necessário, coleta de informação necessária pra modelar ou estimar ruído, escolha de estratégias para manipular campos de dados ausentes, formatação de dados de forma a adequá-los à ferramenta de mineração;
- transformação: localização de características úteis para representar os dados dependendo do objetivo da tarefa, visando a redução do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados, bem como o enriquecimento semântico das informações;
- estimar o modelo (minerar o dado): selecionar os métodos a serem utilizados para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação ou conjunto de representações. Busca pelo melhor ajuste dos parâmetros do algoritmo para a tarefa em questão;
- interpretação do modelo e conclusões: visualização do conhecimento gerado, podendo assim se tirar conclusões e tomar decisões.

A Figura 5 representa as etapas da mineração de dados.

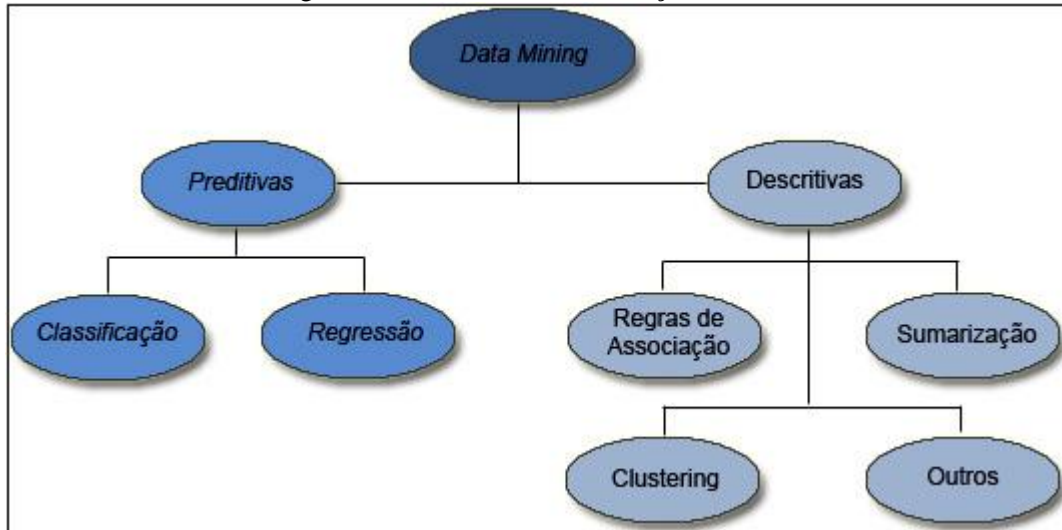
Figura 5 - Etapas da mineração de dados



Fonte: Adaptado de Fayyad, Piatetsky-Shapiro e Smyth (1996).

Algumas técnicas podem ser utilizadas para realizar a tarefa de classificação dos dados. A Figura 6 demonstra algumas dessas técnicas e a forma como elas estão agrupadas.

Figura 6 - Técnicas de classificação de dados



Fonte: Rezende (2003).

De acordo com Semann (2015), dentro da etapa de estimar o modelo conforme a Figura 6, a classificação é o processo responsável por identificar a qual classe um determinado registro pertence. “Classificação é a tarefa de aprender uma função alvo  $f$  que mapeie cada conjunto de atributos  $x$  para um dos rótulos de classes  $y$  pré-determinados” (TANG; STEINBACH; KUMAR, 2009, p.172).

De acordo com Tan, Steinbach e Kumar (2009), as técnicas de classificação são uma abordagem sistemática cujo objetivo é a criação de modelos de classificação resultantes de um conjunto de dados de entrada. “Cada técnica emprega um algoritmo de aprendizagem para identificar um modelo que seja mais apropriado para o relacionamento entre o conjunto de atributos e o rótulo da classe dos dados de entrada” (TANG; STEINBACH; KUMAR, 2009, p.174).

As técnicas de árvore de decisão e redes neurais podem ser destacadas pela ampla utilização na mineração de dados e segundo Castanheira (2008) redes neurais podem ser conceituadas como:

[...] uma técnica que constrói um modelo matemático, de um sistema neural biológico simplificado, com capacidade de aprendizado, generalização, associação e abstração. Assim como no cérebro humano, as redes neurais apresentam uma estrutura amplamente paralelizada, composta por processadores simples (neurônios artificiais) conectados entre si (CASTANHEIRA, 2008, p. 28).

### 2.3.2 Árvore de decisão

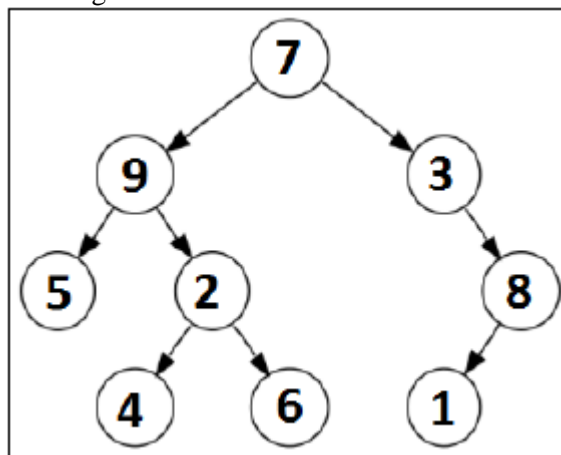
Para Hosokawa (2011, p. 24) “uma árvore de decisão é uma estrutura que pode ser utilizada para, por meio de uma simples regra de decisão, dividir sucessivamente uma grande coleção de registros em conjuntos menores.” Uma árvore de decisão, possui características de uma árvore, que é uma estrutura não linear e possui uma quantidade finita de nodos ou elementos (HALMENSCHLAGER, 2002).

Segundo Tan, Steibach e Kumar (2009), uma árvore de decisão possui uma estrutura composta de:

- a) nodo raiz: é um nodo especial, que é o pai de todas subárvores, cujos nodos, por sua vez, são os nodos filhos do nodo raiz (HALMENSCHLAGER, 2002);
- a) nodos internos: corresponde a um teste do valor de uma das propriedades, e os ramos deste nó são identificados com os possíveis valores do teste (POZZER, 2006);
- b) nodos folha ou terminais: são os nodos que não possuem filhos e podem ser considerados como terminais (HALMENSCHLAGER, 2002).

A Figura 7 é a representação de uma árvore através de um grafo, onde existe apenas um nó raiz, e todos os demais estão de alguma forma conectados a ele, sendo que não é necessário possuir relação de ordem entre os números dos nós (SEMANN, 2015).

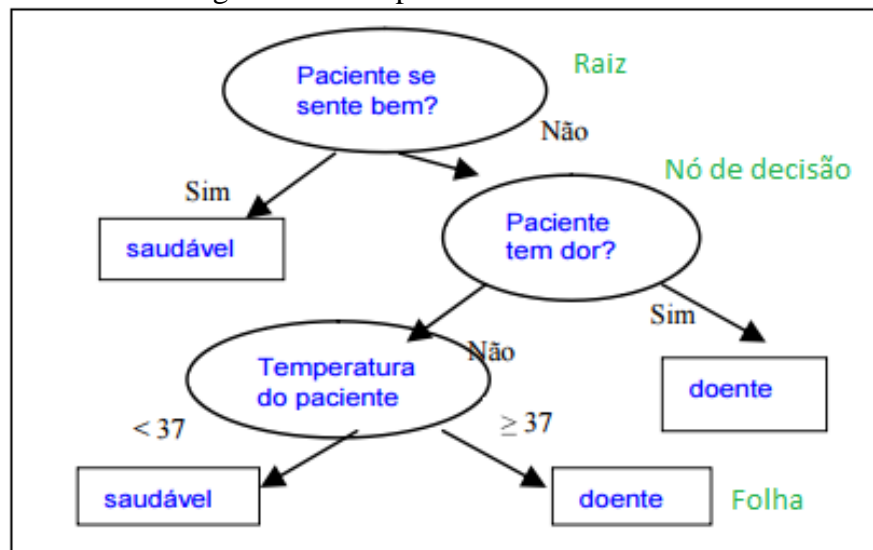
Figura 7 - Grafo de Árvore de decisão



Fonte: Semann (2015).

A Figura 8 mostra um exemplo de árvore de decisão para um simples diagnóstico de um paciente, onde as elipses representam os testes em um dos atributos dos dados do paciente, as setas definem o caminho da árvore, e os retângulos representam o diagnóstico. Para diagnosticar um paciente se faz necessário iniciar pela raiz, seguindo cada teste até alcançar uma folha (SIMÕES, 2008).

Figura 8 - Exemplo de árvore de decisão



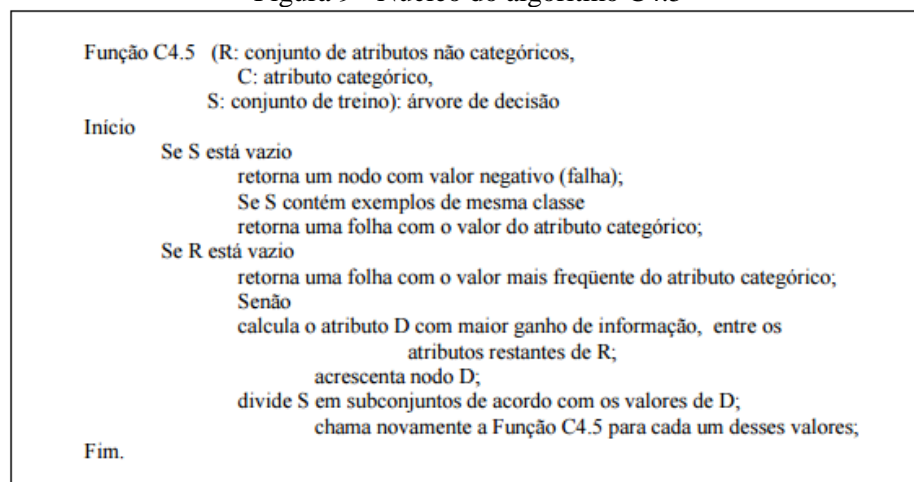
Fonte: Adaptado de Simões (2008).

### 2.3.2.1 Indução na Árvore de Decisão

Indução na árvore de decisão consiste no processo de classificação dos dados e construção da árvore de decisão. De acordo com Tan, Steinbach e Kumar (2009), a partir de um conjunto de atributos é possível construir uma infinidade de árvores de decisão.

A utilização do algoritmo C4.5 pode ser uma das estratégias realizar a indução em árvore de decisão e segundo Halmenschlager (2002, p. 51) esta técnica tem como objetivo “a geração de árvores de decisão com tratamento de atributos contínuos e discretos, construindo uma árvore com número de partições variável e com as folhas sendo indicadas pelos valores do atributo categórico”. A Figura 9 mostra o núcleo de processamento do algoritmo C4.5.

Figura 9 - Núcleo do algoritmo C4.5



Fonte: Halmenschlager (2002).

