



Sistema Especialista em Dispositivos Móveis para Auxílio em Diagnóstico Médico

Trabalho de Conclusão de Curso

Engenharia de Computação

ALLYSON DA COSTA LISBOA DA SILVA

Orientador: Prof. Dr. Fernando Buarque de Lima Neto



Allyson da Costa Lisboa da Silva

Sistema Especialista em Dispositivos Móveis para Auxílio em Diagnóstico Médico

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Engenharia de Computação
Escola Politécnica de Pernambuco
Universidade de Pernambuco

Orientador: Prof. Dr. Fernando Buarque de Lima Neto

Coorientador: Rui Nóbrega de Pontes Filho

Recife - PE, Brasil

Julho de 2018

Allyson da Costa Lisboa da Silva

Sistema Especialista em Dispositivos Móveis para Auxílio em Diagnóstico Médico/
Allyson da Costa Lisboa da Silva. – Recife - PE, Brasil, Julho de 2018-
53 p.

Orientador: Prof. Dr. Fernando Buarque de Lima Neto

Trabalho de Conclusão de Curso – Engenharia de Computação
Escola Politécnica de Pernambuco
Universidade de Pernambuco , Julho de 2018.

1. Diagnóstico Médico. 2. Sistemas Especialistas. 3. Medicina. 4. *Android* 5.
Inteligência Artificial. I. Prof. Fernando Buarque de Lima Neto. II. Universidade de
Pernambuco. III. Escola Politécnica.

Dedico este trabalho aos meus pais que apesar de todos os problemas e dificuldades sempre estiveram ao meu lado e acreditaram no meu potencial.

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 10 de agosto de 2018, às 11:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente **ALLYSON DA COSTA LISBOA DA SILVA**, orientado pelo professor **Fernando Buarque de Lima Neto**, sob título **Sistema Especialista em Dispositivos Móveis para Auxílio em Diagnóstico Médico**, a banca composta pelos professores:

Sérgio Campello Oliveira

- **Fernando Buarque de Lima Neto**

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada Aprovada com Restrições* Reprovada

e foi-lhe atribuída nota: 7,5 (Sete e meio)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 07 dias para entrega da versão final da monografia a contar da data deste documento.

SÉRGIO CAMPELLO OLIVEIRA

FERNANDO BUARQUE DE LIMA NETO

Agradecimentos

Em primeiro lugar, agradeço a Deus por ter me guiado durante toda essa jornada e por ter me dado forças em todos momentos para vencer os obstáculos que a vida me impôs e continuar fazendo o que nasci para fazer. Sem ele, com certeza não estaria aqui hoje.

Agradeço aos meus pais, Verônica da Costa e Aguinaldo Antônio, que nunca deixaram de acreditar no meu potencial e sempre fizeram o possível, e as vezes o impossível, para me proporcionar a melhor educação. Obrigado meus pais por se sacrificarem tanto para que nunca me faltasse nada, por estarem comigo em praticamente todos os momentos da minha vida me dando o apoio necessário para que eu continuasse percorrendo meus sonhos e me dando o amor necessário para que eu crescesse sendo uma pessoa feliz. Vocês são a minha luz e o principal motivo de eu sempre ter buscado ser uma pessoa da qual vocês pudessem se orgulhar. Amo vocês.

Gostaria de agradecer ao meu professor de matemática, Walter Lima, que me acompanhou durante todo o ensino fundamental e ensino médio, me ensinando tudo o que podia ensinar e que já desde essa época enxergava em mim o potencial que eu ainda desconhecia. Agradeço também a duas pessoas de extrema importância em minha vida, Hérica e Bárbara, responsáveis por melhorarem ainda mais a base de conhecimentos que eu possuía, quero que saibam que vocês também possuem uma parcela de participação nessa conquista.

Não posso deixar de agradecer também aos amigos que fiz durante minha trajetória na universidade, que me proporcionaram muitos momentos felizes e fizeram essa jornada ser menos árdua e sofrida, em especial aos meus dois grandes amigos Ericks Rodrigues e Italo Yoshito, que estiveram comigo em momentos felizes e tristes me apoiando sempre que necessário. Vocês todos já fazem parte permanente de minha vida. Por fim, quero agradecer ao meu professor-orientador Fernando Buarque pelo apoio, dedicação, gentileza e paciência durante as etapas deste trabalho. Muito obrigado por ter aceito ser meu orientador, por ter sido não só meu professor durante o curso, mas também meu professor na vida e ter compartilhado seus conhecimentos comigo.

“A ciência humana de maneira nenhuma nega a existência de Deus. Quando considero quantas e quão maravilhosas coisas o homem compreende, pesquisa e consegue realizar, então reconheço claramente que o espírito humano é obra de Deus, e a mais notável”

Galileu Galilei

Resumo

Desde o surgimento da medicina moderna no século XV, a humanidade vêm desenvolvendo métodos e tecnologias capazes de melhorar a eficiência no tratamento de doenças. No entanto, um problema muito comum que vêm se destacando nos últimos anos é o diagnóstico incorreto de doenças, etapa inicial e de grande importância para o tratamento correto delas. Muitas vezes esse problema está associado à inexperiência do profissional avaliador ou de circunstâncias adversas que podem surgir no momento do atendimento ao paciente, como falta de recursos no ambiente de trabalho. Parte desses problemas poderiam ser reduzidos com a utilização de um sistema especialista com domínio na área médica. Sistemas especialistas são sistemas especializados em uma área de conhecimento que utilizam técnicas de inteligência artificial para analisar um grande volume de informações e fornecer uma solução para o problema proposto. Esses sistemas podem ser um grande recurso na realização de diagnósticos corretos, mas que infelizmente são pouco utilizados na prática devido à grande dificuldade de implantação nos hospitais. Por isso, este trabalho apresenta o estudo e descreve a implementação de um sistema especialista, voltado para funcionar em dispositivos móveis, capaz de classificar algumas das doenças transmitidas pelo mosquito *Aedes Aegypti*, sendo elas: chikungunya, dengue e zika. O sistema usa uma base de conhecimentos contendo informações sobre os sintomas apresentados por pacientes diagnosticados anteriormente e utiliza em seu motor de inferência o algoritmo Naive Bayes. Durante os testes, o sistema utilizou 32 amostras para realização do treinamento do algoritmo e 9 amostras para testes, conseguindo obter ao fim uma acurácia de 66,6%. O resultado obtido por ser considerado bom dada a pouca quantidade de amostras utilizadas no treinamento do algoritmo e seu poder de classificação pode aumentar a medida que mais amostras são incluídas em sua base de conhecimentos. Também foram realizados testes com outros algoritmos de classificação como indução baseado em regras e árvore de decisão para fins de comparação, mas que obtiveram resultados inferiores ao naive bayes. Por fim, apresenta-se os benefícios de trazer esse recurso, que é pouco explorado, de forma mais conveniente e acessível para as pessoas, além de mostrar que ele pode ser aplicado em outros domínios do conhecimento.

Palavras-chave: sistemas especialistas, aplicativos, dispositivos móveis, naive bayes, inteligência artificial, diagnóstico médico, saúde, *android*, *smartphone*.

Abstract

Since the emergence of modern medicine in XV century, science has been developing technologies and methods capable of improving efficiency in disease treatment. However, a very common problem standing out in recent years is the incorrect diagnosis, the initial stage and of great importance to the correct treatment. Often this problem is associated with the inexperience of the professional or the adversities that may arise at the time of patient care, such as the lack of resources in the work environment. Some of these problems could be reduced by using a specialist system with mastery in the medical field. Expert systems are specialized systems in an area of knowledge that use artificial intelligence techniques to analyze a large volume of information and provide a solution to the proposed problem. These systems can be a great resource in the realization of correct diagnosis, but that unfortunately are little used in practice due to the great difficulty of implantation in the hospitals. Therefore, this work presents the study and the implementation of a expert system aimed at the use in mobile devices, able to classify some diseases transmitted by *Aedes Aegypti* mosquito, such as: chikungunya, dengue and zika. The system uses a knowledge base with information about the symptoms presented by previously diagnosed patients and uses the Naive Bayes algorithm in its inference engine. During the tests, the system used 32 samples to realize the training of the algorithm and 9 samples for tests, getting an accuracy of 66.6%. The result obtained could considered good given the few samples used in the algorithm training and its classification power can increase as more samples are included in its knowledge base. Also performed tests with other classification algorithms such as rule-based induction and decision tree for comparison purposes, but showing results below naive bayes. Finally, it presents the advantages of bringing this resource, even being little explored, more conveniently and accessible to people, and show that it can be applied in other domains of knowledge.

Keywords: expert systems, mobile devices, naive bayes, artificial intelligence, medical diagnosis, health, *android*, *smartphone*.

