

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS**

**CENTRO DE CIÊNCIAS EXATAS, AMBIENTAIS E DE  
TECNOLOGIAS**

**MARCELO CICERO RIBEIRO DA SILVA**

**APRENDIZAGEM DE MÁQUINA EM APOIO A  
DIAGNÓSTICO EM ORTOPEDIA**

**CAMPINAS**

**2016**

MARCELO CICERO RIBEIRO DA SILVA

APRENDIZAGEM DE MÁQUINA EM APOIO A  
DIAGNÓSTICO EM ORTOPEDIA

Dissertação apresentada ao Programa de Pós Graduação Stricto Sensu em Engenharia Elétrica do Centro de Ciências Exatas, Ambientais e de Tecnologias da Pontifícia Universidade Católica de Campinas como requisito para obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. David Bianchini.

PUC-CAMPINAS

2016

Ficha Catalográfica  
Elaborada pelo Sistema de Bibliotecas e  
Informação - SBI - PUC-Campinas

t303.4833 Silva, Marcelo Cicero Ribeiro da.  
S586a      Aprendizagem de máquina em apoio a diagnóstico em ortopedia / Marcelo Cicero Ribeiro da Silva. - Campinas: PUC-Campinas, 2016.  
132p.

Orientador: David Bianchini.

Dissertação (mestrado) - Pontifícia Universidade Católica de Campinas, Centro de Ciências Exatas, Ambientais e de Tecnologias, Pós-Graduação em Engenharia Elétrica.

Inclui bibliografia.

1. Tecnologia - Serviços de informação. 2. Mineração de dados (Computação). 3. Ortopedia - Diagnóstico. 4. Tecnologia médica. I. Bianchini, David. II. Pontifícia Universidade Católica de Campinas. Centro de Ciências Exatas, Ambientais e de Tecnologias. Pós-Graduação em Engenharia Elétrica. III. Título.

20.ed. CDD - t303.4833

**MARCELO CÍCERO RIBEIRO DA SILVA**

**APRENDIZAGEM DE MÁQUINA EM APOIO A  
DIAGNÓSTICO EM ORTOPEDIA**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro de Ciências Exatas, Ambientais e de Tecnologias da Pontifícia Universidade Católica de Campinas como requisito parcial para obtenção do título de Mestre em Gestão de Redes de Telecomunicações.

Área de Concentração: Engenharia Elétrica.  
Orientador: Prof. Dr. David Bianchini

Dissertação defendida e aprovada em 13 de dezembro de 2016 pela Comissão Examinadora constituída dos seguintes professores:



---

Prof. Dr. David Bianchini  
Orientador da Dissertação e Presidente da Comissão Examinadora  
Pontifícia Universidade Católica de Campinas



---

Prof. Dr. Juan Manuel Adán Coello  
Pontifícia Universidade Católica de Campinas



---

Prof. Dr. Rangel Arthur  
Universidade Estadual de Campinas

Dedico esse trabalho a querida Miriellen, minha esposa e companheira de todas as horas, e reafirmo que, em todos os dias do nosso casamento, eu a escolhi, novamente, como minha esposa.

Dedico também a minha mãe Izildinha (in memoriam) pelos ensinamentos desde pequeno, com você aprendi a ser gente que respeita gente. Aprendi a ter fé, que Deus a proteja sempre e que eu possa sempre sentir e ter esse amor em todos os momentos de minha vida.

# AGRADECIMENTOS

Ao Prof. Dr. David Bianchini pelo apoio, amizade, paciência e orientação, sua sabedoria, direcionamento e conselhos foram fundamentais para o desenvolvimento desse trabalho.

Ao Prof. Dr. Eric Alberto de Mello Fagotto por ter me explicado todo o programa do mestrado três meses antes do processo seletivo que fez com que ficasse mais engajado em dar início a este grande desafio em minha vida.

Aos Profs Dr. David Bianchini, Dr. Eric Alberto de Mello Fagotto, Dr. Marcius Fabius Henriques de Carvalho, Dr. Omar Carvalho Branquinho e Dr. Marcelo Luís Francisco Abbade pelos ensinamentos nas disciplinas do programa do mestrado que agregaram em grande conhecimento na área de engenharia elétrica e gestão.

Ao Prof. Ms. Ranieri Marinho de Souza que no momento em que eu procurava um programa de mestrado me apresentou este programa de altíssima qualidade.

Aos meus colegas de mestrado pelo companheirismo e troca de experiências nas salas de aulas e nos encontros para a entrega de seminários e trabalhos durante o curso

Ao Dr. Henrique da Mota, médico especialista em coluna vertebral, responsável pela base de dados utilizada nesse trabalho por me auxiliar referente às dúvidas sobre as questões médicas.

Ao Dr. Henrique Haidar Jorge, médico ortopedista, professor da PUC Campinas pelos esclarecimentos da área ortopédica assunto desse trabalho.

À Pontifícia Universidade Católica de Campinas pela bolsa concedida durante o período do curso de Mestrado Profissional em Gestão de Redes de Telecomunicações.

Ao Santander por me dar condições financeiras e tempo nas saídas antes do expediente para chegar a tempo hábil nas aulas.

À minha avó Trindade que após o falecimento da minha mãe, não tem me deixado sem apoio e carinho.

Agradeço pela compreensão e incentivo de meus entes queridos, simplesmente maravilhosos.

À minha esposa, pelo carinho e compreensão por todo o tempo que fiquei dedicado ao mestrado deixando os programas de casal como segundo plano.

“Se enxerguei mais longe, é porque estive  
sobre ombros de gigantes”

Isaac Newton

# RESUMO

SILVA, Marcelo Cicero Ribeiro da. Aprendizagem de máquina em apoio a diagnóstico em ortopedia. 2016. 132f. Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-Graduação em Gestão de Redes de Telecomunicações, Pontifícia Universidade Católica de Campinas, Campinas, 2016.

Um dos grandes responsáveis pela mudança de um panorama competitivo é o progresso constante da tecnologia da informação e comunicação (TIC). A maior parte das dificuldades na tomada de decisão é a transformação de dados e informações em conhecimento, principalmente quando as bases de dados dizem respeito à saúde. Com a evolução da tecnologia e do 'aprendizado de máquina' (*machine learning*), já se dispõe de computadores capazes de realizar aprendizado de forma sofisticada, permitindo sua utilização no auxílio nas prescrições de diagnóstico médico, gerando uma segunda opinião para o profissional da medicina e contribuindo, assim, para uma melhor prestação de serviço à comunidade. O objetivo da pesquisa relatada consiste em elaborar um modelo computacional, apoiado em mineração de dados com uso de técnicas de aprendizado de máquina, que, utilizando-se de dispositivos de comunicação integrados às tecnologias de comunicação e informação e que venha oferecer suporte eficiente para o diagnóstico médico na área de ortopedia. A prova do conceito desta proposta utilizará de uma base de dados pública na especialidade da ortopedia (coluna vertebral) e o objetivo específico será o de auxiliar o médico na descoberta das doenças Listese e Hérnia de Disco. Esta aplicação trabalhou com o conceito de descoberta de conhecimento em bases de dados (*Knowledge Discovery in Databases*), para conseguir o resultado desejado. Esse processo a Mineração de Dados que, por meio de algoritmos de classificação, poderá transformar dados em informações úteis ao apoio do profissional médico na elaboração do seu diagnóstico. A pesquisa irá explorar e definir, na ferramenta de *Data Mining WEKA*, o algoritmo mais apropriado, dentre os vários já existentes, que possa oferecer maior acurácia no diagnóstico e que viabilize uma solução tipo *mobile*. A dinâmica estruturada neste trabalho deverá permitir que o sistema seja enriquecido a cada novo paciente tratado e que, com isto, a plataforma se torne mais eficiente e eficaz à medida que se amplie. Espera-se que o modelo computacional elaborado possa se configurar como uma segunda opinião em apoio ao diagnóstico do profissional médico retornando o diagnóstico do paciente. Os resultados obtidos foram satisfatórios obtendo um índice de acuracidade média acima de 86%. Dentre os benefícios acredita-se que será possível auxiliar na formação de novos profissionais auxiliando-os na Residência Médica, na redução de problemas decorrentes de erros médicos e, dessa forma, aumenta-se a eficácia no atendimento com ganhos de tempo e dinheiro.

Palavras Chave: Mineração de Dados, Aprendizado de Máquina, *Weka*, KDD.

# ABSTRACT

SILVA, Marcelo Cicero Ribeiro da. Machine learning in support of medical diagnosis Orthopedics 2016. 132f. Dissertation (Master in Networks and Telecommunications Management) - University of Campinas, 2016.

One of the major responsible to change in a competitive landscape is the steady progress of technology and communication (TIC). With the evolution of technology and 'machine learning', computers are already available to carry out learning in a sophisticated way, improving the prescriptions of medical diagnosis, generating a second opinion for the medical professional and thus, To provide a better service to the community. The objective of this research is to develop a computational model, supported by data mining using machine learning techniques and, using communication devices integrated with communication and information technologies, to provide efficient support for The medical diagnosis in the area of orthopedics. The proof of the concept of this proposal will be used based on a public database in the branch of backbone and the specific objective will be assist the doctor in the discovery of the Diseases Olisthesis and Herniated disk. This application will work with the concept of Knowledge Discovery in Databases to achieve the desired result. This process will use the Data Mining that, through classification algorithms, can transform data into useful information to the support the medical professional in the elaboration of diagnosis. The research will explore and define, in the WEKA Data Mining tool, the most appropriate algorithm, among the several that already exist, that can offer the highest diagnostic accuracy and enable a mobile solution. The dynamics structured in this work should allow that system to be enriched for each new patient treated and, with this, the platform becomes more efficient and effective as it expands. It is expected that the consulted computational model can be configured as a second opinion in support of the diagnosis of the medical professional. The results were satisfactory obtaining an average accuracy index above 86%. Among the benefits it is believed that it will be possible to assist in the graduation of new professionals assisting them in the Medical Residency, and reducing problems in possible medical errors thus, increasing the efficiency during the attendance and saving time and money.

*Keywords: Weka, KDD, Data Mining, Machine Learning.*

# LISTA DE FIGURAS

Figura 1. Mapeamento em 2006, dos locais com potenciais para terem núcleo.....	23
Figura 2. Cenário de 2015, a cor verde indica que o local esta homologado e em uso na RNP .....	24
Figura 3. As especialidades mais demandadas em recursos na corte.....	28
Figura 4. Principais fatores que interferem na ocorrência de erros.....	39
Figura 5. Etapas do processo de KDD.....	43
Figura 6. Fluxo da tarefa de classificação.....	47
Figura 7. Visão geral de um modelo de classificação.....	48
Figura 8. Modelo de classificação aplicado na base pública da coluna vertebral.....	48
Figura 9. Exemplo de árvore de decisão.....	50
Figura 10. Modelo de uma estrutura de uma árvore de decisão.....	50
Figura 11. Regras geradas do algoritmo J48.....	51
Figura 12. Modelo de Árvore gerada do algoritmo J48.....	52
Figura 13. Matriz de Confusão.....	55
Figura 14. Área sob a Curva ROC.....	57
Figura 15. Tela Inicial do weka.....	59
Figura 16. Aba de pré-processamento, exibindo as variáveis (insumos).....	60
Figura 17. Aba que permite selecionar o tipo de algoritmo e forma de teste.....	61
Figura 18. Exemplo de uma árvore de decisão sendo processada.....	62
Figura 19. Tela do modo knowledge flow do weka.....	63
Figura 20. Parâmetros da pelve.....	67
Figura 21. Fluxo de funcionamento da solução móbile.....	70
Figura 22. Apresentação das variáveis da base pública.....	71
Figura 23. Parte do arquivo da base de dados publica.....	73
Figura 24. Tela das regras de árvore de decisão aplicado no algoritmo LMT.....	83
Figura 25. Matriz de confusão do resultado do algoritmo LMT.....	84
Figura 26. Tela principal do aplicativo.....	86
Figura 27. Menu principal do aplicativo.....	87
Figura 28. Dados sobre a coluna vertebral do paciente.....	88
Figura 29. Dados sobre a coluna vertebral do paciente preenchida.....	89
Figura 30. Tela de processando diagnóstico.....	90
Figura 31. Resultado do diagnóstico.....	91
Figura 32. Tela de consultas salvas .....	92
Figura 33. Tela de consultas salvas exibindo a performance.....	93

# LISTA DE TABELAS

Tabela 1. Ranking das condenações no STJ .....	40
Tabela 2. São mostrados exemplos de casos de diagnóstico errado ou tardio .....	40
Tabela 3. Exemplo de matriz de Confusão.....	57
Tabela 4. Trecho da base de dados pública utilizada neste trabalho.....	68
Tabela 5. Resultados dos testes feitos pelo método Holdout .....	80
Tabela 6. Resultados dos testes feitos pelo método validação cruzada .....	82

## LISTA DE GRÁFICOS

Gráfico 1. Expectativa Crescimento da População no Mundo.....	21
Gráfico 2. Expectativa Crescimento da População no Brasil.....	21
Gráfico 3. Quantidade de julgamentos de recurso no STJ.....	27
Gráfico 4. Proporção de pessoas de 18 anos ou mais de idade com problema crônico de coluna, com indicação do intervalo de confiança de 95% segundo as grandes regiões.....	29
Gráfico 5. Índice de Precisão do teste 1.....	81
Gráfico 6. Índice de Precisão do teste 2.....	82

## LISTA DE ABREVIATURAS E SIGLAS

ARFF	=	<i>Attribute-Relation File Format</i>
ABRAHUE	=	Associação Brasileira de Hospitais Universitários
API	=	<i>Application Programming Interface</i>
CEO	=	<i>Chief Executive Officer</i>
CIPNSP	=	Comitê de Implementação do Programa Nacional de Segurança do Paciente.
CREMESP	=	Conselho Regional de Medicina do Estado de São Paulo
CRM	=	<i>Customer relationship Management</i>
FIOCRUZ	=	Fundação Oswaldo Cruz
FINEP	=	Financiadora de Estudos e Projetos
IHC	=	Interação Humano-Computador
KDD	=	<i>Knowledge Discovery in Databases</i>
NAS	=	<i>The National Academies Of Sciences, Engineering and Medicine</i>
ONU	=	Organização das Nações Unidas
PNS	=	Pesquisa Nacional de Saúde
P&D	=	Pesquisa e Desenvolvimento
RNP	=	Rede Nacional de Ensino e Pesquisa
ROC	=	<i>Receiver Operating Characteristic</i>
RUTE	=	Rede Universitária de Telemedicina
STJ	=	Superior Tribunal de Justiça
TCE	=	Taxa Concordância Esperada
TCO	=	Taxa Concordância Observada
TI	=	Tecnologia da Informação
TIC	=	Tecnologia da Informação e Comunicação
WEKA	=	<i>Waikato Environment for Knowledge Analysis</i>

# SUMÁRIO

1	INTRODUÇÃO .....	16
1.1	Objetivos .....	18
1.2	Objetivos específicos .....	19
1.3	Objetivos secundários .....	19
1.4	Organização do Texto .....	20
2	O CENÁRIO ATUAL .....	21
2.1	Problema crônico de coluna .....	29
2.2	Estado da Arte .....	30
3	CIDADES INTELIGENTES PARA A SAÚDE .....	35
3.1	A definição de cidades inteligentes .....	35
3.2	As TIC e as Cidades Inteligentes .....	35
4	O PROBLEMA DO ERRO MÉDICO .....	38
5	O KDD – Descoberta de conhecimento em bases de dados .....	43
5.1	Nomenclatura básica .....	44
5.2	Mineração de Dados .....	45
5.3	Classificação .....	46
5.4	Árvore de Decisão .....	49
5.5	Redes <i>Bayesianas</i> .....	52
5.5.1	Classificação <i>Naive Bayes</i> .....	53
5.6	Método de Avaliação dos classificadores .....	53
5.6.1	Precisão .....	54
5.6.2	Área sob a Curva ROC .....	55
5.6.3	Estatística Kappa .....	57
5.6.4	Medida F .....	58
6	<i>WEKA</i> .....	59
7	ORTOPEDIA .....	65
7.1	Listese .....	65
7.2	Hérnia de disco .....	66
8	PERCURSO METODOLÓGICO .....	69
8.1	Definição do algoritmo mais adequado. ....	69
8.2	Método de comparação entre os algoritmos de Classificação .....	73
8.3	Elaboração da Interface amigável .....	75
8.3.1	Relevâncias do Design .....	76

8.3.2	Objetivos da Avaliação.....	77
8.3.3	Inspeção de Usabilidade.....	78
9	SOFTWARES UTILIZADOS NESTE TRABALHO.....	79
9.1	MYSQL.....	79
9.2	Xamarin.....	79
10	RESULTADOS.....	80
10.1	Resultados e análise dos testes realizados.....	80
10.2	Resultados e análise da Interface.....	85
11	CONCLUSÃO.....	94
11.1	Contribuições.....	96
11.2	Trabalhos Futuros.....	97
12	REFERÊNCIAS.....	98
	APENDICE A: Script para o Aplicativo Dr. 2º Opinião.....	105

## 1 INTRODUÇÃO

As ferramentas de Tecnologia da Informação e Comunicação (TIC), segundo Silva (2004), disponíveis nas organizações, conseguem oferecer desde planejamentos orientados nos processos de compra, produção e venda, até recursos para que o tomador de decisão tenha subsídios informacionais e técnicos importantes no momento da ação. Uma estratégia organizacional deve se concentrar no desenvolvimento de um conjunto de mecanismos que, juntos, explorem os benefícios e minimizem os pontos fracos da estrutura de organização seja ela uma empresa ou uma instituição governamental.

É neste contexto que o da Tecnologia da Informação (TI) a partir do avanço rápido de novas técnicas de armazenamento e coleta de grandes volumes de dados permitiu com que as organizações pudessem armazenar uma expressiva quantidade de dados. É dentro desse quadro, que em especial na área da saúde, se torna significativo obter informações úteis de um banco de dados com grande volume de dados ali depositados, agregando-se valor à medicina aplicando uma “inteligência” para esta base de dados desconhecida. Cabe salientar a fala de John Naisbitt Perizzolo (2005), que conseguiu resumir bem esse estado na seguinte frase: “*estamos afogados em informação, mas sedentos de conhecimento*”. Essas bases de dados muitas vezes não são utilizadas porque os softwares tradicionais de análise não possuem suporte para processar um grande volume de dados tornando-se necessário a criação de novos procedimentos para esta análise. Devido essa necessidade surgiu o campo de estudo determinado KDD (*Knowledge Discovery in Databases*) - Descoberta de Conhecimento em Bases de Dados. Essa área utiliza algoritmos e técnicas que tem por finalidade transformar dados em informações úteis.

Segundo Garcia (2013) ferramentas estatísticas, modelos matemáticos, consultas estruturadas são comumente utilizadas para auxiliar na obtenção de informações a partir das bases de dados. No entanto, estas ferramentas possuem limitações que podem comprometer a precisão da informação gerada. Neste contexto surgiu em 1980 o conceito da descoberta de conhecimento em bases de

dados que reúne técnicas de diferentes áreas como inteligência computacional, reconhecimento de padrões, aprendizado de máquina, métodos estatísticos e banco de dados a fim de identificar padrões nos dados.

Segundo Freitas (2000), o conhecimento a ser descoberto (processo KDD) deve satisfazer três propriedades:

1. Deve ser correto;
2. Deve ser compreensível por usuários humanos;
3. Deve ser interessante, útil, novo.

Dentro deste contexto o presente trabalho buscará mostrar, a partir de um problema médico real, a viabilidade da aplicação de técnicas de *Data Mining*, e irá testar dezesseis diferentes algoritmos de árvore de decisão e classificadores bayesianos e realizar a comparação em termos de precisão de cada um.

Conforme Cruz (2007) a evolução tecnológica facilitou a produção e o acesso à informação. Por isso, a informação atualizada se tornou fator de sobrevivência em qualquer área. Para o indivíduo, a informação contribui para maior conhecimento e ações mais eficazes. Um problema similar diagnosticado serve como parâmetro para outro, o que faz aumentar a confiabilidade e a rapidez dos procedimentos, por isso há necessidade da disseminação e compartilhamento das informações.

É relevante apresentar aqui que pesquisas da Fundação Oswaldo Cruz (FIOCRUZ) subordinada ao Ministério da Saúde, indicam que 10% dos pacientes são vítimas de erros em diversos procedimentos, ou seja, a cada dez pacientes que procuram um médico e cura para o seu problema, um sofre efeito adverso RUTE (2015). O ministro da Saúde em 2013, Alexandre Padilha informa que 66% dessas ocorrências podem ser evitadas RUTE (2015). Nesse mesmo ano foi implantado o CIPNSP (Comitê de Implementação do Programa Nacional de Segurança do Paciente) composto por representantes do governo, sociedade, entidades de classe e universidades que tem como objetivo promover e apoiar iniciativas referentes à segurança do paciente.

Segundo a NAS (*The National Academies Of Sciences, Engineering and Medicine*) a maioria das pessoas que procuram atendimento médico serão diagnosticadas de forma incorreta em algum momento da vida NATIONAL ACADEMIES (2015), é o que afirma em seu relatório publicado em 22 de setembro de 2015, pouco tem sido feito para reduzir erros nessa área, o que pode agravar problemas se o setor não se preocupar com melhorias de forma efetiva. Um dado interessante na pesquisa é que os médicos cometem em torno de doze milhões de falhas de diagnóstico por ano.

A relevância destas questões induz a uma importante reflexão quanto à urgência de se ter na área médica um modelo computacional que possa aprender com os diagnósticos já realizados com sucesso e possa auxiliar o médico com uma 2ª opinião.

De acordo com Silva (2004), a técnica de mineração de dados tem sido aplicada com sucesso em diversos setores, como exemplos a seguir. Na área de Vendas, onde entender o comportamento dos consumidores, localizar características dos mesmos a fim de prever quais clientes serão atingidos em uma campanha de CRM (*Customer relationship Management*); no campo das Finanças, em que compreender o comportamento de fraudes em diversas áreas como cartão de crédito, identificar clientes potenciais que em breve irão encerrar a conta em um banco; também nos Transportes, onde se deve traçar a distribuição de horários entre os caminhos alternativos a fim de não gerar sobrecarga em horários de pico; e por fim, dentre muitos outros campos, salienta-se a Medicina, onde é crucial aprender o comportamento dos pacientes para ter um modelo preditivo em novas consultas.

## **1.1 Objetivos**

O objetivo da pesquisa descrita nesta dissertação consiste em elaborar um modelo computacional, apoiado em mineração de dados com uso de técnica de aprendizado de máquinas (*machine learning*) que utilizando-se de dispositivos de comunicação integrados às tecnologias de comunicação e informação, possa oferecer suporte eficiente para o diagnóstico médico na área de ortopedia.

Com este objetivo espera-se propiciar ao profissional de saúde informação de qualidade de forma a ampliar seu conhecimento sobre diagnósticos e tratamentos em sua área de atuação com o apoio da pesquisa em bases de dados históricas e “alimentadas” com frequência a partir da informação de sinais e sintomas armazenados através do aplicativo proposto neste trabalho, permitindo ao profissional conceder diagnósticos rápidos e precisos.

Dentro deste quadro, o problema desta pesquisa se configura a partir da seguinte questão:

É possível dispor-se de um algoritmo apropriado na ferramenta *WEKA* (*Waikato Environment for Knowledge Analysis*) que exerça o aprendizado de máquina em mineração de dados, de forma que se possa proceder, de forma eficiente ao diagnóstico de Listese e Hérnia de Disco, e que se viabilize como uma solução de acesso móvel?

## **1.2 Objetivos específicos**

O objetivo deste trabalho pode ser dividido em três sub-objetivos:

1. Aplicar técnicas de mineração de dados para descoberta de conhecimento em banco de dados de coluna vertebral utilizando-se de uma base pública.
2. Avaliar os padrões descobertos por meio de medidas de qualidade em mineração de dados.
3. Desenvolver uma *interface* amigável de suporte ao diagnóstico médico, com base nos princípios da interação humano-computador (IHC), com o intuito de propiciar uma segunda opinião a um profissional médico.

## **1.3 Objetivos secundários**

1. Instigar estudos no campo da aprendizagem de máquina com o intuito de aprimorar o uso das tecnologias na área médica.

2. Apresentar novos campos de reflexão aos profissionais da área da saúde, propiciando a eles condições de acompanharem os avanços no campo da tecnologia da informação (TI).

3. Contribuir para que os erros médicos em diagnósticos mal elaborados ou diagnóstico tardios no tratamento do paciente possam ser mitigados.

4. Apresentar uma proposta que contribua com a formação de profissionais médicos, em fase de Residência Médica.

#### **1.4 Organização do Texto**

Este documento está organizado de forma a esclarecer e apresentar todas as fases da pesquisa e desenvolvimento da solução. A dissertação é composta pelos seguintes capítulos:

Capítulo 2: Descreve cenário atual referente ao problema de coluna bem como o estado da arte.

Capítulo 3: Dedicado a atuação da saúde no contexto de cidades inteligentes.

Capítulo 4: Descreve o problema do Erro Médico.

Capítulo 5: Explora o tema de descoberta de conhecimento em banco de dados, abordando a mineração de dados bem como seus métodos de mineração.

Capítulo 6: Descreve o funcionamento da ferramenta *WEKA*.

Capítulo 7: Dedicado ao entendimento da Ortopedia no contexto deste trabalho.

Capítulo 8: Descreve materiais e métodos utilizados para este trabalho.

Capítulo 9: Descreve os softwares utilizados neste trabalho.

Capítulo 10: Discute sobre os resultados obtidos da solução proposta.

Capítulo 11: Trata das considerações finais, apresentando a conclusão, contribuições e trabalhos futuros.

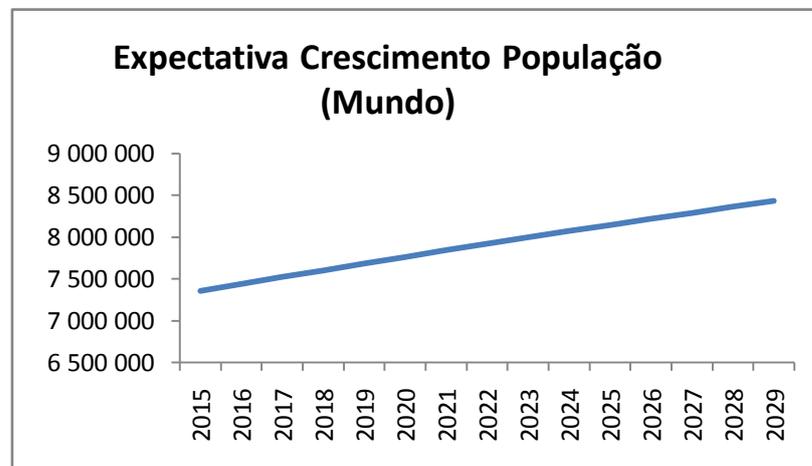
Referências Bibliográficas: Esta seção apresenta o levantamento bibliográfico, com todas as fontes citadas, pertinente ao desenvolvimento deste trabalho.

Apêndices.

## 2 O CENÁRIO ATUAL

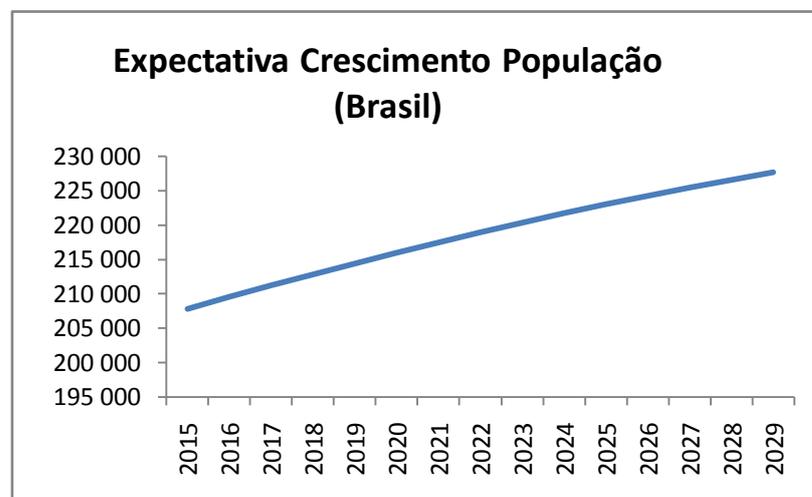
Projeções da Organização das Nações Unidas (ONU) indicam que a população mundial crescerá um bilhão de pessoas nos próximos 13 anos, ultrapassando o patamar de 8,5 bilhões de habitantes, e mais de 65% deles viverão em cidades. No Brasil, nesse mesmo intervalo sairá da faixa de 210 milhões para 228 milhões, um crescimento de aproximadamente 9% na população ONU (2015).