De acordo com Halmenschlager (2002) o algoritmo C4.5 possui as seguintes características:

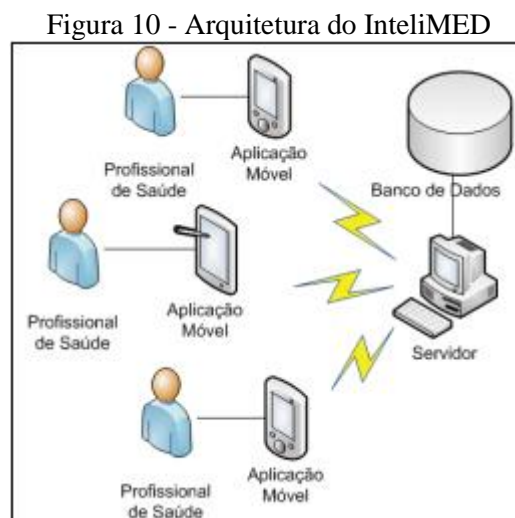
- a) eleição do melhor atributo: utiliza o critério do ganho da informação;
- b) tratamento de atributos discretos: atribui uma ligação distinta a cada valor do atributo ou forma agrupamentos de valores em vários conjuntos;
- c) tratamento de atributos contínuos: utiliza a técnica do teste simples para a partição, escolhendo como ponto de cisão o ponto médio entre os valores;
- d) tratamento de valores desconhecidos: desconsidera os atributos com valores desconhecidos, utilizando aqueles apenas com valores totalmente conhecidos;
- e) determinação da classe associada à folha: é efetuada por atribuição da classe mais provável nesta folha;
- f) método de poda: utiliza a técnica pós-poda baseada no erro, examinando a árvore de forma *button-up* e substituindo uma subárvore por folha.

## 2.4 TRABALHOS CORRELATOS

Foram encontrados dois trabalhos correlatos desenvolvidos no meio acadêmico, o software SADM-DST Desktop proposto por Zoby (2009) e o InteliMED proposto por Morais et al. (2012). Estes trabalhos serão detalhados a seguir.

### 2.4.1 InteliMED - Sistema Móvel de Apoio a Decisão Médica Aplicado ao Diagnóstico de Asma

O software desenvolvido por Morais et al. (2012) é composto por dois módulos: aplicação móvel e servidor. A arquitetura do InteliMED pode ser vista na Figura 10. A visão global da arquitetura mostra que a aplicação está distribuída entre o equipamento móvel e o servidor.



Fonte: Morais et al. (2012).

A aplicação foi desenvolvida na plataforma Android utilizando a linguagem de programação Java e principal funcionalidade do InteliMED é prover o diagnóstico de asma. O aplicativo garante alta disponibilidade, sendo possível utilizar as principais funcionalidades sem acesso ao servidor.

De acordo com Morais et al. (2012), para determinar um diagnóstico através da aplicação móvel, o médico responde um questionário que se tratam das evidências que serão utilizadas pela ferramenta para determinar o diagnóstico da doença respiratória do paciente. O processo de mineração de dados gera uma árvore de decisão, que será percorrida de acordo com as respostas do questionário, e após esse processamento o resultado será exibido ao usuário. Após exibir o diagnóstico, o médico pode confirmar se concorda com o resultado apresentado e ainda é possível inserir comentários a critério do profissional que está aplicando o questionário. A Figura 11 mostra algumas telas do fluxo deste sistema.

Figura 11 - Telas do InteliMED

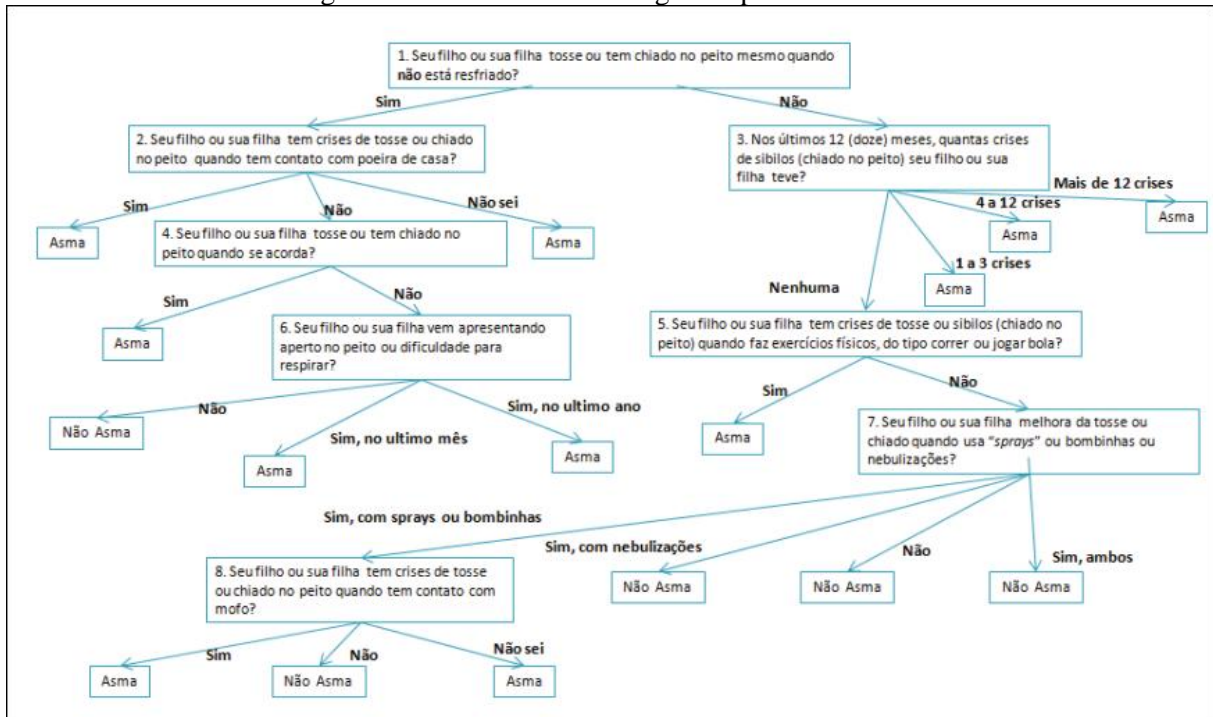


Fonte: Morais et al. (2012).

Na Figura 12 é possível visualizar uma árvore de decisão utilizada para determinar o diagnóstico de asma de um paciente.



Figura 12 - Árvore de decisão gerada pelo InteliMED



Fonte: Morais et al. (2012).

#### 2.4.2 SADM-DST Desktop

O sistema de apoio à decisão médica desenvolvido por Zoby (2009), utiliza o algoritmo de árvore de decisão e tem como objetivo principal extrair as hipóteses diagnósticas de uma base de dados. As principais funcionalidades originaram-se de entrevistas com especialistas em doenças sexualmente transmissíveis (DST) e também com alguns estudantes vinculados às instituições Hospital Universitário Oswaldo Cruz e à Faculdade de Ciências Médicas da Universidade de Pernambuco.

De acordo com Zoby (2009), o sistema foi projetado para rodar localmente pois trata-se de uma aplicação desktop e foi desenvolvido com a linguagem de programação Java. Na Figura 13 é possível visualizar a arquitetura da aplicação que possui quatro camadas, sendo elas:

- arquivos: responsável pela manipulação de arquivos;
- processamento: responsável por agrupar toda a lógica do sistema;
- fachada: tem o papel de atender as solicitações da interface gráfica, funcionando como um repositório de todas as funcionalidades disponíveis no sistema;
- gui: responsável por toda a interface com o usuário, a parte visível do sistema.

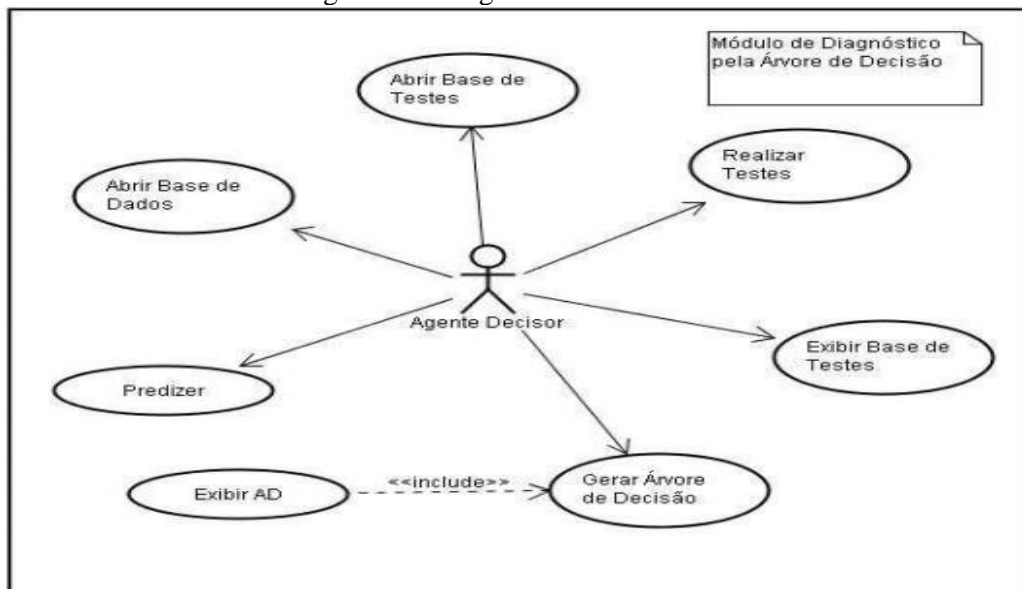
Figura 13 - Camadas da aplicação



Fonte: Zoby (2009).

Na Figura 14 é possível visualizar o diagrama de casos de uso que demonstra as principais funcionalidades da aplicação. O Agente Decisor pode ser um médico ou estudante de medicina.

Figura 14 - Diagrama de casos de uso



Fonte: Zoby (2009).

A principal tela da aplicação onde o usuário irá realizar a interação pode ser visualizada na Figura 15.

Figura 15 - Principal tela do SADM-DST Desktop

Sistema de Apoio ao Diagnóstico Médico de Doenças Sexualmente Transmissíveis - SADM/DST

Base de Dados Algoritmo

Dados Testes

Árvore de Decisão

Corrimento Abundante ou Discreto

- Discreto
  - Presença de Corrimento Uretral
    - Não → Nenhuma
    - Sim → Clamídia
  - Abundante
    - Presença de Corrimento Uretral
      - Não → Nenhuma
      - Sim → Gonorréia

Predição

Presença de Corrimento Uretral: Sim

Coloração do Corrimento: Branco Leitoso

Corrimento Abundante ou Discreto: Abundante

Volume do Corrimento: Sai espontaneamente

Ardor a micção: Sim

Diplococos Gram Negativos: Sim

Saída: **Gonorréia**

Predizer

Base de Dados: dst.data (33)

Presença ...	Coloraçã...	Corriment...	Volume d...	Ardor a m...	Diplococo...	Doença
Sim	Branco Leitoso	Abundante	Sai expona...	Sim	Sim	Gonorréia
Sim	Amarelo	Abundante	Sai expona...	Sim	Sim	Gonorréia
Sim	Mucóide	Discreto	Apenas com...	Sim	Não	Clamídia
Não	Branco Leitoso	Abundante	Sai expona...	Sim	Sim	Nenhuma

Fonte: Adaptado de Zoby (2009).

### 3 DESENVOLVIMENTO

Nesta seção são apresentadas as etapas do desenvolvimento da ferramenta. São apresentados os requisitos, a especificação e a implementação, demonstrando as técnicas e ferramentas utilizadas. Ao final é apresentada a operacionalidade da ferramenta e os resultados obtidos.