Lista de ilustrações

Figura 1 – Fluxograma de arquitetura geral do sistema.	30
Figura 2 – Diagrama de casos de uso do sistema.	31
Figura 3 – Comunicação entre uma aplicação Java e um sistema operacional.	32
Figura 4 – Tela de acesso do sistema.	34
Figura 5 – Modelagem do banco de dados online do sistema.	35
Figura 6 – Modelagem do banco de dados offline do sistema.	35
Figura 7 – Tela do sistema de seleção de sintomas.	37
Figura 8 – Tela de resultados da classificação.	39
Figura 9 – Tela de amostras com avaliação pendente.	40
Figura 10 – Formulário usado para coleta de dados.	42
Figura 11 – Incidência de casos de chikungunya por região.	43
Figura 12 – Incidência de casos de dengue por região.	43
Figura 13 – Incidência de casos de zika por região.	44

Lista de tabelas

Tabela 1 – Resultado apresentado ao usuário para amostra do vírus da Dengue. . .	29
Tabela 2 – Exemplo de tabela de frequência criada contendo apenas 2 sintomas. . .	36
Tabela 3 – Resultados obtidos usando o algoritmo Naive Bayes.	45
Tabela 4 – Matriz de confusão usando o algoritmo Naive Bayes.	45
Tabela 5 – Análise da matriz de confusão usando o algoritmo Naive Bayes.	45
Tabela 6 – Resultados obtidos usando o algoritmo de indução baseado em regras. .	46
Tabela 7 – Matriz de confusão usando o algoritmo de indução baseado em regras.	46
Tabela 8 – Análise da matriz de confusão usando o algoritmo de indução baseado em regras.	46
Tabela 9 – Resultados obtidos usando o algoritmo de árvore de decisão	47
Tabela 10 – Matriz de confusão usando o algoritmo de árvore de decisão.	47
Tabela 11 – Análise da matriz de confusão usando o algoritmo de árvore de decisão.	47
Tabela 12 – Comparação de resultados entre os algoritmos utilizados.	47

Lista de abreviaturas e siglas

BD	Banco de dados.
C	Linguagem de programação imperativa.
FN	Falso negativo.
FP	Falso positivo.
GB	Gigabyte.
GHz	Gigahertz.
IA	Inteligência artificial.
IDE	<i>Integrated Development Enviroment</i> (Ambiente de Desenvolvimento Integrado).
JDK	<i>Java Development Kit</i> (Kit de Desenvolvimento Java).
JRE	<i>Java Runtime Environment</i> (Ambiente de Execução Java).
JVM	<i>Java Virtual Machine</i> (Máquina Virtual Java).
NEG	Negativo.
POS	Positivo.
SE	Sistema especialista.
SO	Sistema operacional.
SQL	<i>Structured Query Language</i> (Linguagem de Consulta Estruturada).
SDK	<i>Software Development Kit</i> (Kit de Desenvolvimento de <i>Software</i>).
TB	Terabyte.
TF	Tabela de Frequência.
VN	Verdadeiro negativo.
VP	Verdadeiro positivo.

Sumário

1	INTRODUÇÃO	15
1.1	Motivação e Caracterização do Problema	15
1.2	Estrutura do Documento	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	Heurísticas	18
2.1.1	Heurística de representatividade	19
2.1.2	Heurística de disponibilidade	19
2.1.3	Heurística de âncora	20
2.1.4	Heurística de simulação	20
2.2	Processo de Diagnóstico Médico	20
2.2.1	O que é diagnóstico médico?	20
2.2.2	Criação do diagnóstico médico	21
2.2.3	Influências nas hipóteses médicas	22
2.3	Inteligência Artificial	23
2.3.1	Histórico e definição	23
2.3.2	Sistemas especialistas	24
2.4	Naive Bayes	27
2.4.1	Funcionamento do algoritmo	27
2.4.2	Vantagens e desvantagens	28
3	DESENVOLVIMENTO DO APLICATIVO	29
3.1	Breve Explicação sobre o Sistema	29
3.2	Ferramentas e Recursos Utilizados	31
3.3	Detalhamento do Sistema	33
3.3.1	Treinamento	34
3.3.2	Execução	36
3.3.3	Avaliação	38
4	RESULTADOS	41
4.1	Aquisição de Amostras	41
4.2	Testes e Resultados	44
5	CONCLUSÃO	49
6	TRABALHOS FUTUROS	50

REFERÊNCIAS 51

1 Introdução

O objetivo deste trabalho é apresentar um sistema especialista desenvolvido para dispositivos móveis, a ser utilizado por profissionais da área de saúde, para auxílio no diagnóstico de enfermidades. O sistema usará informações sobre os sintomas do paciente, previamente identificados pelo médico, e técnicas de inteligência artificial para realizar os diagnósticos. Espera-se que seu uso resulte na diminuição da taxa de erros de diagnósticos em hospitais do país.

1.1 Motivação e Caracterização do Problema

Atualmente existe uma quantidade muito grande de doenças espalhadas por todo o globo terrestre, desde doenças hereditárias e congênitas à doenças comuns adquiridas no convívio social [1]. Com o surgimento da medicina moderna no século XV, a humanidade adotou um método padrão para identificação de doenças e posterior tratamento das mesmas [2].

Este método ficou conhecido como diagnóstico médico, que basicamente é o processo de entrevistar o doente a respeito dos sintomas manifestados pelo mesmo, a fim de descobrir que tipo de enfermidade o assola. Desde então, a medicina só avançou, agregando cada vez mais conhecimentos científicos aliados à tecnologia, a fim de melhorar esse processo de identificação de doenças para reduzir a taxa de mortalidade das pessoas.

Apesar de todos os avanços conseguidos ao longo do tempo, infelizmente no Brasil os casos de diagnósticos médicos realizados erroneamente aumentaram abruptamente nos últimos anos [3]. Muitos desses erros estão frequentemente atrelados à inexperiência do avaliador ou as condições e circunstâncias adversas de um determinado lugar que podem induzir o profissional ao erro.

Na medicina o termo falso positivo se refere a casos em que um paciente é diagnosticado com uma doença, quando na verdade ele não a possui [4]. Além dos casos de falsos positivos existentes no domínio da saúde, há também os casos onde uma pessoa é diagnosticada com uma enfermidade diferente da que possui. Em ambos os casos, pode-se provocar um tratamento equivocado que pode resultar em consequências graves para o paciente.

No Brasil esses problemas são amplificados devido às condições precárias, e as vezes desumanas, dos estabelecimentos de saúde no qual muitos pacientes são internados e tratados. Parte desses problemas poderiam ser reduzidos se houvesse, e fosse amplamente utilizado, um sistema especialista (SE) com domínio na área de saúde capaz de auxiliar os

profissionais nesses diagnósticos. Sistemas especialistas são sistemas que utilizam técnicas de inteligência artificial (IA), desenvolvidos por engenheiros de *software* com auxílio de um profissional especialista em uma área de atuação, capazes de analisar uma grande quantidade de informações utilizando recursos computacionais para fornecer uma solução para o problema que está sendo tratado [5].

Apesar dos benefícios desses sistemas, é muito raro vê-los sendo utilizados de maneira rotineira na prática em redes hospitalares do país. Parte disso deve-se ao fato de que o sistema depende de recursos computacionais que nem sempre se dispõe em alguns hospitais. Uma alternativa a esse problema seria o planejamento e desenvolvimento de um SE que funcione de forma independente em dispositivos móveis, já que o uso de *smartphone's* têm aumentado significativamente nos últimos anos [6]. Além da conveniência e facilidade de usar um *smartphone*, é fato que alguns utilizados diariamente por muitos usuários já possuem recursos superiores há muitos computadores pessoais disponíveis no mercado.

Apenas no Brasil, já são mais de 70 milhões de pessoas a possuírem *smartphone's* para realizar diversos tipos de atividades [7]. As pessoas estão praticamente o tempo inteiro utilizando seu dispositivo para alguma coisa, seja para verificar sua localização em um lugar ou consultar seu perfil em redes sociais. Aproveitando-se disso, muitas empresas já estão explorando o potencial desse novo recurso criando aplicativos responsáveis pela comunicação direta com elas e permitindo a resolução de problemas com uma maior facilidade [6].

O principal objetivo deste trabalho é o desenvolvimento de um sistema especialista com domínio na área médica que possa ser utilizado por profissionais, experientes ou não, para auxiliar em diagnósticos médicos de pacientes que apresentem os sintomas das doenças Chikungunya, Dengue ou Zika, todas provenientes do mosquito *Aedes Aegypti*. O sistema funcionará em dispositivos móveis e utilizará uma base de conhecimentos que contém informações sobre os sintomas apresentados por pacientes diagnosticados anteriormente.

Em seu motor de inferência será utilizado o algoritmo Naive Bayes devido à velocidade ao trabalhar com grandes volumes de dados e resultados melhores a medida que a base de conhecimentos aumenta. Esse sistema será capaz de aumentar a taxa de sucesso no diagnóstico de pacientes em hospitais que ainda não contam com um sistema semelhante ou que possuam profissionais inexperientes no ofício.

Outros fins buscados no desenvolvimento deste trabalho são:

- Estudo dos elementos necessários para concepção do diagnóstico médico;
- Desenvolvimento de um sistema especialista voltado para dispositivos móveis que seja capaz de realizar o diagnóstico médico de forma satisfatória;

- Explorar as funcionalidades de um sistema já consolidado, mas agora disponível de forma mais acessível e conveniente;
- Avaliação e revisão do sistema para verificar sua acurácia e possíveis problemas durante sua utilização.