Gráfico 1: Expectativa Crescimento da População no Mundo



Fonte: ONU (2015)

Gráfico 2: Expectativa Crescimento da População no Brasil



Fonte: ONU (2015)

Segundo Weiss, Bernardes e Consoni (2014) as perspectivas globais sobre a dramática aglomeração urbana representam importantes desafios para a gestão pública. Tais desafios, decorrentes da pressão por efetividade no atendimento às demandas sociais, exigem novas abordagens para seu planejamento, projeto, financiamento, execução e operação.

Nesse cenário de crescimento urbano o papel das TIC's é fundamental para uma prestação de serviços públicos para proporcionar maior eficiência na gestão das cidades, com isso contribuindo para uma estruturação de Cidades Inteligentes. Em especial na área da saúde, com a evolução das TIC's, já existe o desenvolvimento de estruturas informacionais com o objetivo de prestação de serviço na área médica, que buscam integrar os órgãos públicos e privados, secretarias e hospitais a fim de levar o conhecimento para um número maior de profissionais da área da medicina, a exemplo desse propósito temos a RUTE (Rede Universitária de Telemedicina) que é uma iniciativa do Ministério da Ciência e Tecnologia, apoiada pela Financiadora de Estudos e Projetos (FINEP) e pela Associação Brasileira de Hospitais Universitários (ABRAHUE) e coordenada pela Rede Nacional de Ensino e Pesquisa (RNP), que visa apoiar o aprimoramento de projetos em telemedicina já existentes e incentivar o surgimento de futuros trabalhos. A iniciativa provê a infra-estrutura de serviços de comunicação, assim como parte dos equipamentos de informática e comunicação para os grupos de pesquisa, promovendo integração e conectividade e disseminando atividades de P&D (Pesquisa e Desenvolvimento) das instituições participantes. A utilização de serviços avançados de rede deverá promover o surgimento de novas aplicações e ferramentas que explorem mecanismos inovadores na educação em saúde, na colaboração à distância para pré-diagnóstico e na avaliação remota de dados de atendimento médico.

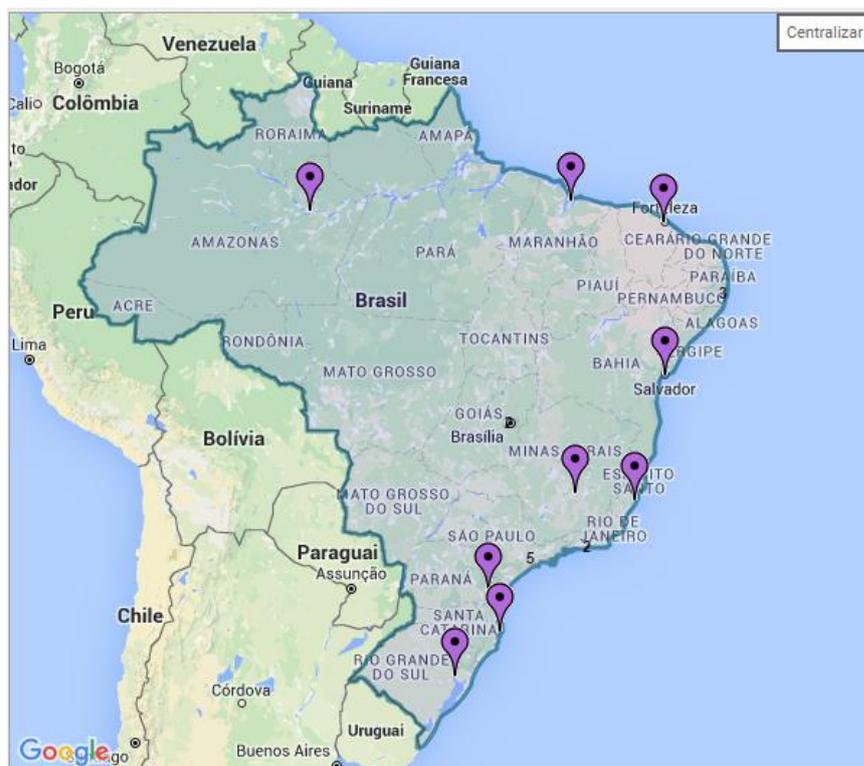
A RUTE possibilita, em um primeiro momento, a utilização de aplicativos que demandam mais recursos de rede e o compartilhamento dos dados dos serviços de telemedicina dos hospitais universitários e instituições de ensino. Em um segundo momento, a RUTE leva os serviços desenvolvidos nos hospitais universitários do país a profissionais que se encontram em cidades

distantes, por meio do compartilhamento de arquivos de prontuários, consultas, exames e segunda opinião.

A RNP disponibiliza serviços de TIC para atender à área da saúde. A RUTE possui 120 unidades em todo o Brasil e cerca de 60 grupos de interesse especial, compostos por profissionais da saúde, que conduzem projetos colaborativos em pesquisa, inovação, gestão, educação e assistência.

Para vermos a evolução dos núcleos (infra-estrutura de suporte às atividades de Telemedicina em hospitais e afins) ao longo do tempo, avalie-se a Figura 1 e a Figura 2.

Figura 1: Mapeamento em 2006, dos locais com potenciais para terem núcleo.



Fonte: RUTE (2015)

Figura 2: Cenário de 2015, a cor verde indica que o local está homologado e em uso na RNP



Fonte: RNP(2015)

Uma questão para reflexão, no contexto atual, se expressaria da seguinte forma: “Como as Tecnologias da Informação e Comunicação (TICs) têm evoluído no suporte à gestão pública e serviços urbanos a fim de atender às demandas de diversos pilares (educação, transporte, saúde e afins)”. Realizar-se pesquisas e criação de modelos computacionais que ajudem a minimizar essa necessidade e interesses do serviço público é uma tarefa árdua que, no entanto, não pode ser deixada em segundo plano.

Esse trabalho tem por fim auxiliar as demandas da área da saúde, fazendo com que se tenha um serviço prestado eficiente e que se possa utilizar os dados que temos em nosso banco de dados para transformar estes dados em informações úteis, ou seja, um serviço prestado a toda uma rede de comunicação existente. Por exemplo, entre departamentos médicos e hospitais sejam públicos ou privados, auxiliando o governo no enfrentamento dos desafios e ao mesmo tempo melhorando a qualidade de vida no setor da saúde.

Portanto a busca por ferramentas que possam extrair informações de grande volume de dados em bases de dados e que possam padronizar e gerar informações úteis afim dos profissionais da saúde utilizar como meio para tomada de decisão faz com que a mineração de dados seja um tema relevante de pesquisa. Mineração de Dados é um campo voltado para a ampliação de métodos de extração de padrões segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), uma das áreas que foram beneficiadas com a descoberta em banco de dados é a área de aprendizado de máquina (*machine learning*).

Conforme Brinks e Richards (2014) o campo de aprendizado de máquina é considerado um ramo da área de Inteligência Artificial, sendo um campo especializado no estudo e construção de sistemas que sejam capazes de aprender de forma automatizada a partir de dados, ou seja, o campo de aprendizado de máquina tem como objetivo desenvolver ferramentas que possam aprender a encontrar padrões de acordo com conhecimentos passados, o campo de aprendizado de máquina e mineração de dados possuem conceitos similares, o que faz que possam ser utilizadas em conjunto. De acordo com Zhou (2003) sem as técnicas de análise de dados avançadas doadas pelo campo de aprendizado de máquina, a mineração de dados seria como buscar uma agulha em um palheiro.

Uma das técnicas originárias da área de aprendizado de máquina é o algoritmo de Árvores de Decisão, segundo Witten e Frank (2000) é o algoritmo mais utilizado em desenvolvimento de mineração de dados devido ser um modelo onde sua representação de conhecimento é naturalmente compreendida.

Perante isso, compreende-se ser possível a contribuição com desenvolvimento de aplicações oriundas da Tecnologia da Informação (TI) na medicina com intuito de aprimorar a qualidade na prescrição do diagnóstico auxiliando os médicos em sua decisão final.

O estudo busca, em termos objetivos, aplicar as técnicas de aprendizado de máquina em conjunto com a mineração de dados para resolver problemas na área da saúde identificando padrões de sintomas e gerando informações úteis para o diagnóstico de Hérnia de Disco e Listese.

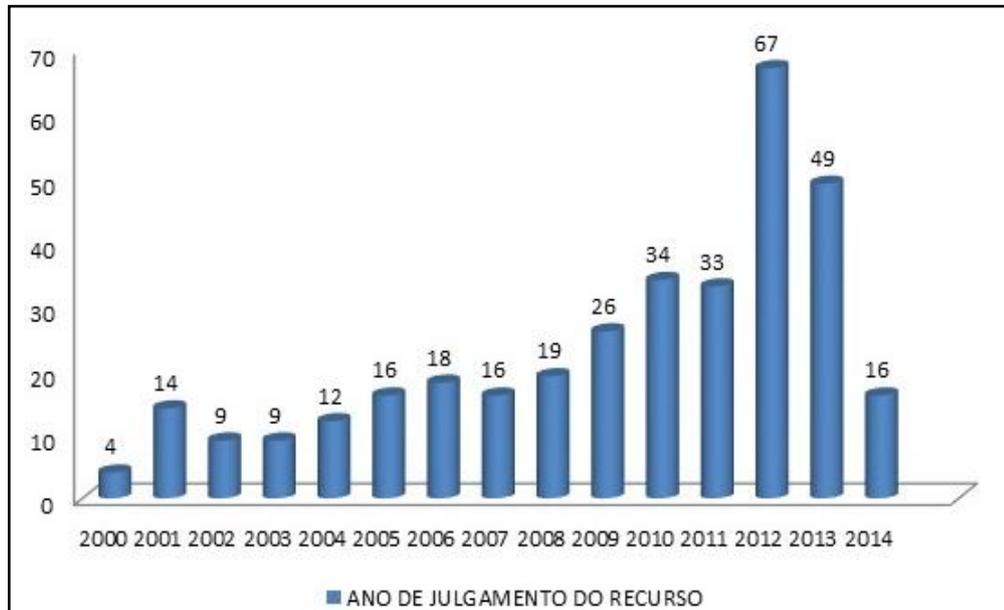
A Sociedade Brasileira de Direito Médico e Bioética – ANADEM é uma entidade sem fins lucrativos que promove o debate sobre problemas relacionados ao exercício profissional da medicina trazendo soluções não só no campo jurídico, mas em todas as áreas de interesse do médico associado, seus relatórios e estatísticas mostra-se que deve-se ter um olhar especial para o campo da medicina.

O Superior Tribunal de Justiça - STJ é uma corte especial, destinada a uniformizar o entendimento da Legislação Federal. Portanto, nem todas as demandas conseguem chegar ao Superior Tribunal de Justiça. Para que o recurso especial consiga subir ao exame dessa Corte Superior, se faz necessário que tenha havido o ferimento ou negativa de vigência de alguma Lei Federal ou ainda que o Tribunal Estadual tenha aplicado alguma Lei local, em prejuízo de outra Lei da República ou tenha interpretado alguma Lei Federal de forma diversa de outro Tribunal. Dos quase trinta milhões de processos inaugurados todos os anos perante o Poder Judiciário, pouco mais de trezentos mil chegam à Corte Superior, ou seja, insignificantes 1,07%. ANADEM (2015).

O número de recursos decorrentes de ações indenizatórias em virtude de supostos erros médicos, de 2000 a 2012, cresceu assustadores 1.600% junto ao Superior Tribunal de Justiça, desde a passagem do milênio, segundo informações estatísticas da própria Corte de 2011 para 2012, o crescimento foi de exatos 100% ANADEM (2015).

No gráfico 3 pode-se observar a evolução da quantidade de julgamentos do recurso no STJ.

Gráfico 3: Quantidade de julgamentos de recurso no STJ



\* 2014 – 1º trimestre

Fonte: ANADEM(2015)

O Estado do Rio de Janeiro é o campeão de recursos perante o Superior Tribunal de Justiça. O Estado detém 25,69% dos recursos. Em segundo lugar vem São Paulo, com 19,27%. Na terceira posição está o Rio Grande do Sul, com 15,92%, os três estados juntos somam mais de 60% dos recursos do Brasil.

Na Figura 3 pode-se observar as especialidades mais demandadas em recursos na corte.

Figura 3: As especialidades mais demandadas em recursos na corte.



Fonte: ANADEM(2015)

A Ginecologia/Obstetrícia ocupa o primeiro lugar, com 27,14%. A vice campeã é a Traumato-Ortopedia, com 15,71%. Em terceiro lugar, vem empatadas a Cirurgia Plástica e a Cirurgia Geral, com 10% cada. Depois vem a Neurocirurgia com 7,14%. A Pediatria, a Oftalmologia e a Anestesiologia, vem em seguida, empatadas com 5,71% cada. A hematologia, a Clínica Médica e a Otorrinolaringologia estão presentes em 2,85% das demandas. Já a Cardiologia, a Angiologia e a Medicina Intensiva aparecem, cada uma delas, em 1,43% dos processos. ANADEM (2015)

Portanto as duas especialidades com maior número de processos no STJ devido algum procedimento incorreto pelo médico são Ginecologia e **Ortopedia** que representam 42,85% do total, dentro deste quadro escolheu-se apenas destes dois temas relevantes, expressos na estatística em questão, e definiu-se a ortopediacomo a especialidade objeto de estudo neste trabalho.

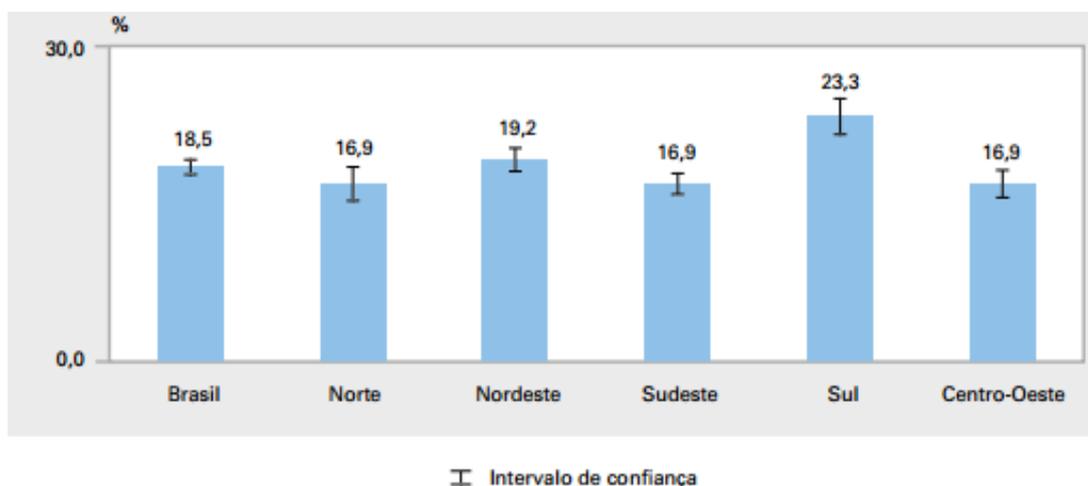
## 2.1 Problema crônico de coluna

As dores e os problemas musculoesqueléticos podem acometer grande parcela da população, tendo impacto econômico e na qualidade de vida dos indivíduos acometidos. Entre os problemas crônicos de coluna, os problemas lombares crônicos são os mais comuns. Pesquisa Nacional de Saúde (2013)

Segundo a Pesquisa Nacional de Saúde - PNS foram estimadas 27 milhões de pessoas de 18 anos ou mais de idade (18,5%) que referiram problema crônico de coluna no Brasil. O Sul do país mostrou proporção de casos diagnosticados de problema crônico de coluna superior à média nacional, 23,3%.

No Gráfico 4 pode-se observar a proporção de pessoas de 18 anos ou mais de idade com problema crônico de coluna separado por região do Brasil.

Gráfico 4: Proporção de pessoas de 18 anos ou mais de idade com problema crônico de coluna, com indicação do intervalo de confiança de 95% segundo as grandes regiões.



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional de Saúde 2013.

No Brasil, as mulheres tiveram maior proporção (21,1%) de diagnóstico médico de problemas crônicos de coluna do que os homens (15,5%). Conforme aumentava o grupo de idade, maior era a proporção estimada do indicador: 8,7% das pessoas de 18 a 29 anos, 19,9% das pessoas de 30 a 59 anos e, com patamares estatisticamente equivalentes, 26,6% das pessoas de 60 a 64 anos, e

de 28,9% das pessoas de 65 a 74 anos de idade e 28,5% para 75 anos ou mais de idade. A proporção foi significativamente maior entre as pessoas sem instrução e fundamental incompleto (24,6%) do que nas demais categorias de escolaridade. Por cor ou raça, este indicador apontou para maior percentual entre indivíduos brancos (19,3%) em relação a pretos (17,9%) e pardos (17,7%). (PNS, 2013)

## **2.2 Estado da Arte**

No campo da Tecnologia da informação, são muitos os trabalhos que se voltam para mitigar os problemas atuais no campo da medicina. Em especial, considerando a mineração de dados, que é o modelo computacional proposto neste trabalho, este visa diminuir o impacto na sociedade com diagnóstico errado ou tardio. O processo de KDD – *Knowledge Discovery in Databases* - “Descoberta em Banco de Dados” é uma técnica já utilizada em muitas áreas para fins comerciais.

Trabalhos já realizados deixam clara a importância do KDD também no campo da saúde, cabe ressaltar aqui algumas das contribuições que antecederam esta pesquisa, e que enriqueceram este campo de trabalho.

No campo da fisioterapia, o trabalho denominado “Mineração de dados aplicado na fisioterapia”, (CARVALHO, D. R. (et al)2012), observa-se que os autores adotaram o KDD para aplicação em um conjunto de dados de acompanhamento fisioterapêutico de pacientes, bem como sua utilidade na tomada de decisão. A base de dados utilizada foi disponibilizada por uma clínica de fisioterapia, nesse estudo foram extraídos três grandes grupos de tarefas de mineração: associação, classificação e agrupamento. O resultado permitiu descobrir padrões em diversos formatos, os quais evidenciaram as possíveis relações entre variáveis, com isso ampliou o entendimento das especificidades de pacientes atendidos pela clínica.

Nesse trabalho foi possível com a aplicação dos algoritmos de mineração de dados descobrirem vários comportamentos entre as variáveis, mostraram a importância de ter todas as variáveis preenchidas no momento do

cadastro do paciente afim de que possam ser aprendidas pelo modelo de mineração e que possa ampliar as análises pelos profissionais da clínica.

No campo da diabetes, o trabalho denominado “Uso de Técnicas de mineração de dados para a identificação automática de beneficiários propensos ao diabetes mellitus tipo 2”, Carvalho (2015), propuseram um modelo baseado em técnicas de mineração de dados a fim de identificar de forma automática beneficiários com propensão a doenças crônicas, para tal os autores aplicaram todas as fases do KDD como identificação das variáveis, seleção das variáveis a serem utilizadas, mineração de dados e validação das regras de descoberta por especialistas, para testar o modelo proposto foi realizado um teste para reconhecer indivíduos com propensão ao diabetes mellitus tipo 2. No processo de KDD foram selecionadas doze variáveis, uma base com 43.375 beneficiários (coletados por uma operadora de saúde), sendo descobertas 843 regras com taxa de acuracidade em 88,9%. O algoritmo utilizado foi o J48 da árvore de decisão da ferramenta WEKA, foram selecionadas 6 das 843 regras para serem avaliadas por quatro especialistas. A avaliação concluiu pela eficácia do modelo.

Os resultados foram satisfatórios, os autores testaram somente o algoritmo J48 onde é utilizado em muitos trabalhos semelhantes, não houve comparativos com demais algoritmos. Porém além de retornar um bom índice de precisão, foi possível os autores criarem um portal para a operadora de saúde onde funciona de forma automática selecionando os beneficiários para encaminhamento aos diversos programas de prevenção de doença e promoção da saúde. Ao final na base utilizada conseguiram identificar 5,7% de beneficiários, número esse que sem aplicação de um processo de KDD não seria possível essa identificação tão importante e ágil.

No campo da oncologia, o trabalho denominado “Mineração de dados para descoberta de conhecimento na área de oncologia”, Scheunemann e Preto (2015) expõe o conceito e técnicas de mineração de dados com aplicação do algoritmo de classificação denominado TAN com a descoberta não supervisionada, utilizou a ferramenta WEKA com o intuito de descobrir conhecimento útil da especialidade médica de oncologia na base de dados de uma Casa de Saúde do Vale do Taquari– RS. Para compor a base participaram

pacientes com diferentes diagnósticos na área da oncologia, em um período de 2011 a 2014 totalizando 3.267 pacientes. Com esse trabalho chegou-se a conclusão que os cânceres de esôfago, pele, mama e colo de útero são os tipos de câncer que os pacientes mais buscam por tratamento cirúrgico.

O autor alcançou o seu objetivo que era ter informação útil sobre uma base massiva de pacientes com doenças na área da oncologia, os gestores da Casa de Saúde ficaram satisfeitos com o resultado, pois a partir dessas informações puderam tomar decisões visando o bem estar do paciente.

No campo do traumatismo cranioencefálico, temos um trabalho denominado “Avaliação de Métodos de *Data Mining* e Regressão Logística aplicados na análise de traumatismo cranioencefálico grave”, Garcia (2015), a proposta de Garcia nesse trabalho foi à comparação de técnicas de mineração de dados com um modelo de regressão logística aplicado em base de dados com informações sobre traumatismo cranioencefálico para identificar uma melhor técnica para predição do desfecho óbito. A conclusão foi que o algoritmo *Naive Bayes* dos classificadores *Bayesianos* se destacou com relação ao método de regressão logística e outros métodos de *data mining* com acuracidade de 80,2% na classificação do óbito, foram considerados uma base com 748 pacientes.

A autora foi pioneira no tema no Brasil, conforme cita em seu trabalho as técnicas de *data mining* são aplicadas em algumas especialidades médicas, porém voltado para traumatismo cranioencefálico ao que tudo indica é a primeira pesquisa, com o seu trabalho foi possível realizar a predição do óbito.

O trabalho denominado “Mineração de Regras de Associação aplicada a dados da secretaria municipal de saúde de Londrina-PR”, Silva (2004), Glauco em seu trabalho aplicou técnicas de descoberta de conhecimento vinculando dados sobre a situação socioeconômica do paciente com os procedimentos realizados nas internações hospitalares a que foi submetido, construiu um protótipo para extração de regras de associação, para tal utilizou-se uma base de dados da secretaria municipal de saúde de Londrina-PR. Foram utilizadas nessa base 56.539 pacientes, após o processo de pré-seleção do *KDD* chegou-se ao número de 55.734 casos. O algoritmo utilizado foi o Apriori, que permitiu

identificar associações que fugissem do padrão esperado pelo especialista. Alguns dos resultados satisfatórios foram que 88,85% das safenectomias interna radical são realizadas em pessoas do sexo feminino com mais de 35 anos. Aprincípio, com esse resultado, foi considerado que era por questões de estética, porém ao ter um detalhamento e aprofundamento pelo especialista chegou-se à conclusão que era devido à atividade física (ocasionada pelo trabalho no lar). Outro fator interessante descoberto é que 78,78% das cirurgias múltiplas são realizadas em pessoas do sexo masculino, porém a base de dados é de sua grande maioria do sexo feminino. Portanto esse resultado foi encaminhado para o setor de avaliação e controle para melhor estudo, foi confirmado com o estudo que o caráter de internação que consome a maioria dos leitos é de urgência ou emergência.

O autor mostrou que é possível através de técnicas de mineração de dados realizar-se descobertas em banco de dados, transformando dados em conhecimento, afim de que possa despertar no âmbito médico curiosidades que sejam investigadas e/ou modelos de classificações de padrões que permitam uma classificação de novos casos de forma precisa e rápida.

O trabalho denominado "*Applications of data mining techniques in healthcare and prediction of heart attacks*", Srinivas (2010), os autores afirmam que há uma falta de ferramentas de análise eficazes para descobrir relacionamentos e tendências desconhecidas em dados e que conhecimento valioso pode ser descoberto da aplicação de técnicas de mineração de dados na área da saúde. No seu trabalho de pesquisa examinaram brevemente o uso potencial de técnicas de mineração de dados de classificação com base em árvore de decisão, Naïve Bayes e Redes Neurais Artificiais para enorme volume de dados de saúde. Usando perfis médicos, tais como idade, sexo, sangue, pressão e de açúcar no sangue puderam prever a probabilidade de pacientes terem uma doença cardíaca. Em suas comparações demonstrou ter o melhor resultado retornando descobertas que podem prever uma doença cardíaca.

O tema de pesquisa dos autores é relevante, segundo a literatura escrita em seu trabalho a doença cardíaca mata uma pessoa a cada 34 segundos nos EUA. Sendo assim a preocupação em tentar salvar vidas e antecipar

diagnósticos de forma precisa e ágil é um tema que temos pesquisadores de forma global aplicando o processo de KDD na descoberta do conhecimento.

No campo da astronomia, temos um trabalho denominado “Árvores de decisão em classificação de dados astronômicos”, Ruiz (2009), é um estado da arte para observarmos a mineração de dados, porém com um olhar para uma base de dados robusta. Nesse trabalho os autores tinham o desafio científico de separar estrelas de galáxias, utilizarem-se da técnica de árvore de decisão, a amostra final foi composta por 43.289 estrelas e 542.400 galáxias. Foram selecionados os parâmetros fotométricos considerados relevantes na distinção entre estrelas e galáxias. O método foi um sucesso com índice de acuracidade acima de 99% utilizando a ferramenta *WEKA* que se mostrou eficaz.

### **3 CIDADES INTELIGENTES PARA A SAÚDE**

#### **3.1 A definição de cidades inteligentes**

Segundo Washburn e Sindhu (2009), a utilização de TIC faz com que a infraestrutura crítica e os serviços de uma cidade - que incluem a administração da cidade, educação, saúde, segurança pública, imobiliário, transporte e serviços urbanos - sejam mais inteligentes, interconectados e eficientes.

De acordo com Arup (2010), a atividade da cidade é monitorada com sensores, que providenciam dados para um modelo organizacional que são em seguida utilizados por meios informáticos de maneira a poderem alterar a operação da infraestrutura da cidade ou informar os cidadãos.

Conforme Di Maio (2011) as cidades usarão o poder das redes de comunicação ubíquas, redes de sensores altamente distribuídas e sistemas de gestão inteligentes para resolver desafios presentes e futuros e criar novos serviços. A Cidade Inteligente pode ser vista como o resultado de um processo de transformação tecnológico intensivo que cobre a maior parte da sua informação e dos seus serviços.

Com estas definições pode-se concluir que uma cidade inteligente é aquela que busca soluções baseadas nas TICs para suprimir necessidades públicas através de vários setores que englobam uma cidade como educação, transporte, saúde, urbanismo, energia e afins.

#### **3.2 As TIC e as Cidades Inteligentes**

As Tecnologias de Informação e Comunicação (TIC) têm hoje papel preponderante na nossa sociedade e a sua influência ultrapassa largamente os limites a que inicialmente estiveram confinadas, com um impacto muito considerável na vida e no bem-estar das pessoas. Hoje se passa de uma era da Sociedade Industrial para a Sociedade da Informação, como já explorou Castells (2004).

É importante salientar que as TIC são hoje ubíquas e tão acessíveis como as infraestruturas de comunicações clássicas. Segundo Fabrega e Sajuria (2013) o poder computacional e de análise de dados presentemente disponível para uma grande quantidade da população têm impactos enormes na forma como a sociedade se organiza, criando oportunidades incríveis para micro e pequenas empresas através da Internet.