#### 3.1 REQUISITOS

A seguir são mostrados os Requisitos Funcionais (RF) e Não Funcionais (RNF) atendidos pela ferramenta, apresentados respectivamente no Quadro 1 e no Quadro 2, que para cada RF foi relacionado o seu caso de uso.

Quadro 1 - Requisitos funcionais

<b>Requisitos funcionais (RF)</b>	<b>Caso de uso (UC)</b>
RF01: permitir o cadastro de conexão da base de dados com informações clínicas	UC01
RF02: permitir o cadastro das configurações da árvore	UC02
RF03: permitir o cadastro das configurações dos atributos da árvore	UC03
RF04: permitir executar mineração de dados	UC04
RF05: permitir o cadastro das configurações de usuário	UC05
RF06: permitir executar avaliação clínica	UC06
RF07: permitir visualizar árvore com diagnóstico geral	UC07
RF08: permitir visualizar árvore com o diagnóstico da avaliação clínica	UC08

Fonte: elaborado pelo autor.

Quadro 2 - Requisitos não funcionais

<b>Requisitos não funcionais (RNF)</b>
RNF01: utilizar o ambiente de desenvolvimento Netbeans e a linguagem Java
RNF02: utilizar o banco de dados MySQL

Fonte: elaborado pelo autor.

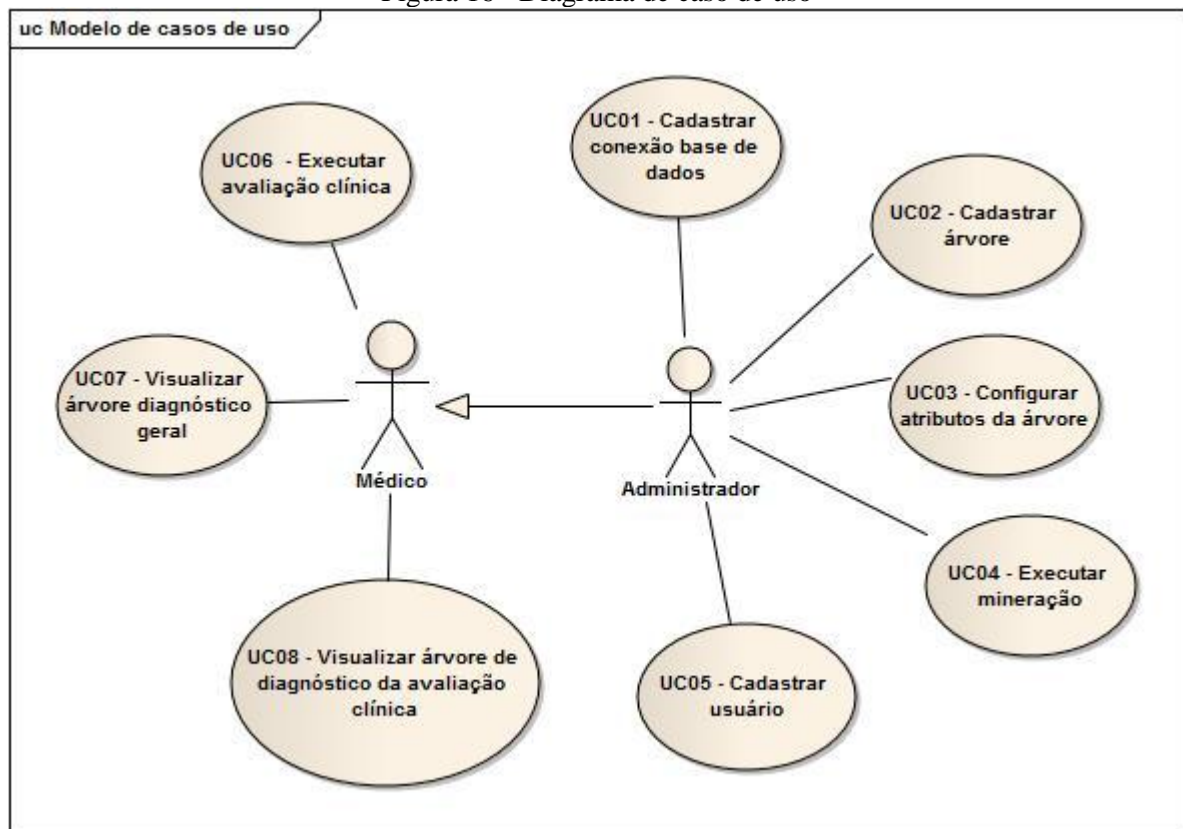
#### 3.2 ESPECIFICAÇÃO

Nesta seção é apresentada a especificação da ferramenta. A especificação foi desenvolvida seguindo o Unified Modeling Language (UML), sendo os diagramas desenvolvidos através da ferramenta Enterprise Architect.

##### 3.2.1 Diagrama de casos de uso

A Figura 16 demonstra o diagrama de caso de uso da ferramenta, na qual o administrador é responsável por todos os cadastros, bem como executar a mineração para gerar as árvores. O médico deve executar a avaliação clínica e visualizar as árvores de diagnóstico.

Figura 16 - Diagrama de caso de uso

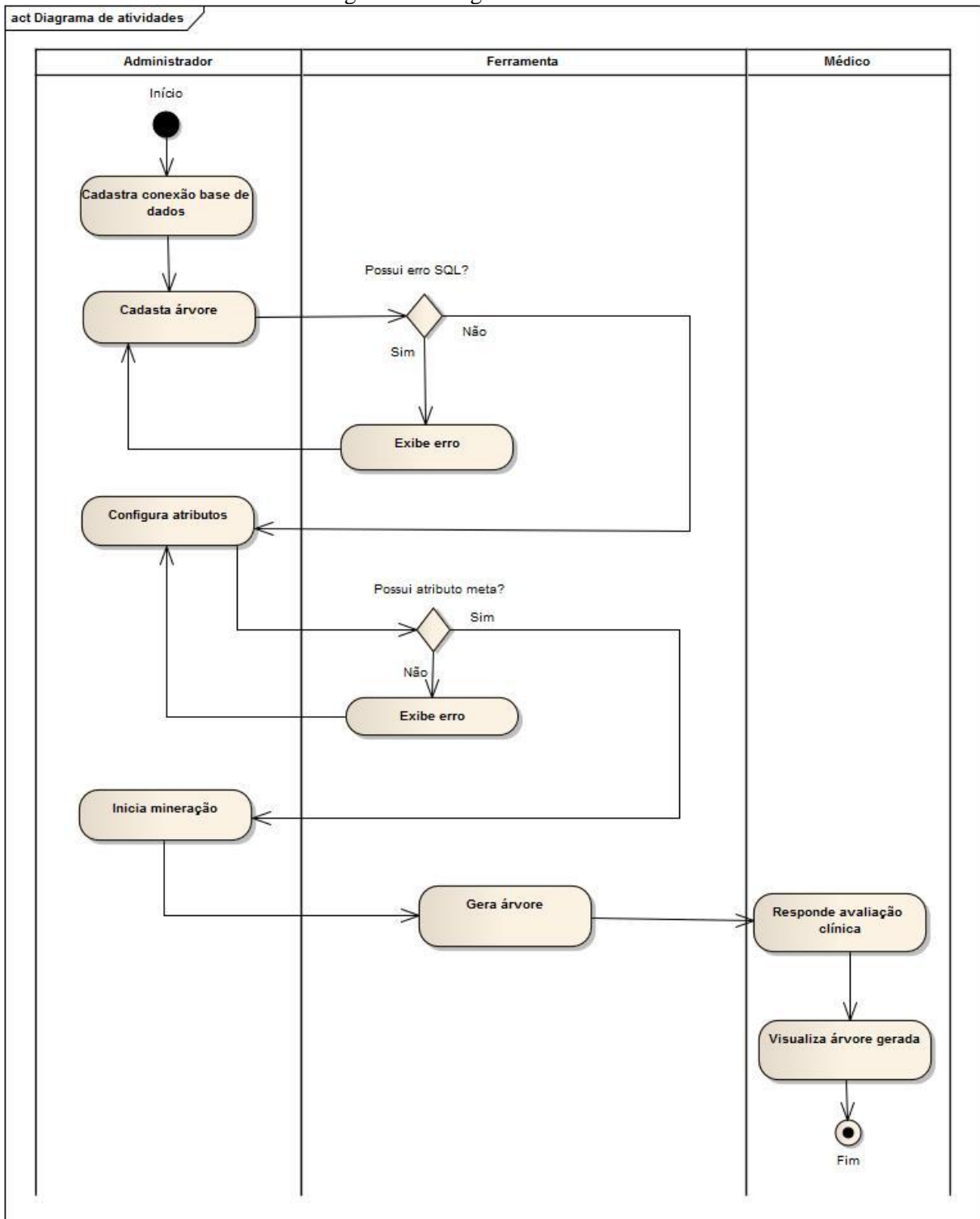


Fonte: elaborado pelo autor.

### 3.2.2 Diagrama de atividades

Nesta seção é apresentado o diagrama de atividades. A Figura 17 demonstra o processo para indução da árvore e visualização das árvores geradas. O administrador deve cadastrar a conexão do banco de dados da base de informações clínicas, as configurações da árvore e também configurar os atributos da árvore (definir classe meta, tipo atributo, etc.). A árvore corresponde ao diagnóstico pretendido de uma doença específica (doença pulmonar, doença cardíaca, doença cardiovascular, etc.), e os atributos representam os fatores predominantes que caracterizam a doença.

Figura 17 - Diagrama de atividades



Fonte: elaborado pelo autor.

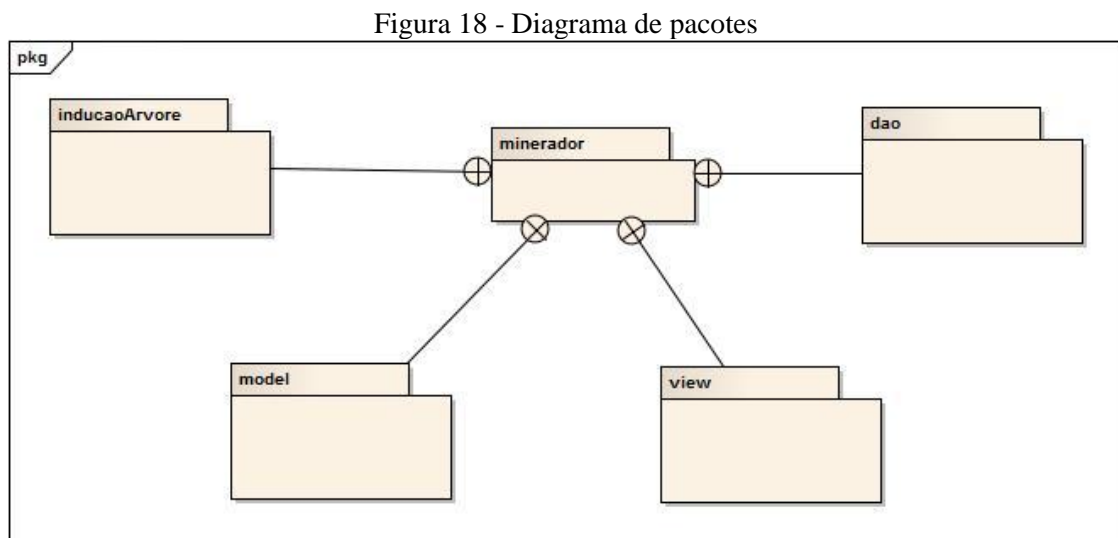
Após a configuração da árvore e dos atributos, o administrador inicia o processo de mineração. Na sequência a ferramenta executa a indução na árvore através do algoritmo C4.5 e cria uma estrutura de árvore para apresentar o resultado da mineração de dados. A árvore resultante exibe todas as doenças do diagnóstico pretendido, juntamente com a combinação de

fatores que podem caracterizar cada doença. Por exemplo: a possibilidade da doença cardíaca angina é de 90% quando o paciente é fumante, sedentário e hipertenso.