1.2 Estrutura do Documento

Este documento está dividido em 6 capítulos. O Capítulo 1 refere-se à introdução ao assunto e apresentação dos aspectos do problema a ser resolvido, assim como os objetivos do trabalho. No Capítulo 2 são apresentados as fundamentações teóricas, explicando o que são heurísticas e sua importância, como é realizado o processo de identificação de enfermidades por parte de um médico, o que é inteligência artificial e por fim apresentando a teoria por trás do algoritmo de classificação utilizado no trabalho.

No Capítulo 3 são descritos os processos realizados durante o desenvolvimento deste trabalho, bem como as especificações necessárias para sua replicação e o ambiente onde a aplicação foi implementada. No Capítulo 4 são apresentados os testes realizados no trabalho e os resultados obtidos. No Capítulo 5 são apresentadas as conclusões sobre o que foi produzido. No Capítulo 6 são apresentados os trabalhos futuros que pode-se realizar a partir do que foi construído.

2 Fundamentação Teórica

Este capítulo descreve as fundamentações teóricas que dão o embasamento necessário para concepção da abordagem proposta. A Seção 2.1 apresenta os conceitos e importância das heurísticas, enquanto na seção 2.2 são explicados os processos realizados para a criação de um diagnóstico médico. A seção 2.3 detalha o que é inteligência artificial e suas aplicações no mundo real e na seção 2.4 é descrito o funcionamento do algoritmo utilizado no desenvolvimento do sistema especialista objetivado neste trabalho.

2.1 Heurísticas

Os seres vivos inteligentes são famosos por tomarem centenas de decisões diárias importantes durante suas vidas, sejam elas desde as mais simples, como o que comer ou beber, até as mais complexas como em quem votar nas eleições presidenciais. Todas essas decisões, por mais simples que pareçam ser, podem resultar em uma consequência inevitável se não bem conduzida, principalmente quando o ser vivo em questão é o ser humano. A decisão de cada pessoa na sociedade é multiplicada por um conjunto de fatores aleatórios ocorrendo no momento da decisão, que podem, no fim de um ciclo (dia, mês ou ano) resultar em uma decisão boa ou ruim.

Há muitas decisões a serem feitas e nem todas são fáceis de fazer, isto porque a decisão de uma única pessoa, seja essa decisão boa ou ruim, pode afetar de maneira positiva ou negativa a vida de várias outras pessoas. No entanto, é possível resolver decisões muito importantes de forma mais fácil do que espera-se. Isto é possível porque a mente dos seres vivos utiliza atalhos mentais chamados heurísticas.

Heurísticas são estratégias cognitivas que o subconsciente de um ser vivo utiliza para reformular um problema e transformá-lo em um mais simples e fácil de resolver, ou seja, é basicamente um artifício que a mente utiliza para tornar mais fácil o processo de tomada de decisões [8].

Explicando melhor, ao considerar a seguinte questão, “Qual curso escolher na hora de prestar um vestibular?”, muitas pessoas simplesmente não conseguem tomar uma decisão, isto é muito comum dado que esta é uma questão complexa de se resolver e que exige uma profunda pesquisa sobre o assunto e reflexão sobre as vantagens e dificuldades de cada curso. Nesse momento as heurísticas na mente são acionadas e começam a trabalhar de forma automática, sem o consentimento e conhecimento da pessoa. Em vez de responder à questão “Que curso fazer?”, as heurísticas “quebram” a questão e a transforma em “Quais cursos oferecem melhores oportunidades profissionais após a formação?”

Apesar de a segunda questão ser bastante diferente da questão original, ambas possuem relação e a resposta para a segunda questão é dada como resposta automática para a questão original. Nota-se agora que a segunda questão está atrelada a um grupo genérico baseado em impressões superficiais que permite uma maior facilidade para avaliação. Essa estratégia é um atalho usado quando é muito difícil atender a todos os problemas a serem resolvidos utilizando a estratégia racional, e prosseguir ou não com esse raciocínio é um problema que a mente consciente simplesmente decide não processar.

O problema dessas heurísticas é que elas são acionadas justamente quando trabalha-se com questões onde não há dados suficientes para responder à questão de forma racional, o que pode acarretar algumas vezes em respostas pouco precisas e que podem induzir uma pessoa a um erro. Sabendo o que são heurísticas fica evidente sua importância, principalmente na área de saúde onde os profissionais lidam o tempo todo com decisões que podem mudar a vida de uma pessoa para sempre. As heurísticas também são utilizadas em IA, dando base para desenvolvimento de diversos sistemas de classificação e previsão, sendo utilizado nesta abordagem como um dos elementos presentes no motor de inferência do sistema desenvolvido.

Dentre os tipos de heurísticas existentes, as usadas com mais frequências são: heurística de representatividade, heurística de disponibilidade, heurística de âncora e heurística de simulação [9].

2.1.1 Heurística de representatividade

Consiste em realizar inferências a respeito das chances de algo pertencer a uma determinada categoria. Esta inferência é feita a partir de características superficiais. É importante ressaltar que mesmo que haja informação suficiente para realização de uma inferência não significa que ela seja verdadeira [9].

Um exemplo desse tipo de heurística é quando uma pessoa conhece outras duas e anteriormente ela possuía a informação de que uma dessas duas trabalha com desenvolvimento de software. Se uma das pessoas que ela acabou de conhecer estiver vestindo uma camisa que seja relacionado a área de tecnologia, esta pessoa pode deduzir que quem está com a camisa citada é quem trabalha com desenvolvimento de *software*.

2.1.2 Heurística de disponibilidade

Consiste em estimar as chances de algo ocorrer, a frequência de um evento ou a associação entre dois acontecimentos. Essa estimativa é realizada através da lembrança de casos que vêm à mente através da experiência adquirida ao longo do tempo. Um exemplo desse tipo de heurística é quando alguém pergunta se há mais homens ou mulheres atuando na área de desenvolvimento de software. Nesse caso, se a pessoa a responder teve mais

contato com mulheres nessa área, então será esta a resposta que ela dará [9].

2.1.3 Heurística de âncora

Consiste em tomar algo de referência quando não há experiência sobre o problema a ser resolvido. Nesse caso é tomado um ponto como referência, e em seguida são feitos ajustes intuitivos nesse ponto para que o mesmo se aproxime de uma situação que se tenha conhecimento [9].

Um exemplo desse tipo de heurística é quando alguém pergunta a renda média de um desenvolvedor de software. Nesse caso, a pessoa a responder pode analisar sua renda anual trabalhando como desenvolvedor e avaliar se ela está acima ou abaixo da média. Em seguida, realiza-se alguns ajustes e infere-se intuitivamente o valor que pode ser a renda média de um desenvolvedor de software.

2.1.4 Heurística de simulação

Consiste em estimar as chances de algo acontecer com base na facilidade com que imagina-se sua ocorrência. Quanto mais fácil for imaginar essa ocorrência, mais provável será acreditar em seu acontecimento [9]. Um exemplo desse tipo de heurística é quando em um sorteio uma pessoa deixa de ganhar o prêmio pela diferença de um dígito no bilhete que possui. Para essa pessoa é fácil imaginar a situação de ter ganho o prêmio se alguma coisa tivesse mudado.

2.2 Processo de Diagnóstico Médico

2.2.1 O que é diagnóstico médico?

O diagnóstico médico é uma das principais atividades no cotidiano de um profissional da saúde, é por meio dele que se conseguem informações a respeito da saúde atual de um determinado paciente. Este processo consiste em basicamente três etapas, sendo elas a anamnese, os exames físicos e os exames laboratoriais complementares.

A anamnese consiste em uma entrevista com o paciente a fim de relembrar os fatos anteriores ao surgimento da doença, bem como suas possíveis causas. Quando bem realizada é responsável por 85% do diagnóstico em clínicas médicas, restando dessa forma apenas 10% para os exames físicos e 5% para exames complementares laboratoriais [10].

Durante a anamnese o profissional responsável recolhe informações como dados de identificação do paciente, principal queixa que levou o paciente a procurar um atendimento clínico, registros de quando, como e onde a doença se manifestou, histórico do paciente, histórico familiar, dentre outras informações.

Os exames físicos por sua vez compreendem os métodos para identificar os sinais de uma enfermidade, onde eles consistem em inspeção do corpo do paciente, palpação, auscultação e percussão [11].

Por fim, a última etapa consiste na realização de exames laboratoriais para auxiliar o raciocínio do médico após a obtenção das informações prévias na anamnese e nos exames físicos. Após as três etapas terem sido realizadas, é feita uma síntese das informações para definir quais possíveis doenças estão acometendo o paciente e posteriormente planejar o tratamento baseado no quadro de saúde apresentado.

Durante a síntese de informações para definir as possíveis doenças que um paciente possui, um profissional da saúde normalmente utiliza os conhecimentos médicos obtidos durante a formação acadêmica e também sua experiência profissional. Essa síntese de informações exige um raciocínio complexo e elaborado, pois é necessário vislumbrar centenas de possibilidades para se chegar à conclusão correta.

No decorrer da seção, serão apresentadas as metodologias empregadas no processo de síntese de informações para obtenção do diagnóstico correto, bem como práticas habituais utilizadas por diversos médicos que atuam na área.