Nesse contexto é que surge no tema relevante das Cidades Inteligentes o campo de pesquisa da aplicação das Tecnologias da Informação e da Comunicação para melhorar a qualidade de vida das pessoas.

Dentre as muitas definições de Cidades Inteligentes (*Smarts Cities*), cabe ressaltar aqui os autores Nan e Pardo (2010) que as apresentam como aquelas cidades que trazem consigo o objetivo de melhoria na qualidade de serviços que são prestados aos cidadãos e que ao estabelecer sistemas integrados, apoiados em TIC's, não fazem destes um fim em si mesmo, mas etapas que irão permitir o fornecimento de serviços e informações compartilhadas para uma melhor qualidade de vida.

Esses autores salientam também a importância da dimensão humana, onde a criatividade se torna um elemento chave para o sucesso das Cidades Inteligentes, visto a gama de frentes de trabalho e inovação que se tornam possíveis. Dentre eles, e que se torna elemento de especial atenção nesse trabalho, os sistemas especializados voltados para a área Saúde com o desenvolvimento de soluções específicas para seus múltiplos setores. Como exemplo é possível ampliar a interação entre os diversos atores (governos, hospitais, pacientes) de forma efetiva; propiciar uso mais amplo de recursos como armazenamento em nuvem, de aplicações em banco de dados e mineração de dados, dentre outros.

Já salientou Goodspeed (2014) que cidades inteligentes constituem linhas de negócio de empresas como a IBM (*International Business Machines*), a Siemens e a CISCO que vêem as cidades como um mercado importante e onde existem oportunidades significativas de melhorias de eficiência através das TIC. Conforme citação de Goodspeed (2014), percebe-se a importância do tema e

o interesse de grandes empresas em diversos setores empresariais e na indústria das TIC.

Viver na cidade é a realidade de 84,36% dos brasileiros” (Instituto Brasileiro de Geografia e Estatística - IBGE, 2014). Essa população tem nas cidades diversas facilidades da vida urbana (como cultura, lazer, transporte público, saúde e oportunidades de emprego), mas, por vezes as cidades não estão preparadas para receber tantas pessoas. Esta realidade tende a piorar, visto que, de acordo com a Organização das Nações Unidas - ONU (2012), dois terços da população mundial viverá em áreas urbanas em 2030.

Com essas estatísticas tende-se a ter um volume populacional maior nos centros urbanos, o que se faz necessário um crescimento planejado em vários âmbitos de uma cidade inteligente como transporte, educação, mobilidade urbana e saúde, tema deste trabalho, sendo assim este desafio de planejar e atender essas áreas se faz presente nas grandes cidades e precisam ser mais inteligentes no sentido de retornar para a população um atendimento eficiente, este trabalho focará na questão da saúde a fim de obter através de um aplicativo *mobile* que possa auxiliar o especialista médico em seus pré-diagnósticos obtendo um atendimento eficiente uma vez que será uma 2ª opinião para o médico.

#### 4 O PROBLEMA DO ERRO MÉDICO

O erro médico é caracterizado pelo dano corporal causado ou provocado no paciente decorrente da ação ou omissão do médico no exercício profissional, sem a intenção de produzi-lo, ou seja, sem a vontade direcionada em cometer o dano, resultante, portanto, de negligência, imprudência ou imperícia. ANADEM (2015)

O Manual de Orientação Ética Disciplinar do Conselho Federal de Medicina define o erro médico como sendo “a falha do médico no exercício da profissão”. É o mau resultado ou resultado adverso decorrente da ação ou da omissão do médico, por inobservância de conduta técnica, estando o profissional no pleno exercício de suas faculdades mentais.

Ainda com todos os progressos da medicina o problema do erro médico continua a persiste na prática médica, podendo resultar em tragédias, prorrogação do tempo de internação e aumento dos custos hospitalares afirma Carvalho (2002).

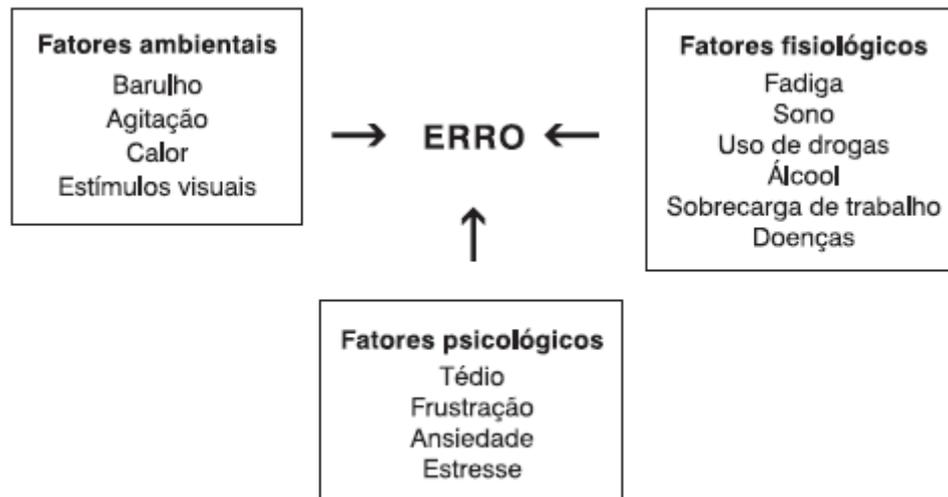
Grande recorrência das vezes a causa do erro médico é o desgaste psicológico (tédio, ansiedade, estresse), fatores ambientais (barulho, calor) e fisiológicos (fadiga, sono).

Segundo Carvalho (2002) uma das explicações para um grande número de erros observados na prática médica é justamente a ausência de mecanismos que diminuam a sua ocorrência, ou que interceptem o erro antes de chegar ao consumidor final (o paciente), ou seja, trabalha-se com a premissa de que o profissional de saúde não comete erros e, portanto, não se criam mecanismos de prevenção e correção.

O cenário nos hospitais principalmente os públicos são alarmantes, temos superlotação e a fila de atendimento é enorme, os médicos às vezes estão a muito tempo trabalhando chegando a passar de 24 horas, com todo esse *stress* faz necessário ter algum suporte ao médico para que este possa se orientar e realizar um diagnóstico com eficiência.

Na figura 4 observa-se os principais fatores que interferem na ocorrência de erros.

Figura 4: Principais fatores que interferem na ocorrência de erros



Fonte: Carvalho (2002)

A dificuldade em se aceitar o erro impede a ação de tomada de consciência em relação ao problema. Erros são ocorrentes, porém minimizá-los requer, primeiramente, a aceitação de sua existência. O impacto emocional é sempre grande (medo, culpa, raiva, vergonha e humilhação) sendo, na maioria das vezes, vivenciado de forma solitária afirma Carvalho (2002).

A falha médica é um grande problema de saúde pública, sendo necessária uma importância maior nesse tema, de um processo dinâmico de quantificar sua ocorrência para acompanhamento de sua evolução e a definição de ações que tratem a convivência médica com a sua ocorrência.

Na Tabela 1 é mostrado o Ranking (TOP 5) das condenações no Superior Tribunal de Justiça.

Tabela 1: Ranking das condenações no STJ

Fato Médico	Dano ao Paciente	Valor da condenação
Esquecimento de compressa cirúrgica de 45 cm no ventre da paciente durante cesariana.	– Septicemia– Perda auditiva– Colostomia– Histerectomia parcial	875.000,00 (danos morais) 400.000,00 (danos estéticos)
Cirurgia para correção de fratura de clavícula, resultando em Encefalopatia Pós-Anóxica ou Lesão Anóxica Cerebral.	– Paraplegia.– Estado vegetativo– Coma irreversível.	R\$ 830.000,00
Mulher que vinha fazendo tratamento para engravidar e que veio a suspendê-lo, por se constatar ser ela portadora de miomatose uterina – Histerectomia total, não precedida de exames específicos para verificação de possível gravidez – Sonogação, pelo cirurgião, do exame histopatológico, o qual revelou, além do mioma uterino, gravidez uterina com embrião vivo medindo 4,5 cm.	– Histerectomia- Ooforectomia– Salpingectomia– Aborto do embrião	750 salários mínimos
Retirada do colar de proteção cervical, sem proceder a exames de radiografia	– Tetraplegia irreversível.	R\$ 520.000,00
Aplicação de soro glicosado em paciente diabético.	Óbito.	R\$ 380.000,00

Fonte: ANADEM (2005)

Tabela 2 – São mostrados exemplos de casos de diagnóstico errado ou tardio

Fato Médico	Dano ao Paciente	Valor da
Diagnóstico tardio de câncer.	Óbito.	200 salários mínimos
Prescrição de oxigenoterapia para neonato em intensidade e período injustificados, falta de monitoramento adequado, periódico e eficiente, por médicos e enfermeiras, e, ainda, ausência de informação e recomendação de acompanhamento oftalmológico especializado.	Cegueira.	R\$ 106.400,00
Gravidez tubária diagnosticada equivocadamente como cisto ovariano.	Óbito.	R\$ 130.000,00
Recém-nascido que após apresentar quadro de icterícia, foi acometido pela doença de Kernicterus – Circunstâncias que levam a concluir pela alta prematura da criança e falta de orientação – Ausência de controle do quadro icterico – Tratamento que, embora correto, foi dispensado tardiamente.	Óbito.	R\$ 100.000,00
Liberção prematura da paciente, sem repetição de exames ou encaminhamento para atendimento especializado mesmo diante da persistência do quadro de dor torácica, e a despeito do tabagismo e do histórico familiar de problemas cardíacos como fatores de risco. Tratamento do quadro clínico como sendo de dor muscular, quando em verdade anunciava problema cardíaco, que evoluiu poucas horas depois para um infarto agudo do miocárdio, culminando com parada cardiorrespiratória.	Estado vegetativo.	R\$ 100.000,00
Comprometimento encefálico não detectado e paciente dispensado com diagnóstico de síndrome de abstinência alcoólica.	Óbito.	R\$ 100.000,00
Falso diagnóstico de síndrome de Down em ultrassonografia com Translucência Nucal.	Constrangimento	R\$ 6.000,00

Fonte: ANADEM (2015)

Levantamento do Sindicato Médico do Rio Grande do Sul aponta que um em cada cinco médicos possui um processo em andamento no estado, entre cíveis e criminais, relacionados ao alegado erro médico.

Em 02/04/2015 o CREMESP (Conselho Regional de Medicina do Estado de São Paulo) ministrou uma palestra na Santa Casa de Misericórdia de SP com o tema "A prevenção do erro médico" palestra proferida na Santa Casa de Misericórdia de São Paulo por Bráulio Luna Filho, presidente do CREMESP. (CREMESP, 2015).

Para Luna Filho (CREMESP, 2015), o erro médico é um fenômeno multifatorial e a principal causa de denúncias no CREMESP, com cerca de 15 demandas por dia.", o que gera aproximadamente 4.000 denúncias/ano. Segundo o presidente da Casa, "dos 3.800 processos atualmente em andamento, 60% estão relacionados aos erros médicos."

O exame do Conselho Regional de Medicina do Estado de São Paulo (CREMESP) de 2015, que registrou 48,13% de reprovação entre os participantes, revelou elevado nível de desconhecimento dos recém-formados em medicina em procedimentos considerados básicos. O exame foi aplicado entre 2.726 alunos egressos de 30 cursos de medicina do Estado e consistiu de 120 questões de múltipla escolha. CREMESP (2015)

Segundo Luiz Bin (conselheiro corregedor do CREMESP), o desempenho dos alunos nessas questões está diretamente ligado à qualidade do aprendizado nas faculdades, cada vez mais numerosas no Estado diante de um número insuficiente de professores capacitados para a didática.

"Não existe a formação do professor da universidade, porque o doutorando, o mestrando demora tempo. Para se formar um professor de faculdade de medicina vai-se aí em torno de cinco a dez anos e as faculdades estão sendo abertas quase que mensalmente", critica. CREMESP (2015)

"Os alunos aprendem com médicos da região onde é aberta a faculdade com um pouquinho mais de segurança, mas que não têm a capacidade didática de passar os conhecimentos deles. Isso vai ao encontro desse nível de

erros que a gente está vendo aqui, porque são alunos que foram orientados por professores que não são professores na verdade."

Segundo o artigo 196 da constituição federal de 1988 "A saúde é direito de todos e dever do Estado, garantido mediante políticas sociais e econômicas que visem à redução do risco de doença e de outros agravos e ao acesso universal e igualitário às ações e serviços para sua promoção, proteção e recuperação".

Pelo artigo temos a clareza da importância da saúde no cenário brasileiro, é dever do estado minimizar o risco de doenças para tal a TIC pode surgir como uma forma de criação de soluções que possam auxiliar os atores dos órgãos público em vários âmbitos, realizar-se pesquisas na área da saúde a fim de descobrir-se conhecimento em dados desconhecidos é sinal de avanço na medicina agregando qualidade e ao mesmo tempo buscando seguir a constituição federal.

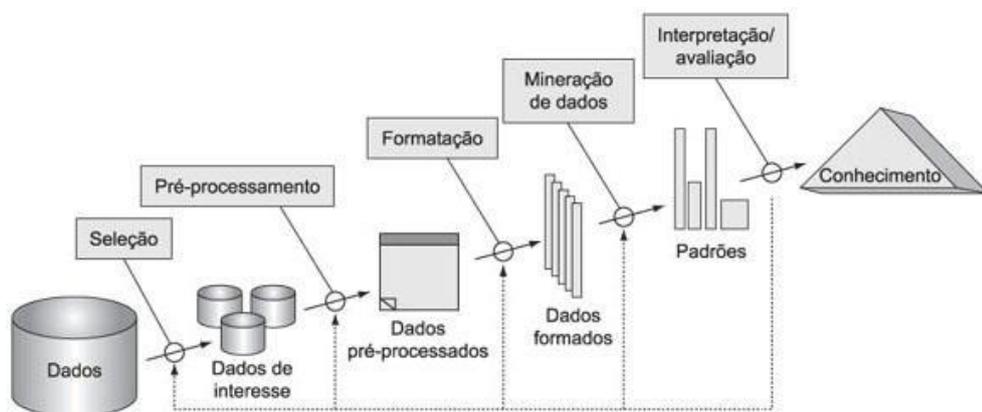
## 5 O KDD – Descoberta de conhecimento em bases de dados

A descoberta de conhecimento em bases de dados, do inglês *Knowledge Discovery In Databases* (KDD), segundo Fayyad, Piatetsky-Shapiro e Smyth (1996) é um processo automático de identificação de características e relacionamento nos dados, os quais são transformados em conhecimento útil e compreensível.

É importante frisar que o KDD é o processo de descoberta de conhecimento útil a partir de dados e a Mineração de dados é uma etapa fundamental neste processo. A mineração de dados é por sua vez a aplicação de algoritmos para extração de padrões de dados.

Na figura 5 pode-se observar todo o processo de KDD e suas fases.

Figura 5: Etapas do processo de KDD



Fonte: FAYYAD et al. (1996)

O processo KDD é um conjunto de atividades que compõem 5 etapas:

1. Seleção: Nesta etapa são decididos quais os conjuntos de dados que serão importantes para a tarefa de análise da base de dados.

2. Pré-Processamento: Nesta etapa é feita a limpeza de dados e ajuste nas informações ausentes, errôneas e inconsistentes nas bases de dados, a fim de ter uma qualidade dos dados.

3. Formatação ou Transformação: Nesta etapa é feita a transformação dos dados, serão analisados os dados e reorganizá-los para que sejam interpretados por um software de mineração de dados.

4. Mineração de Dados: Nesta etapa faz-se com que os dados sejam transformados em conhecimento através de algoritmos específicos.

5. Interpretação ou Avaliação: Nesta etapa são conhecidas as regras geradas pelo processo de Mineração de Dados para a descoberta de conhecimento, após esse aprendizado poderão surgir padrões, relacionamento e descoberta de novos fatos.

É importante ressaltar que para esse trabalho não será necessário seguir as etapas 1, 2 e 3 uma vez que a base disponibilizada já se encontra com 6 atributos selecionados, não possuem dados faltantes ou com inconsistências e já está em um formato para leitura de um software de Mineração de Dados.

Nesse trabalho o foco está na identificação de padrões válidos, por isso se realiza uma avaliação comparativa entre diferentes métodos de KDD, a fim de avaliar o resultado por meio de critérios de avaliação aos quais se constituem em medidas de qualidade.

## **5.1 Nomenclatura básica**

Inicialmente, adota-se a seguinte nomenclatura para as informações em uma base de dados conforme cita Lopes (2007).

➤ Atributos são as informações que se pode ter em cada uma das instâncias. Um atributo descreve uma característica ou um aspecto que as instâncias podem possuir. Imaginando uma base de dados como uma tabela, os atributos são habitualmente as colunas da tabela;

➤ Instâncias são conjuntos de valores, até um para cada atributo, que definem as ocorrências de uma base de dados. Note-se que uma instância pode

ter alguns atributos sem valores associados. Imaginando a base de dados como uma tabela, as instâncias são habitualmente as linhas da tabela;

➤ Valores são as informações que uma determinada instância pode ter para cada atributo. Os tipos destes valores podem ser valores numéricos ou categóricos. Imaginando uma base de dados como uma tabela, os valores são as informações contidas em cada uma das células da tabela;

➤ Classes são as possíveis categorias em que as instâncias podem ser agrupadas. Na verdade, a classe representa a informação que se quer extrair da base de dados, sendo, portanto, o objetivo da aprendizagem supervisionada determinar a classe correta para cada instância. Em algumas bases de dados, as classes podem ser definidas por um único atributo que é chamado de atributo preditor, pois o seu valor define, ou prediz, qual será a classe na qual cada instância será classificada

## 5.2 Mineração de Dados

A mineração de dados é a etapa principal do processo de KDD, sendo responsável pela busca no conjunto de dados, dos padrões que podem originar conhecimento útil segundo Dasu e Johnson (2003) e Hand, Manilla e Smyth (2001).

Segundo Carvalho (2005) e Tan, Steinbach e Kumar (2009) *Data Mining* faz uso de técnicas automáticas de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações que, devido ao volume de dados, não seriam facilmente descobertas a olho nu pelo ser humano.

Segundo Carvalho (2005), a mineração de dados pode ser realizada de três diferentes formas em função do nível de conhecimento que se tenha do problema estudado. Se há pouco conhecimento, faz-se a descoberta não supervisionada; se há suspeita de alguma relação interessante, faz-se a testagem em hipótese; se há muito conhecimento, faz-se a modelagem matemática da relação (descoberta supervisionada).

Na base pública que se usa neste trabalho temos a descoberta supervisionada, ou seja, temos pacientes já diagnosticados.

Na fase de mineração de dados, dentro do processo de KDD, necessita-se definir a técnica e o algoritmo a ser utilizado da tarefa proposta. Uma vez escolhido o algoritmo a ser utilizado, deve-se implementá-lo e adaptá-lo ao problema proposto. Para finalizar esse processo deve-se executar o algoritmo a fim de obter resultados que serão analisados na fase de interpretação e avaliação do resultado.

A mineração de dados difere das técnicas comum de estatísticas por que ao invés de verificar padrões utiliza os próprios dados para descobrir tais padrões. De acordo com Cabral (2015) aproximadamente 5% de todas as relações podem ser encontradas por esses métodos estatísticos, a mineração de dados pode descobrir outras relações anteriormente desconhecidas: os 95% restante.

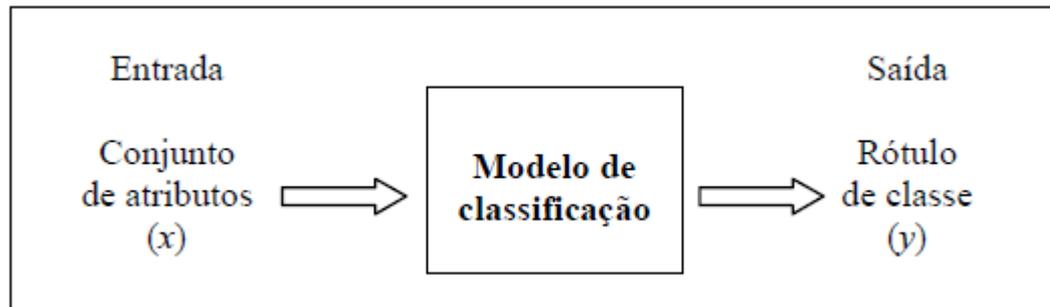
Neste capítulo serão abordados algoritmos de Classificação com foco em Árvores de Decisão e classificadores Bayesianos, os demais algoritmos não serão abordados por não serem objeto comum de pesquisa de acordo com os artigos estudados.

### **5.3 Classificação**

A classificação é uma tarefa que consiste na organização de objetos em uma categoria pré-definida, sendo a aprendizagem de uma função alvo  $f$ , também chamada de modelo de classificação, que mapeia cada conjunto de atributos  $x$  a um único rótulo  $y$ , denominada de classe ou objeto de saída conforme definição de Tan, Steinbach e Kumar (2009).

Na figura 7 observa-se de forma sucinta o fluxo de uma tarefa de Classificação.

Figura 6: Fluxo da Tarefa de Classificação



Fonte: Tan, Steinbach e Kumar (2009)

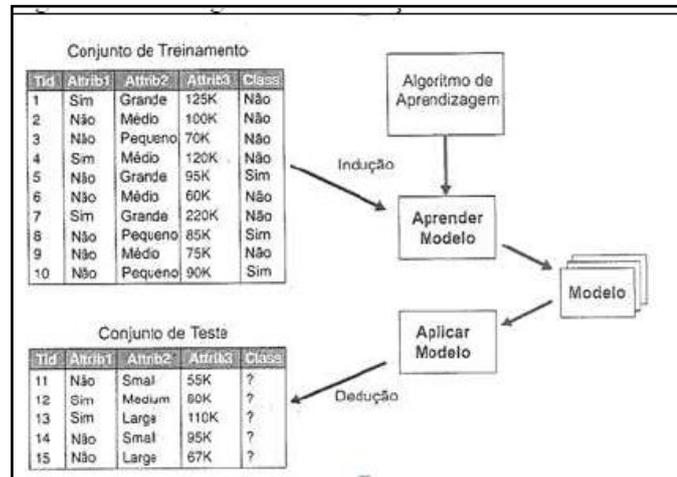
Essa tarefa, de acordo com Tang e Maclennan (2005) é uma das mais populares em *data mining*, sendo aplicada em uma diversidade de problemas. Um classificador encontra propriedades comuns entre um conjunto de registros pertencentes a uma base de dados e os classifica em diferentes classes conforme um modelo. A tarefa é de aprendizado supervisionado, pois os registros têm suas classes pré-definidas a partir do grupo inicialmente analisado. Tan et al. (2008).

A classificação constrói modelos, partindo do conjunto de dados de entrada, por meio de métodos que incluem classificadores de árvores de decisão, baseado em regras, redes neurais artificiais, máquinas de vetor de suporte e classificadores Bayesianos. Cada um desses métodos emprega um algoritmo de aprendizagem a fim de identificar um modelo mais adequado para o relacionamento entre o conjunto de atributos e o rótulo da classe. O modelo originado pelo algoritmo de aprendizagem deve ser bem adaptado aos dados de entrada e prever corretamente as classes de registros desconhecidos. Tan, Steinbach e Kumar (2009).

A base utilizada neste trabalho é de aprendizado supervisionado conforme definição de Tan et al. (2008) por possuir as classes pré-definidas e a técnica utilizada é de classificação que conforme definição de Tang e Maclennan (2005) e outros trabalhos semelhantes é a técnica mais popular em mineração de dados.

Na figura 7 observa-se uma visão geral de um modelo de classificação

Figura 7: Visão Geral de um modelo de classificação

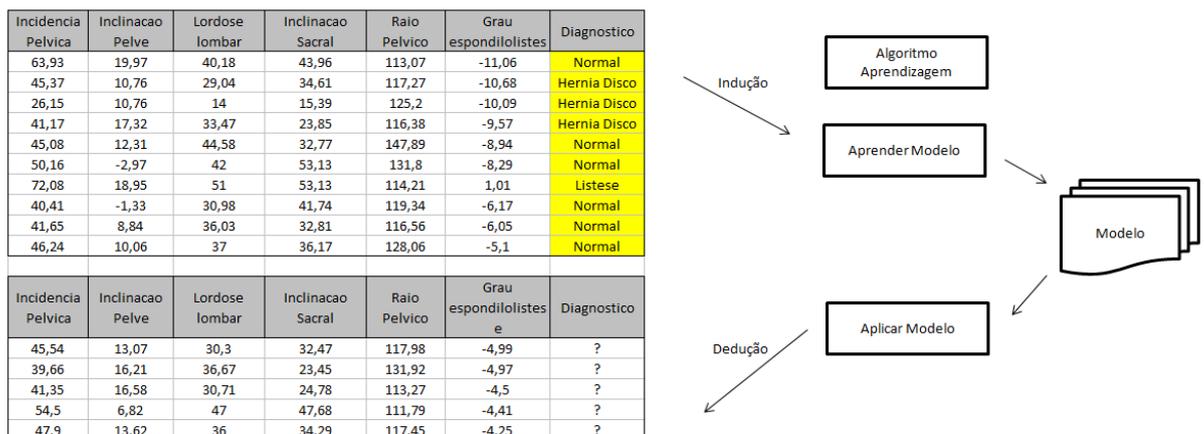


Fonte: Tan, Steinbach e Kumar (2009)

A avaliação do desempenho de um modelo de classificação é baseada na contagem dos registros de testes que foram classificados de forma correta e incorreta pelo modelo.

Na Figura 8, temos uma figura similar em termos de fluxo à figura 7 com dados da base pública da coluna vertebral.

Figura 8: Modelo de Classificação aplicado na base pública da coluna vertebral



## 5.4 Árvore de Decisão

A abordagem “dividir para conquistar” a partir de um conjunto de casos independentes leva a uma representação chamada árvore de decisão. Para classificar um exemplo desconhecido, este percorre a árvore de acordo com os valores dos atributos testados em nós sucessivos, e, quando uma folha é atingida, a instância é classificada de acordo com a classe atribuída à folha.

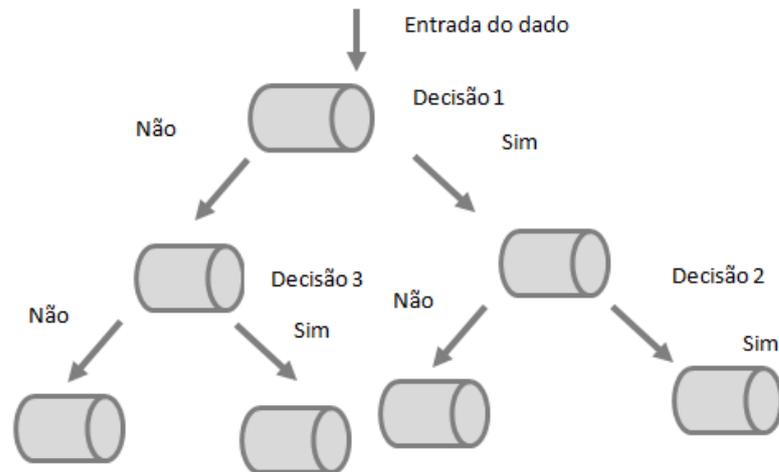
Segundo Santos (2009) define essa técnica da seguinte forma: as árvores de decisão, como o próprio nome indica, são constituídas por estruturas em árvores que representam um conjunto de decisões. Os algoritmos dessa técnica permitem gerar regras de classificação dos dados, baseadas nas informações armazenadas na base de dados.

A procura por esses algoritmos pelos trabalhos do estado da arte se dá pelo fato de ser uma técnica simples, não são necessárias configurações de parâmetros e geralmente tem um bom grau de assertividade, porém para que a previsão seja satisfatória é fundamental que os dados que serão trabalhados estejam com qualidade, para estes dados estarem com qualidade é necessário que as etapas do KDD sejam cumpridas, ou seja, os atributos sejam escolhidos por um especialista na área, que seja feita a limpeza de dados ocultos ou inconsistentes.