A partir do momento que a mineração foi executada, o médico pode executar uma avaliação clínica para determinar o diagnóstico de um paciente. Nesta avaliação serão apresentadas algumas perguntas, que correspondem aos atributos da árvore configurada, ou seja, são os fatores que caracterizam uma determinada doença. Como opção de resposta para cada pergunta, são apresentados os valores distintos resultantes da consulta de cada atributo. Após responder o questionário será exibida a árvore resultante da mineração, mostrando a possibilidade de cada doença de acordo com os sintomas do paciente.

### 3.2.3 Diagrama de pacotes

O objetivo desta seção é apresentar o diagrama de pacotes e de classes implementados para o desenvolvimento da ferramenta. O projeto foi separado em pacotes, com o objetivo de organizar as classes por solução. A Figura 18 apresenta um diagrama de pacotes da solução proposta.



Fonte: elaborado pelo autor.

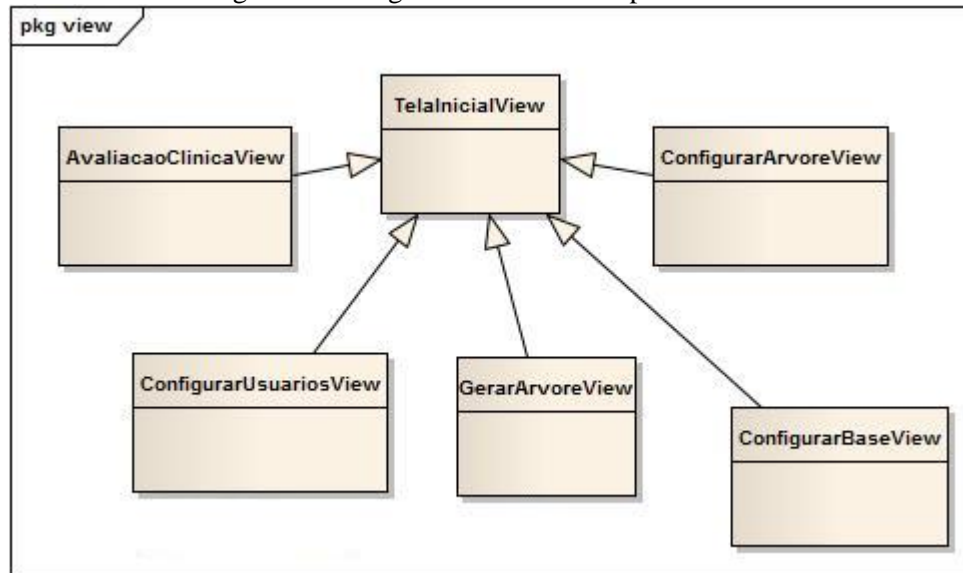
#### 3.2.3.1 Pacote view

As classes do pacote `view` são apresentadas na Figura 19. A classe `TelaInicialView` corresponde ao frame principal da ferramenta onde são exibidas as opções de menu Avaliação clínica, Gerar Árvore, Configurar base, Configurar árvore e Usuários.

A classe `ConfigurarBaseView` corresponde ao cadastro das informações da base de dados que será utilizada na mineração que pode ser acessado pelo menu Configurar base. A classe `ConfigurarUsuariosView` corresponde cadastro de usuários que pode ser acessado

pelo menu `Usuários`. A classe `ConfigurarArvoreView` corresponde ao cadastro da árvore de diagnóstico e dos atributos e pode ser acessado pelo menu `Configurar árvore`. A classe `AvaliacaoClinicaView` corresponde a avaliação clínica que o médico irá realizar com o paciente e pode ser acessada pelo menu `Avaliação clínica`.

Figura 19 - Diagrama de classes do pacote `view`



Fonte: elaborado pelo autor.

### 3.2.3.2 Pacote `dao`

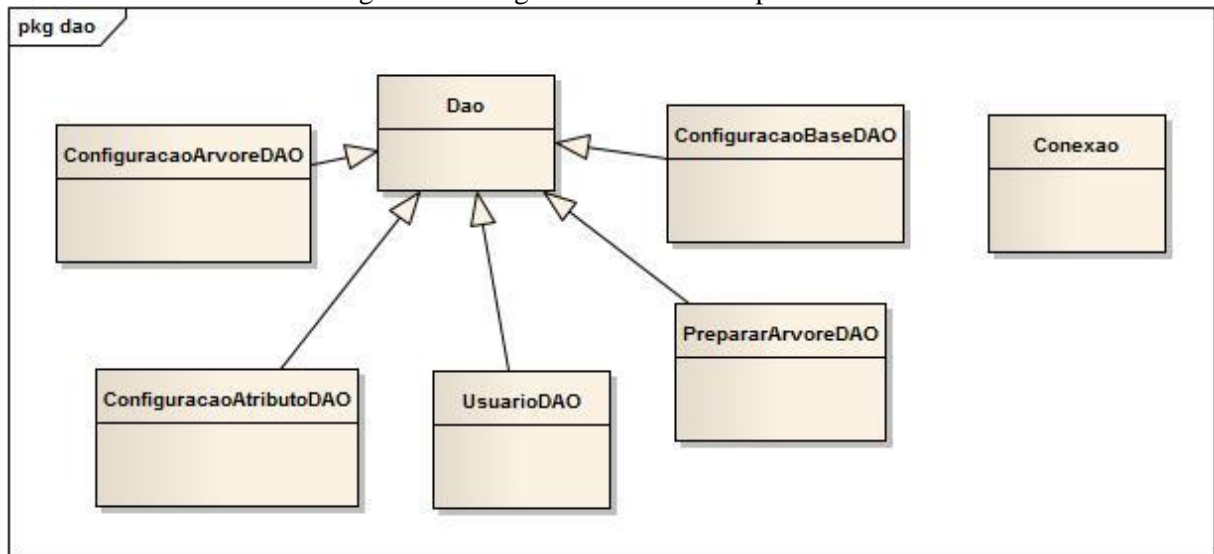
As classes do pacote `dao` são apresentadas na Figura 20. As classes `ConfiguracaoArvoreDAO`, `ConfiguracaoAtributoDAO`, `UsuarioDAO` e `ConfiguracaoBaseDAO` são responsáveis pelas operações de inclusão, edição, exclusão e seleção de registros no banco de dados utilizado pela ferramenta.

A classe `Conexao` é responsável por gerenciar o ciclo de vida das conexões utilizadas pela ferramenta, tanto para as operações de cadastro como o processo de mineração.

A classe `PrepararArvoreDAO` é responsável pela operação de consulta dos valores dos atributos da árvore de diagnóstico.



Figura 20 - Diagrama de classes do pacote dao



Fonte: elaborado pelo autor.

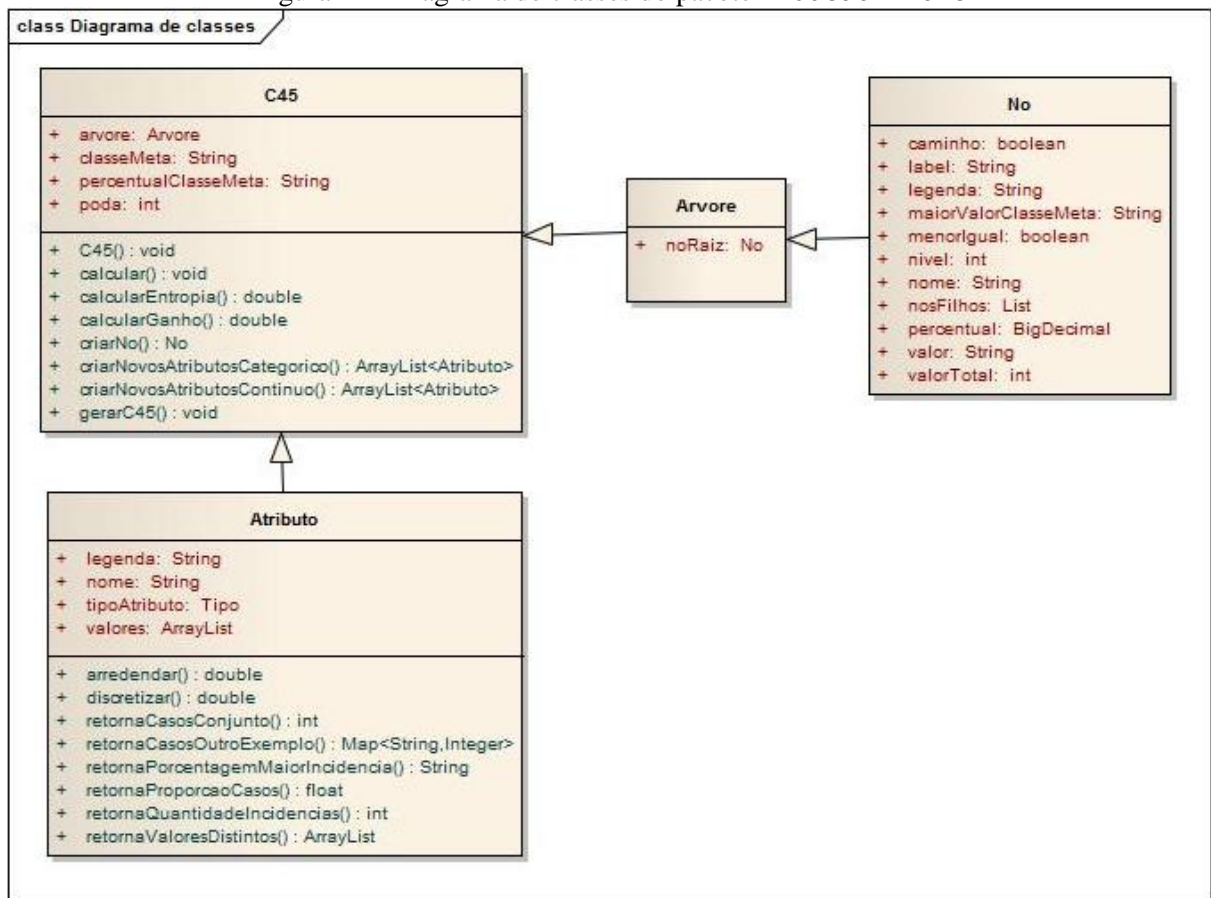
### 3.2.3.3 Pacote `inducacaoArvore`

As classes do pacote `inducacaoArvore` são apresentadas na Figura 21 e são responsáveis pela mineração de dados da aplicação. A classe `C45` é responsável pela indução da árvore de decisão onde é realizada a entropia da classe meta e o ganho de cada atributo de acordo com o nível atual da árvore. O processo de indução é iniciado através do método `Calcular` que é responsável por definir a classe meta como nó raiz da árvore resultante e chamar o método recursivo `exectaInducacaoArvoreC45`.