2.2.2 Criação do diagnóstico médico

Durante a entrevista com o paciente, o profissional utiliza de seu conhecimento acadêmico e sua experiência para criar as primeiras hipóteses a serem avaliadas dentro de um conjunto de centenas de possibilidades. Cada nova informação obtida a respeito da doença permite que o profissional formule novas hipóteses e descarte outras, dessa forma direcionando a entrevista para a hipótese mais provável [10].

O sucesso na formulação do diagnóstico médico depende muito da experiência do profissional, isto porque a cada nova evidência informada pelo paciente, o médico adapta as perguntas seguintes à hipótese que acredita ser a correta. Dessa forma, ao fim da entrevista as etapas de semiotécnica e solicitação de exames complementares também estarão diretamente associados à primeira impressão do médico.

Considerando a importância das primeiras impressões a respeito da enfermidade, as hipóteses descartadas são tão importantes quanto as hipóteses aceitas, isto porque uma hipótese indevidamente descartada pode acarretar em consequências graves para o paciente. Um raciocínio muito comum é o qual alguém não avalia outras possibilidades, fixando-se rapidamente em apenas uma hipótese que acredita ser a correta. O profissional analisa o quadro geral e enxerga apenas as referências que quer ver, influenciado pelas heurísticas de disponibilidade e simulação, acabando por negligenciar informações importantes que indicam outras hipóteses.

Por exemplo, após um período de tempo atendendo diversos casos de pacientes

com gripe em uma determinada região, um médico experiente acaba por diagnosticar os sintomas de um novo paciente também como gripe, quando na verdade o paciente está infectado com o vírus da dengue que a princípio tem sintomas semelhantes à de uma gripe. Um erro que pode levar o paciente a óbito por recomendar medicações contraindicadas.

2.2.3 Influências nas hipóteses médicas

No mundo todo, pesquisas são realizadas todos os dias acerca das melhores práticas para tomadas de decisões em diagnósticos médicos, esses estudos levaram os pesquisadores a associar diagnósticos satisfatórios em três conjuntos diferentes e igualmente importantes, sendo eles: o conjunto de características pessoais e estilo de prática do médico, o conjunto de contexto de atuação e o conjunto de incentivo financeiro [12].

Conjunto de características pessoais e estilo de prática do médico

Nesse conjunto destaca-se a importância do conhecimento, treinamento e experiência do profissional da área de saúde. Isto implica que na grande maioria dos casos, profissionais que estão atuando há mais tempo na área possuem uma maior probabilidade de prestar um auxílio de alta qualidade ao paciente, visto que o profissional em questão estará mais familiarizado com as evidências que presencia na área em que atua [12].

Nesse caso, como a experiência está diretamente ligada ao sucesso no diagnóstico, é de senso comum que profissionais especialistas como cardiologistas possuam mais conhecimento e mais sucesso em diagnósticos de sua área que um clínico geral, salvo às exceções.

Conjunto de contexto de atuação

Nesse conjunto destaca-se a importância dos recursos que o profissional dispõe no momento do atendimento, isto é, medicações, acessórios médicos, informações sobre consultas anteriores. Tudo isso pode afetar na atuação do profissional, visto que durante uma consulta o profissional pode precisar utilizar um material específico que acabe por não dispôr naquele momento [12].

Conjunto de incentivo financeiro

Nesse conjunto destaca-se a importância na prática de profissionais da saúde. Isto porque em alguns casos, a postura de atendimento do profissional da área muda de acordo com o tipo de remuneração que o mesmo recebe. Basicamente um profissional pode ser remunerado pelo seu empregador por número de atendimentos realizados ou por remuneração fixa mensal [12].

No primeiro caso, um maior número de atendimentos realizados no período de um mês implica em uma remuneração maior, o que pode resultar em certo descaso por parte do profissional no momento do atendimento, pois geralmente ele tende a realizar atendimentos mais rápidos.

No segundo caso não há esse problema, pois a remuneração ao fim do mês será a mesma independente do número de atendimentos que o profissional tenha realizado durante esse período, o que faz com que ele dedique um tempo maior no atendimento do paciente.

2.3 Inteligência Artificial

2.3.1 Histórico e definição

O termo Inteligência Artificial (IA) refere-se à um ramo de pesquisa da ciência da computação que tem como objetivo criar mecanismos computacionais que sejam capazes de simular o pensamento humano [13]. Seu estudo iniciou após a segunda guerra mundial, com o artigo "*Computing Machinery and Intelligence*" do matemático Alan Turing [14], e desde então seu desenvolvimento vem avançando rapidamente.

Inicialmente a IA objetivava apenas uma forma de simular o raciocínio humano para resolver problemas complexos de forma rápida, e dessa forma aumentar a produtividade onde quer que fosse aplicada. No entanto, com o passar do tempo os pesquisadores enxergaram nesse meio a capacidade de se criar uma tecnologia não só capaz de fazer isso, mas de criar algo que pudesse ter emoções, criatividade, consciência e comunicação, assim como todos os humanos [13].

No campo da pesquisa, os cientistas dividem a IA em 2 tipos, sendo elas: IA forte e IA fraca [15]. A IA forte investiga maneiras de criar formas de inteligência capazes de possuir autoconsciência, enquanto a IA fraca lida com a criação de métodos para resolução de problemas não determinísticos. Uma forma de descobrir se uma tecnologia é de fato autoconsciente é realizando o teste de Turing [16], no qual consiste em colocar um humano para ter uma conversa com uma máquina e ao fim da conversa, o mesmo não conseguir distinguir se estava falando com outro humano ou não.

Enquanto o desenvolvimento da IA forte progride lentamente, as técnicas baseadas em IA fraca dominam o mercado tecnológico contendo a maior parte das aplicações desenvolvidas atualmente. Uma das aplicações frequentemente utilizadas é o processamento digital de imagens, que consiste em desenvolver algoritmos capazes de manipular imagens para melhorar seu aspecto visual e fornecer outros recursos para sua interpretação [17]. Outras aplicações de inteligência artificial são: planejamento automatizado, raciocínio baseado em casos, jogos, robótica e reconhecimento de padrões.

- Planejamento automatizado: Atuam fazendo o processo de escolha e organização de ações para alcançar, da melhor forma possível, objetivos pré-definidos [18].
- Sistemas especialistas: Têm como objetivo resolver problemas adaptando soluções observadas em problemas anteriores [19].
- Diagnósticos: Sistemas de diagnóstico baseados em análise probabilística capazes de executar tarefas em um nível equivalente à um profissional especialista em diversas áreas. Um exemplo disso é o sistema de diagnóstico de etiologia para dores abdominais agudas, desenvolvido na Inglaterra e utilizado amplamente em diversas enfermarias de emergência do país. Seu uso reduziu em 50% as taxas de apendicite perfurada e em 22% a incidência de cirurgias abdominais desnecessárias [20].
- Robótica: Englobam o uso de computadores, robôs e computação para desenvolvimento de sistemas compostos por partes mecânicas que podem ser controlados manualmente ou automaticamente por circuitos elétricos. Utilizada atualmente por muitas fábricas e indústrias para obter sucesso na redução de custos e aumento de produtividade [21].
- Reconhecimento de padrões: Tem como objetivo a classificação de objetos, dentro de uma série de categorias, que podem variar entre imagens, sinais, gestos, caracteres ou qualquer outro tipo de categoria que precise ser classificada [22].

2.3.2 Sistemas especialistas

Nos anos 70, durante o desenvolvimento de tecnologias que permitissem a resolução de problemas recorrentes do mundo real, os pesquisadores de IA notaram que para que isso fosse possível era necessário abandonar a estratégia de algoritmizar os problemas de forma tradicional e começar a desenvolver formas de representar o funcionamento do pensamento humano. Dessa forma, eles chegaram a conclusão que o sucesso para isso ocorrer dependia mais do conhecimento sobre o problema do que das regras no algoritmo.

Com isso surgiram os sistemas especialistas, sistemas inteligentes capazes de funcionar de forma similar ao raciocínio de um profissional especializado em uma área de conhecimento, que tem como objetivo realizar a classificação de problemas de seu domínio [5]. Eles utilizam uma base de conhecimento de regras e/ou fatos, e heurísticas para realizar a classificação dos problemas aos quais são submetidos.

O principal elemento de um SE é o motor de inferência, ele é o responsável por realizar o processamento necessário entre a base de conhecimento e as heurísticas para resolver o problema. O motor de inferência pode realizar o processamento das heurísticas do sistema de duas diferentes formas: progressiva e regressiva.

Durante o processamento progressivo, apenas os dados que podem identificar o problema são apresentados pelo usuário, a partir daí o sistema busca na base de conhecimento os fatos e regras que melhor se encaixam às informações fornecidas. Enquanto no processamento regressivo, tanto os dados do problema quanto a resposta para ele são fornecidos ao sistema, que por sua vez inicia uma busca para provar se a hipótese fornecida é de fato a resposta real para o problema.

De forma geral, um SE possui um domínio de aplicação muito restrito ao problema ao qual foi designado, onde o classificador de seu motor de inferência é baseado em modelos estatísticos. Para que um SE seja considerado efetivo, além de classificar corretamente os problemas, o mesmo também deve ser capaz de oferecer sugestões e aprender com novos exemplos inseridos pelo usuário.