Para Garcia (2000), as árvores de decisão também consistem de: nodos (nós), que representam os atributos, e de arcos (ramos), provenientes desses nodos e que recebem os valores possíveis para esses atributos (cada ramo descendente corresponde a um possível valor desse atributo). Nas árvores existem nodos folha (folha da árvore), que representam as diferentes classes de um conjunto de treinamento, ou seja, cada folha está associada a uma classe. Cada percurso na árvore (da raiz à folha) corresponde a uma regra de classificação.

Na figura 9 observa-se um exemplo do fluxo de uma árvore de decisão.

Figura 9: Exemplo de Árvore de decisão

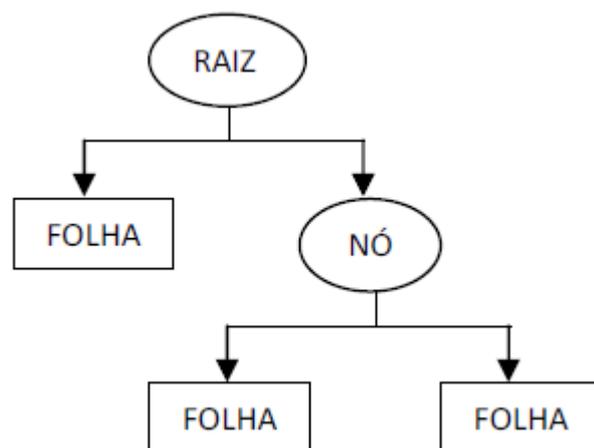


Fonte: adaptado de Carvalho(2005)

Na construção de uma árvore de decisão adota-se o particionamento recursivo de dados, dividindo-se o conjunto de treinamento em duas ou mais partições conforme o conjunto de valores de cada atributo, isso se repete até que todos ou pelo menos a maioria dos exemplos em cada uma das partições pertença a um rótulo de classe segundo Goldschmidt e Passos (2005).

Na figura 10 Observa-se um modelo de estrutura de uma árvore de decisão.

Figura 10: Modelo de uma estrutura de uma árvore de decisão



Podemos observar na figura 11 um exemplo de regra de uma árvore de decisão aplicado à base de dados pública da coluna vertebral, e nesse fluxo conseguimos de forma clara entender as diversas decisões que são tomadas ao longo do algoritmo, a fim de encontrar a folha onde teremos a classe do diagnóstico. Conforme a base se amplie obtendo novas variáveis (pacientes) esse fluxo será atualizado de acordo com o novo “aprendizado”, sendo assim a precisão no diagnóstico aumentará em função de novos aprendizados.

O fluxo da Figura 11 foi construído a partir do resultado do algoritmo J48 da árvore de decisão da ferramenta Weka, observa-se logo no início da árvore que o campo grau de listese é comparado se é menor do que 15.779697, caso seja maior o diagnóstico retornará Listese como a doença desse paciente, ao longo da árvore são aplicadas várias condições estas geradas pelo próprio algoritmo.

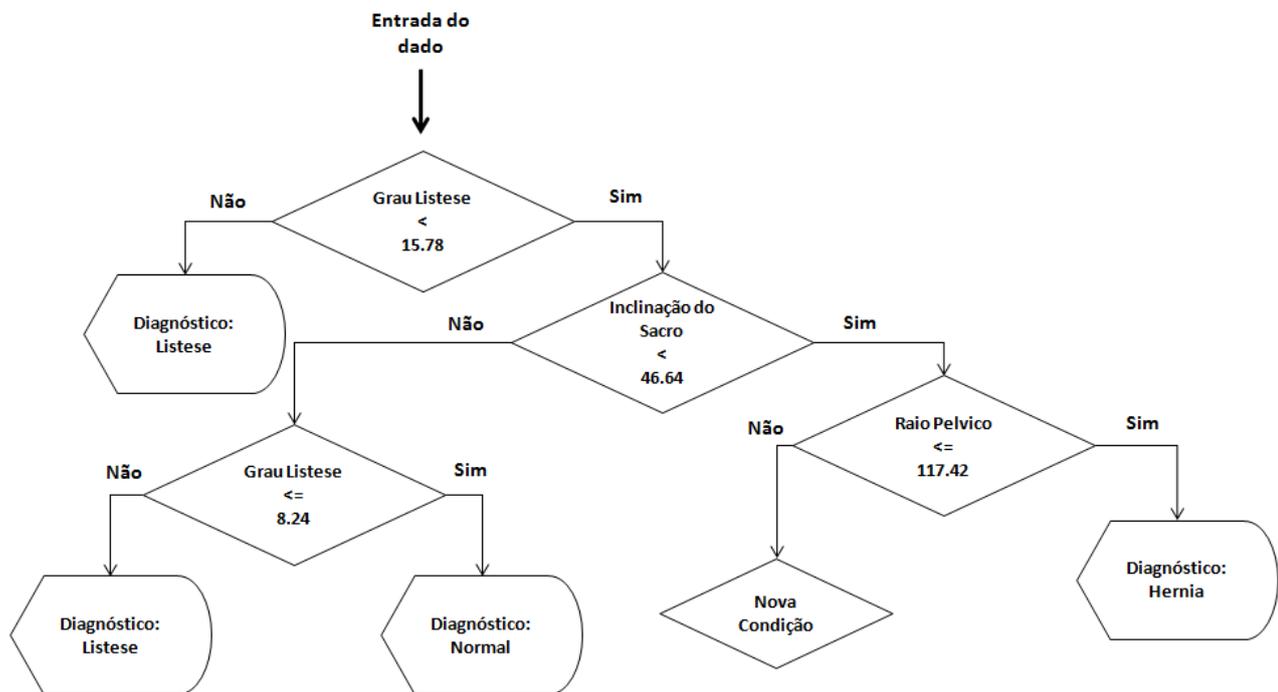
#### Classifier output

```

grau_listese <= 15.779697
|  inclinacao_sacral <= 46.636577
|  |  raio_pelvico <= 117.422259: Hernia (46.0/12.0)
|  |  raio_pelvico > 117.422259
|  |  |  inclinacao_sacral <= 28.131342
|  |  |  |  inclinacao_pelve <= 17.114312
|  |  |  |  |  inclinacao_pelve <= 14.930725
|  |  |  |  |  |  grau_listese <= 0.75702: Normal (4.0/1.0)
|  |  |  |  |  |  grau_listese > 0.75702: Hernia (5.0)
|  |  |  |  |  |  inclinacao_pelve > 14.930725: Normal (4.0)
|  |  |  |  |  |  inclinacao_pelve > 17.114312: Hernia (10.0)
|  |  |  |  inclinacao_sacral > 28.131342
|  |  |  |  inclinacao_pelve <= 12.306951: Normal (33.0)
|  |  |  |  inclinacao_pelve > 12.306951
|  |  |  |  |  grau_listese <= 5.074353: Normal (24.0/3.0)
|  |  |  |  |  grau_listese > 5.074353
|  |  |  |  |  |  grau_listese <= 8.235294: Hernia (6.0)
|  |  |  |  |  |  grau_listese > 8.235294: Normal (5.0/1.0)
|  |  |  inclinacao_sacral > 46.636577
|  |  grau_listese <= 8.235294: Normal (21.0/1.0)
|  |  grau_listese > 8.235294: Spondylolisthesis (4.0/1.0)
grau_listese > 15.779697: Spondylolisthesis (148.0/3.0)

```

Figura 12: Modelo de Árvore gerada do algoritmo J48



Na árvore da figura 12 fica nítido o direcionamento que o algoritmo aplica em tempo de execução, faz-se várias condições ao longo da árvore até que seja encontrado a folha onde será feito o diagnóstico, nesse caso fica claro o por que do nome Árvore de Decisão.

Os resultados gerados pelas árvores de decisão, do nó raiz até as folhas, podem ser visualizados por meio de regras de classificação. As condições de teste formam os antecedentes da regra, enquanto que o rótulo da classe na folha da árvore constitui o consequente.

O uso de árvore de decisão ocorre através de algoritmos que particionam, recursivamente o conjunto de treinamento em subconjuntos. Essa pesquisa fará um comparativo com 12 algoritmos do método de árvore de decisão e 4 algoritmos do método *Bayesiano* será validado por métodos de avaliação.

## 5.5 Redes *Bayesianas*

Os classificadores *Bayesiano* são classificadores estatísticos com base no teorema de Bayes, que preveem a probabilidade de uma tupla pertencer a

uma determinada classe. Os classificadores *Naive Bayes* assumem independência condicional de classe, o que significa que o efeito do valor de um atributo em uma determinada classe é independente dos valores dos outros atributos segundo Vieira (2011).

### 5.5.1 Classificação *Naive Bayes*

O classificador *Naive Bayes* produz uma estimativa da probabilidade. Para cada valor de classe, ele estima a probabilidade de uma determinada tupla pertencer a essa classe. Desenvolvido por Pearl (1995), as redes bayesianas, também conhecidas como Redes de *Bayes*, são uma alternativa estatística pertencente à família de modelos probabilísticos gráficos que representam um conjunto de variáveis aleatórias nos nodos e suas dependências condicionais nas arestas entre os nós, combinando os princípios da teoria gráfica, da teoria da probabilidade, das ciências da computação e da estatística. As Redes de *Bayes* especificam conjuntos de distribuições de probabilidades condicionais que permitem independências condicionais de classe a serem definidas entre grupo de variáveis. As Redes de *Bayes* também usam métodos estatísticos *Bayesiano*, oferecendo uma abordagem eficiente e de princípios conforme Vieira 2011.

## 5.6 Método de Avaliação dos classificadores

O grau de relevância das informações adquiridas no processo de *data mining* pode ser observado por meio das avaliações de desempenho do algoritmo classificador, geralmente representada pela taxa de erros resultante da classificação conforme Witten e Frank (2005).

A análise de desempenho empregada nos classificadores pode ser realizada por meio de uma matriz de confusão conforme é exibido na Tabela 2, resultante do processo de classificação pelos métodos de testes *holdout* ou *crossvalidation* de *k* partes.

No método *holdout* os dados originais são particionados em dois conjuntos, sendo um de treinamento e o outro de teste. O modelo de classificação

é induzido a partir do conjunto de treinamento e tem o seu desempenho avaliado no conjunto de teste. Neste método a proporção dos dados reservados para treinamento e para teste fica em 50% para cada conjunto ou 2/3 para treinamento e 1/3 para teste conforme Tan, Steinbach e Kumar (2009); Witten, Frank, Hall (2011).

No método *cross-validation*, validação cruzada de  $k$  partes, segmentam-se os dados em  $k$  partições de tamanho igual. Durante cada execução, uma das partições é escolhida para teste, enquanto as demais são utilizadas para treinamento. O processo se repete  $k$  vezes até que cada partição seja usada para teste uma vez conforme Tan, Steinbach e Kumar (2009).

### 5.6.1 Precisão

A maior parte da análise da avaliação começa a partir de uma matriz de confusão (também chamada de tabela de contingência), que exibe a quantidade de classificações corretas e incorretas de cada classe, podendo ser vista na Figura 13. Os verdadeiros positivos (TP) e verdadeiros negativos (TN) são as classificações corretas. Um falso positivo (FP) ocorre quando o resultado está previsto incorretamente como positivo, quando na verdade é negativo. Um falso negativo (FN) ocorre quando o resultado é incorretamente previsto como negativo quando realmente é positivo. A taxa de verdadeiros positivos é igual à TP dividido pelo número total de positivos, o que é TP + FN, e a taxa de falsos positivos é igual ao FP dividido pelo número total de negativos, FP + TN. A acurácia é o número de classificações corretas dividido pelo número total de classificações. A taxa de erro é um menos a acurácia definida pela equação  $Acurácia = (TP+TN) / (TP+TN+FP+FN)$ .

Figura 13: Matriz de confusão

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	<b>VP</b> Verdadeiro Positivo	<b>FP</b> Falso Positivo
	negativos	<b>FN</b> Falso Negativo	<b>VN</b> Verdadeiro Negativo

Fonte: Souza (2009)

Portanto ao aplicarmos a equação da Acurácia tem-se como resultado um índice que quanto maior mais preciso é o resultado, portanto busca-se o algoritmo que apresente a maior acurácia, o resultado dessa equação a própria ferramenta *WEKA* disponibiliza.

### 5.6.2 Área sob a Curva ROC

A área sob a curva ROC (*Receiver Operating Characteristic*) foi desenvolvida no contexto de detecção de sinais eletrônicos e problemas com radares, durante a segunda guerra mundial, com o objetivo de quantificar as habilidades dos operadores dos radares em distinguir um sinal de um ruído. A área sob a curva ROC em meados de 70 foi amplamente disseminada dentro da pesquisa biomédica, com o objetivo de auxiliar a classificação de indivíduos em doentes e não doentes.

Para entender melhor a área sob a curva ROC se faz necessário compreender dois conceitos: o da sensibilidade e o da especificidade. A sensibilidade é definida como a probabilidade de o teste sob investigação fornecer um resultado positivo, dado que indivíduo detém a doença. A especificidade é definida como a probabilidade de o teste fornecer um resultado negativo, dado que o indivíduo não tem a doença. A eficiência é a média aritmética entre a sensibilidade e a especificidade.

$$\text{Sensibilidade} = TP / (TP + FN)$$

$$\text{Especificidade} = TN / (TN + FP)$$

$$\text{Eficiência} = (\text{Sensibilidade} + \text{Especificidade}) / 2$$

É importante salientar que essas duas medidas não são calculadas sobre os mesmos indivíduos, ou seja, no cálculo da sensibilidade utilizam-se apenas os indivíduos doentes e no caso da especificidade utilizam-se os não doentes.

O resultado da equação da eficiência deve ser levado em consideração, quanto maior o valor da eficiência, a chance de termos casos de pacientes doentes sendo diagnosticados como normal ou pacientes normais sendo diagnosticados com alguma doença é menor.

Segundo Camara (2009) usa-se a seguinte estimativa para se avaliar apreciação de um teste usando curva ROC:

0,9 – 1,0 = Excelente

0,8 – 0,9 = Bom

0,7 – 0,8 = Regular

0,6 – 0,7 = Ruim

0,5 – 0,6 = Reprovado

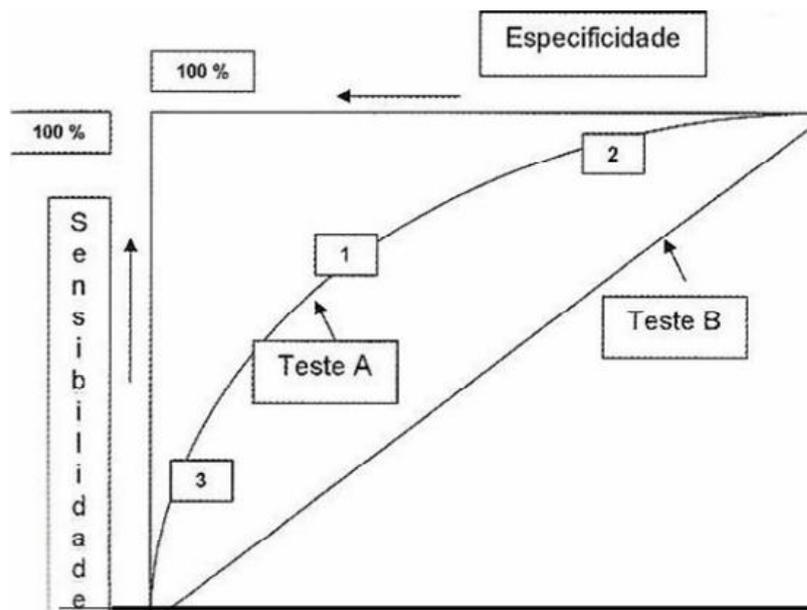
Com essa classificação dos índices acima, chega-se à conclusão que os algoritmos testados necessitam atingir uma acurácia acima de 0,9 para serem considerados satisfatórios uma vez que a aplicação desse estudo faz-se na área da saúde, tema delicado dentro do conjunto de cidades inteligentes.

Alguns exames de alguns pacientes possuem valores medidos de forma numérica, sendo assim a sensibilidade e a especificidade dependem de onde coloca o ponto do corte entre os resultados positivos e negativos.

Uma forma de apresentar a correlação entre a sensibilidade e especificidades é através da área sob a curva ROC.

A Curva ROC permite a comparação entre exames diagnósticos, no exemplo da Figura 14 temos um exemplo, o Teste A tem melhor acurácia que o Teste B, O ponto 1 confere maior valor de sensibilidade e especificidade; o ponto 2 confere maior sensibilidade, porém menor especificidade e já o ponto 3, maior especificidade, porém, menor sensibilidade.

Figura 14: Área sob a Curva ROC



### 5.6.3 Estatística Kappa

A porcentagem total de casos corretamente classificados reflete uma avaliação simples de um classificador, a mesma avaliação pela área sob a curva ROC. Um classificador contando com seleção aleatória de casos com frequência classifica alguns casos corretamente, o índice Kappa é utilizado para controlar essas instâncias que podem ter sido corretamente classificadas apenas por acaso, conforme Tighe, Laduzenski, Edwards, Ellis, Boezaart, Aygtug (2011).

A estatística Kappa é usualmente utilizada para distinguir entre a confiabilidade dos dados coletados e a sua validade. A pontuação *Kappa* média de um algoritmo para ser satisfatório são na faixa de 0,6-0,7. Conforme Othman e Yau (2007). É calculada através da taxa de concordância observada (TCO) e da taxa de concordância esperada (TCE) se as respostas das duas ocasiões fossem

estatisticamente independentes. Observa-se o exemplo da Tabela 3. Nesse caso, a taxa de concordância observada é igual à somas dos valores nos quais as duas ocasiões classificaram negativos como negativos e positivos como positivos, ou seja, é a soma dos valores da diagonal principal (24+17) dividida pelo total de classificações (50). Isso resulta em 0,82. A taxa de concordância esperada se calcula a probabilidade dos casos negativos em relação ao total de casos tanto para a ocasião 1 como para a ocasião 2, ou seja, será igual a 27/50 multiplicado por 30/50. Faz-se o mesmo agora para os casos positivos, e tem-se 23/50 multiplicado por 20/50. Depois, somam-se essas duas quantias e tem-se 0,51 como resultado. Finalmente, o valor de Kappa será igual a 0,6327 de acordo com a equação  $Kappa=(TCO-TCE)/(1-TCE)$ .

Tabela 3: Exemplo de matriz de Confusão

		Ocasião 1		Total
		Negativo	Positivo	
Ocasião 2	Negativo	24	3	27
	Positivo	6	17	23
		30	20	50

#### 5.6.4 Medida F

A medida F é usada porque, apesar da precisão e do *recall* serem métricas válidas, uma pode ser otimizada em detrimento da outra. A medida F somente produz um resultado elevado quando tanto a precisão quanto o recall estão balanceados, tornando portanto esta medida bastante significativa.

A precisão é a proporção de resultados relevantes que estavam corretos sendo representado pela equação:  $Precisão = TP / (FP+TP)$

Já o Recall é a proporção de resultados relevantes que foram identificados corretamente representado pela equação:  $Recall = TP / (FN+TP)$

Por fim, a medida F é derivada dos valores de precisão e recall representado pela equação:  $Medida\ F = (2 * Recall * Precisão) / (Recall + Precisão)$ .

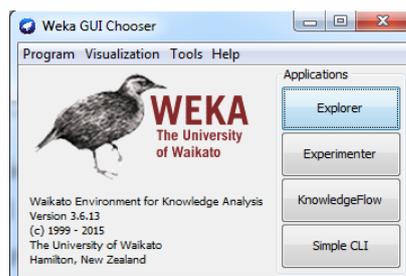
## 6 WEKA

*WEKA* é uma ferramenta de mineração de dados que por sua vez é uma coleção do estado da arte de algoritmos de aprendizagem de máquina e de ferramentas de pré-processamento. Ela foi construída para que se possam realizar experimentos de forma rápida e flexível. A ferramenta auxilia em todo o processo de mineração de dados, incluindo a preparação da entrada de dados, aplicação das técnicas existentes da ferramenta e resultados da aprendizagem. Esse conjunto de ferramentas é acessado através de uma *interface* comum para que os usuários consigam comparar diversos métodos e identificar aquele que se adéqua ao problema em questão.

O *Weka* foi desenvolvido na Universidade de *Waikato*, na Nova Zelândia e o seu nome significa *Waikato Environment for Knowledge Analysis*. Conforme Witten e Frank (2005). A ferramenta é implementada em Java e é distribuída sob os termos da *General Public License*. O *Weka* disponibiliza as implementações de algoritmos de aprendizado de máquina a fim de que estas possam ser aplicadas sobre as bases de dados em estudo. Entre as técnicas de Mineração de Dados disponíveis estão os de regressão, classificação, associação, agrupamento e afins. O formato tradicional de arquivo que o *Weka* entende é o ARFF (*Attribute-Relation File Format*).

A tela inicial do *WEKA*, vista na Figura 15, mostra as quatro aplicações que compõem a ferramenta, que são o modo *Explorer*, o *Experimenter*, o *Knowledge Flow* e o *SimpleCLI*. A versão utilizada é a 3.6 (estável)

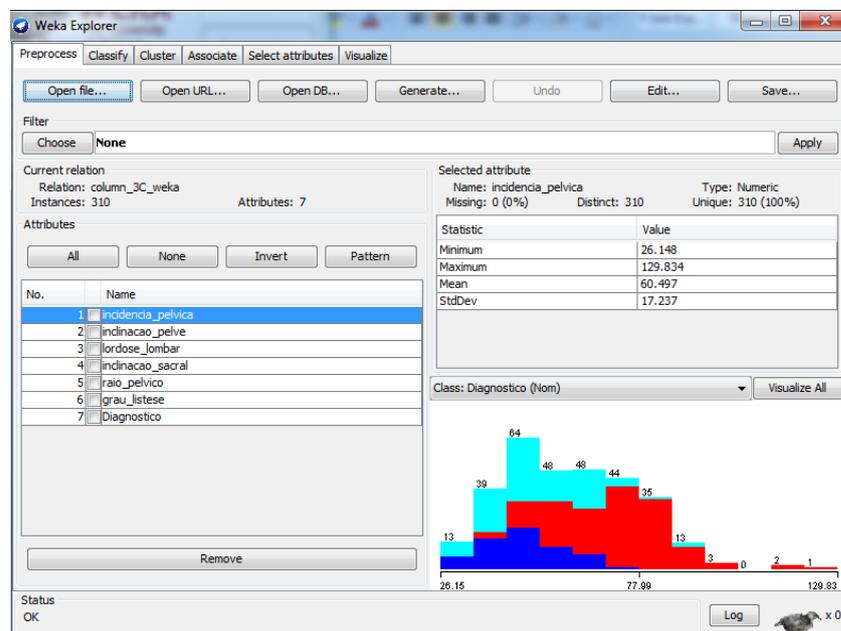
Figura 15: Tela Inicial do *Weka*



O modo *Explorer* é utilizado para aplicar os métodos de Mineração de Dados e de pré-processamento sobre uma base de dados específica. Ao utilizar o

*Explorer*, e selecionar determinado conjunto de dados, o *WEKA* mostra ao usuário quantos e quais são os atributos que definem cada instância da base e seus tipos (nominais, numéricos e afins). Além disso, para cada atributo são mostradas informações como o valor mínimo e máximo do atributo, sua média bem como o seu desvio padrão. Também é informado se existem instâncias com atributos ausentes, quantos valores de atributos são distintos ou únicos. Na Figura 16 são mostradas as informações citadas, que ficam na aba de pré-processamento do modo *Explorer* do *WEKA*.

Figura 16: Aba de pré-processamento, exibindo as variáveis (insumos)

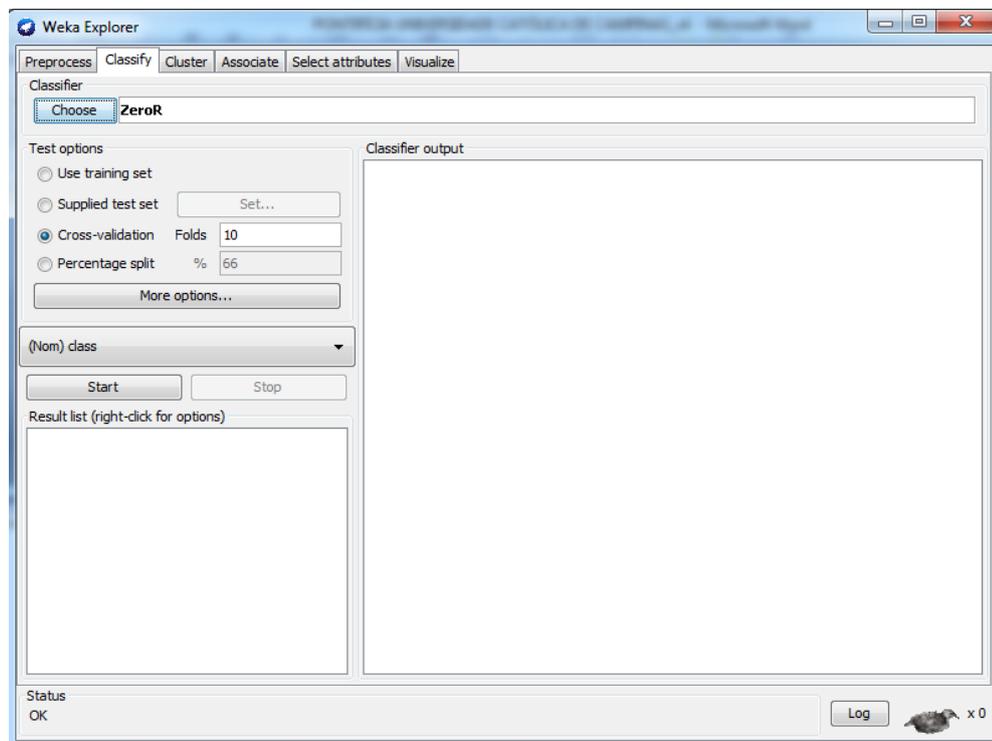


Veja que é possível realizar-se carga de dados através de arquivo ARFF através do botão “*Open File*”, e também é possível realizar a leitura de dados que estejam em algum tipo de banco de dados através do botão “*Open DB*”

Das demais abas do modo *Explorer*, para as tarefas de classificação e de associação de bases de dados, utilizam-se as abas *Classify* e *Associate*. Na Figura 17 visualiza-se a *Interface* disponibilizada pelo *WEKA* para aplicação de algoritmos de classificação sobre a base de dados escolhida previamente na aba *Preprocess*. Para tanto, na seção *Classifier* (no canto superior esquerdo da tela) é feita a escolha do algoritmo a ser empregado. Clicando duas vezes sobre o nome do algoritmo escolhido, é possível alterar os parâmetros de configuração

existentes a fim de refinar a classificação dos dados. Logo abaixo, é possível escolher qual tipo de teste que será aplicado para determinar a avaliação do desempenho do algoritmo. Pode-se escolher entre usar o próprio corpo de dados de treino como dados de teste, indicar o uso de um arquivo com instancias para testes em separado, usar o *cross-validation* nessa opção é utilizado toda a base como treino e teste, porém em momentos diferentes já que o algoritmo é processado  $K$  vezes de acordo com a sua escolha, o padrão é 10 ou ainda determinar um percentual de divisão da base de dados carregado na aba *Preprocess* entre dados para treino e dados para teste.