A classe `Atributo` representa os atributos da árvore de diagnóstico e executa algumas operações complementares no processo de indução da árvore. O método `retornaValoresDistintos` irá filtrar as informações distintas dentro da lista de valores do atributo.

A classe `No` representa cada elemento da árvore resultante do processo de mineração de dados e ela possui apenas atributos de identificação. A classe `Arvore` possui uma referência para o nó raiz, que é definido no método `Calcular` da classe `C45` e a partir desta referência será criada a árvore que será apresentada para o usuário.

Figura 21 - Diagrama de classes do pacote inducaoArvore



Fonte: elaborado pelo autor.

### 3.3 IMPLEMENTAÇÃO

A seguir são mostradas as técnicas e ferramentas utilizadas e a operacionalidade da implementação.

#### 3.3.1 Técnicas e ferramentas utilizadas

A ferramenta foi implementada na linguagem Java para as operações de mineração de dados. Para a parte visual foi utilizada a tecnologia Java Swing, que possui bibliotecas gráficas de fácil utilização.

O Netbeans foi utilizado para codificação, pois ele possui um kit de ferramentas para criar interfaces do usuário, e possibilita o desenvolvimento rápido e fácil de aplicações desktop.

Foi utilizado o MySQL para as operações de banco de dados, pois trabalha com o modelo relacional, possui um bom desempenho, é pouco exigente nos recursos de hardware, e é amplamente utilizado por aplicações com diversos níveis de complexidade.

### 3.3.1.1 Execução da mineração das árvores

No Quadro 3 pode-se visualizar o trecho de código responsável por iniciar o processo de mineração. Inicialmente é construída uma lista com todas as árvores configuradas, e a primeira operação realizada dentro do laço de repetição é a obtenção dos valores dos atributos da árvore. Esta etapa consiste em executar o SQL da árvore e vincular o resultado com os atributos, conforme o cadastro previamente realizado. A execução do método `getRegistros` correspondente a etapa de selecionar os dados do processo macro de mineração de dados, conforme descrito na Figura 5.

Quadro 3 - Mineração de todas as árvores

```
public void executaMineracaoTodasArvores() {
    listaArvores = configuracaoArvore.getRegistros();
    for (ConfiguracaoArvoreVO arvore : listaArvores) {
        listaRetorno =
prepararArvore.getValores(arvore.getNrSequencia());
        inserirLog("Preparando atributos da árvore -> " +
arvore.getNmArvore());
        criarAtributosMineracao();
        inserirLog("    Executando o algoritmo C4.5");
        ArvoreGerada.getInstace().addArvore(executarC45(arvore));
    }
}
```

Fonte: elaborado pelo autor.

No Quadro 4 é possível visualizar o trecho de código que cria os atributos que serão utilizados no método `executarC45`. Este processo consiste em percorrer todos os atributos da árvore e vincular a lista de valores resultantes da consulta SQL. No final do processo o atributo será adicionado numa lista, onde estão os atributos preparados para mineração.

Quadro 4 - Preparação dos atributos para mineração

```

private void criarAtributosMineracao() {
    listaAtributosMineracao.clear();
    String valoresExibir = "";
    for (AtributoRetorno retorno : listaRetorno) {
        valoresExibir = "  Atributo=" + retorno.getNmAtributo()
            + " Tipo=" + retorno.getIeTipoAtributo()
            + " Meta=" + retorno.getIeMeta();
        inserirLog(valoresExibir);
        inserirLog("          Valores=" +
retorno.getValores().toString());

        Atributo novo = new Atributo();
        novo.setNome(retorno.getNmAtributo());
        novo.setTipoAtributo(retorno.getIeTipoAtributo().equals("CO")
? Tipo.Contínuo : Tipo.Categorico);

        if (retorno.getIeTipoAtributo().equals("CA")) {
            ArrayList<String> lista = new ArrayList<String>();

            lista.addAll(retorno.getValores());
            novo.setValores(lista);
        } else if (retorno.getIeTipoAtributo().equals("CO")) {
            ArrayList<Integer> lista = new ArrayList<Integer>();
            lista.addAll(retorno.getValores());
            novo.setValores(lista);
        }
        listaAtributosMineracao.add(novo);
    }
}

```

Fonte: elaborado pelo autor.

O trecho de código utilizado para iniciar a indução na árvore pode ser visualizado no Quadro 5. No método `executarC45` é criada uma instância da classe `C45`, e a partir dessa instância é chamado o método `Calcular` passando como parâmetro a classe meta, e a lista de atributos da árvore de diagnóstico. Após o término da execução do método `Calcular` é possível obter a árvore resultante do processo de mineração, através do método `getArvore` da classe `C45`. No final da execução será atualizada a data de geração da árvore para que o administrador da ferramenta tenha um indicativo visual da última atualização.

Quadro 5 - Função que inicia a indução

```

private ArvoreGeradaVO executarC45(ConfiguracaoArvoreVO
configuracaoArvore) {
    ArvoreGeradaVO arvoreGerada = null;
    try {
        C45 c45 = new C45();
        c45.poda = configuracaoArvore.getNrPoda();
        String atributoMeta = prepararArvore.getAtributoMeta(
            configuracaoArvore.getNrSequencia()).getNmAtributo();
        c45.calcular(atributoMeta, listaAtributosMineracao);
        c45.getArvore();
        inserirLog("    Gerando árvore"); inserirLog("\n");

        GerarTree gerarTree = new GerarTree();
        JTree arvore = gerarTree.gerarArvore(c45.getArvore());

        arvoreGerada = new ArvoreGeradaVO();
        arvoreGerada.setNmArvore(configuracaoArvore.getNmArvore());
        arvoreGerada.setNrSeqArvore(
            configuracaoArvore.getNrSequencia());
        arvoreGerada.setNmUsuario(getNmUsuario());
        arvoreGerada.setDtAtualizacao(new Date());
        arvoreGerada.setArvore(arvore);
    } catch (Exception e) {
        inserirLog("    Falha ao gerar árvore! "); inserirLog("\n");
    }
    return arvoreGerada;
}

```

Fonte: elaborado pelo autor.

### 3.3.1.2 Indução da árvore

As etapas pré-processamento e transformação do processo de mineração de dados não são executadas pela ferramenta e o administrador precisa executar estas tarefas manualmente. A ferramenta irá executar etapa *estimar o modelo* através do algoritmo C4.5, que consiste em selecionar os métodos a serem utilizados para localizar padrões nos dados, seguida da efetiva busca por padrões de interesse numa forma particular de representação ou conjunto de representações.

No Quadro 6 pode-se visualizar o trecho de código responsável por implementar o algoritmo C4.5 para a indução da árvore. Inicialmente é realizado o cálculo do ganho de informação sobre a classe meta e o ganho sobre cada atributo, determinando assim o atributo que irá compor o nível.

Quadro 6 - Indução da árvore

```

private void exectaInducaoArvoreC45(String atributoMeta, List<Atributo>
atributos, No no) {

    Atributo valoresClasseMeta =
    buscarAtributoMeta(classeMeta.getNome(), atributos);
    if (valoresClasseMeta.retornaValoresDistintos().size() <= 1) {
        return;
    }
    double valorEntropia = calcularEntropia(valoresClasseMeta);

    Atributo atributoMaiorGanho = new Atributo();
    double maiorGanho = 0;

    ArrayList<Atributo> atributosTemp =
    obterListaAtributosDiferentes(classeMeta.getNome(), atributos);
    if (atributoMeta.equals("Outras palavras") ||
    atributoMeta.equals("Sem comentário")) {
        atributosTemp =
    obterListaAtributosDiferentes(classeMeta.getNome(), atributos, no);
    }

    for (Atributo atributo : atributosTemp) {
        double ganhoAtr = calcularGanho(valorEntropia,
valoresClasseMeta, atributo);
        if (ganhoAtr > maiorGanho) {
            maiorGanho = ganhoAtr;
            atributoMaiorGanho = atributo;
        }
    }
}

```

Fonte: elaborado pelo autor.

### 3.3.1.3 Gerar árvore para apresentação dos resultados da mineração

No Quadro 7 pode-se visualizar o trecho de código que cria uma árvore resultante do processo de indução. A árvore criada nesta rotina será apresentada como resultado para uma avaliação clínica. A execução do método `gerarArvore` corresponde a etapa de interpretação do modelo do processo macro de mineração de dados, conforme descrito na Figura 5.

Quadro 7 - Criação da árvore para apresentação

```

public JTree gerarArvore(Arvore arvore) {
    root = new DefaultMutableTreeNode(
    String.valueOf(arvore.getNoRaiz().getPercentual().toString() + "%"));
    List<No> nosFilhos = arvore.noRaiz.getNosFilhos();
    for (No no : nosFilhos) {
        idNode++;
        gerarArvoreNodeFilhos(no);
    }
    return tree = new JTree(root);
}

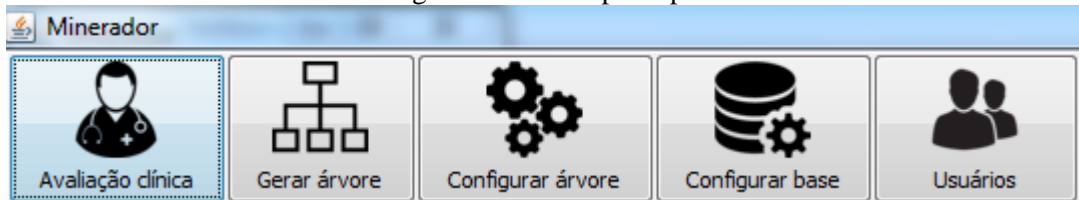
```

Fonte: elaborado pelo autor.

### 3.3.2 Operacionalidade da implementação

Nesta seção são apresentados os detalhes da implementação dos processos que compõem a ferramenta. Na Figura 22 é apresentado o menu com todas as funcionalidades da ferramenta.

Figura 22 - Menu principal

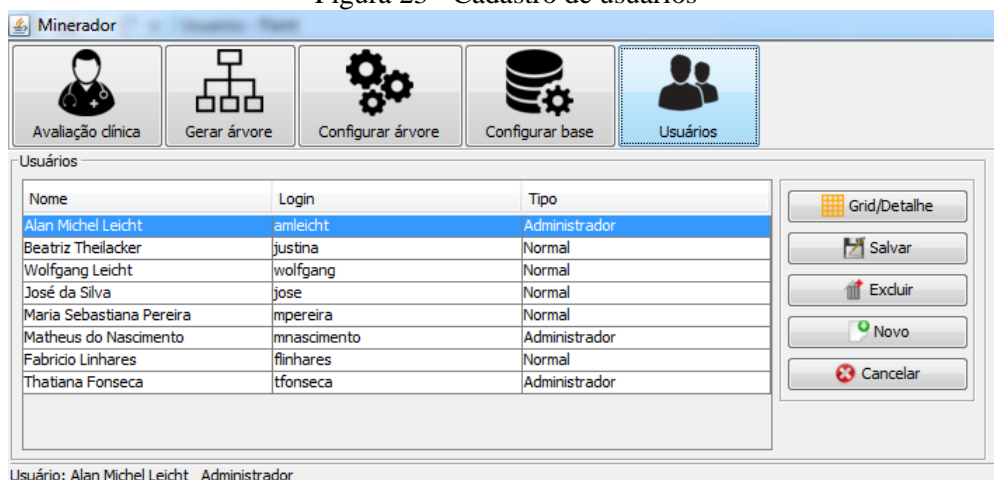


Fonte: elaborado pelo autor.