Dentre as aplicações de IA disponíveis, os SE são os mais adequados no que se refere ao desenvolvimento de sistemas para auxílio em diagnóstico médico, e por este motivo foi escolhido para ser utilizado neste trabalho.

Fases de desenvolvimento

O processo de desenvolvimento de um SE é dividido comumente em fases, onde cada uma delas deve ser cuidadosamente executada para a criação de um SE eficaz. Essas fases podem ser divididas em: identificação, conceituação, formalização, implementação, avaliação e revisão.

Identificação

Fase responsável pela identificação dos recursos e características do problema, bem como os objetivos desejados ao fim do desenvolvimento. Durante a identificação dos recursos, deve-se obter a base de conhecimento a ser utilizada no sistema, isto é possível consultando um especialista na respectiva área de conhecimento e buscando livros, artigos e todo tipo de conteúdo relacionado. Enquanto na identificação das características ressaltam-se os principais elementos do problema, como quais problemas o sistema resolverá, que tipo de situação pode ocorrer no momento da utilização do sistema, quais dados são mais importantes para a resolução de um caso, etc. Por fim, a fase de identificação também é responsável por identificar os recursos computacionais e financeiros para execução do projeto.

Conceituação

Nesta fase definem-se os principais conceitos e mecanismos necessários para resolução do problema. O engenheiro e especialista responsáveis pelo sistema decidem quais conceitos são mais adequados para que o sistema consiga atingir melhores resultados durante sua execução.

Formalização

Fase responsável por transformar os conceitos escolhidos em regras formais que simulem o raciocínio humano durante a utilização do sistema.

Implementação

Fase responsável por utilizar dos conhecimentos e recursos coletados durante as fases anteriores para o desenvolvimento do sistema propriamente dito. Nesta fase, o engenheiro responsável algoritmizará as regras criadas que simulam o raciocínio de um profissional especialista no domínio da aplicação.

Avaliação

Fase responsável pela realização de testes para verificação da eficácia do sistema e identificação de possíveis *bugs* que o sistema pode apresentar para posterior reparo.

Revisão

Fase responsável por realizar o acompanhamento do sistema, isto é, observar o sistema continuamente para possíveis mudanças em alguns recursos notados durante a fase de teste e avaliação.

Vantagens e desvantagens

- Vantagens

- Não é influenciado por emoções ou outras condições adversas;
- Velocidade na classificação de problemas;
- Redução de custos com funcionários;
- Sem perdas no conhecimento adquirido ao longo do tempo;
- Integração com outras ferramentas.

- Desvantagens

- Não conseguem classificar problemas mais genéricos;
- Podem gerar insegurança com outros funcionários que atuam na mesma área de conhecimento do sistema;
- Não possuem senso comum, podendo fornecer resultados que impactem negativamente alguém.

2.4 Naive Bayes

O algoritmo Naive Bayes é uma técnica de classificação, baseado em probabilidades e inspirado no teorema de Bayes, utilizado para prever uma classe de um determinado conjunto de dados [23]. O teorema de Bayes foi criado pelo matemático Thomas Bayes e descreve a probabilidade de um evento ocorrer baseado em um conhecimento a priori (probabilidade inicial de um evento ocorrer) que pode ter relação com o evento. O algoritmo tem como objetivo encontrar a probabilidade a posteriori de um evento, isto é, a probabilidade de um evento A ocorrer, dado que um evento B ocorreu, conforme é visto em 2.1.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2.1)$$

Onde:

- $P(A)$ é a probabilidade do evento A ocorrer (probabilidade a priori);
- $P(B)$ é a probabilidade do evento B ocorrer;
- $P(A|B)$ é a probabilidade do evento A ocorrer, dado que B ocorreu (probabilidade a posteriori);
- $P(B|A)$ é a probabilidade do evento B ocorrer, dado que A ocorreu.

A palavra *naive* presente no nome do algoritmo significa ingênuo e foi dada por causa de uma peculiaridade que o mesmo possui. No algoritmo, todos os atributos analisados são tratados de forma independente, isso quer dizer que mesmo que alguns atributos de um problema tenham relação para sua ocorrência, eles não serão tratados dessa forma, por isso a origem do nome. Mesmo sendo um algoritmo relativamente simples de ser implementado, suas aplicações são as mais variadas e o mesmo consegue ter um bom desempenho onde é aplicado.

2.4.1 Funcionamento do algoritmo

O algoritmo Naive Bayes consiste em criar uma tabela de frequência (TF) baseado nas amostras de eventos disponíveis em sua base de conhecimentos, onde a mesma terá a quantidade de ocorrências desses eventos associados às suas causas. Em seguida o algoritmo utiliza a tabela para realizar operações matemáticas que resultarão ao fim do processo nas probabilidades de ocorrência de um novo evento apresentado.

Durante o treinamento do algoritmo pode ocorrer casos em que falte dados na TF criada, quando isso acontece é utilizado um método conhecido por correção laplaciana. Este

método consiste em modificar a tabela adicionando 1 registro na informação faltante, isso é feito porque o algoritmo utiliza os valores da tabela para realizar operações matemáticas que envolvem multiplicações e caso haja uma informação da amostra analisada com valor 0 (zero), a probabilidade do evento ao qual essa amostra está associada também tenderá a 0 (zero), impossibilitando dessa forma sua classificação. No entanto, ao adicionar informações inexistentes na tabela, o processo de classificação pode ser influenciado caso a base de dados utilizada seja pequena.

2.4.2 Vantagens e desvantagens

Vantagens

- Fácil implementação;
- Facilidade e velocidade em prever um novo conjunto de dados apresentados;
- Necessários poucos exemplos no conjunto de treino para obter classificação com boa acurácia;
- Quando há independência nos atributos, seu desempenho é superior a outros modelos de classificação.

Desvantagens

- Se o exemplo a ser classificado possuir uma característica que não foi observada durante o treinamento, o algoritmo atribuirá probabilidade zero à ela e não será possível classificar o exemplo. Para esses casos, pode-se usar a correção laplaciana;
- Quando há forte dependência nos atributos, o desempenho do algoritmo tende a ser menor.

3 Desenvolvimento do Aplicativo

Este capítulo descreve como o aplicativo foi desenvolvido, bem como as ferramentas e técnicas de classificação utilizadas. Na seção 3.1, descreve-se de forma geral a abordagem proposta, informando as técnicas utilizadas. Na seção 3.2, detalha-se as informações das ferramentas utilizadas e linguagem de programação, além dos recursos necessários para replicação da presente abordagem. Na seção 3.3, descreve-se o detalhamento do funcionamento do sistema, bem como suas etapas de execução.

3.1 Breve Explicação sobre o Sistema

O sistema foi desenvolvido baseando-se na teoria sobre SE, abordada na seção 2.3.2, e o tipo de SE criado é um baseado em casos, visto que o algoritmo utilizado usa amostras inteiras no processo de treinamento e classificação. O algoritmo escolhido para ser usado como classificador do sistema foi o Naive Bayes, algoritmo apresentado na seção 2.4, e sua escolha deve-se à velocidade ao tratar grandes volumes de dados e obtenção de bons resultados quando trabalha-se com bases pequenas, além de apresentar as respostas de classificação no formato de probabilidades para cada doença

Esta última característica permite ao utilizador do sistema analisar de forma mais crítica se deve ou não aceitar a resposta final dada pelo sistema. Ao fim de cada classificação o sistema apresenta as probabilidades de cada doença e dessa forma o usuário pode decidir, com base em seu raciocínio, se avalia corretamente ou incorretamente a resposta fornecida, caso os resultados da classificação sejam muito próximos. Um exemplo de resultado de classificação pode ser visto na Tabela 1.

Tabela 1 – Resultado apresentado ao usuário para amostra do vírus da Dengue.

Doença	Probabilidade
Chikungunya	27.75%
Dengue	39.16%
Zika	33.07%

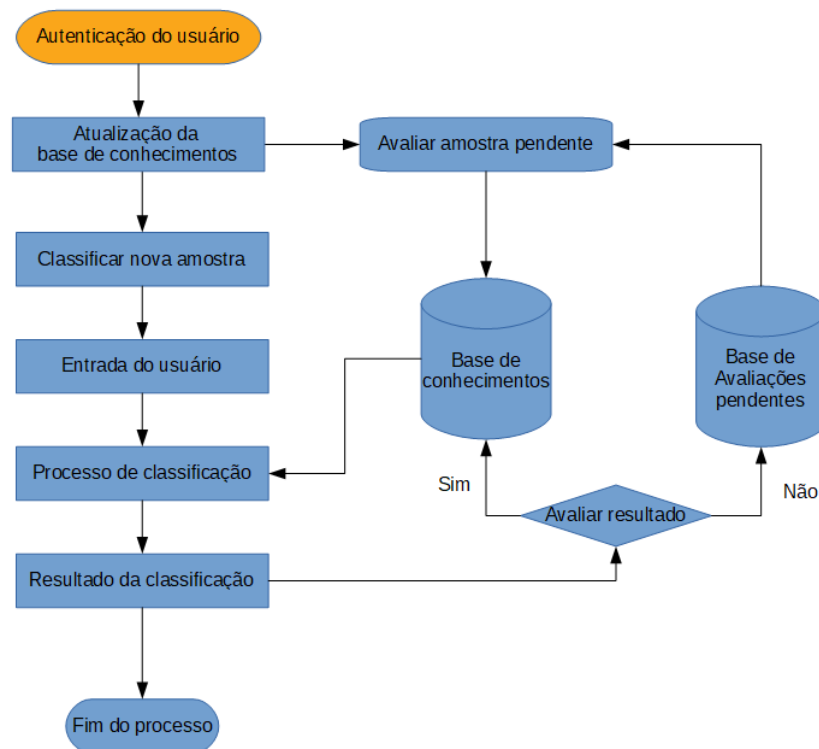
O funcionamento do sistema está dividido basicamente em três fases: treinamento, execução e avaliação. Durante a fase de treinamento o sistema consulta a base de conhecimentos para criar uma tabela de frequência (TF) que será utilizada em posteriores classificações. Na fase de execução o usuário interagirá com o sistema, informando os sintomas do paciente, previamente identificados pelo médico durante a entrevista, e solicitará que ele execute a classificação para verificar qual doença pode ter contaminado o

paciente. Na fase de avaliação o sistema informará ao usuário os resultados da classificação e solicitará que ele informe se a avaliação está correta ou não.