Figura 17: Aba que permite selecionar o tipo de algoritmo e forma de teste

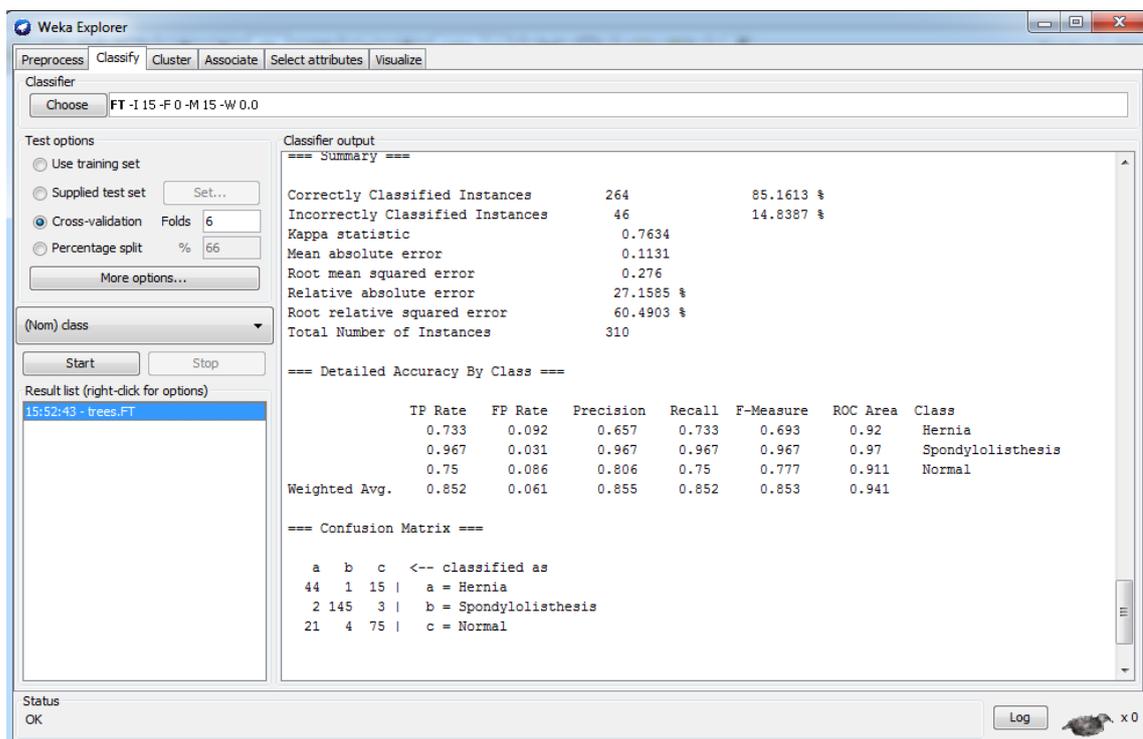


Após realizar as escolhas desejadas, o usuário deverá clicar no botão *Start*, aparece no painel à direita, a saída correspondente à execução da tarefa de classificação da base de dados configurada. Nela é possível verificar os resultados que algumas métricas de desempenho geraram, como por exemplo a matriz de confusão.

Podemos observar na figura 18 que foi escolhido o algoritmo FT com *Cross Validation* com folder de 10, ou seja, o algoritmo será processado 10 vezes

e em cada rodada usará parte da base como treino e parte da base como teste alternando-se a cada rodada garantindo que todas as instancias uma hora será treino e em outro momento será teste. Nesse caso em cada rodada do algoritmo é aplicado 279 instancias como treino e 31 instancias como teste, ou seja, a base total é dividida por 10 e 9/10 é utilizada como treino e 1/10 será utilizada como teste, ao final faz-se a média entre os processamentos.

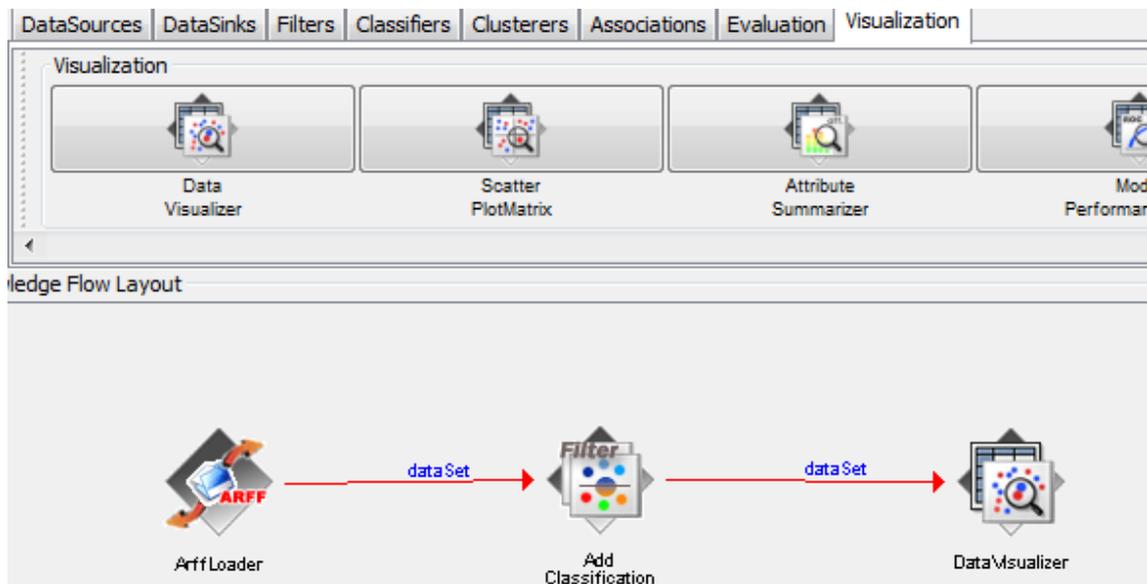
Figura 18: Exemplo de uma árvore de decisão sendo processada



Ao realizar os estudos sobre algoritmos de classificação, quando se faz necessário realizar uma comparação entre diferentes algoritmos, o *Weka* oferece um ambiente gráfico apropriado que pode ser acionado na tela inicial pelo botão *Experimenter*. Ao entrar no *Experimenter*, o usuário visualiza a aba *Setup*, na qual é possível configurar onde será salvo o experimento, (arquivo .exp), qual será o tipo de experimento (divisão da base de dados para treino e teste ou *cross-validation*), qual ou quais as bases de dados a serem experimentadas e quais os algoritmos de classificação (ou regressão) que serão comparados. Após essa configuração, o usuário deve ir para a aba *Run*, na qual é disparado o experimento, ou seja, os algoritmos selecionados serão aplicados sobre as bases de dados alvo, armazenando os resultados de desempenho percebidos. A aba

*Run* mostra um relatório da execução do experimento, indicando a presença ou não de erros. Por fim, na aba *Analyse*, clica-se no botão *Experiment* (no canto superior direito da tela) para que sejam carregados os resultados de desempenho do experimento que foi concluído na aba *Run*. Sobre as demais opções da tela inicial do *Weka*, o *KnowledgeFlow* é responsável por permitir que os filtros de pré-processamento e os algoritmos de classificação e associação sejam aplicados à base de dados através de uma *interface* gráfica, na qual os blocos responsáveis pela filtragem e pela Mineração de Dados são conectados por setas indicando o fluxo do experimento como pode ser visualizado na figura 19.

Figura 19: Tela do Modo Knowledge Flow do Weka



Quanto ao *SimpleCLI*, este se trata de um terminal para a execução dos experimentos já citados, apenas permitindo ao usuário entrar com linhas de comandos, sem *interface* gráfica.

Após a demonstração dos diferentes tipos de serviços providos pela ferramenta *WEKA* através da sua *interface* gráfica, é importante ressaltar que os algoritmos dos filtros, dos classificadores, associadores, entre outros métodos de Mineração de Dados e afins, são disponibilizados via uma API, do inglês *Application Programming Interface*. Sendo assim, é possível fazer uso dessa API para implementar aplicativos capazes de executar as diferentes etapas do

processo de descoberta de conhecimento de base de dados. A API é escrita em linguagem *Java*.

## 7 ORTOPEDIA

Segundo Malavolta (2013) a medicina baseada em evidências apresenta destaque na literatura médica atual. Fundir experiência clínica com resultados estatísticos é o objetivo dessa prática científica. Por meio dela, os médicos apresentam maior embasamento para atuar e podem obter melhores resultados clínicos e reduzir custos.

Contudo, Malavolta (2013) contextualiza que algumas áreas médicas apresentam características que impossibilitam a elaboração desse tipo de estudo com seu rigor máximo. A ortopedia é uma disciplina cirúrgica que apresenta obstáculos próprios para a elaboração de um ensaio clínico de alta qualidade, por esse motivo, existe um grande número de pesquisas feitas de maneira observacional que geram dados com evidências discutíveis, que podem até aumentar as dúvidas já existentes.

Portanto esse trabalho busca ser uma pesquisa com o intuito de dar um suporte ao médico ortopedista a fim de ampliar o seu conhecimento acadêmico e experiências com a descoberta de informações em bases desconhecidas.

A ortopedia é a especialidade clínica e cirúrgica dedicada à prevenção, diagnóstico e tratamento de moléstias e lesões do sistema musculoesquelético, os distúrbios musculoesqueléticos correspondem a um conjunto de desordens inflamatórias e/ou degenerativas que podem resultar principalmente em dor e incapacidade funcional sendo considerado hoje um dos mais significativos e dispendiosos problemas de saúde pública em populações de todo o mundo, conforme Marras (2009).

Neste trabalho a base de dados pública (<http://archive.ics.uci.edu/ml/datasets/Vertebral+Column>) trabalhada envolve a especialidade da Ortopedia as doenças Listese e Hérnia de Disco.

### 7.1 Listese

O termo Espondilolistese mais conhecido como Listese é definido como uma translação de uma vértebra sobre a outra em sentido anterior ou posterior.

No adulto isso ocorre na coluna lombar, como resultado de um defeito na arquitetura óssea, trauma ou processo degenerativo, conforme Tebet (2014).

Contudo Tebet (2014) frisa que a listese tem sido uma patologia de difícil compreensão para ortopedistas, neurocirurgiões e pediatras por causa da grande variedade de formas anatômicas e clínicas existentes.

A incidência de listese na população geral é de cerca de 6%, com uma proporção de homens:mulheres de 2:1, a incidência em crianças até 6 anos é de 2,6% já em adultos é de 5,4%, conforme Tebet (2014).

Como já citado em capítulo anterior a população do Mundo atualmente segundo a ONU é 7.4 Bilhões, ou seja, 444 milhões (6% de 7,4Bilhões) sofrem com a listese, número expressivo que merece uma atenção para um diagnóstico com qualidade.

O diagnóstico é feito com radiografias da coluna lombar nas incidências de frente e perfil em ortostatismo, a ressonância magnética é o exame de escolha para visualizar o disco no nível da deformidade. Exames de imagem mais avançados como tomografia computadorizada por emissão de fóton único são mais sensíveis e fornecem maiores detalhes, o exame radiográfico dos pacientes deve incluir a **radiografia panorâmica da coluna** e visualizar as cabeças femorais, para que sejam possíveis a avaliação angular da junção lombossacra e o equilíbrio sagital, conforme Tebet (2014).

## 7.2 Hérnia de disco

Gabriel (2001) descreve a hérnia de disco lombar como sendo uma protrusão do núcleo discal que pode comprimir raízes nervosas, gerando sintomas sensitivo-motores.

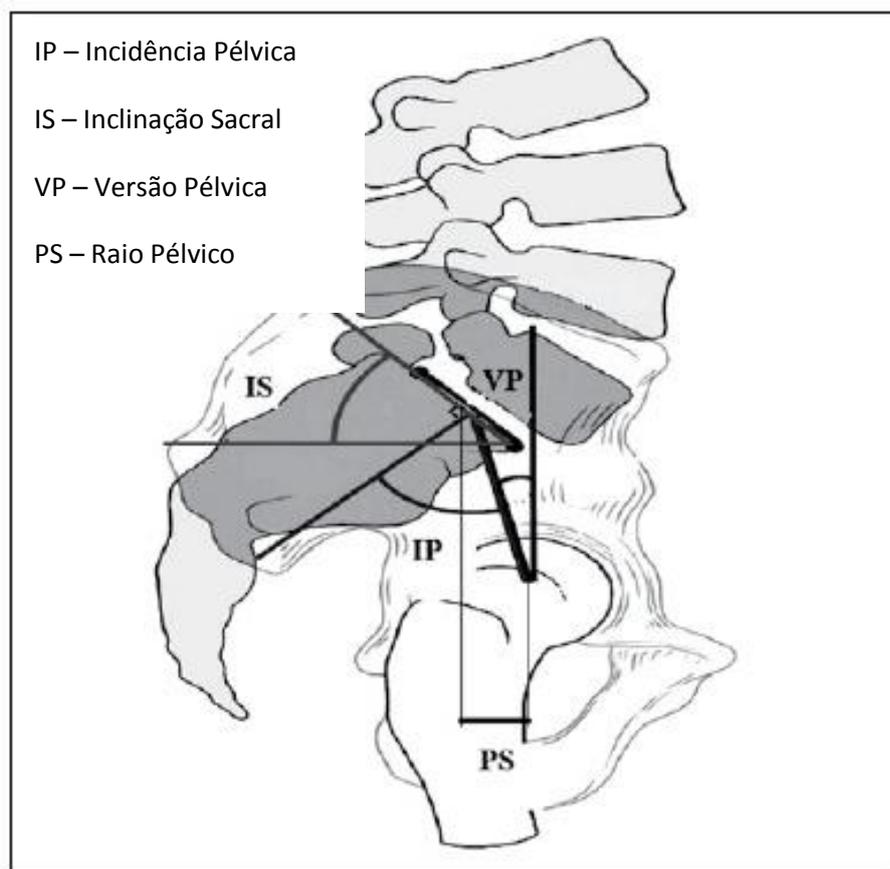
A hérnia de disco provoca sintomas como dor na região lombar, que pode irradiar para o membro inferior, levando seu portador a ter dificuldade em realizar atividades laborais, diminuindo assim a qualidade de vida do mesmo conforme Falavigna (2010). Ela acontece nos discos intervertebrais, que são estruturas finas e achatadas que servem de “amortecedores” entre as vértebras.

Cada disco é feito de uma parte interna gelatinosa e macia cercada por um revestimento externo duro e fibroso. Quando parte desta estrutura sai de sua posição normal, ocorre ruptura do anel fibroso, com subseqüentemente deslocamento da massa centrado do disco nos espaços intervertebrais, e conseqüentemente comprimem as raízes nervosas, conforme Alfieri (2015).

A base pública deste trabalho com 310 instâncias detém de 6 atributos que em conjunto formam o diagnóstico como Normal, Listese ou Hérnia de Disco.

Na Figura 20, pode-se observar 4 dos 6 atributos da base de dados

Figura 20: Parâmetros da Pelve



Fonte: HENNEMAN, A.O. (et al.;2012)

**Incidência pélvica (IP):** o ângulo entre uma linha perpendicular ao ponto médiado platô sacral, e uma linha que liga este ponto até o eixo central da cabeçafemoral.

**Inclinação sacral (IS):** o ângulo entre uma linha no platô superior de S1 e uma linha horizontal (paralela ao solo).

**Versão pélvica (VP):** o ângulo entre uma linha que liga o ponto médio do platô sacral até ao eixo da cabeça femoral, e uma linha vertical (perpendicular ao solo).

**Projeção de S1 (PS) = Raio Pélvico:** a distância entre o eixo bicoxofemoral e a projeção para este nível do ponto médio do platô superior sacral.

Na Tabela 4 pode-se observar a composição da base de dados.

Tabela 4: Trecho da base de dados pública utilizada neste trabalho

Incidencia Pelvica	Inclinacao Pelve	Lordose lombar	Inclinacao Sacral	Raio Pelvico	Grau espondilolistese	Diagnostico
63,93	19,97	40,18	43,96	113,07	-11,06	Normal
45,37	10,76	29,04	34,61	117,27	-10,68	Hernia Disco
26,15	10,76	14	15,39	125,2	-10,09	Hernia Disco
41,17	17,32	33,47	23,85	116,38	-9,57	Hernia Disco
45,08	12,31	44,58	32,77	147,89	-8,94	Normal
50,16	-2,97	42	53,13	131,8	-8,29	Normal
72,08	18,95	51	53,13	114,21	1,01	Listese
40,41	-1,33	30,98	41,74	119,34	-6,17	Normal
41,65	8,84	36,03	32,81	116,56	-6,05	Normal
46,24	10,06	37	36,17	128,06	-5,1	Normal

## **8 PERCURSO METODOLÓGICO**

Esse trabalho foi desenvolvido com o intuito de evidenciar a possibilidade de aplicação de conceitos da área da Tecnologia da Informação na Medicina, a fim de acrescentar a esta última, progressos nos seus métodos de diagnóstico e de extração de informação úteis. Será estruturada em duas etapas, a primeira correspondendo à questão da escolha do algoritmo mais adequado aos objetivos e propostos seus respectivos testes, e em segundo lugar a elaboração do acesso móvel para tornar a proposta de fácil utilização pelos seus futuros usuários.

### **8.1 Definição do algoritmo mais adequado.**

A disponibilização de bases para apoio à medicina é uma tendência esperada segundo apresentam os autores Zhang, Williams, Pengetnze (2015); Viceconti, Hunter, Hose (2015); Perez, Poon, Merrified, Wong, Yahng (2015), e as TICs permitindo fácil acesso, torna-se a busca por bons serviços algo de grande interesse.

A predição de diagnóstico é realizada aplicando algoritmos de classificação em uma ferramenta de mineração de dados que são responsáveis em arquitetar um modelo que representa como esses dados se relacionam a partir de uma base de dados histórica com diagnósticos já realizados por médicos.

Algoritmos de classificação fazem com que os algoritmos tenham a capacidade de aprender como os diagnósticos foram gerados e, no futuro, ao precisar diagnosticar novos pacientes poderá usar esse modelo de aprendizado.

Portanto, é necessário que as instâncias no processo de KDD sejam analisadas por um ou mais profissionais da área médica. Propõe-se o desenvolvimento de um aplicativo com interface amigável contendo o módulo de entrada de dados derivados de exames, agregando uma segunda opinião ao médico. Ao passo que temos evolução no conhecimento medicinal, esse aplicativo servirá como um aprendizado pelo profissional da saúde para atualização do seu conhecimento sobre a especialidade ortopédica.

Na Figura 21 observa-se o fluxo sugerido do atendimento médico.

Figura 21: Fluxo de funcionamento da solução *mobile*



A classificação segundo Tan, Steinbach e KUMAR (2013) é uma das tarefas cognitivas humanas mais realizadas no auxílio à compreensão do ambiente em que se vive. Esse contexto faz também da classificação uma das tarefas mais utilizadas do *data mining*.

Neste trabalho usaremos uma base de dados pública, de forma que as questões éticas de se trabalhar com dados oriundos de seres humanos estão sanadas pelos organizadores da mesma.

Os responsáveis doadores desse conjunto de dados são Guilherme de Alencar Barreto e Ajalmar Rêgo da Rocha Neto, do Departamento de Engenharia Teleinformática, da Universidade Federal do Ceara, Fortaleza, Brasil, e Henrique Antônio Fonseca da Mota Filho do Hospital Monte Klinikum, em Fortaleza, Ceara, Brasil. Esse conjunto de dados biomédicos foi construído pelo Dr. Henrique da Mota, durante um período de residência médica no Grupo de Pesquisa Aplicada em Ortopedia (GARO) do Centro Médico-Cirúrgica de readaptação des Massues, Lyon, França. Os dados foram organizados em duas tarefas de classificação diferentes, mas relacionadas. A primeira tarefa consiste em classificar os pacientes como pertencentes a uma de três categorias: normal (100 pacientes), hérnia de disco (60 pacientes) ou Espondilolistese (150 pacientes).

A classificação dos casos/exemplos contidos na base de dados é feita analisando-se 6 atributos:

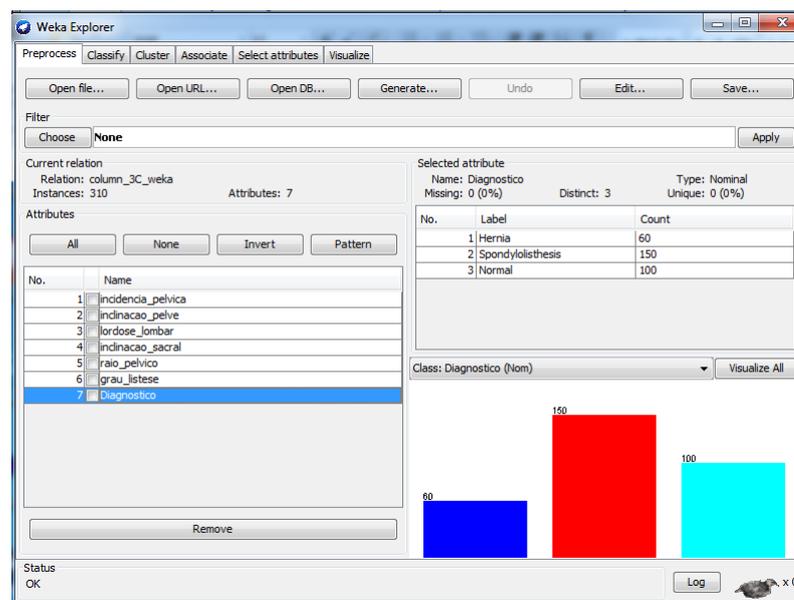
1. Incidência Pélvica
2. Inclinação da Pelve
3. Ângulo de Lordose Lombar
4. Inclinação Sacral
5. Raio Pélvico
6. Grau de Espondilolistese

Assim, a segunda tarefa consiste em classificar os pacientes como pertencentes a uma de duas categorias: Normal (100 pacientes) ou anormal (210 pacientes). Neste trabalho, foi utilizada a versão do conjunto de dados com 3 categorias. Cada paciente é representado no conjunto de dados por seis atributos derivadas da forma e da orientação da coluna vertebral lombar e da pelve.

A base de dados usada pode ser acessada no endereço que se segue: <http://archive.ics.uci.edu/ml/datasets/Vertebral+Column>

A Figura 22 exhibe as variáveis da base pública

Figura 22: Apresentação das variáveis da base pública



Escolher uma ferramenta de *data mining* não é foco deste trabalho, portanto, optou-se por analisar a ferramenta utilizada e citada em trabalhos semelhantes (estado da arte).

Dessa forma, considerou-se as literaturas existentes Scheunemann (2015), Carvalho (2012) e dentre estes ressalta-se a dissertação de Cruz (2007) que traz um detalhamento das ferramentas mais utilizadas, e onde foram contabilizadas 159 ferramentas de mineração de dados, o que mostra ser um tema relevante pela quantidade de *softwares* disponíveis.

Pode-se ressaltar aqui que o autor classificou as ferramentas utilizando-se de alguns critérios que estão a seguir:

- A versão** - final (F) ou beta (B);
- A licença** - comercial (C), *freeware* e *shareware* (F) ou pública (P);
- A disponibilidade** – se é ou não disponibilizada uma versão de demonstração (Demo) ou a ferramenta é totalmente operacional para *download*;
- A aplicação** - acadêmica (A) ou comercial (C);
- A arquitetura** - *Stand alone* (S), *Client/Server* (C/S) ou Processamento Paralelo (PP).

A ferramenta *WEKA* possui versão final, sua licença é pública, é disponível uma versão de demonstração, sua aplicação é acadêmica e sua arquitetura é *Stand Alone*, ou seja, é uma ferramenta auto-suficiente, para seu funcionamento não depende de outros *software*, além de que novos algoritmos podem ser agregados à ferramenta, ampliando sua variedade de métodos de mineração de dados.

Por essas características, principalmente por ser uma ferramenta *open source*, além de ser uma coleção do estado da arte a ferramenta *WEKA* foi escolhida para auxiliar na extração de conhecimento da base pública da coluna vertebral.

O *Weka* se encontra disponível no endereço que se segue: (<http://www.cs.waikato.ac.nz/ml/weka>).

Para a aplicação de técnicas de mineração de dados necessita-se que os dados a serem utilizados estiveram de forma organizada. Os arquivos contendo a base de dados podem estar em alguma estrutura de dados, planilha ou banco de dados. O *WEKA* possui um formato para a organização dos dados, seu nome é *ARFF*. Neste arquivo devem estar presentes uma série de informações, dentre elas: domínio do atributo, valores que os atributos podem representar e atributo classe. O arquivo *ARFF* é dividido em duas partes, a primeira contém uma lista de todos os atributos, onde se deve definir o tipo do atributo e/ou os valores que ele pode representar. Os valores devem estar entre chaves ({}), separados por vírgulas. A segunda é composta pelas instâncias presentes nos dados, os atributos de cada instância devem ser separados por vírgula, e aqueles que não contêm valor, o valor deve ser representado pelo caractere '?'. As informações presentes no arquivo *ARFF* são especificadas utilizando marcações. Por exemplo, o nome do conjunto de dados é especificado através da marcação @relation, @attribute para os atributos e os dados em si são definidos através da marcação @data.

Na figura 23 observa-se um trecho do arquivo que contém a base de dados pública.

Figura 23: Parte do arquivo da base de dados publica

```
column_3C_weka - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
@relation column_3C_weka
@attribute pelvic_incidence numeric
@attribute pelvic_tilt numeric
@attribute lumbar_lordosis_angle numeric
@attribute sacral_slope numeric
@attribute pelvic_radius numeric
@attribute degree_spondylolisthesis numeric
@attribute class {Hernia, Spondylolisthesis, Normal}
@data
84.5856071,30.36168482,65.47948563,54.22392228,108.0102185,25.11847846,Spondylolisthesis
89.83467631,22.63921678,90.56346144,67.19545953,100.5011917,3.040973261,Normal
41.35250407,16.57736351,30.70619135,24.77514057,113.2666746,-4.497957556,Hernia
56.10377352,13.10630665,62.63701952,42.99746687,116.2285032,31.17276727,Normal
72.64385013,18.92911726,67.99999999,53.71473287,116.9634162,25.38424676,Spondylolisthesis
67.53818154,14.65504222,58.00142908,52.88313932,123.6322597,25.9702063,Normal
```

## 8.2 Método de comparação entre os algoritmos de Classificação

Para escolher o melhor algoritmo de classificação foram selecionados doze algoritmos baseados no método de árvore de decisão e quatro algoritmos

baseados no método classificadores bayesianos. Esses algoritmos foram comparados a fim de escolher o algoritmo com a maior acurácia.

A lista de algoritmos do método de árvore de decisão é a seguinte:

1. Best-Firts Decision Trees
2. Decision Stump
3. Functional Trees
4. J48
5. J48 graft
6. Logistic Model Tree
7. Naive-Bayes Trees
8. Random Forest
9. Random Trees
10. SimpleCart
11. Logic-Boost Alternating Decision Tree
12. REPTree (Fast Decision Tree)

A lista de algoritmos do método de classificadores Bayesianos é a seguinte:

1. NaiveBayes
2. NaiveBayesUpdateable
3. NaiveBayesSimple
4. BayesNet

Os critérios utilizados para a comparação entre os algoritmos serão:

1. Índice de Precisão
2. Índice de estatística Kappa
3. Medidor F
4. Curva ROC

Neste trabalho foram utilizados dois métodos para separar a base de dados como treino (dados reais) e teste (simulando que algumas instâncias são novos pacientes).

1. *cross-validation* (Validação Cruzada) com o parâmetro  *folds* 10 em função de trabalhar com todas as instâncias como treinamento e teste utilizando-se dessa forma todo o conjunto de dados.

2. *Percentage Split* (Porcentagem de divisão) com o parâmetro 66% da base como treino e 33% como teste, para o processamento desse método a base foi ordenada de forma randômica para evitar uma desproporção da quantidade de dados de uma mesma classe aparecer somente no conjunto de dados de treino e não de teste.

### 8.3 Elaboração da Interface amigável

O estudo da Interface entre o Ser Humano e o Computador (IHC) envolve muitas disciplinas, dentre estas disciplinas as principais são Rocha, Baranauskas(2003) :

Psicologia: Social: Estudo da natureza e as causas do comportamento humano na conjuntura social; Organizacional: Provê aos Designers o conhecimento sobre as estruturas organizacionais e sociais sobre como a introdução de computadores influencia nas práticas de trabalho; Cognitiva: Entender o comportamento humano e os processos mentais; Ergonomia: Seu objetivo é conceder e fazer o design de diversas ferramentas e artefatos para diferentes ambientes de trabalho; Sociologia, Filosofia e Antropologia: O conjunto destas 3 disciplinas tem por objetivo entender o comportamento do que acontece no momento de uso da tecnologia, ou seja, analisam o comportamento quando as pessoas se comunicam com as máquinas.

Linguística: Visa explorar a estrutura da linguagem natural da concepção de interfaces, principalmente para facilitar o acesso e consulta a bases de dados.