#### 3.3.2.1 Cadastros

A funcionalidade de cadastro de usuários pode ser vista na Figura 23, onde é possível definir o perfil de acesso do usuário. Somente os usuários que possuem perfil de administrador têm acesso a todas as funcionalidades da ferramenta, e todos os outros usuários terão acesso somente ao menu Avaliação clínica.

Figura 23 - Cadastro de usuários



Fonte: elaborado pelo autor.

A configuração da base de dados utilizada para buscar os registros clínicos pode ser visualizada na Figura 24, onde é possível realizar um teste de conexão para confirmar se os dados de acesso foram cadastrados corretamente.

Figura 24 - Tela de configuração da base de dados

Minerador

Configurar base

Servidor  
localhost

Base  
hospital

Usuário  
root

Senha  
••••••

Conexão realizada com sucesso!

Usuário: Alan Michel Leicht Administrador

Fonte: elaborado pelo autor.

A Figura 25 demonstra o cadastro da estrutura da árvore.

Figura 25 - Configuração da árvore

Minerador

Configurar árvore

Árvore Atributos

Código	Nome
1	Doença cardíaca
6	Doença pulmonar
7	Doença cardiovascular

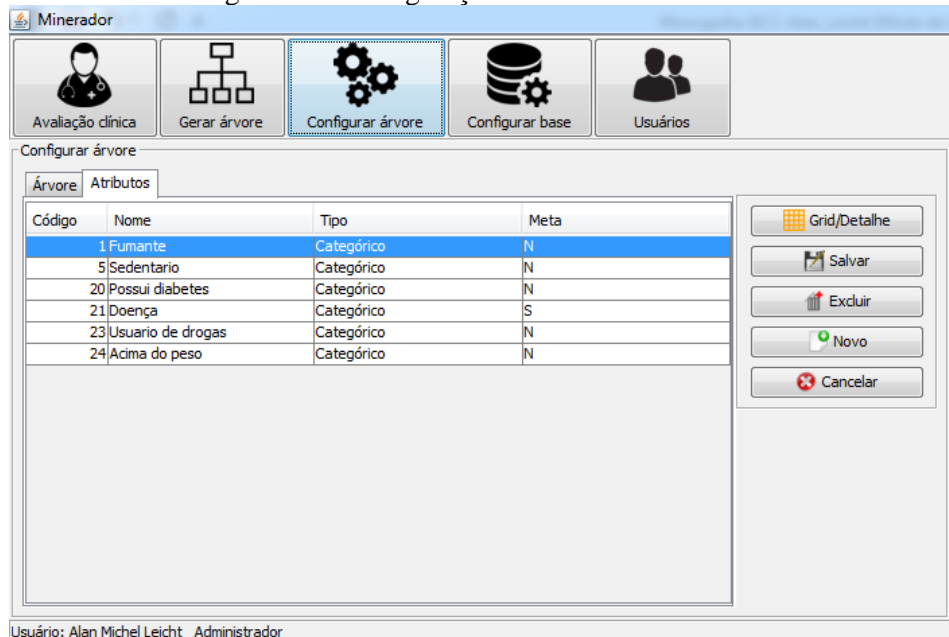
Usuário: Alan Michel Leicht Administrador

Fonte: elaborado pelo autor.

Para que seja possível a geração da árvore se faz necessário realizar a configuração dos atributos da árvore (Figura 26). Os atributos são carregados a partir da consulta SQL informada no cadastro da árvore. Deve-se informar o atributo que será classe meta, o tipo de cada atributo (Figura 27).

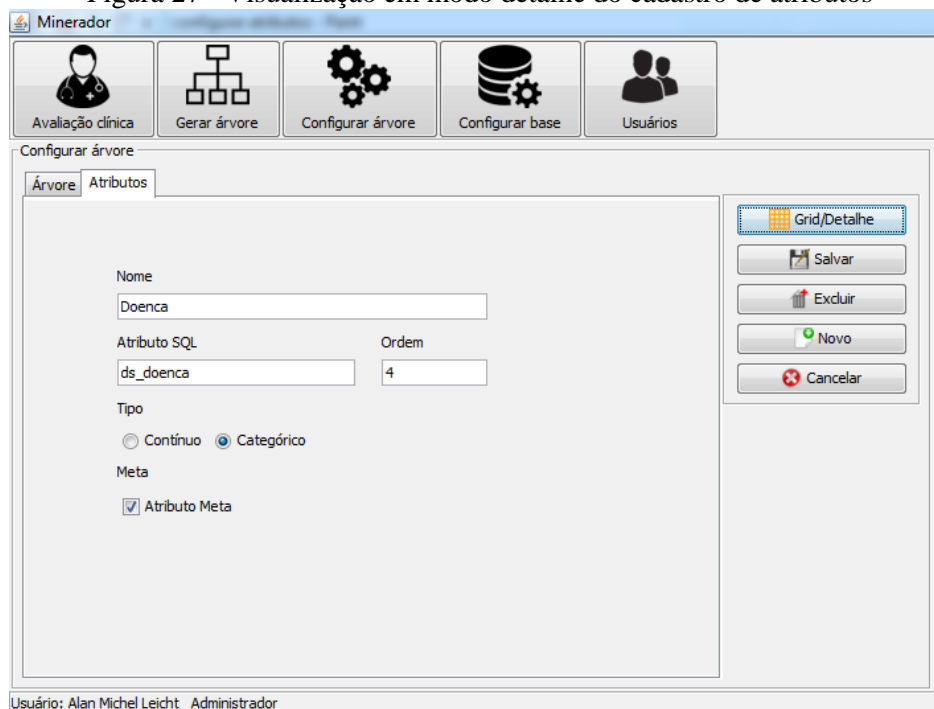


Figura 26 - Configuração dos atributos da árvore



Usuário: Alan Michel Leicht Administrador  
 Fonte: elaborado pelo autor.

Figura 27 - Visualização em modo detalhe do cadastro de atributos



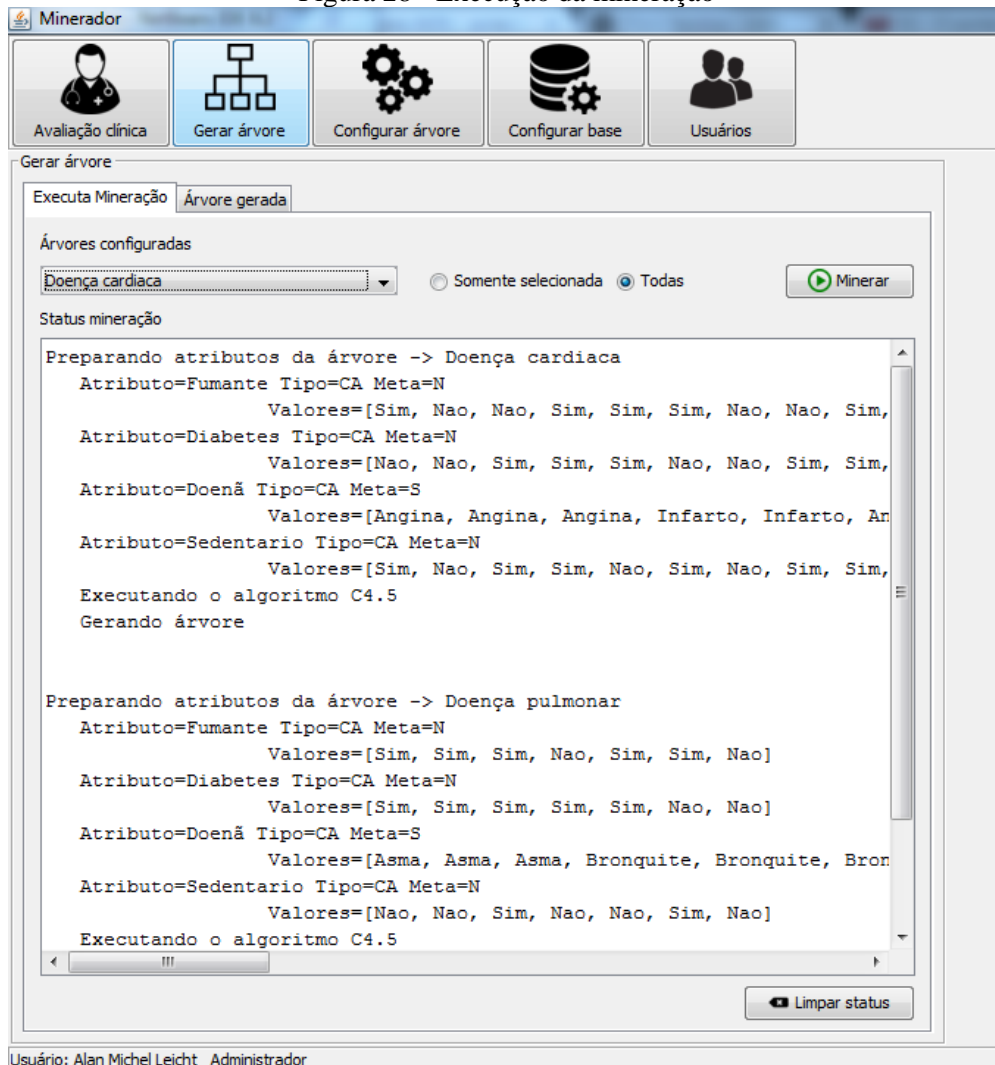
Usuário: Alan Michel Leicht Administrador  
 Fonte: elaborado pelo autor.

A Figura 28 demonstra o processo de execução da mineração de dados. O administrador pode optar por executar a mineração de todas as árvores configuradas, ou somente de uma árvore específica. O `Status mineração` exibe os logs do processo de mineração, que são divididos em três etapas:

- preparar atributos: será executado o SQL configurado na árvore e os valores serão vinculados com os atributos;

- b) indução da árvore: nesta etapa será executado o algoritmo C4.5;
- c) gerar árvore resultante: o resultado da indução será transformado em uma árvore para que seja apresentado ao usuário que está realizando a avaliação clínica.

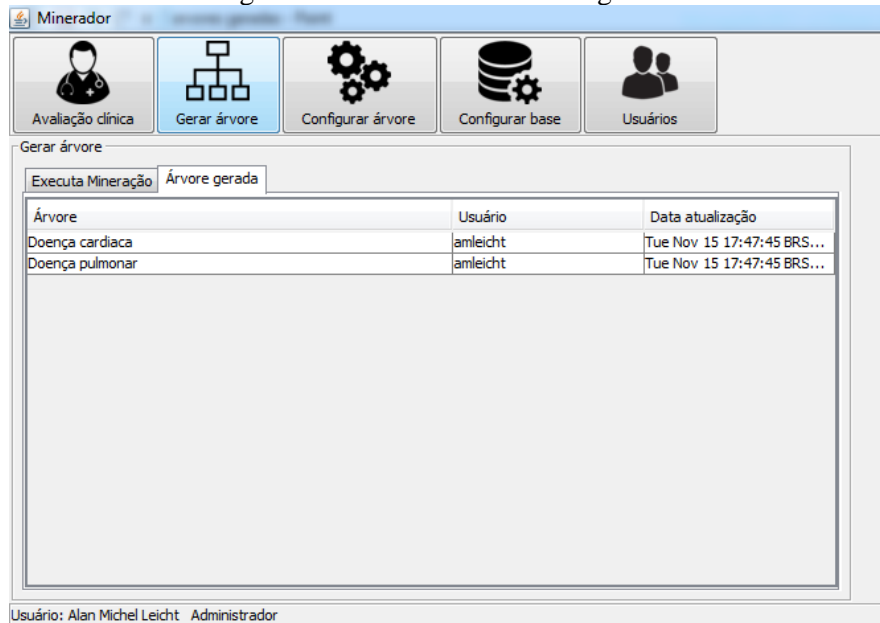
Figura 28 - Execução da mineração



Fonte: elaborado pelo autor.