A Figura 1 demonstra de modo geral o funcionamento do sistema, onde pode ser visto que inicialmente o usuário precisa acessá-lo usando suas credenciais, neste caso *login* e senha, e em seguida ele poderá atualizar a base de dados do sistema para posterior utilização. Com a base de dados já atualizada, o usuário poderá classificar uma nova amostra ou avaliar uma amostra com classificação pendente, onde ao optar pela primeira opção ele terá que entrar com os sintomas do paciente que foram previamente identificados.

Após a inserção dos sintomas é possível iniciar o processo de classificação, onde o sistema consultará a base de conhecimentos para a coleta de informações e acionará o algoritmo Naive Bayes para realizar a classificação da amostra. Por fim o sistema apresentará os resultados ao usuário e solicitará que ele avalie-os, caso ele opte por não realizar o *feedback* no momento, os resultados serão armazenados em uma base de conhecimentos provisória e disponível para avaliação depois.

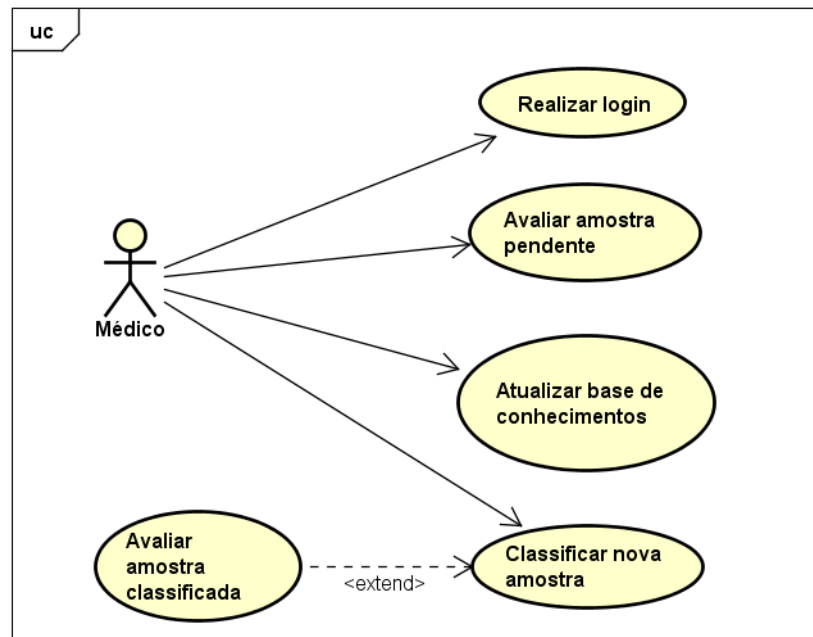
Figura 1 – Fluxograma de arquitetura geral do sistema.



Fonte: Autoria própria (2018)

A Figura 2 ilustra os diagramas de casos de uso do sistema, onde é possível ver as funcionalidades disponíveis citadas anteriormente.

Figura 2 – Diagrama de casos de uso do sistema.



powered by Astah

Fonte: Autoria própria (2018)

3.2 Ferramentas e Recursos Utilizados

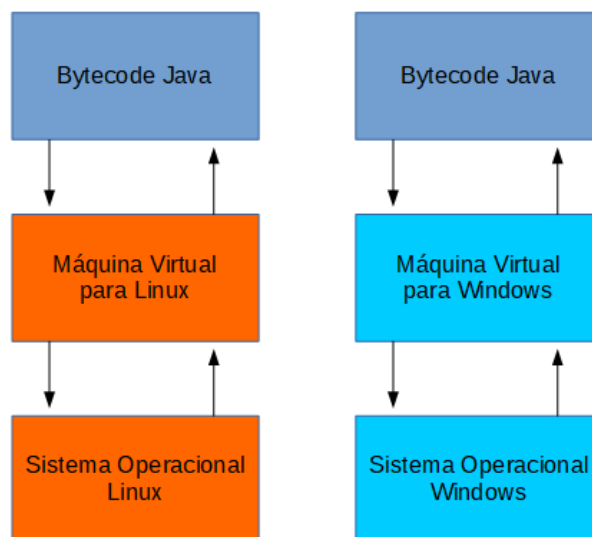
Para desenvolvimento do sistema foi utilizado um computador do tipo *desktop*, e um *smartphone* rodando o sistema operacional (SO) *Android* para realização dos testes com as seguintes configurações:

- Computador
 - Processador: AMD Phenom II X4 965 Black Edition;
 - Memória RAM: 16GB;
 - HD: 1 Terabyte;
 - Placa de Vídeo: Nvidia GeForce GTX 750Ti.
- Smartphone
 - Processador: Cortex-A7 MediaTek MT6580M, 1.3 GHz Quad Core;
 - Memória RAM: 1GB;
 - Memória Interna: 8GB;
 - Versão do SO: Android 6.0.

A linguagem de programação utilizada para desenvolvimento do projeto foi a Java, uma linguagem de programação orientada a objetos originalmente criada pela Sun e atualmente mantida pela Oracle [24]. A linguagem Java foi criada inicialmente para aplicações *desktop* robustas, mas atualmente é utilizada para desenvolvimento *web*, *desktop* e *mobile*. Diferente de algumas linguagens, ela utiliza o conceito de máquina virtual, isto é, utiliza um ambiente que funciona como uma interface entre o SO utilizado e a aplicação desenvolvida.

Além de muitas outras funcionalidades, essa interface é responsável por realizar a comunicação entre os dois pontos sem que haja a necessidade do usuário se preocupar em ter de desenvolver sua aplicação para cada SO disponível. A aplicação desenvolvida executa sem envolvimento algum do SO, apenas comunicando-se com a máquina virtual, que neste caso é a *Java Virtual Machine* (JVM), responsável por realizar essa comunicação entre ambas as partes. A figura 3 ilustra o processo de comunicação realizado entre uma aplicação Java e o SO utilizando uma máquina virtual. É importante ressaltar que para desenvolvimento de qualquer aplicação que utilize a linguagem Java é necessário ter instalado uma versão do *Java Development Kit* (JDK), responsável por rodar as aplicações durante a fase de desenvolvimento, em seu ambiente de trabalho.

Figura 3 – Comunicação entre uma aplicação Java e um sistema operacional.



Fonte: Autoria própria (2018)

Para que a aplicação rode em qualquer SO, é necessário que ele tenha instalado uma versão da *Java Runtime Environment* (JRE), pacote de instalação que contém o ambiente de execução para aplicações Java, como a JVM e outras bibliotecas utilizadas. Este ambiente já vem instalado por padrão em dispositivos com o SO android.

A IDE (*Integrated Development Environment*) utilizada para desenvolvimento do projeto foi o Android Studio, IDE criada pela Google para desenvolvimento de aplicativos

do SO android.

Para o banco de dados (BD), foram utilizados 2 serviços diferentes, um para armazenamento do banco de forma *online*, dessa forma disponível para todos os usuários que utilizem o sistema em dispositivos diferentes, e outro para armazenamento *offline* para que não seja necessário estar conectado o tempo inteiro à internet ao realizar a classificação de novas amostras.

Para o BD *online* foi utilizado o *MySQL*, um sistema de gerenciamento de banco de dados que utiliza a linguagem SQL (*Structured Query Language*) como interface e é usada para armazenamento de grandes quantidades de informações [25]. Enquanto que para o BD *offline* foi utilizado o *SQLite*, uma biblioteca em linguagem C que implementa um banco de dados SQL e que está presente nativamente em sistemas operacionais *android* [26].