Design: Tem oferecido o conhecimento mais que evidente, por exemplo, na área de design gráfico. Essa área propiciou o aperfeiçoamento das telas e de seu efeito final.

Engenharia: É onde as soluções dos problemas encontrados aparecem. A engenharia utiliza os resultados e constatações da pesquisa para a produção de soluções.

Ciência da Computação: Provê o conhecimento técnico e oferece idéias de como aproveitar em sua plenitude todo o potencial da interface.

Inteligência Artificial: Visa desenvolver sistemas computacionais, que tomem certas decisões semelhantes à que os Seres Humanos tomariam em certas circunstâncias.

### 8.3.1 Relevâncias do Design

Muitos programadores, ao desenvolver qualquer tipo de *software* não se preocupam com o *design* das interfaces. Normalmente são interfaces que apenas eles próprios entendem, mas sem nenhum tipo de parâmetro ou preocupação com usuário comum. Segundo Rocha e Baranauskas (2003), P.25 “*Se pensarmos na complexidade da maioria dos sistemas computacionais, veremos que o potencial de se ter uma precária IHC é bastante alto.*”

Observando o comportamento das pessoas com suas experiências de sucessos e fracassos, Rocha e Baranauskas (2003) propõe quatro diretrizes que o design deve levar em consideração, são eles:

- Visibilidade:

Consiste em deixar bem sinalizado e bastante clara a utilização de partes da interface, conduzindo o usuário a tomar as ações mais corretas.

- Bom modelo conceitual:

Segundo Rocha e Baranauskas (2003, P.26) “*Um bom modelo conceitual permite prever o efeito de ações. Sem um bom modelo conceitual opera-se sob o comando, cegamente.*”

- Bons mapeamentos:

Serve para mostrar o relacionamento entre os controles e suas respostas, ou seja, naturalmente consegue-se entender o que aquele controle faz e mais, nos bons mapeamentos, os controles executam exatamente o que os usuários esperam deles.

### 8.3.2 Objetivos da Avaliação

Segundo Rocha e Baranauskas (2003) podemos dizer que a avaliação tem três grandes objetivos: avaliar a funcionalidade do sistema, avaliar o efeito da interface junto ao usuário e identificar problemas específicos do sistema.

A funcionalidade do sistema é um fator extremamente importante. Além de atender todos os requisitos do usuário, o sistema precisa disponibilizar tudo de maneira fácil e eficiente para o uso. A avaliação, neste caso, envolve também o desempenho do usuário junto ao sistema, medindo a eficiência do sistema na execução da tarefa pelo usuário.

Junto com a funcionalidade, deve-se medir também o impacto do *design* junto ao usuário, ou seja, a usabilidade do sistema (itens de avaliação citados no item anterior).

E por último, como todo teste, o objetivo é também identificar problemas específicos com o *design*. Deve-se testar o sistema de forma a encontrar aspectos do *design* que quando usados podem causar resultados inesperados ou confusão entre os usuários. Obviamente estes testes envolvem não somente a usabilidade, como também a funcionalidade do sistema.

Os métodos de avaliação desses objetivos, de acordo com Rocha e Baranauskas (2003) podem ser feitos em duas dimensões:

- Se usuários reais estão ou não envolvidos e;
- Se a *interface* está ou não implementada.

Estes autores, nessas dimensões destacam dois grupos de métodos de avaliação:

- Inspeção de usabilidade:

Sem envolver usuários e podendo ser usados em qualquer fase do desenvolvimento de um sistema.

- Testes de usabilidade:

Métodos de avaliação centrados no usuário que incluem métodos experimentais ou empíricos, métodos observacionais e técnicas de questionamento.

Este trabalho utilizará a inspeção de usabilidade, descrito acima, como método de avaliação, uma vez que não foi possível durante o desenvolvimento do aplicativo estar com o médico da PUC para auxiliar no desenvolvimento.

### 8.3.3 Inspeção de Usabilidade

A inspeção de usabilidade objetiva encontrar problemas de usabilidade de uma interface com o usuário. Com base nos problemas identificados, fazer recomendações no sentido de eliminá-los e conseqüentemente melhorar a usabilidade.

Como a definição de inspeção de qualidade conforme Rocha e Baranauskas (2003) tem-se: a inspeção de usabilidade como um conjunto de métodos baseados em ter avaliadores inspecionando ou examinando aspectos relacionados à usabilidade de uma interface de usuário.

Dessa maneira, fica claro que as inspeções devem ser feitas na fase de desenvolvimento da interface, evitando que os problemas de usabilidade sejam identificados tardiamente.

## **9 SOFTWARES UTILIZADOS NESTE TRABALHO**

### **9.1 MYSQL**

Esse módulo da solução tem por objetivo prover a *interface* de entrada e saída de dados para o usuário que se utiliza de dispositivos móveis *Android*. Os *softwares* utilizados nesse projeto foram todos *open source* a fim de disponibilizar uma solução com baixo custo.

O MySQL tem como função neste projeto armazenar os registros da base pública contendo as 310 instâncias e disponibilizá-los para serem utilizados em tempo real pela aplicação web, assim como também manter um histórico de novos diagnósticos.

### **9.2 Xamarin**

Para criação do Aplicativo Dr. 2º Opinião, foi utilizada a plataforma *Xamarin*, *software* para desenvolvimento de aplicativos móveis para os sistemas operacionais *IOS*, *Android* e *Windows*.

A *interface* do aplicativo proposto também buscou atender as questões de usabilidade já citadas em capítulos anteriores.

## 10 RESULTADOS

### 10.1 Resultados e análise dos testes realizados.

Neste capítulo, são reportados os resultados dos testes realizados com os algoritmos de classificação sendo testados 16 algoritmos sendo 12 algoritmos do método árvore de decisão e 4 algoritmos do método Classificadores Bayesianos, com o objetivo de escolher o melhor algoritmo baseado nos indicadores de precisão, curva ROC, índice Kappa e Medida F.

Resultado dos testes com os algoritmos de classificação utilizando o método de aprendizado usando a base de dados pública: o treinamento dos dados e a validação da classificação foram realizados pela própria ferramenta Weka pelo método *Holdout*, conhecido como método Treino-Teste. Nesse método 2 / 3 dos dados são destinados para o treinamento e 1 / 3 dos dados são usados para validação do modelo conforme Olson e Delen (2008).

1. Resultados dos testes feitos com 12 algoritmos do método de árvore de decisão e 4 algoritmos do método de classificadores Bayesianos utilizando o método *Holdout*.

Na tabela 5 observa-se o resultado dos algoritmos.

Tabela 05 - Resultados dos testes feitos pelo método *Holdout*

Nº	Classificadores	Algoritmo	Instancias Total	Instancias Correta	Instancias Incorreta	Precisão	Estatística Kappa	Medida F	Curva ROC
1	Arvores de Decisão	FT	105	88	17	84,4%	0,74	0,84	0,97
2	Arvores de Decisão	RandomTree	105	88	17	84,2%	0,74	0,84	0,89
3	Bayesianos	NaiveBayesSimple	105	88	17	83,4%	0,74	0,83	0,96
4	Bayesianos	NaiveBayes	105	88	17	83,3%	0,74	0,83	0,96
5	Bayesianos	NaiveBayesUpdateable	105	88	17	83,3%	0,74	0,83	0,96
6	Arvores de Decisão	RandomForest	105	86	19	83,0%	0,71	0,81	0,96
7	Arvores de Decisão	LMT	105	86	19	82,7%	0,71	0,82	0,96
8	Arvores de Decisão	NBTree	105	85	20	82,6%	0,69	0,80	0,93
9	Arvores de Decisão	LADTree	105	86	19	82,3%	0,71	0,81	0,93
10	Arvores de Decisão	SimpleCart	105	86	19	81,7%	0,71	0,82	0,88
11	Arvores de Decisão	BFTree	105	83	22	79,6%	0,67	0,78	0,91
12	Arvores de Decisão	REPTree	105	83	22	78,9%	0,66	0,78	0,88
13	Arvores de Decisão	J48graft	105	81	24	76,1%	0,63	0,77	0,88
14	Arvores de Decisão	J48	105	81	24	76,1%	0,63	0,77	0,88
15	Bayesianos	BayesNet	105	74	31	69,2%	0,53	0,69	0,90
16	Arvores de Decisão	DecisionStump	105	77	28	60,5%	0,58	0,65	0,91

### Descrição dos campos da tabela 04

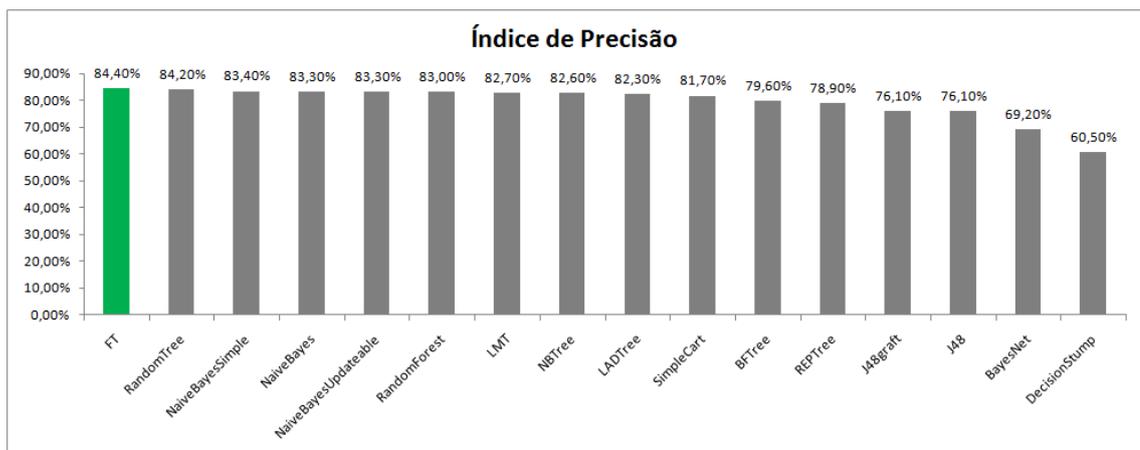
Instâncias = Quantidade de instâncias, que foram usadas como teste (105 aproximadamente 33% da base).

Instâncias Corretas = Quantidade de pacientes que foram diagnosticados de forma correta.

Instâncias Incorretas = Quantidade de pacientes que foram diagnosticados de forma incorreta.

Precisão, Estatística Kappa, Medida F e Área sob a Curva ROC: Índice fornecido pela ferramenta *Weka*

Gráfico5: Índice de Precisão do teste 1



Como já citado, no gráfico 5 observa-se o algoritmo FT como melhor índice na precisão, além de obter índices melhores nos quesitos: Estatística Kappa, Medida F e Curva ROC.

O gráfico 5 evidenciou que no teste 1 o melhor algoritmo foi FT com precisão de 84,4%.

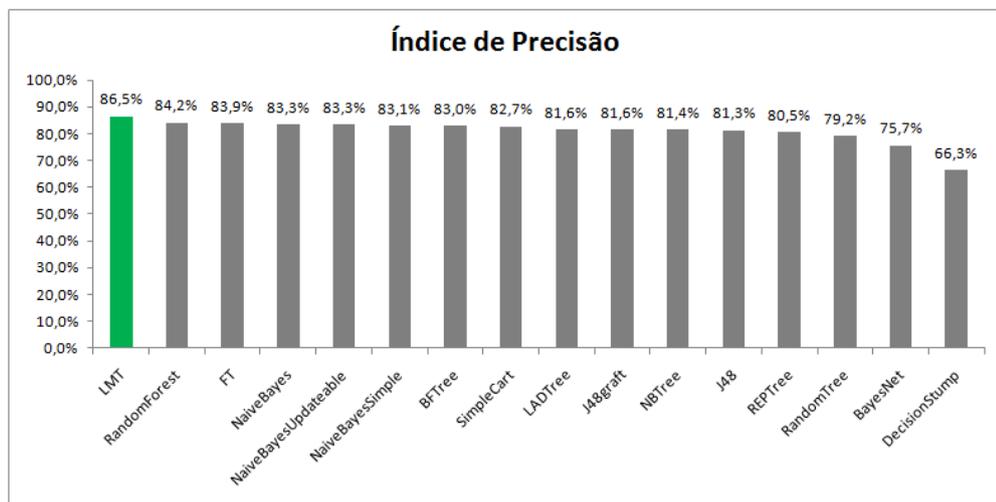
Resultados dos testes feitos com 12 algoritmos do método de árvore de decisão e 4 algoritmos do método de classificadores Bayesianos utilizando o método validação cruzada.

Tabela 06: Resultados dos testes feitos pelo método validação cruzada

Nº	Classificadores	Algoritmo	Instancias Total	Instancias Correta	Instancias Incorreta	Precisão	Estatística Kappa	Medida F	Curva ROC
1	Arvores de Decisão	LMT	310	268	42	86,5%	0,78	0,86	0,97
3	Arvores de Decisão	RandomForest	310	261	49	84,2%	0,75	0,84	0,95
2	Arvores de Decisão	FT	310	260	50	83,9%	0,74	0,84	0,93
14	Bayesianos	NaiveBayes	310	258	52	83,3%	0,73	0,83	0,95
16	Bayesianos	NaiveBayesUpdateable	310	258	52	83,3%	0,73	0,83	0,95
15	Bayesianos	NaiveBayesSimple	310	258	52	83,1%	0,73	0,83	0,95
7	Arvores de Decisão	BFTree	310	258	52	83,0%	0,73	0,83	0,89
5	Arvores de Decisão	SimpleCart	310	256	54	82,7%	0,72	0,83	0,91
11	Arvores de Decisão	LADTree	310	253	57	81,6%	0,71	0,82	0,94
10	Arvores de Decisão	J48graft	310	254	56	81,6%	0,71	0,82	0,90
4	Arvores de Decisão	NBTree	310	251	59	81,4%	0,70	0,81	0,92
9	Arvores de Decisão	J48	310	253	57	81,3%	0,70	0,81	0,90
8	Arvores de Decisão	REPTree	310	250	60	80,5%	0,69	0,81	0,91
6	Arvores de Decisão	RandomTree	310	244	66	79,2%	0,66	0,79	0,85
13	Bayesianos	BayesNet	310	232	78	75,7%	0,60	0,75	0,92
12	Arvores de Decisão	DecisionStump	310	240	70	66,3%	0,62	0,70	0,87

Na Tabela 06 observa-se que foram testadas 310 instâncias devido o método validação cruzada utilizar todas as instâncias da base, as demais colunas seguem a mesma descrição da Tabela 05.

Gráfico 06: Índice de Precisão do teste 2

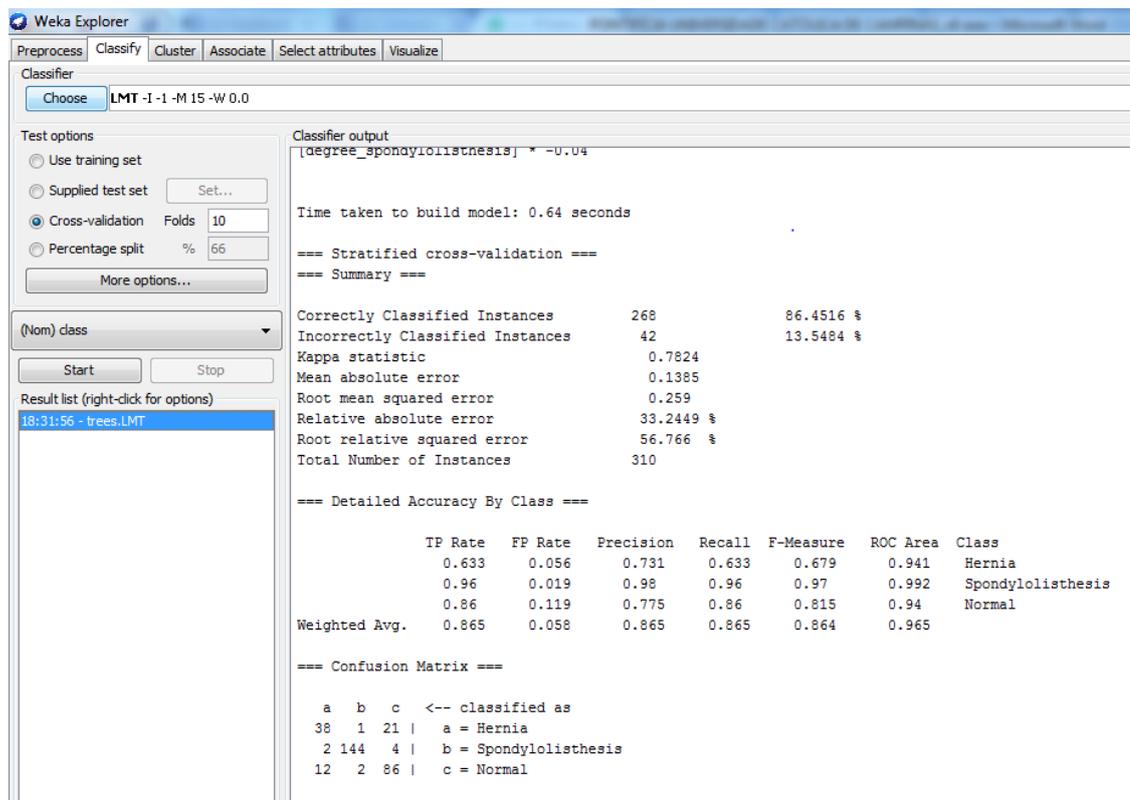


No Gráfico 06 observa-se que o algoritmo LMT (*Logistic Model Trees*) tem grande destaque obtendo um índice de 86,5%, importante salientar que no propósito deste trabalho quanto maior o índice de acerto melhor resultado estará sendo apresentado para o profissional da medicina, os índices que validam os algoritmos foram satisfatórios por apresentarem percentuais próximos ou acima de 0,80 como Estatística Kappa 0,78, Medidor F 0,86 e Curva ROC 0,97.

Comparando-se os dois classificadores FT (melhor do teste 1) e LMT (melhor do teste 2), nota-se que o LMT tem maiores índices de precisão e demais indicadores de validação, portanto, a princípio, o LMT é o melhor classificador para as atividades de classificação sobre a base de dados da coluna vertebral.

Na figura 24 temos o resultado do algoritmo LMT aplicado no teste 2.

Figura 24: Tela das regras de árvore de decisão aplicado no algoritmo LMT.



Podemos observar que a Tabela 04 evidencia que das 310 instâncias o resultado foi 268 instâncias classificadas corretamente e 42 instâncias classificadas de forma incorreta.

Ao analisamos a Figura 25 que exibe a Matriz de Confusão para o teste com o Algoritmo LMT podemos observar que:

**Hérnia de Disco:** Dos 60 casos de hérnia de disco, o modelo classificou 38 casos como hérnia de disco e dos 22 casos classificados de forma incorreta, 21 casos foram diagnosticados como normais.

**Listese:** Dos 150 casos de Listese, o modelo classificou 144 casos como Listese, 2 casos como Hernia de Disco e 4 casos como Normal.

**Normal:** Dos 100 casos como Normal, o modelo classificou 86 casos como normal, 2 casos como Listese e 12 casos como hérnia de disco.

Figura 25: Matriz de Confusão do resultado do algoritmo LMT

```

=== Confusion Matrix ===
      a    b    c  <-- classified as
    38    1   21 |   a = Hernia
      2 144    4 |   b = Spondylolisthesis
     12    2   86 |   c = Normal
  
```

O índice de precisão médio foi de 86,5% para os três diagnósticos (Normal, Listese e Hérnia de Disco), a precisão para o diagnóstico de hérnia de disco foi de 73,1%, o índice ficou mais baixo que o índice geral devido aos ângulos de alguns pacientes com hérnia de disco em alguns casos serem próximos dos ângulos de uma pessoa normal. A precisão para o diagnóstico de Listese foi de 98%, ou seja, o modelo de aprendizado construído para essa base de dados conseguiu aprender melhor a diagnosticar casos de Listese, a precisão para o diagnóstico Normal foi de 77,5%. A base pública utilizada para obtenção destes índices possui 310 instâncias, não foi encontrado na literatura qual o tamanho de quantidade de instancias necessária para que o resultado seja satisfatório. Na literatura há casos de sucesso conforme já citado no estado da arte deste trabalho Carvalho (2015); Garcia (2015); Silva (2004) quando é alcançado índices de precisão acima de 80% ao que indica que o modelo aplicado esta dentro dessa margem do resultado do estado da arte.

Como foi relatado pelo Dr. Henrique Haidar que à casos de hérnia de disco e pacientes normais terem ângulos semelhantes, ou seja, insumos semelhantes, ao analisarmos a Listese do ponto de vista do paciente ter a doença e ser diagnosticado como normal o resultado do modelo foi de 4 casos em um

total de 150 casos. Com esses resultados o modelo se mostrou bastante eficiente no auxílio do diagnóstico de Listese e para o diagnóstico Normal ou Hérnia de Disco os índices ficaram acima de 70%, índices que podem levar ao médico uma 2ª opinião auxiliando principalmente quando eles não estiverem confiantes no seu diagnóstico.

## 10.2 Resultados e análise da Interface

O desenvolvimento da interface atendendo aos requisitos apontados levou a elaboração de um visual simples e objetivo, o processo de mineração de dados é o módulo responsável por acessar a base de dados pública e aplicar sobre a base de dados a técnica de mineração de dados escolhida conforme apresentado no capítulo anterior. Cabe salientar que o processo de mineração de dados não será responsável em definir qual o algoritmo deve ser utilizado, será aplicado o algoritmo já testado e validado utilizando a ferramenta *Weka*.

O objetivo desse módulo de processamento da mineração de dados é que seja uma aplicação Java (devido ao *Weka* ser implementado em Java e disponibilizar uma *API* que permita através de uma aplicação realizar a chamada desse *plugin* sincronizando a aplicação com a ferramenta *Weka*) hospedada em um servidor contendo a base de dados. Neste trabalho o servidor será simulado por um computador pessoal e a base de dados será armazenada em local específico.

O aplicativo desenvolvido tem como função atender aos requisitos apontados e principalmente permitir ao médico ortopedista realizar um diagnóstico para as doenças Listese e Hérnia de disco através da entrada de parâmetros resultante de **Raio X panorâmico** do paciente.

Na Figura 26 observa-se a tela principal do aplicativo, o código fonte esta no apêndice A.

Figura 26 – Tela principal do aplicativo



Conforme mostrado na Figura 27, existe um menu para que o médico e o administrador do aplicativo possam manusear, tem-se as seguintes opções:

- Realizar diagnóstico:** Realizar um novo diagnóstico
- Visualizar as regras:** Visualizar as regras do diagnóstico (trabalho futuro)
- Consultas salvas:** Consultar diagnósticos realizados
- Configurações:** Definir o endereço IP do servidor
- Sobre o App:** Exibir versão do Aplicativo e licenças.

Figura 27: Menu Principal do Aplicativo



Conforme Figura 28, ao escolher realizar diagnóstico, será aberta uma tela solicitando a entrada obrigatória de 6 campos numéricos, estes parâmetros são resultantes de medição realizada no exame de Raio X panorâmico.

- Grau de incidência pélvica
- Grau de inclinação pélvica
- Ângulo de lordose lombar
- Grau de declive sacral
- Raio pélvico
- Grau de espondilolistese

Figura 28: Dados sobre a coluna vertebral do paciente

Diagnóstico VOLTAR

Angulo de lordose lombar:

\_\_\_\_\_

Grau de declive sacral:

\_\_\_\_\_

Raio pélvico:

\_\_\_\_\_

Grau de espondilolistese:

\_\_\_\_\_

DIAGNOSTICAR PACIENTE

Após o preenchimento de todos os campos como exibe a Figura 30, é necessário clicar no botão 'Diagnosticar Paciente', os dados do exemplo são o resultado de um exame de Raio X panorâmico realizado pelo autor deste trabalho portador de Hérnia de Disco.

Figura 29: Dados sobre a coluna vertebral do paciente preenchida

The figure displays two sequential screenshots of a mobile application interface for patient diagnosis. Both screens have a blue header with the title 'Diagnóstico' and a 'VOLTAR' button. The left screenshot, taken at 18:36, shows the data entry screen with the following fields and values:

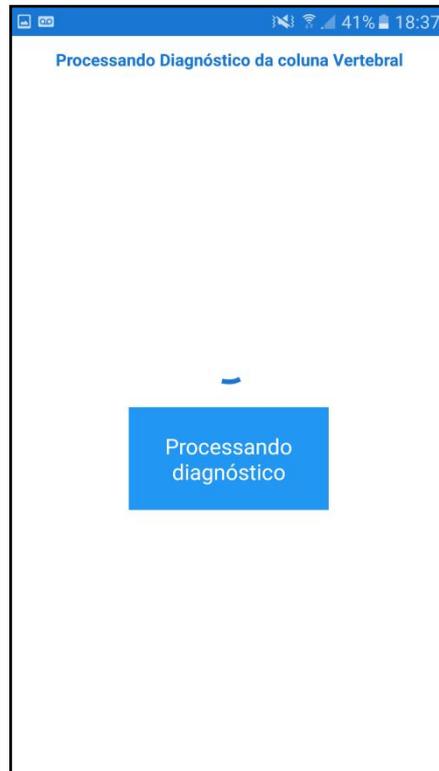
- Dados sobre a coluna vertebral do paciente
- Grau de incidência pélvica: 55
- Grau de inclinação pélvica: 24
- Angulo de lordose lombar: 33
- Grau de declive sacral: 30
- Raio pélvico:

The right screenshot, taken at 18:37, shows the same data after processing, with a 'DIAGNOSTICAR PACIENTE' button at the bottom. The values are:

- Angulo de lordose lombar: 33
- Grau de declive sacral: 30
- Raio pélvico: 116
- Grau de espondilolistese: 0

No momento em que for clicado o botão Diagnosticar Paciente, o aplicativo fará a chamada ao algoritmo LMT e nesse intervalo uma tela de processando diagnóstico (utilizará o classificador gerado pelo algoritmo LMT) ficará sendo exibida ao médico conforme Figura 30. Esse retorno demora menos de 5 segundos.

Figura 30: Tela de processando diagnóstico

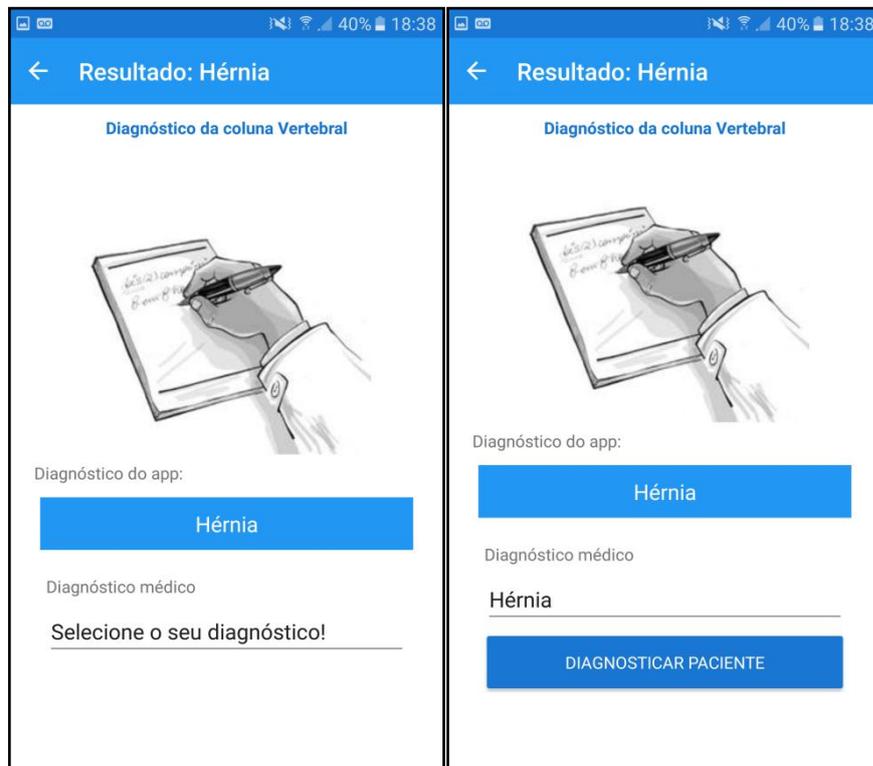


Após o processamento, o retorno será o diagnóstico do paciente podendo ser:

- Normal
- Espondilolistese
- Hérnia

Observa-se pela Figura 31 que o diagnóstico é do aplicativo, para registrar o diagnóstico o médico deverá escolher o seu diagnóstico, podendo confirmar o diagnóstico do aplicativo ou alterar seu diagnóstico. Seja qual for a decisão do médico o aplicativo deixará esse registro armazenado e a qualquer momento será possível consultar qual o índice de acerto do aplicativo baseado na quantidade de diagnósticos do aplicativo similares ao diagnóstico do médico. Com esse resultado será possível aumentar a credibilidade do uso do aplicativo.