O administrador da ferramenta pode acompanhar o status de atualização de cada árvore, conforme a Figura 29, sendo que no grid é possível visualizar o usuário responsável de a data de atualização de cada árvore.

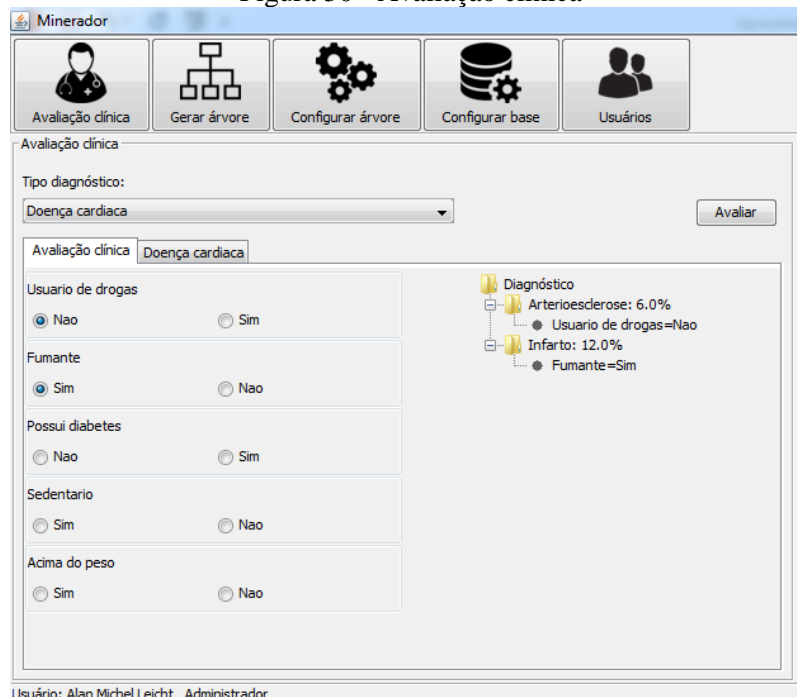
Figura 29 - Grid com as árvores geradas



Fonte: elaborado pelo autor.

Após realizado o processo de geração das árvores, que consiste em executar a mineração de dados e criar uma árvore resultado, será possível executar uma avaliação clínica. A Figura 30 demonstra a avaliação clínica que o médico irá realizar com um paciente. As perguntas exibidas na avaliação, são os atributos configurados na árvore, e a opções representam todos os valores distintos obtidos de cada atributo. Será exibida uma árvore somente com os nós correspondentes às repostas da avaliação, para que o médico consiga realizar o diagnóstico da doença do paciente.

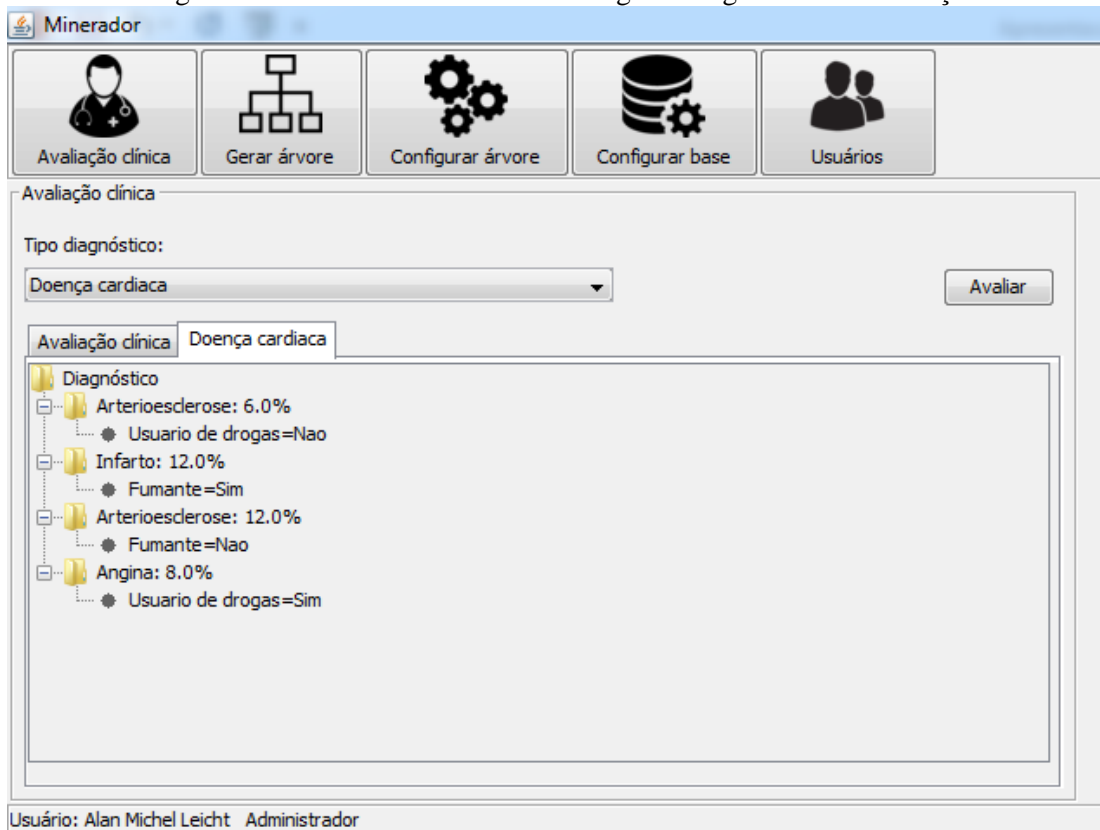
Figura 30 - Avaliação clínica



Fonte: elaborado pelo autor.

A Figura 31 exibe uma árvore completa com o diagnóstico geral de uma doença, ou seja, serão apresentados todos os fatores que determinam o diagnóstico de uma determinada doença de acordo com o resultado da mineração realizada.

Figura 31 - Árvore resultante com o diagnóstico geral de uma doença



Fonte: elaborado pelo autor.

### 3.4 ANÁLISE DOS RESULTADOS

A solução desenvolvida apresentou uma ferramenta de mineração de dados que permite a interação com o usuário, auxiliando-o no processo de tomada de decisão clínica através de uma árvore de decisão sendo capaz de sugerir um diagnóstico para vários tipos de doença.

Em testes simulando as atividades do usuário médico foi possível executar uma avaliação clínica e visualizar a árvore resultante de uma forma muito simples. Esse processo é simplificado porque o questionário da avaliação é criado de acordo com os atributos configurados na árvore de diagnóstico e as respostas possíveis, são obtidas após o processo de mineração, logo o maior esforço é realizado pela ferramenta. Foram realizados testes simulando as atividades do administrador da ferramenta com dois usuários que não possuem conhecimento em sistemas de mineração de dados. Apesar da amostra não ser expressiva, os

resultados obtidos foram positivos, haja vista que os usuários conseguiram configurar uma base e executar o processo de mineração.

Durante os testes com os usuários e desenvolvimento, foram identificados alguns problemas e dificuldades. A primeira a ser destacada foi a definição dos atributos para construir uma árvore de qualquer diagnóstico de doença. Como é necessário um conhecimento específico, seria necessário o auxílio de profissionais da saúde nesse processo de configuração. A participação de um médico neste processo seria importante para fazer o conhecimento e contextualização com os dados antes de iniciar a mineração, o que poderia auxiliar no entendimento dos resultados e geração de informações mais relevantes, bem como possibilitar o ajuste de parâmetros para o melhoramento dos mesmos.

Outra dificuldade foi a obtenção de uma base de dados com registros clínicos reais. Nenhuma instituição forneceu uma base com informações de tratamentos clínicos realizados em pacientes, porque tratam-se de informações sigilosas e não poderiam ser utilizadas neste trabalho. Mesmo com uma base de dados real seria necessário conhecer muitos detalhes da modelagem, para realizar o agrupamento das tabelas e a extração das informações necessárias para a execução da mineração. Neste caso seria necessário uma parceria com algum hospital disposto a ajudar no desenvolvimento da pesquisa. Como solução alternativa foi criada uma base própria com registros de tratamentos clínicos.

A ferramenta foi desenvolvida na linguagem Java, logo não existe uma relação de dependência com um sistema operacional, essa característica é importante pois permite que a aplicação seja executada em diferentes sistemas operacionais. O Quadro 8 apresenta de forma comparativa algumas características em relação aos trabalhos correlatos e ao trabalho apresentado nesta pesquisa.

Quadro 8 - Comparativo com os trabalhos correlatos

	Ferramenta Leicht (2016)	Morais et al. (2012)	Zoby (2009)
Linguagem	Java	Java	Java
Plataforma	Desktop	Móvel	Desktop
Utiliza o algoritmo C4.5	Sim	Sim	Não
Permite o diagnóstico de várias doenças	Sim	Não	Não
Técnica de classificação utilizada	Árvore de decisão	Árvore de decisão	Árvore de decisão

Fonte: elaborado pelo autor.

Analisando o Quadro 8 é possível visualizar que todos os trabalhos utilizaram a linguagem de programação Java e também a mesma técnica de classificação. Sendo que, os trabalhos correlatos não possuem um usuário administrador para realizar toda a configuração

e executar a mineração de dados, enquanto a ferramenta deste trabalho é desenvolvida levando em consideração a existência de dois grupos de usuários, os administradores e os médicos. Neste contexto, os usuários médicos são responsáveis apenas por executar a avaliação clínica e analisar os resultados.

A principal diferença entre o trabalho desenvolvido e os correlatos é que o mesmo permite a criação de diversos tipos de árvores possibilitando a realização do diagnóstico de mais de um tipo de doença. Essa característica é importante pois torna a ferramenta mais dinâmica e de fácil adaptação a diferentes cenários.

## 4 CONCLUSÕES

O principal objetivo do foi disponibilizar uma ferramenta que possa dar suporte às equipes médicas durante o processo de definição do diagnóstico de um paciente, detectando os padrões em uma massa de dados, e indicando quais doenças são mais caracterizadas pelos sintomas informados. O objetivo proposto foi alcançado, mas a utilização de um banco de dados com informações reais de tratamentos clínicos poderia ter contribuído para uma maior confiabilidade dos resultados apresentados, tendo em vista que foi necessário criar uma base própria com registros clínicos.