O banco *MySQL* é responsável por armazenar as informações referentes às amostras obtidas, enquanto que o *SQLite* é responsável por armazenar as informações das amostras recebidas do BD *online*, da TF após ela ser criada e das amostras que ficarem com avaliação pendente. As versões dos *softwares* utilizados estão listados a seguir:

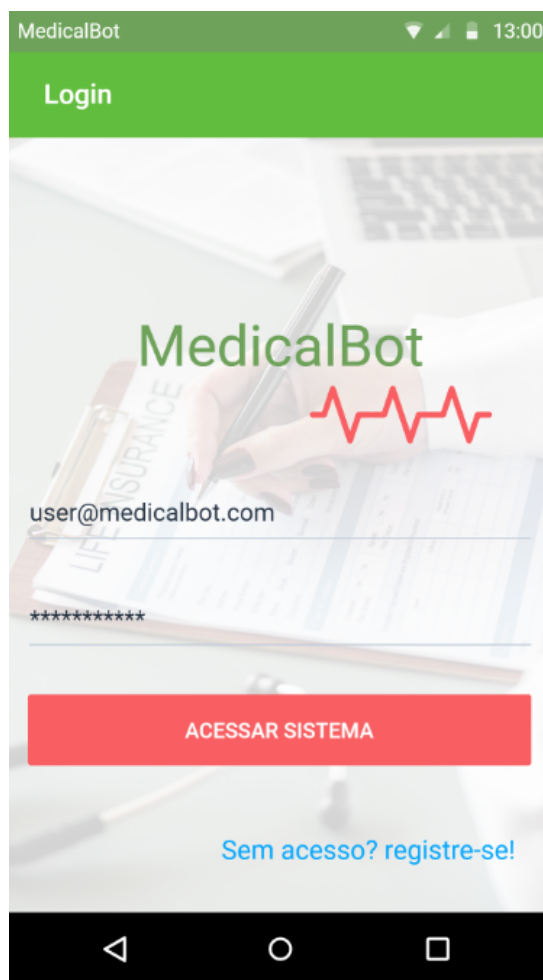
- Versões utilizadas:
 - Versão do Java (JDK): 8.0;
 - Versão da JRE: 8.0;
 - Versão do Android Studio: 3.0;
 - Versão mínima para execução da aplicação: Android 5.0.

Também foi utilizado para fins de comparação de resultados entre algoritmos a biblioteca *scikit-learn*, uma biblioteca de aprendizado de máquina de código aberto para a linguagem de programação Python que inclui vários algoritmos de classificação, regressão e agrupamento [27].

3.3 Detalhamento do Sistema

Antes de começar a utilizar o sistema é necessário que o usuário insira suas credenciais de acesso e aguarde sua autenticação, isso foi feito para evitar que usuários não certificados utilizem o sistema por conta própria e realize ações irreversíveis. Essas credenciais mencionadas são o *login* (neste caso é usado o *email*, mas também pode ser o número de identificação do profissional) e a senha do usuário. A tela inicial do sistema pode ser vista na Figura 4.

Figura 4 – Tela de acesso do sistema.

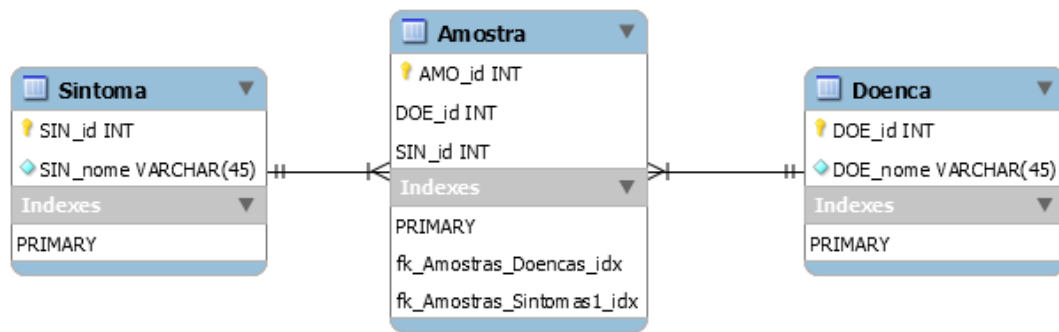


Fonte: Autoria própria (2018)

3.3.1 Treinamento

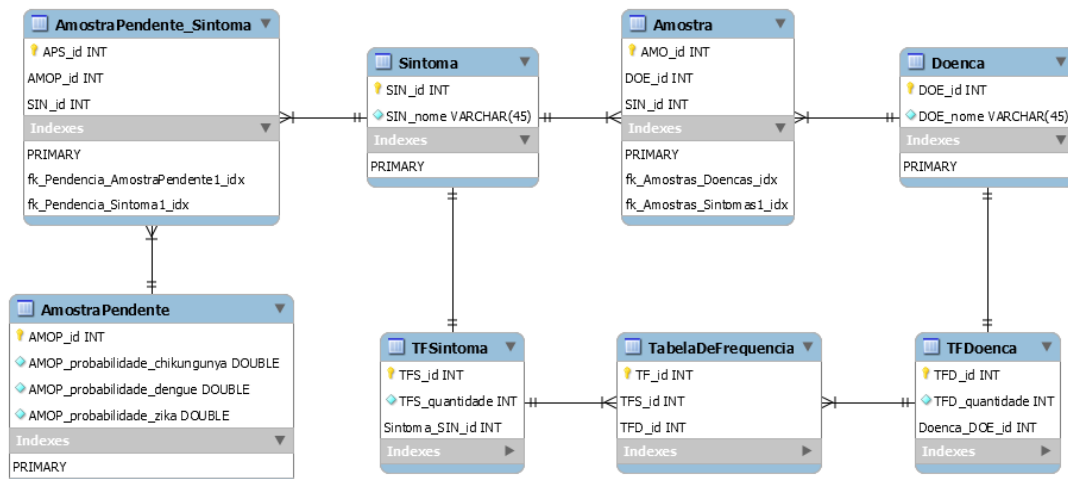
Inicialmente o sistema precisará ser treinado com uma base de conhecimentos previamente armazenada. Para este trabalho, que tem como objetivo a classificação de algumas das doenças transmitidas pelo mosquito *Aedes Aegypti*, o sistema possui três possíveis resultados finais, sendo eles: Chikungunya, Dengue e Zika. As Figuras 5 e 6 ilustram as modelagens dos bancos de dados *online* e *offline*, respectivamente, que o sistema utiliza.

Figura 5 – Modelagem do banco de dados online do sistema.



Fonte: Autoria própria (2018)

Figura 6 – Modelagem do banco de dados offline do sistema.



Fonte: Autoria própria (2018)

Como já mencionado, o funcionamento do sistema divide-se em três fases: treinamento, execução e avaliação. Na fase de treinamento, o algoritmo Naive Bayes consulta cada registro da base de conhecimentos e usa essas informações para criar uma TF, isto é, uma tabela que contém o número de ocorrências de cada doença associada a seus respectivos sintomas. Após criar a TF, o sistema armazena-a no banco de dados para usos posteriores. Um exemplo de TF criada por meio desse método por ser observado em 2.

Tabela 2 – Exemplo de tabela de frequência criada contendo apenas 2 sintomas.

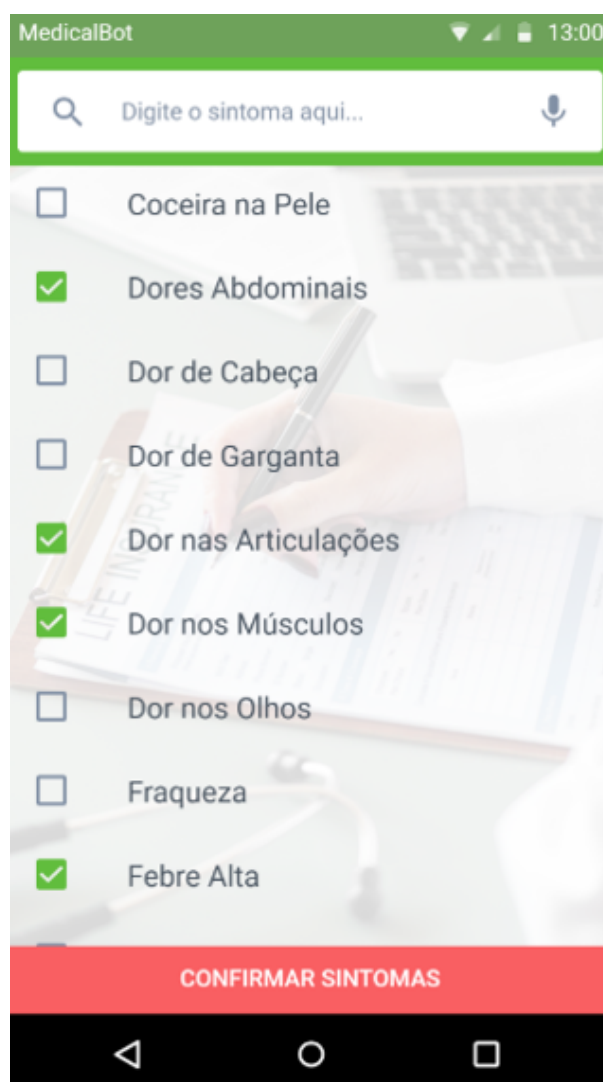
Amostras: 15	A (POS): 7	A (NEG): 9	B (POS): 7	B (NEG): 9
Chikungunya: 5	1	4	3	2
Dengue: 8	5	3	1	7
Zika: 3	1	2	3	0

A primeira coluna da tabela destina-se ao número de ocorrências das doenças, enquanto que todas as outras destinam-se ao número de ocorrências dos sintomas identificados de forma positiva (POS) ou negativa (NEG). É possível notar que a primeira linha da tabela possui uma função diferente das outras, a mesma tem como objetivo registrar a contagem total das doenças e sintomas identificados, isto é necessário porque alguns desses valores são utilizados no processo de classificação. Para este trabalho, a TF do sistema conta com as três doenças já mencionadas e 16 possíveis sintomas que alguém pode apresentar.