Figura 31: Resultado do diagnóstico



No menu principal, ao escolher a opção consultas salvas, será possível resgatar diagnósticos realizados, do mais recente para o mais antigo como mostra a Figura 32.

Observa-se que é possível no cabeçalho desta tela observar quantos diagnósticos já foram realizados e o índice de assertividade baseado na confirmação do médico. No caso da figura 32, tivemos 10 simulações com dados reais de pacientes com hérnia de disco e o aplicativo acertou 70% dos casos, índice coerente com a precisão apresentada pela ferramenta *WEKA*.

Figura 32: Tela de consultas salvas

Consultas		VOLTAR
Total de consultas: 10 Total de acertos: 70%		
Hérnia	11/05/2016 18:59:31	
Hérnia	11/05/2016 18:58:50	
Hérnia	11/05/2016 18:57:57	
Hérnia	11/05/2016 18:56:55	
Hérnia	11/05/2016 18:56:09	
Hérnia	11/05/2016 18:55:26	
Hérnia	11/05/2016 18:54:33	
Hérnia	11/05/2016 18:50:52	
Hérnia	11/05/2016 18:49:54	
Hérnia	11/05/2016 18:49:07	

Após 30 simulações de diagnóstico (10% da base) sendo 10 registros como Normal, 10 registros como Hérnia e 10 registros como Listese, veja pela Figura 33 que o aplicativo está com um índice de 83,33% de acerto, valor próximo do índice informado pelo *WEKA* de 86,5% de precisão. Ao estratificar estes índices o diagnóstico Normal obteve um índice de precisão de 77% *versus* 80% do índice do aplicativo. O diagnóstico Hérnia obteve índice de 73% de precisão *versus* 70% do índice do aplicativo e por fim o diagnóstico para Listese está com índice de precisão em 98% no *WEKA* e obtemos 100% no aplicativo.

Figura 33: Tela de Consultas Salvas exibindo a *performance*

## 11 CONCLUSÃO

A utilização de alguma metodologia da área de aprendizagem de máquina na contribuição para a medicina mostrou-se relevante. Isso por que a rotina atual praticada nos hospitais, clínicas, consultórios geram uma enorme massa de dados que são armazenados em banco de dados, podendo-se utilizar de técnicas de mineração de dados que é um método fundamental para extrair conhecimento que estão escondidos nesses bancos de dados.

As pesquisas na área de aprendizagem de máquina são inúmeras, a comunidade acadêmica tenta descobrir soluções em busca de agregar valor aos sistemas atuais. Pelas pesquisas de trabalhos semelhantes e com o resultado do atual trabalho podemos concluir que as máquinas têm um grande poder de aprendizado contribuindo para tomada de decisões.

A motivação principal deste trabalho foi à criação de um aplicativo para entrada de dados sobre a coluna vertebral de pacientes através de Raio X panorâmico, aplicando técnicas de mineração de dados dessa entrada e disponibilização de diagnóstico prévio que sirva como uma 2<sup>o</sup> opinião do médico ortopedista, com isso contribuindo em melhores diagnósticos nas doenças de Listese e Hérnia de disco.

Pela análise de diversos tipos de métodos de mineração de dados, conclui-se que o mais apropriado para extração de conhecimento sobre bases selecionadas são as de classificação utilizada nesse trabalho devido ser uma base de dados de estágio supervisionado ou seja, as bases já estão prontas (com diagnósticos) para serem utilizadas em qualquer ferramenta de mineração de dados, pois com essa técnica é permitido aprender a diagnosticar doenças utilizando um modelo de aprendizagem atualizado com a entrada de diagnóstico de novos pacientes.

A partir dos resultados do Capítulo 9, conclui-se que o melhor classificador de árvore de decisão foi o algoritmo LMT, por apresentar índices de precisão, estatística Kappa, medidor F, curva ROC satisfatórios de acordo com a literatura. Este último indicador curva ROC apresentou índice no valor de 0,96 sendo um índice excelente e sendo um indicador muito utilizado na área médica.

Se compararmos os índices de precisão do algoritmo LMT na ferramenta *WEKA* versus a simulação com 30 casos reais no aplicativo obtivemos índices bem próximos, no geral a ferramenta *WEKA* gera uma precisão de 86,5% e o resultado do aplicativo foi de 83,3% de precisão.

Ao estratificar por tema temos 73% de precisão para diagnóstico de hérnia de disco, onde o aplicativo acertou 70%, temos 98% de precisão para diagnóstico de listese, onde o aplicativo acertou 100%, temos 77% de precisão para diagnóstico normal, onde o aplicativo acertou 80%, os casos de diagnóstico normal que o aplicativo errou foram diagnosticados como hérnia que segundo o Dr. Henrique professor da PUC Campinas informou que é devido o resultado dos exames de pacientes normais e com hérnia de disco serem similares.

Por esse comparativo pode-se concluir que o modelo de aprendizado construído obteve um desempenho melhor nos casos de listese, onde 2 fatores podem ter contribuído, a quantidade de pacientes com listese é maior nessa base de dados e os ângulos do Raio X panorâmico de quem tem listese se destacam com relação aos pacientes com hérnia e normais.

Por fim o objetivo de utilização de técnicas de mineração de dados sobre a base de dados publica a fim de obter novos conhecimentos foi realizado com êxito. A base de dados continuara com os mesmos atributos mantendo o sigilo do paciente sendo assim poderá ser utilizada para futuras pesquisas por outros profissionais de tecnologia ou da área da saúde.

Foi possível a construção de um aplicativo mobile capaz de auxiliar a elaboração de diagnósticos sobre doenças de listese e hérnia de disco a ser executada em dispositivos *Android*, com isso conclui-se que a utilização da *API* fornecida pelo *WEKA* para que fossem aplicados os algoritmos no módulo de processamento de mineração de dados em Java atenderam as expectativas propostas para este trabalho.

O grande foco deste trabalho reside na confiabilidade dos seus resultados, o aplicativo foi capaz de diagnosticar poucos pacientes de forma equivocada, das 30 simulações, 25 foram diagnosticadas com êxito e 5 diagnósticos não foram satisfatórios, sendo que destes 5 casos apenas 2 deveria ter um resultado como normal e foi avaliado como hérnia de disco, conclui-se que

a plataforma se torne mais eficiente e eficaz à medida que se amplie com a entrada de novos pacientes. O aplicativo propõe ser uma 2ª opinião ao médico e jamais tomar a decisão final e substituir a decisão do médico.

### 11.1 Contribuições

Esse trabalho tem por fim auxiliar as demandas da área da saúde, fazendo com que se tenha um serviço prestado eficiente e que se possa utilizar os dados que estão presentes em banco de dados para transformar estes dados em informações úteis, ou seja, que configurem um serviço a ser prestado por meio de toda uma rede de comunicação já existente no país.

Partindo desse contexto, tem-se que a Tecnologia da Informação através de metodologia de mineração de dados possa contribuir para a relação do homem com as Tecnologias da Informação e da Comunicação (TIC's) de forma positiva, realizando sinergia entre novas descobertas de conhecimento e infraestrutura de redes já existentes em hospitais, postos de saúde e afins, com isso encontrando soluções digitais em diversos setores inclusive no setor médico agregando valor na medicina, contribuindo para que a saúde seja um dos quesitos eficiente dentro do conceito de Cidades Inteligentes.

Este estudo permitiu verificar que com o uso de *softwares open-source* fosse possível a criação de um aplicativo para auxílio do médico ortopedista como uma 2ª opinião.

Essa proposta pode ser ampliada, e em estudos posteriores poderiam levá-la tanto para outras doenças ortopédicas, como para outras especialidades médicas.

Demonstrou-se um estudo que possa ser explorado na comunidade acadêmica tanto por profissionais de TIC's quanto profissionais da saúde, principalmente os residentes que estão em fase de descobertas de conhecimento.

## 11.2 Trabalhos Futuros

Como ações futuras, planeja-se a substituição do servidor de banco de dados de local (notebook) para um servidor WEB para que o aplicativo possa ser utilizado de forma mundial.

Deseja-se aprimorar o aplicativo, implementando a visualização das regras em forma de árvore para que o médico possa visualizar as condições que levaram para determinado diagnóstico, da forma que foi construído este aplicativo o aprendizado ocorre a cada novo diagnóstico sendo importante a realização de estudos que possibilitem efetuar a atualização automática do mapa de visualização da árvore.

Agregar novas doenças e novas especialidades a fim de aumentar a cobertura de diagnósticos para a comunidade médica.

Atualmente no mercado já existem aplicativos que usam da infraestrutura de redes de telecomunicações para benefício a comunidade, a exemplo temos o aplicativo Dr. Uber da medicina. Visualiza-se aqui que com esta plataforma é possível chamar o médico para uma visita onde quer que o paciente esteja permitindo que mães com filhos pequenos, idosos, pessoas que necessitam de cuidados especiais ou que possuem dificuldades de locomoção e até mesmo quem não deseja enfrentar filas de hospitais e consultórios possam contar com o serviço personalizado.

Nessa plataforma existe um módulo de sintomas a ser enviado ao médico que irá prestar o atendimento, com a sinergia com o aplicativo Dr 2º Opinião seria possível o médico realizar a visita com um pré-diagnóstico, o autor desse trabalho esteve em contato com o CEO (*Chief Executive Officer*) do Dr. Uber da medicina que se mostrou entusiasmado com uma futura parceria.

## 12 REFERÊNCIAS

ALFIERI, F.M.; LUTTING, F.T; PIERGENTILE, B. *Comparação da flexibilidade, intensidade da dor e funcionalidade de portadores de hérnia discal submetidos à hidrocinésioterapia versus cinesioterapia clássica*, 2015; Journal life style.

ARUP. *Smart Cities: Transforming the 21st century city via the creative use of technology* Disponível em <[http://publications.arup.com/publications/s/smart\\_cities](http://publications.arup.com/publications/s/smart_cities)>. Acesso em: 31 jan.2016

BERTOLINI, A.V.A.G; CHIAPPIN, M.A.; MAYOLO, V.R; D'ARRIGO, F.P; BARCELLOS, P.F.P; DIAS, D.T.A. *Soluções Business Intelligence open source no suporte à estratégia organizacional*. 2015. Revista Inteligência Competitiva. ISSB: 2236-210x.

BUCKLE, P. W.; DEVEREUX, J. J. *The nature of work-related neck and upper limb musculoskeletal disorders*. Appl. Ergon., Oxford, v. 33, n.3, p. 207-217, 2002.

BRINK, H.; RICHARDS, J. *Real World Machine Learning*. [S.l.]: Manning Publications C.O, 2014.

CABRAL, P. A. *O uso do data mining na descoberta de marcas indentitárias do Piauí: Um estudo de caso da Rede Social TWITTER*, Dissertação (Mestrado em Comunicação) – Universidade Federal do Piauí. 2015.

CAMARA, F.P. *Psiquiatria e estatística V: Validação de procedimentos diagnóstica pela curva ROC*. 2009. Disponível em <<http://www.polbr.med.br/ano09/cpc0409.php>>. Acesso em: 31 jan.2016.

CARVALHO, L. A. V. –*Datamining: a mineração de dados no marketing, medicina, economia, engenharia e administração* - Ciência Moderna – RJ. 2005.

CARVALHO, M; VIEIRA, A. A. *Erro médico em pacientes hospitalizados*. Jornal de Pediatria. 2002. Vol. 78, nº 4

CARVALHO, D.R; MOSER, A.D; SILVA, V.A; DALLAGASSA, M.R., *Mineração de Dados Aplicado à Fisioterapia*. 2012. Scielo ISSN 1980-5918, vol. 25 nº 3

CARVALHO, D.R; MOSER, A.D; SILVA, V.A; DALLAGASSA, M.R, *Uso de Técnicas de mineração de dados para a identificação automática de beneficiários propensos ao diabetes mellitus tipo 2*. 2015. Informação & Informação. ISSN: 1981-8920. Vol. 20 nº3

CASTELLS, M. *The Network Society: A Cross-Cultural Perspective*. 2004. Editora Amazon.

CONSELHO REGIONAL DE MEDICINA DO ESTADO DE SÃO PAULO – CREMESP. Disponível em

<<http://www.cremesp.org.br/?siteAcao=NoticiasC&id=3582>>Acesso em: 18 mai.2016.

CRUZ, A. J. R. *Data Mining via Redes Neurais Artificiais e Máquinas de Vectores de Suporte*. 2007. 123 f. Dissertação (Mestrado em Sistemas de Informação) – Universidade do Minho, Lisboa, 2007.

DASU, T.; JOHNSON, T. *Exploratory data mining and data cleaning*. New Jersey: Wiley Publishing, 2003.

DI MAIO, A. *Ten Questions on Smart Government : What It Is , What It Is Not , Why It Is Important*. 2011. Gartner, ID: G00213208

FABREGA, J.; SAJURIA, J. (2013). *The emergence of political discourse on digital networks: the case of the occupy movement*. In COINs13.

FALAVIGNA, A.; et al. *Qual a relevância dos sinais e sintomas no prognóstico de pacientes com hérnia de disco lombar*. Revista COLUNA/COLUMNA, p.186-192, set. 2010.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. *From Data Mining to Knowledge Discovery in Databases*. 1996. AI magazine Vol. 17 nº 3.

FREITAS, A. A. *Uma Introdução a Data Mining. Informática Brasileira em Análise*. CESAR - Centro de Estudos e Sistemas Avançados do Recife. Ano II, n. 32, mai./jun. 2000. Instituto privado de Inovação.

GABRIEL, M.R.S. *Fisioterapia em traumatologia, ortopedia e reumatologia*. Ed. Revinter, PP, 2001. ISBN 8573094842, 9788573094848, 402 páginas.

GARCIA, M. C. M.; MARTINS, E. T.; AZEVEDO, F. M. *Agrupamento Fuzzy e Regressão Logística Aplicados na Análise de Traumatismo Cranioencefálico Grave*. 2013, In: Conferencia Iberoamericana de computação aplicada 2013/International Association for development of the information Society (IADIS), Porto Alegre/RS.

GOLDSCHIMIDT, R; PASSOS, E. *Data Mining: Um guia prático*. Rio de Janeiro: Campus, 2005

GOODSPEED, R. (2014). *Smart cities: moving beyond urban cybernetics to tackle wicked problems*. Cambridge Journal of Regions, Economy and Society. doi:10.1093/cjres/rsu013

HAND, D.; MANNILA, H.; SMYTH, P. *Principles of data mining* Cambridge: MIT Press, 2001.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Perfil dos municípios brasileiros 2013. Rio de Janeiro, 2014. Disponível em: <[ftp://ftp.ibge.gov.br/Perfil\\_Municipios/2013/munic2013.pdf](ftp://ftp.ibge.gov.br/Perfil_Municipios/2013/munic2013.pdf)>. Acesso em: 4 dez. 2015.

LOPES, L. Aprendizagem de máquina baseada na combinação de classificadores em bases de dados da área de saúde. 121 f. Dissertação (Mestrado em tecnologia da Saúde) – PUC Parana, Curitiba, 2007.

MARRAS, W. S. et al. *National occupational research agenda (NORA) future directions in occupational musculoskeletal disorder health research*. Appl. Ergon., Oxford, v. 40, n. 1, p. 15–22, 2009.

MALAVOLTA, E.A; et al. *Análise crítica das publicações científicas da Revista Brasileira de Ortopedia no período de 2006 a 2010*. 2012

NATIONAL ACADEMIES ,2015,

Disponível em: <<http://www.nationalacademies.org/hmd/Reports/2015/Improving-Diagnosis-in-Healthcare.aspx>>, Acesso em: 20 mai.2016

OLSON, D. L; DELEN, D. *Advanced Data Mining Techniques*. Springer, 2008.

OTHMAN, M.F.B; YAU, T.M.S. *Comparison of Different Classification Techniques Using WEKA for Breast Cancer*. 3rd Kuala Lumpur International Conference on Biomedical Engineering, vol. 15, pp. 520-523, 2007

ORGANIZAÇÃO DAS NAÇÕES UNIDAS – ONU.(2015), <<http://esa.un.org/unpd/wpp/Download/Standard/Population/>>, Acesso em: 15 mai.2016

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Rio+20: O futuro que queremos. Fatos sobre as cidades. Rio de Janeiro, 2012. Disponível em:<<http://www.onu.org.br/rio20/cidades.pdf>> Acesso em: 4 dez. 2015.

Park, CA, USA: *American Association for Artificial Intelligence*, 1996. Cap. *From Data Mining to Knowledge Discovery: An Overview*, p. 1–34. ISBN 0-262-56097-6. Disponível em: <[http://web.cacs.louisiana.edu/~cmeps566/References/R4\\_1\\_34.pdf](http://web.cacs.louisiana.edu/~cmeps566/References/R4_1_34.pdf)>, Acesso em: 08 mai.2016

PEREZ J.;PONN C.C.Y.; MERRIFIELD R.D.; WONG S.T.C.; Yang G.; *Big Data for Health*, 2015, IEEE Journal of biomedical and health Informatics, Vol. 19, n. 04

PERIZZOLO, C. B. *Sistema de identificação espacial em ambientes de consumo utilizando mineração de dados*, Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Santa Catarina, Florianópolis, 2005.

PESQUISA NACIONAL DE SAÚDE – PNS (2013). Disponível em: <<ftp://ftp.ibge.gov.br/PNS/2013/pns2013.pdf>>, Acesso em: 15 mai.2016.

REDE BRASIL ATUAL(2013) Disponível em: <<http://www.redebrasilatual.com.br/saude/2013/04/ministerio-da-saude-anuncia-acoes-contra-erros-medicos-1>>, Acesso em: 15 mai.2016.

RUTE - REDE UNIVERSITARIA DE TELEMEDICINA (2015) Disponível em <<http://rute.rnp.br/>>, Acesso em: 11 mai.2016

ROCHA, V.H; BARANAUSKAS, C.C. *Design e Avaliação de interfaces Humano-Computador*. Campinas: NIED/Unicamp, 2003.

RUIZ, R.S.R; VELHO, H.F.C; SANTOS, R.D.C, *Árvores de decisão em classificação de dados astronômicos, 2009*, Laboratório de Computação e Matemática Aplicada, LAC, INPE 12245-970-000, São José dos Campos, SP

SANTOS, M.Y. ; RAMOS, R. *Business Intelligence – Tecnologias da Informação na Gestão de Conhecimento*, FCA, 2a Edição, 2009.

SCHEUNEMANN, F., PRETO, F. *Mineração de dados para descoberta de conhecimento na área de oncologia, 2015*. Revista destaques acadêmicos, vol.7 n.4 CETEC/UNIVATES.

HENNEMAN, S.A.; et AL. *Incidência Pélvica: um parâmetro fundamental para definição do equilíbrio sagital da coluna vertebral*. 2012, Scielo Vol. 11 n. 3

SRINIVA S, K.; RANI, B.K.; GOVRDHAN, A; *Applications of data mining techniques in healthcare and prediction of heart attacks, 2010*. (IJCSSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255

SILVA, G.C.; *Mineração de Regras de Associação Aplicada a Dados da Secretaria Municipal de Saúde de Londrina - PR, 2004*, Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

SINDICATO MÉDICO DO RIO GRANDE DO SUL – SIMERS (2015). Disponível em: <<http://www.simers.org.br/2015/11/processos-por-alegado-erro-medico-crescem-140-no-stj-saiba-como-se-proteger/>>, Acesso em: 18mai.2016.

SISTEMA NACIONAL DE AUDITORIA Disponível em: <<http://sna.saude.gov.br/noticias.cfm?id=5012>>, Acesso em: 20 mai.2016

SOCIEDADE BRASILEIRA DE DIREITO MÉDICO E BIOÉTICA – ANADEM, Disponível em: <<http://anadem.org.br/decisoes-sobre-erro-medico-no-stj/>>, Acesso em: 20 mai.2016

SOUZA, C; *Análise de poder discriminativo através de curvas roc, 2009* Disponível em: <<http://crsouza.com/2009/07/analise-de-poder-discriminativo-atraves-de-curvas-roc/>>, Acesso em: 20 mai.2016

TANIAR, D. *Data mining and knowledge discovery technologies*. New York: IGI Global, 2008.

TAN, P-N; STEINBACH, M.; e KUMAR, V. - *Introdução ao Data Mining - Mineração de Dados*. – Ciência Moderna Ltda - RJ, 2009.

TANG, Z. H.; MACLENNAN, J. *Data mining with SQL Server 2005*. Indianapolis: Wiley, 2005. Editora Wiley.

TEBET, M.A. *Conceitos atuais sobre equilíbrio sagital e classificação da espondilolise e espondilolistese*, 2014, Revista Brasileira de Ortopedia, Editora Elsevier.

TIGHE, P; LADUZENSKI, S; EDWARDS, D; ELLIS, N; BOEZAART, AP; AYGUG, H. “*Use of machine learning theory to predict the need for femoral nerve block following ACL repair*”, 2011, US National Library of Medicine – National Institutes of Health.

VICECONTU M.; HUNTER P.;HOSE R.; *Big Data, Big Knowledge: Big Data for Personalized Healthcare*, 2015, IEEE J Biomed Health Inform.

VIEIRA, J.A.P.; *Algorithm development for physiological signals analysis and cardiovascular disease diagnosis – A data mining approach*, Dissertação (Mestrado em Ciência da Computação) – Universidade de Coimbra Faculdade de Ciências e Tecnologia, Coimbra, 2011.

WASHBURN, D., SINDHU, U. (2009). *Helping CIOs Understand “Smart City” Initiatives*. Disponível em: [http://www-935.ibm.com/services/us/cio/pdf/forrester\\_help\\_cios\\_smart\\_city.pdf](http://www-935.ibm.com/services/us/cio/pdf/forrester_help_cios_smart_city.pdf), Acesso em: 28 mar.2016.

WEISS M.C., BERNARDES R,C; CONSONI F. *Cidades inteligentes como nova prática para o gerenciamento dos serviços e infraestruturas urbanos: a experiência da cidade de Porto Alegre*, 2014, Scielo, *Brazilian Journal of Urban Management* , SSN 2175-3369.

ZHANG, W.; WILLIAMS, M; PENGETNZE, Y. *Predicting Asthma-Related Emergency Department Visits Using Big Data*, 2015, IEEE J Biomed Health Inform. Vol. 19 n.4.

ZHOU, Z. *Three perspectives of data mining*. Artif. Intell., Elsevier Science Publishers Ltd., Essex, UK, v. 143, n. 1, p. 139–146, jan. 2003. ISSN 0004-3702.

WITTEN, I. H.; FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. 2nd. ed. San Francisco, CA: Morgan Kaufmann, 2005. ISBN 0-12-088407-0.

## APÊNDICE A – Script para o Aplicativo Dr. 2º Opinião

Código deste Apêndice foi desenvolvido em Java SE e Xamarin SDK.

Inicialização para utilização do Xamarin

### Home

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Home : ContentPage
    {
        public Home()
        {
            InitializeComponent();
        }
    }
}
```

Neste código é configurado o algoritmo a ser processado no aplicativo como padrão esta o LMT algoritmo escolhido neste trabalho

### Configurações

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Home : ContentPage
    {
        public Home()
        {
            InitializeComponent();
        }
    }
}using Newtonsoft.Json;
using System;
```

```

using System.Collections.Generic;
using System.Linq;
using System.Net.Http;
using System.Text;
using System.Text.RegularExpressions;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Configuracoes : ContentPage
    {
        public Configuracoes()
        {
            InitializeComponent();

            String[] itens = { "LMT", "J48", "HoeffdingTree", "RandomFlorest" };

            int i = 0;
            foreach (String item in itens)
            {

                picker.Items.Add(item);

                if (item.Equals(Application.Current.Properties["Algoritmo"].ToString()))
                    picker.SelectedIndex = i;

                i++;
            }

            picker.SelectedIndexChanged += async (e, a) =>
            {
                try
                {
                    Application.Current.Properties["Algoritmo"] =
picker.Items[picker.SelectedIndex];
                    await Application.Current.SavePropertiesAsync();
                    await DisplayAlert("Sucesso" , "Algoritmo alterado com sucesso!",
"Fechar");
                }
                catch (Exception)
                {
                    await DisplayAlert("Erro", "Erro ao alterar o algoritmo\nPor favor,
tente novamente!", "Fechar");
                }
            }
        }
    }
}

```

```

};

ip.Text = Application.Current.Properties["ipServidor"].ToString();

ip.TextChanged += (e, a) =>
{
    if (a.NewTextValue.Length > 0)
    {
        String novaString = a.NewTextValue.Replace(",", ".");
        novaString = Regex.Replace(novaString, @"[^0-9\.]+", String.Empty);
        var splits = novaString.Split('.');
        if (splits.Count() > 0 && splits.Count() < 5)
        {
            novaString = String.Empty;

            foreach (string ips in splits)
            {
                if (ips.Length <= 3)
                {
                    if (novaString.Equals(String.Empty))
                        novaString += ips;
                    else
                        novaString += "." + ips;
                }
                else
                {
                    if (novaString.Equals(String.Empty))
                        novaString += ips.Substring(0, 3);
                    else
                        novaString += "." + ips.Substring(0, 3);
                }
            }
        }
        else
        {
            novaString = "" + splits[0] + "." + splits[1] + "." + splits[2] + "." +
splits[3];
        }

        ip.Text = novaString;
    }
};
ipsss.Clicked += salvaIP;
}

```

```

private async void salvaIP(object sender, EventArgs e)
{
    if (ip.Text.Length > 0)
    {

        try
        {

var url = "http://" + ip.Text + ":8099/";

HttpClient http = new HttpClient();

            http.DefaultRequestHeaders.Add("user-agent", "Mozilla/5.0
(compatible; MSIE 10.0; Windows NT 6.2; WOW64; Trident/6.0)");

            http.BaseAddress = new Uri(url);

HttpResponseMessage retorno = await http.GetAsync("");

            retorno.EnsureSuccessStatusCode();

            var resposta = await retorno.Content.ReadAsStringAsync();

var res = JsonConvert.DeserializeObject<Dictionary<String,
String>>(resposta.Trim());

            if (res["servidor"].Equals("ok"))
            {
                Application.Current.Properties["ipServidor"] = ip.Text;
await Application.Current.SavePropertiesAsync();
                await DisplayAlert("Sucesso", "IP Registrado com
sucesso", "Fechar");
            } else
            {
                await DisplayAlert("Erro", "Não foi possível localizar o servidor
neste IP\nPor favor, verifique o endereço IP", "Fechar");
            }

        }
        catch (Exception)
        {
            await DisplayAlert("Erro", "Não foi possível localizar o servidor neste
IP\nPor favor, verifique o endereço IP", "Fechar");
        }
    }
}

```

```

    }
  }
}

```

Neste código é possível consultar diagnósticos já realizados.