O processo de mineração de dados foi realizado através da técnica de árvore de decisão implementada com o algoritmo C4.5. A partir do momento que a mineração foi executada, o médico pode iniciar uma avaliação clínica para determinar o diagnóstico de um paciente. Nesta avaliação serão apresentadas algumas perguntas, que correspondem aos atributos da árvore configurada, ou seja, são os fatores que caracterizam uma determinada doença. Como opção de resposta para cada pergunta, serão apresentados os valores distintos resultantes da mineração de dados. Após responder o questionário será exibida a árvore resultante da mineração

A árvore de decisão resultante do processo de mineração exibirá todas as doenças do diagnóstico pretendido, juntamente com a combinação de fatores que podem caracterizar cada doença. Por exemplo: a possibilidade da doença cardíaca angina é de 90% quando o paciente é fumante, sedentário e hipertenso. A árvore gerada é apresentada em dois formatos, para que o médico possa ter uma visão geral de todos os fatores que podem estar relacionados ao diagnóstico de determinada doença, e outra visão apenas com os fatores respondidos na avaliação clínica, representado apenas a realidade de um paciente.

A ferramenta possibilita a criação de várias árvores de diagnóstico e este é um dos diferenciais mais importantes em relação aos trabalhos correlatos. Essa característica torna a solução dinâmica e de fácil adaptação de em diferentes cenários. Por exemplo, o médico pode fazer uma avaliação clínica de doença cardíaca, e com uma pequena interação pode alterar para uma avaliação de doença pulmonar.

O processo de diagnóstico da doença de um paciente, é complexo e a qualidade das informações obtidas é importante, e a partir delas que o médico estabelece relações entre os sintomas e a doença e os possíveis tratamentos, com base nos conhecimentos adquiridos ao longo da profissão. Assim sendo, a ferramenta deverá ser utilizada como complemento aos

conhecimentos e habilidades dos médicos, visando a agilidade e a diminuição de falhas no atendimento do paciente

Dentre os objetivos alcançados, vale destacar ainda que, com a possibilidade do médico poder realizar uma avaliação clínica durante o processo de tomada de decisão, há um ganho de segurança para o paciente, tendo em vista que o volume de informações necessárias para a análise é grande e falhas humanas poderiam ocorrer.

#### 4.1 EXTENSÕES

São sugeridas as seguintes extensões para que a ferramenta seja melhorada em trabalhos futuros:

- a) migrar para a plataforma Web;
- b) criar modo de visualização da árvore gerada em formato de gráfico;
- c) adicionar funcionalidade para agendar a execução da mineração de dados;
- d) permitir acesso ao uma base que esteja na nuvem.



## REFERÊNCIAS

- ABREU, Mirhelen Mendes de et al. Apoios de decisão: instrumento de auxílio à medicina baseada em preferências. **Bras Reumatol**, São Paulo, v. 46, n. 4, p.266-272, jun. 2006. Disponível em: <<http://www.scielo.br/pdf//rbr/v46n4/31822.pdf>>. Acesso em: 23 ago. 2016.
- ANDRADE, Pedro T. B. **Desenvolvimento e implementação de um sistema especialista para auxiliar o acompanhamento e o diagnóstico médico na área de pediatria**. 2005. 83 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais, Belo Horizonte.
- CARVALHO, Luis A. V. **Dataming: a mineração de dados no marketing, medicina, economia, engenharia e administração**. 2. Ed. São Paulo: Érica, 2001.
- CASTANHEIRA, Luciana Gomes. **Aplicação de técnicas de mineração de dados em problemas de classificação de padrões**. 2008. 95 f. Dissertação (Mestrado) - Curso de Programa de Pós-graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, 2008. Disponível em: <<http://www.ppgee.ufmg.br/defesas/349M.PDF>>. Acesso em: 13 out. 2016.
- CHIMIESKI, Bruno F. **Solução de auxílio ao diagnóstico e à pesquisa médica baseada em mineração de dados utilizando interface android**. 2013. 146 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Pontífca Universidade Católica do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <<http://repositorio.pucrs.br/dspace/bitstream/10923/3199/1/000445055-Texto+Completo-0.pdf>>. Acesso em: 25 ago. 2016.
- CÔRTEZ, Sérgio C.; PORCARO, Rosa M.; LIFSCHITZ, Sérgio. **Mineração de Dados: funcionalidades, técnicas e abordagens**. [Rio de Janeiro], 2002. Disponível em <[ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02\\_10\\_cortes.pdf](ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf)>. Acesso em: 07 set. 2016.
- FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, Brasília, v. 17, n. 1, 1996. Disponível em: <<http://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>>. Acesso em: 21 set. 2016.
- GONÇALVES, Lóren P. F. **Mineração de dados em supermercados: o caso do supermercado “TAL”**. 1999. 36 f. Dissertação (Mestrado em Administração) Universidade Federal do Rio Grande do Sul, Porto Alegre.
- HALMENSCHLAGER, Carine. **Um algoritmo para indução de árvores e regras de decisão**. 2002. 112 f. Dissertação (Mestrado) - Curso de Programa de Pós-graduação em Computação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/2755/000325797.pdf?sequence=1>>. Acesso em: 01 set. 2016.
- HOSOKAWA, Eric O. **Técnica de árvore de decisão em mineração de dados**. 2011. 40 f. Monografia (Tecnólogo em Processamento de Dados) – Faculdade de Tecnologia de São Paulo, São Paulo.
- LEITE, Corrêa J.; PAES, Libânia R. A. **O uso da informática no processo de tomada de decisão médica: um estudo de caso em hospitais de cardiologia na cidade de São Paulo**. [São Paulo], 2003. Disponível em <[http://www.anpad.org.br/diversos/trabalhos/EnANPAD/enanpad\\_2003/ACT/2003\\_ACT1910.pdf](http://www.anpad.org.br/diversos/trabalhos/EnANPAD/enanpad_2003/ACT/2003_ACT1910.pdf)>. Acesso em: 21 set. 2016.

- MOURA, Angelo O.; ANDRADE, Diego S.; KIKUCHI, Wagner k. Inteligência artificial aplicada no auxílio a tomadas de decisões na área da saúde. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 2., 2005, Resende. **Anais...** Resende: [s.n.], 2005. p. 1096-1102. Disponível em: <[http://www.aedb.br/seget/artigos05/293\\_Inteligencia%20Artificial%20Aplicada%20no%20Auxilio%20a%20Tomadas%20de%20Decisoes%20na%20Area%20de%20Saude.pdf](http://www.aedb.br/seget/artigos05/293_Inteligencia%20Artificial%20Aplicada%20no%20Auxilio%20a%20Tomadas%20de%20Decisoes%20na%20Area%20de%20Saude.pdf)>. Acesso em: 29 set. 2016.
- MORAIS, Dyego C. L. et al. Sistema móvel de apoio a decisão clínica aplicada ao diagnóstico de asma – InteliMED. In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO, 8., 2012, Recife. **Anais...** Recife: [s.n.], 2012. p. 528-539. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2012/0051.pdf>>. Acesso em: 01 out. 2016.
- RÉA NETO, Álvaro. **Raciocínio clínico: o processo de decisão diagnóstica e terapêutica.** [Curitiba], 1998. Disponível em <<http://www.scielo.br/pdf/ramb/v44n4/1898.pdf>>. Acesso em: 19 set. 2016.
- POZZER, Cesar Tadeu. **Aprendizado por árvores de decisão.** 2006. Disponível em: <[http://www-usr.inf.ufsm.br/~pozzzer/disciplinas/pj3d\\_decisionTrees.pdf](http://www-usr.inf.ufsm.br/~pozzzer/disciplinas/pj3d_decisionTrees.pdf)>. Acesso em: 01 nov. 2016.
- REZENDE, Solange Oliveira. **Sistemas inteligentes: fundamentos e aplicações.** Barueri: Editora Manole Ltda, 2003. 525 p, il.
- ROCHA, Álvaro; VASCONCELOS, José B.; MOREIRA, Rui. **Educação Superior em Informática: oportunidade para Portugal.** [Porto], 2005. Disponível em <<http://www.lbd.dcc.ufmg.br/colecoes/wim/2005/0013.pdf>>. Acesso em: 20 set. 2016.
- SEMANN, Rafael. **Ferramenta para predição de dados proeminentes de sistemas RH.** 2015. 69 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade Regional de Blumenau, Blumenau, 2015. Disponível em: <<http://dsc.inf.furb.br/tcc/index.php?cd=6&tcc=1697>>. Acesso em: 01 set. 2016.
- SILVA, Marcelino P. S. **Mineração de Dados: conceitos, aplicações e experimentos com Weka.** [S.l.], 2004. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 20 set. 2016.
- SIGULEM, Daniel et al. **Sistemas de apoio a decisão em medicina.** [São Paulo], [2010?]. Disponível em: <[http://www.virtual.epm.br/material/tis/curr-med/sad\\_html/sistema.htm](http://www.virtual.epm.br/material/tis/curr-med/sad_html/sistema.htm)>. Acesso em: 23 set. 2016.
- SIMÕES, Adriana C. A. **Mineração de dados baseada em árvores de decisão para análise do perfil de contribuintes.** 2008. 140 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2008. Disponível em: <<http://repositorio.ufpe.br:8080/bitstream/handle/123456789/1476/acas.pdf?sequence=2&isAllowed=y>>. Acesso em: 20 ago. 2016.
- TANG, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao dataming.** Tradução Acauan P. Fernandes. Rio de Janeiro: Ciência Morena, 2009.
- ZOBY, Eduardo. **Sistema de apoio à decisão para o diagnóstico médico de doenças sexualmente transmissíveis (SADM – DST).** 2006. 57 f. TCC (Graduação) - Curso de Engenharia da Computação, Escola Politécnica de Pernambuco, Recife, 2006.

## APÊNDICE A – Detalhamento dos casos de uso

Nesta seção são apresentados os detalhamentos dos casos de uso, com descrição, ator, pré-condição e cenário. O pré-condição e cenário. O Quadro 9 apresenta o caso de uso 04 (Executar mineração), o Fonte: elaborado pelo autor.

Quadro 10 apresenta o caso de uso 06 (Permitir executar avaliação clínica), o Quadro 11 apresenta do caso de uso 07 (Permitir visualizar árvore com diagnóstico geral).

Quadro 9 - Caso de uso: Executar mineração

Número	UC04
Caso de Uso	Executar mineração
Ator	Administrador
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> <li>1. O usuário entra na ferramenta.</li> <li>2. O usuário acessa o menu Gerar árvore.</li> <li>3. O usuário seleciona uma árvore configurada</li> <li>4. O usuário utiliza a opção Minerar para iniciar o processo de mineração de dados</li> </ol>

Fonte: elaborado pelo autor.

Quadro 10 - Caso de uso: Permitir executar avaliação clínica

Número	UC06
Caso de Uso	Permitir executar avaliação clínica
Ator	Médico
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> <li>1. O usuário entra na ferramenta.</li> <li>2. O usuário acessa o menu Avaliação clínica.</li> <li>3. O usuário responde o questionário.</li> </ol>

Fonte: elaborado pelo autor.

Quadro 11 - Caso de uso: Permitir visualizar árvore com diagnóstico geral

Número	UC07
Caso de Uso	Permitir visualizar árvore com diagnóstico geral
Ator	Médico
Pré-condição	Nenhuma
Cenário Principal	<ol style="list-style-type: none"> <li>1. O usuário entra na ferramenta.</li> <li>2. O usuário acessa o menu Avaliação clínica.</li> <li>3. O usuário acessa a aba correspondente ao diagnóstico pretendido.</li> <li>4. O usuário consulta a árvore com todas as informações do diagnóstico pretendido.</li> </ol>

Fonte: elaborado pelo autor.