### Consultas

```

using DataMiningApp.Models;
using Newtonsoft.Json;
using Plugin.DeviceInfo;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Net.Http;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Consultas : ContentPage
    {
        public Consultas()
        {
            InitializeComponent();

            if (!Application.Current.Properties.ContainsKey("ipServidor") ||
                Application.Current.Properties["ipServidor"].ToString().Equals(String.Empty))
            {
                Device.BeginInvokeOnMainThread(async () =>
                {
                    Boolean escolha = await DisplayAlert("Atenção", "Ainda não foi
                    definido o endereço do servidor!\n\nDeseja ir em configurações e acrescentar o
                    endereço do servidor?", "Sim, eu quero!", "Não, obrigado!");
                    await this.Navigation.PopModalAsync();
                    if (escolha)
                    {
                        var tool = new ToolbarItem { Text = "Voltar" };

                        tool.Clicked += async (e, a) =>
                        {
                            await App.m.Navigation.PopModalAsync();
                        };

                        Configuracoes pgc = new Configuracoes();
                        pgc.ToolbarItems.Add(tool);
                        var pgNavc = new NavigationPage(pgc);
                    }
                });
            }
        }
    }
}

```

```

        await App.m.Navigation.PushModalAsync(pgNavc);
    }
    });
}

    inicializa();
}

private async void inicializa()
{
    List<WekaMOD> listaResultado = new List<WekaMOD>();

    try
    {
        var url = "http://" + Application.Current.Properties["ipServidor"].ToString()
+ ":8099";
        HttpClient http = new HttpClient();

        http.DefaultRequestHeaders.Add("user-agent", "Mozilla/5.0 (compatible;
MSIE 10.0; Windows NT 6.2; WOW64; Trident/6.0)");
        url += "/WekaConsulta?client=" + CrossDeviceInfo.Current.Id;
        http.BaseAddress = new Uri(url);
        HttpResponseMessage retorno = await http.GetAsync("");
        retorno.EnsureSuccessStatusCode();
        var resposta = await retorno.Content.ReadAsStringAsync();

        listaResultado =
        JsonConvert.DeserializeObject<List<WekaMOD>>(resposta);
        if (listaResultado.Count > 0)
        {
            List<WekaMOD> lista = new List<WekaMOD>();
            int acertos = 0;
            foreach (var weka in listaResultado)
            {
                if (weka.classe.Equals(weka.doapp))
                {
                    acertos++;
                }
                weka.classe = (weka.classe.Equals("Hernia")) ? "Hérnia" :
(weka.classe.Equals("Spondylolisthesis")) ? "Espondilolistese" : "Normal";
                lista.Add(weka);
            }
        }
    }
}

```

```

        resultado.Children.Clear();

        resultado.BackgroundColor = Color.FromHex("#2196F3");

        Label label1 = new Label();

        label1.HorizontalTextAlignment = TextAlignment.Center;

        label1.TextColor = Color.White;

        label1.FontSize = Device.GetNamedSize(NamedSize.Small,
typeof(Label));

        label1.Text = String.Format("Total de consultas: {0}\nTotal de acertos:
{1}%", lista.Count, (float)(acertos*100)/lista.Count);

        resultado.Children.Add(label1);

        listar.ItemsSource = lista;

    }

}
catch (Exception)
{

}

}

private void Editar(object sender, EventArgs e)
{

    var x = ((MenuItem)sender).BindingContext as WekaMOD;

    var tool = new ToolbarItem { Text = "Voltar" };
    tool.Clicked += voltar;
    EditarResultado pg = new EditarResultado(x);
    pg.ToolbarItems.Add(tool);
    var pgNav = new NavigationPage(pg);

```

```

pgNav.Disappearing += (xxx, xx) => {
    listar.ItemsSource = null;
    inicializa();
};

App.m.Navigation.PushModalAsync(pgNav);
}

private async void Apagar(object sender, EventArgs e)
{
    try
    {
        var x = ((MenuItem)sender).BindingContext as WekaMOD;

        Dictionary<String, String> b = new Dictionary<string, string>();
        b["id"] = CrossDeviceInfo.Current.Id.ToString();
        b["idm"] = "" + x.id;

        var url = "http://" + Application.Current.Properties["ipServidor"].ToString()
+ ":8099/WekaConsulta";

        HttpClient http = new HttpClient();

        http.DefaultRequestHeaders.Add("user-agent", "Mozilla/5.0 (compatible;
MSIE 10.0; Windows NT 6.2; WOW64; Trident/6.0)");

        HttpContent formulario = new FormUrlEncodedContent(b);

        HttpResponseMessage retorno = await http.PostAsync(new Uri(url), formulario);

        retorno.EnsureSuccessStatusCode();

        var resposta = await retorno.Content.ReadAsStringAsync();

        var res = JsonConvert.DeserializeObject<Dictionary<String, String>>(resposta);

        if (res["status"].Equals("1"))
        {
            await DisplayAlert("Sucesso!", "Item deletado!", "Fechar");

            var i = listar.ItemsSource as List<WekaMOD>;

            listar.ItemsSource = null;

            inicializa();
        }
    }
}

```

```

    }
    else
    {
        await DisplayAlert("Erro!", "Não foi possível apagar os dados, por favor tente novamente!", "Fechar");
    }

}
catch (Exception)
{
    await DisplayAlert("Erro!", "Não foi possível apagar os dados, por favor tente novamente!", "Fechar");
}

}

private void abrir(object sender, ItemTappedEventArgs e)
{

    var tool = new ToolbarItem { Text = "Voltar" };
    tool.Clicked += voltar;
    MostrarResultado pg = new MostrarResultado(e.Item as WekaMOD);
    pg.ToolbarItems.Add(tool);
    var pgNav = new NavigationPage(pg);
    App.m.Navigation.PushModalAsync(pgNav);
}

private void voltar(object sender, EventArgs e)
{
    App.m.Navigation.PopModalAsync();
}
}
}
}

```

Neste código é realizado o diagnóstico médico

### **Diagnóstico**

```

using DataMiningApp.Models;
using Newtonsoft.Json;
using System;
using System.Collections.Generic;

```

```

using System.Linq;
using System.Net.Http;
using System.Text;
using System.Threading;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Diagnostico : ContentPage
    {
        private List<String> resultado;

        private bool naoestalendo;
        public Diagnostico()
        {

            InitializeComponent();

            naoestalendo = true;

            limparCampos();

            if (!Application.Current.Properties.ContainsKey("ipServidor") ||
                Application.Current.Properties["ipServidor"].ToString().Equals(String.Empty))
            {
                Device.BeginInvokeOnMainThread(async () =>
                {
                    Boolean escolha = await DisplayAlert("Atenção", "Ainda não foi
                    definido o endereço do servidor!\n\nDeseja ir em configurações e acrescentar o
                    endereço do servidor?", "Sim, eu quero!", "Não, obrigado!");
                    await this.Navigation.PopModalAsync();
                    if (escolha)
                    {
                        var tool = new ToolbarItem { Text = "Voltar" };

                        tool.Clicked += async (e, a) =>
                        {
                            await App.m.Navigation.PopModalAsync();
                        };

                        Configuracoes pgc = new Configuracoes();
                        pgc.ToolbarItems.Add(tool);
                        var pgNavc = new NavigationPage(pgc);
                        await App.m.Navigation.PushModalAsync(pgNavc);
                    }
                });
            }
        }
    }
}

```

```

    }
}

private static Dictionary<String, String> dadosParaEnvio = new
Dictionary<string, string>();

private async void diagnosticarFuncao(Object o, EventArgs a)
{
    Boolean erro = false;

    if (graudeincidenciapelvica.Text.Length > 0)
    {
        dadosParaEnvio["pelvic_incidence"] =
graudeincidenciapelvica.Text.Replace('.', ',');
    }
    else
    {
        erro = true;
    }

    if (graudeinclinacaopelvica.Text.Length > 0)
    {
        dadosParaEnvio["pelvic_tilt"] = graudeinclinacaopelvica.Text.Replace('.',
',');
    }
    else
    {
        erro = true;
    }

    if (angulodelordoselombar.Text.Length > 0)
    {
        dadosParaEnvio["lumbar_lordosis_angle"] =
angulodelordoselombar.Text.Replace('.', ',');
    }
    else
    {
        erro = true;
    }

    if (graudedecivesacral.Text.Length > 0)
    {
        dadosParaEnvio["sacral_slope"] = graudedecivesacral.Text.Replace('.',
',');
    }
    else

```

```

    {
        erro = true;
    }

    if (raiopelvico.Text.Length > 0)
    {
        dadosParaEnvio["pelvic_radius"] = raiopelvico.Text.Replace('.', ',');
    }
    else
    {
        erro = true;
    }

    if (graudeespondilolistese.Text.Length > 0)
    {
        dadosParaEnvio["degree_spondylolisthesis"] =
graudeespondilolistese.Text.Replace('.', ',');
    }
    else
    {
        erro = true;
    }

    if (erro)
    {
        await DisplayAlert("Atenção:", "Por favor, preencha todos os campos",
"Fechar");
    }
    else
    {
        await executaFuncao();
    }
}

private async Task executaFuncao()
{
    if (naoestalendo)
    {
        await this.Navigation.PushModalAsync(new Loading());
    }
}

```

```

resultado = new List<string>();
String sugestao = String.Empty;
List<RetornoMOD> listaResultado = new List<RetornoMOD>();

try
{
    var url = "http://" + Application.Current.Properties["ipServidor"].ToString()
+ ":8099/WekaApi";

    HttpClient http = new HttpClient();

    dadosParaEnvio["alگو"] = Application.Current.Properties["Algoritmo"].ToString();

    http.DefaultRequestHeaders.Add("user-agent", "Mozilla/5.0 (compatible; MSIE
10.0; Windows NT 6.2; WOW64; Trident/6.0)");

    url += "?" + string.Join("&", dadosParaEnvio.Select((x) => x.Key + "=" +
x.Value.ToString()));

    http.BaseAddress = new Uri(url);

    HttpResponseMessage retorno = await http.GetAsync("");

    retorno.EnsureSuccessStatusCode();

    var resposta = await retorno.Content.ReadAsStringAsync();

    listaResultado =
JsonConvert.DeserializeObject<List<RetornoMOD>>(resposta);

    resultado.Add("Hérnia");
    resultado.Add("Espondilolistese");
    resultado.Add("Normal");

    int v = 0;
    foreach (var r in listaResultado)
    {
        if (r.numberCorrect.Equals("1.0"))
        {
            sugestao = resultado[v];
        }
        v++;
    }
} catch (Exception)
{
}
}

```

```

if (resultado.Count > 0)
{
    await Navigation.PopModalAsync();

    WekaMOD mc = new WekaMOD();
mc.lumbar_lordosis_angle = dadosParaEnvio["lumbar_lordosis_angle"];
    mc.pelvic_incidence = dadosParaEnvio["pelvic_incidence"];
    mc.pelvic_radius = dadosParaEnvio["pelvic_radius"];
    mc.pelvic_tilt = dadosParaEnvio["pelvic_tilt"];
    mc.sacral_slope = dadosParaEnvio["sacral_slope"];
mc.degree_spondylolisthesis = dadosParaEnvio["degree_spondylolisthesis"];
mc.classe = sugestao;

    Resultado pgs;

    RetornoMOD resultados = listaResultado.SingleOrDefault(x =>
x.pctCorrect.Equals("1.0"));

    pgs = new Resultado(sugestao, resultado, mc, resultados);

pgs.Title = " Resultado: " + sugestao;

    await Navigation.PushAsync(pgs);
limparCampos();

}
else
{
    var x = await DisplayAlert("Erro ao processar:", "Não foi possível
concluir a pesquisa\nGostaria de tentar novamente?", "Tentar novamente",
"Cancelar");
    if (x)
    {
        naoestalendo = false;
        await executaFuncao();
    }
    else

```

```

        {
            limparCampos();
            await this.Navigation.PopModalAsync();
        }

    }

}

private void limparCampos()
{
    angulodelordoselombar.Text = "";
    graudedelordoselombar.Text = "";
    graudeespondilolistese.Text = "";
    graudeincidenciapelvica.Text = "";
    graudeinclinacaopelvica.Text = "";
    raiopelvico.Text = "";
}
}
}

```

Neste código permite ao médico escolher o seu diagnóstico

### **Editar Conteúdo**

```

using DataMiningApp.Models;
using Newtonsoft.Json;
using Plugin.DeviceInfo;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Net.Http;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class EditarResultado : ContentPage
    {
        public EditarResultado(WekaMOD weka)
        {
            Title = weka.classe;
            InitializeComponent();
            this.weka = weka;

            List<String> dados = new List<string>();

```

```

dados.Add("Hérnia");
    dados.Add("Espondilolistese");
dados.Add("Normal");
    int i = 0;
    foreach (String d in dados)
    {
        picker.Items.Add(d);
        if (d.Equals(weka.classe))
        {
            picker.SelectedIndex = i;
        }
        i++;
    }
    String x = (weka.classe.Equals("Hérnia")) ? "Hernia" :
(weka.classe.Equals("Espondilolistese")) ? "Spondylolisthesis" : weka.classe;

    if (!x.Equals(weka.doapp))
    {
        String x2 = (weka.doapp.Equals("Hernia")) ? "Hérnia" :
(weka.doapp.Equals("Spondylolisthesis")) ? "Espondilolistese" : weka.doapp;
        StackLayout v = new StackLayout { Padding = new Thickness(15, 10,
15, 10), BackgroundColor = Color.FromHex("2196F3") };
        v.Children.Add(new Label { Text = x2, HorizontalTextAlignment =
TextAlignment.Center, TextColor = Color.White, FontSize =
Device.GetNamedSize(NamedSize.Large, typeof(Label)) });
        doapp.Padding = new Thickness(25, 5, 25, 0);
        doapp.Children.Add(new Label { Text = "" });
        doapp.Children.Add(new Label { Text = "Diagnóstico do app:",
FontAttributes = FontAttributes.Bold, TextColor = Color.Black });
        doapp.Children.Add(v);

        classe.FontSize = Device.GetNamedSize(NamedSize.Large,
typeof(Label));
        classe.HorizontalTextAlignment = TextAlignment.Center;
        classe.TextColor = Color.White;

        diagmed.Text = "Diagnósticado pelo médico como:";
    }
    else
    {
        doapp.Padding = new Thickness(0);
        doapp.Children.Clear();
    }

    graudeespondilolistese.Text = weka.degree_spondylolisthesis;
    raiopelvico.Text = weka.pelvic_radius;
    graudedecivesacral.Text = weka.sacral_slope;
    angulodelordoselombar.Text = weka.lumbar_lordosis_angle;
    graudeinclinacaopelvica.Text = weka.pelvic_tilt;

```

```

        graudeincidenciapelvica.Text = weka.pelvic_incidence;
        classe.Text = weka.classe;

        picker.SelectedIndexChanged += (e, a) =>
        {

            botao.IsVisible = true;

        };
    }

    private WekaMOD weka;

    private async void botaoSalvar(object sender, EventArgs e)
    {
        List<String> resultado = new List<string>();
        resultado.Add("Hernia");
        resultado.Add("Spondylolisthesis");
        resultado.Add("Normal");

        try
        {
            Dictionary<String, String> b = new Dictionary<string, string>();
            b["id"] = "" + CrossDeviceInfo.Current.Id;
            b["classe"] = resultado[picker.SelectedIndex];
            b["degree_spondylolisthesis"] =
weka.degree_spondylolisthesis.Replace(',', '.');
            b["lumbar_lordosis_angle"] = weka.lumbar_lordosis_angle.Replace(',',
'.');
            b["pelvic_incidence"] = weka.pelvic_incidence.Replace(',', '.');
            b["pelvic_radius"] = weka.pelvic_radius.Replace(',', '.');
            b["pelvic_tilt"] = weka.pelvic_tilt.Replace(',', '.');
            b["sacral_slope"] = weka.sacral_slope.Replace(',', '.');

            var url = "http://" + Application.Current.Properties["ipServidor"].ToString()
+ ":8099/WekaApi";

            HttpClient http = new HttpClient();

            http.DefaultRequestHeaders.Add("user-agent", "Mozilla/5.0 (compatible;
MSIE 10.0; Windows NT 6.2; WOW64; Trident/6.0)");

            HttpContent formulario = new FormUrlEncodedContent(b);

            HttpResponseMessage retorno = await http.PostAsync(new Uri(url), formulario);

            retorno.EnsureSuccessStatusCode();

```



```

private Boolean trava = false;

private async void enviarDados(object s, EventArgs a)
{

    if (trava == false) { Botao.IsEnabled = false; trava = true; }

    Dictionary<String, String> campo = new Dictionary<String, String>();
    Dictionary<String, String> erro = new Dictionary<String, String>();

    var erros = false;

    if (Nome.Text != null && !Nome.Text.Equals(String.Empty) &&
Nome.Text.Length > 1)
    {
        campo["Nome"] = Nome.Text;
    }
    else
    {
        erro["Nome"] = "Nome";
        erros = true;
    }

    if (Email.Text != null && !Email.Text.Equals(String.Empty) &&
Email.Text.Length > 1)
    {
        Regex email = new Regex(@"^[A-Za-z0-9]([_\.\/]?\[a-zA-Z0-9\])*\@[A-
Za-z0-9]([_\.\/]?\[a-zA-Z0-9\])*\.[A-Za-z]{2,}$");

        if (email.IsMatch(Email.Text))
        {
            campo["Email"] = "Email";
        }
        else
        {
            erro["Email"] = "Email";
            erros = true;
        }
    }
    else
    {
        erro["Email"] = "Email";
        erros = true;
    }

    if (Problema.Text != null && !Problema.Text.Equals(String.Empty) &&
Problema.Text.Length > 1)

```

```

{
    campo["Problema"] = Problema.Text;
}
else
{
    erro["Problema"] = "Problema";
    erros = true;
}

if (erros == true)
{
    conta = 0;
    pisca(erro);
}

}

else
{
    await enviarAgora();

    await Task.Delay(2000);
    trava = false;
    Botao.IsEnabled = true;
}

}

private async Task enviarAgora()
{
    try
    {
        Dictionary<String, String> post = new Dictionary<String, String>();

        post["Nome"] = Nome.Text.Trim();
        post["Email"] = Email.Text.Trim();
        post["Descricao"] = Problema.Text.Trim();
        post["AppNome"] = App.nome;
        post["VersaoApp"] = App.versao;
        post["Os"] = "" + CrossDeviceInfo.Current.Platform;
        post["VersaoOs"] = "" + CrossDeviceInfo.Current.Version;
        post["ModeloCel"] = "" + CrossDeviceInfo.Current.Model;
        post["Uid"] = "" + CrossDeviceInfo.Current.Id;

        HttpClient cliente = new HttpClient();
        HttpContent conteudo = new FormUrlEncodedContent(post);
    }
}

```

```

int tempoMaximo = 10000;

cliente.Timeout = TimeSpan.FromMilliseconds(tempoMaximo);
CancellationTokensource cts = new CancellationTokensource();
cts.CancelAfter(tempoMaximo);

Uri url = new
Uri("http://app.orcamento.fagnerabsynth.com.br/informarProblema.php");

HttpResponseMessage retorno = await cliente.PostAsync(url, conteudo,
cts.Token);

retorno.EnsureSuccessStatusCode();

String r = await retorno.Content.ReadAsStringAsync();

if (r.Equals("alerta(\\"Dados enviados com sucesso\\",\\"Muito
obrigado\\");window.history.back(-1)"))
{
    Nome.Text = "";
    Email.Text = "";
    Problema.Text = "";

    await DisplayAlert("Sucesso", "Os dados foram enviados com sucesso!\n\nMuito
obrigado!", "Fechar");

    await Navigation.PopAsync();

}
else
{
    await DisplayAlert("Erro:", "Erro ao enviar os dados.\n\nPor favor,
tente novamente!", "Fechar");

}
}
catch (Exception)
{
    await DisplayAlert("Erro:", "Erro ao enviar os dados.\n\nPor favor, tente
novamente!", "Fechar");
}

```

```

    }

    private Int32 conta;
    private async void pisca(Dictionary<string, string> erro)
    {
        conta++;

        if (conta < 6)
        {
            var Cor = Color.Default;
            String erroMensagem = null;

            if (conta % 2 == 0) { Cor = Color.FromHex("#990000"); } else { Cor =
            Color.Default; }

            foreach (var t in erro)
            {
                if (erroMensagem != null) { erroMensagem += "\n"; }

                if (t.Value == "Problema")
                {
                    Problema.BackgroundColor = Cor;
                    erroMensagem += "A \"Descrição\" é obrigatória";
                }
                else if (t.Value == "Email")
                {
                    Email.BackgroundColor = Cor;
                    erroMensagem += "O \"Email\" é inválido";
                }
                else
                {
                    Nome.BackgroundColor = Cor;
                    erroMensagem += "O \"Nome\" é inválido";
                }
            }
            if (conta == 1)
            {
                await DisplayAlert("Atenção", erroMensagem, "Fechar");
            }

            await Task.Delay(600);
            var erros = erro;
            pisca(erros);
        }
        else
        {
            trava = false;
            Botao.IsEnabled = true;
        }
    }

```

```

    }
}
}

```

Neste código será carregado o diagnóstico médico

### **Carregando**

```

using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Loading : ContentPage
    {
        private int i;

        public Loading()
        {
            i = 0;
            BackgroundColor = Color.White;
            InitializeComponent();

            inicia();
        }

        private async void inicia()
        {
            if (i > 3)
                i = 0;

            String pontos = String.Empty;

            for (int m = 0; m < i; m++)
                pontos += ".";

            loading.Text = "Processando\ndiagnóstico" + pontos;

            await Task.Delay(500);
            i++;
        }
    }
}

```

```

        inicia();
    }
}
}

```

Neste código é apresentado o Menu

### **Menu**

```

using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Menu : ContentPage
    {
        public Menu()
        {
            InitializeComponent();
            List<String> dadosLista = new List<string>();
            dadosLista.Add("Realizar diagnóstico");
            dadosLista.Add("Visualizar as regras");
            dadosLista.Add("Consultas salvas");
            dadosLista.Add("Configurações");
            dadosLista.Add("Sobre o App");
            lista.ItemsSource = dadosLista;
            lista.ItemSelected += selecionado;
        }

        private void selecionado(object sender, SelectedItemChangedEventArgs e)
        {
            App.m.IsPresented = false;

            var tool = new ToolbarItem { Text = "Voltar" };
            tool.Clicked += voltar;

            lista.SelectedItem = -1;

            switch (e.SelectedItem as String)
            {
                case "Realizar diagnóstico":

                    Diagnostico pg = new Diagnostico();
                    pg.ToolbarItems.Add(tool);

```

```
        var pgNav = new NavigationPage(pg);
        App.m.Navigation.PushModalAsync(pgNav);

break;

        case "Visualizar as regras":

Regras pgs = new Regras();
        pgs.ToolbarItems.Add(tool);
        var pgNavs = new NavigationPage(pgs);
        App.m.Navigation.PushModalAsync(pgNavs);

        break;

        case "Consultas salvas":

        Consultas pga = new Consultas();
        pga.ToolbarItems.Add(tool);
        var pgNava = new NavigationPage(pga);
        App.m.Navigation.PushModalAsync(pgNava);

        break;

        case "Configurações":

        Configuracoes pgc = new Configuracoes();
        pgc.ToolbarItems.Add(tool);
        var pgNavc = new NavigationPage(pgc);
        App.m.Navigation.PushModalAsync(pgNavc);

        break;

        case "Sobre o App":

        SobreoApp pgb = new SobreoApp();
        pgb.ToolbarItems.Add(tool);
        var pgNavb = new NavigationPage(pgb);
        App.m.Navigation.PushModalAsync(pgNavb);

        break;

    }

}
```

```

        private void voltar(object sender, EventArgs e)
        {
            App.m.Navigation.PopModalAsync();
        }
    }
}

```

Neste código é apresentado o resultado final do processamento do diagnóstico

### Mostrar Resultado

```

using DataMiningApp.Models;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class MostrarResultado : ContentPage
    {
        public MostrarResultado(WekaMOD e)
        {
            Title = e.classe;
            InitializeComponent();
            String c = e.classe;
            if (e.classe.Equals("Hernia")) { c = "Hérnia"; }
            if (e.classe.Equals("Spondylolisthesis")) { c = "Espondilolistese"; }

            classe.Text = c;
            graudeespondilolistese.Text = e.degree_spondylolisthesis;
            raiopelvico.Text = e.pelvic_radius;
            graudedecivesacral.Text = e.sacral_slope;
            angulodelordoselombar.Text = e.lumbar_lordosis_angle;
            graudeinclinacaopelvica.Text = e.pelvic_tilt;
            graudeincidenciapelvica.Text = e.pelvic_incidence;
            update.Text = e.update.ToString();
        }
    }
}

```

### Regras

```

using System;
using System.Collections.Generic;

```

```

using System.Linq;
using System.Text;
using System.Threading.Tasks;

using Xamarin.Forms;

namespace DataMiningApp.Views
{
    public partial class Regras : ContentPage
    {
        public Regras()
        {
            InitializeComponent();
        }
    }
}

```

### Resultado

```

using DataMiningApp.Models;
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using Xamarin.Forms;
using Plugin.DeviceInfo;
using System.Net.Http;
using Newtonsoft.Json;

namespace DataMiningApp.Views
{
    public partial class Resultado : ContentPage
    {
        private WekaMOD weka;

        public Resultado(String result, List<String> dados, WekaMOD ret,
RetornoMOD retorno)
        {
            Title = result;
            InitializeComponent();

            weka = ret;

            resultado.Text = result;

            int i = 0;

```

```

picker.Items.Add("Selecione o seu diagnóstico!");
picker.SelectedIndex = 0;
    foreach (String d in dados)
    {

        picker.Items.Add(d);

        i++;
    }

    picker.SelectedIndexChanged += (e, a) =>
    {
        if (picker.SelectedIndex >= 1)
        {
            botao.IsVisible = true;
        }
    };

    Span cx = new Span();
cx.FontAttributes = FontAttributes.None;
    cx.Text = "Diagnóstico do app: ";

FormattedString c = new FormattedString();
    c.Spans.Add(cx);

    dig.FormattedText = c;

    botao.Clicked += botaoSalvar;

}

private async void botaoSalvar(object sender, EventArgs e)
{
    List<String> resultado = new List<string>();
    resultado.Add("");
    resultado.Add("Hernia");
    resultado.Add("Spondylolisthesis");
    resultado.Add("Normal");

    var appclasse = (weka.classe.Equals("Hérnia")) ? "Hernia" :
(weka.classe.Equals("Normal")) ? "Normal" : "Spondylolisthesis";
    try
    {
        Dictionary<String, String> b = new Dictionary<string, string>();
        b["id"] = "" + CrossDeviceInfo.Current.Id;
    }
}

```

```

        b["classe"] = resultado[picker.SelectedIndex];
        b["degree_spondylolisthesis"] =
weka.degree_spondylolisthesis.Replace(',','.');
        b["lumbar_lordosis_angle"] = weka.lumbar_lordosis_angle.Replace(',','
');
        b["pelvic_incidence"] = weka.pelvic_incidence.Replace(',','.');
        b["pelvic_radius"] = weka.pelvic_radius.Replace(',','.');
        b["pelvic_tilt"] = weka.pelvic_tilt.Replace(',','.');
        b["sacral_slope"] = weka.sacral_slope.Replace(',','.');
        b["app"] = appclasse;

        var url = "http://" + Application.Current.Properties["ipServidor"].ToString()
+ ":8099/WekaApi";

        HttpClient http = new HttpClient();

        http.DefaultRequestHeaders.Add("user-agent", "Mozilla/5.0 (compatible;
MSIE 10.0; Windows NT 6.2; WOW64; Trident/6.0)");

        HttpContent formulario = new FormUrlEncodedContent(b);

        HttpResponseMessage retorno = await http.PostAsync(new Uri(url), formulario);

        retorno.EnsureSuccessStatusCode();

        var resposta = await retorno.Content.ReadAsStringAsync();

        var res = JsonConvert.DeserializeObject<Dictionary<String, String>>(resposta);

        if (res["status"].Equals("1"))
        {
            await DisplayAlert("Sucesso!", "Diagnóstico salvo com sucesso!",
"Fechar");
            await App.m.Navigation.PopModalAsync();
        }
        else
        {
            await DisplayAlert("Erro!", "Não foi possível salvar os dados, por favor
tente novamente!", "Fechar");
        }
    }
    catch (Exception)
    {
        await DisplayAlert("Erro!", "Não foi possível salvar os dados, por favor
tente novamente!", "Fechar");
    }
}
}

```

```
protected override void OnDisappearing()
{
    base.OnDisappearing();
}
}
}
```