3.3.2 Execução

Após a criação da TF, o sistema segue para a fase de execução, onde o usuário poderá escolher classificar uma nova amostra. Ao escolher realizar uma nova classificação, o usuário será apresentado a uma tela onde estarão listados, em ordem alfabética, todos os possíveis sintomas relacionados às doenças registradas na base de conhecimentos. Dentre os sintomas apresentados, o usuário precisará escolher os sintomas que foram previamente identificados durante a consulta com o paciente. O formato de lista de sintomas foi adotado para melhor usabilidade do aplicativo, mas é possível também informar o sintoma de forma textual na barra de busca localizada no topo da tela, conforme pode ser visto na Figura 7.

Figura 7 – Tela do sistema de seleção de sintomas.



Fonte: Autoria própria (2018)

Após o usuário selecionar e confirmar todos os sintomas apresentados pelo paciente, o sistema iniciará o processo de classificação, onde ele consultará em sua TF o número de ocorrências de cada sintoma para gerar as probabilidades de ocorrência de cada doença. Para exemplificar esta etapa, suponha que o usuário indicou no sistema os sintomas dor de cabeça, febre alta e náuseas. Nesse momento, o algoritmo Naive Bayes consulta a TF para colher as informações referentes ao número de ocorrências positivas dos sintomas listados e negativas dos sintomas não-listados, e usa estes dados para realizar o cálculo demonstrado em 3.1.

$$P(x) = \frac{X}{T} \cdot \frac{A}{X} \cdot \frac{B}{X} \cdot \frac{C}{X} \cdot \frac{N_1}{X} \cdot \frac{N_2}{X} \cdot \frac{N_3}{X} \dots \frac{N_n}{X} \quad (3.1)$$

Onde:

- $P(x)$ é a probabilidade da doença analisada, neste caso x ;
- X é o número de amostras, registradas na base de conhecimentos, da doença x ;
- T é o número total de amostras registradas na base de conhecimentos;
- A é o número de ocorrências positivas do sintoma “dor de cabeça” relacionado à doença x ;
- B é o número de ocorrências positivas do sintoma “febre alta” relacionado à doença x ;
- C é o número de ocorrências positivas do sintoma “náuseas” relacionado à doença x ;
- N_i , com i variando de 1 à n , são os números de ocorrências negativas, relacionados à doença x , de cada sintoma não informado na etapa de seleção de sintomas;

Dessa forma o algoritmo consegue obter a probabilidade de ocorrência para uma doença, em seguida esse processo é repetido para todas as doenças registradas no sistema.

3.3.3 Avaliação

Ao fim do processo de classificação, o sistema apresentará ao usuário uma tela informando a probabilidade de cada doença que o paciente pode ter contraído realçando a doença vencedora, isto é, a que possui maior probabilidade dentre elas. A tela em questão pode ser visualizada na Figura 8.

Na mesma tela o usuário terá a oportunidade de avaliar a classificação realizada pelo sistema, onde é possível informar se o resultado está correto (ou ajudou a sanar as possíveis dúvidas) ou incorreto. No caso de classificação incorreta, o sistema avançará para mais uma etapa, onde será perguntado ao usuário qual doença seria a correta dentre as que o sistema possui em sua base de conhecimentos.

Figura 8 – Tela de resultados da classificação.



Fonte: Autoria própria (2018)

Se o usuário não optar por avaliar o resultado, a amostra será armazenada em uma base de conhecimentos provisória e poderá ser consultada depois para que a avaliação final seja feita, conforme pode ser visto na Figura 9. Todas as avaliações finais, sejam elas positivas ou negativas, serão adicionadas à base de conhecimentos do sistema para serem utilizadas em futuras classificações, permitindo dessa forma que o sistema tenha uma retroalimentação, assim como outros sistemas baseados em IA.

Figura 9 – Tela de amostras com avaliação pendente.



Fonte: Autoria própria (2018)

4 Resultados

Este capítulo descreve o processo de aquisição da base de conhecimentos utilizada para treinamento e testes e os resultados obtidos durante o desenvolvimento do projeto.

4.1 Aquisição de Amostras

Durante o período de 29 de Junho de 2018 à 09 de Julho de 2018 foram realizadas as aquisições de amostras para a base de conhecimentos que seria utilizada no sistema proposto. Essa aquisição foi feita por meio da coleta de informações, que foram fornecidas por voluntários, por meio de um formulário online criado utilizando o serviço do Google, onde cada voluntário respondeu 12 perguntas no total, sendo 3 para cada doença estudada (Chikungunya, Dengue, Febre Amarela e Zika). Isso foi feito pois não havia disponível, de forma oficial para uso, uma base de dados contendo casos de pessoas infectadas pelos vírus já citados. Apesar da doença febre amarela ter sido incluída no processo de aquisição de amostras, não houve voluntários que apresentaram a doença e por isso ela não foi incluída durante a fase de treinamento e testes do sistema.

A Figura 10 mostra a primeira etapa de preenchimento de dados do formulário usado para criação da base de conhecimentos informal.

Figura 10 – Formulário usado para coleta de dados.

Questionário sobre arboviroses

Formulário de pesquisa para projeto de TCC sobre diagnóstico de doenças transmitidas pelo mosquito *Aedes aegypti*. Se você já teve ou conhece alguém que teve e possa responder as perguntas abaixo, por favor responda. Esse trabalho ajudará muitas pessoas no futuro. Nenhum dado será compartilhado e informações mais pessoais como "Nome" e "Sexo" não são obrigatórias serem informadas.

***Obrigatório**

Nome

Sua resposta

Cidade *

Sua resposta

Sexo

Masculino

Feminino

Outro

PRÓXIMA

Página 1 de 10

Nunca envie senhas pelo Formulários Google.

Fonte: Autoria própria (2018)

Disponível em: < http://bit.ly/formulario_de_coleta > (Acessado em 12/08/2018)

Além das perguntas sobre a ocorrência de alguma das doenças estudadas, os voluntários inicialmente respondiam de forma opcional o nome, sexo e região onde residiam. As perguntas foram elaboradas a partir de uma pesquisa sobre doenças transmitidas pelo mosquito *Aedes Aegypti* [28], e foram dispostas da seguinte forma para facilitar na coleta de dados.

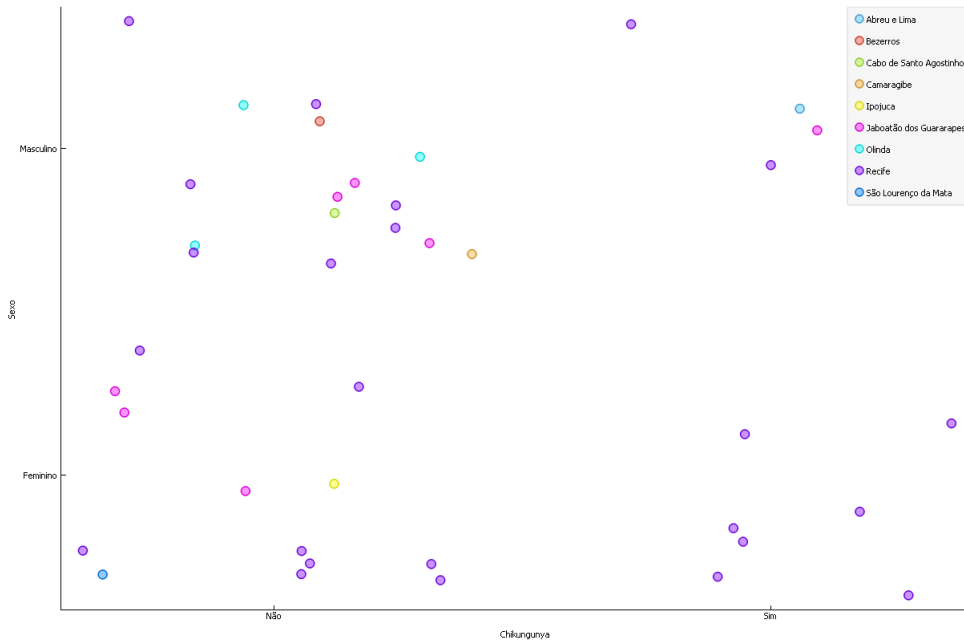
- Perguntas:

1. Você já teve esta Doença?
2. Quais sintomas você apresentou enquanto esteve doente?
3. Por quanto tempo você apresentou os sintomas informados?

Na pergunta de número 2 dispôs-se um total de 16 sintomas diferentes que poderiam ser marcados pelo voluntário, sendo eles: coceira na pele, dores abdominais, dor de cabeça, dor na garganta, dor nas articulações, dor nos olhos, dor nos músculos, fraqueza, febre baixa (37° à 38°), febre alta (39° à 40°), icterícia, inchaço no corpo, manchas vermelhas, náuseas, vermelhidão nos olhos e vômito [28].

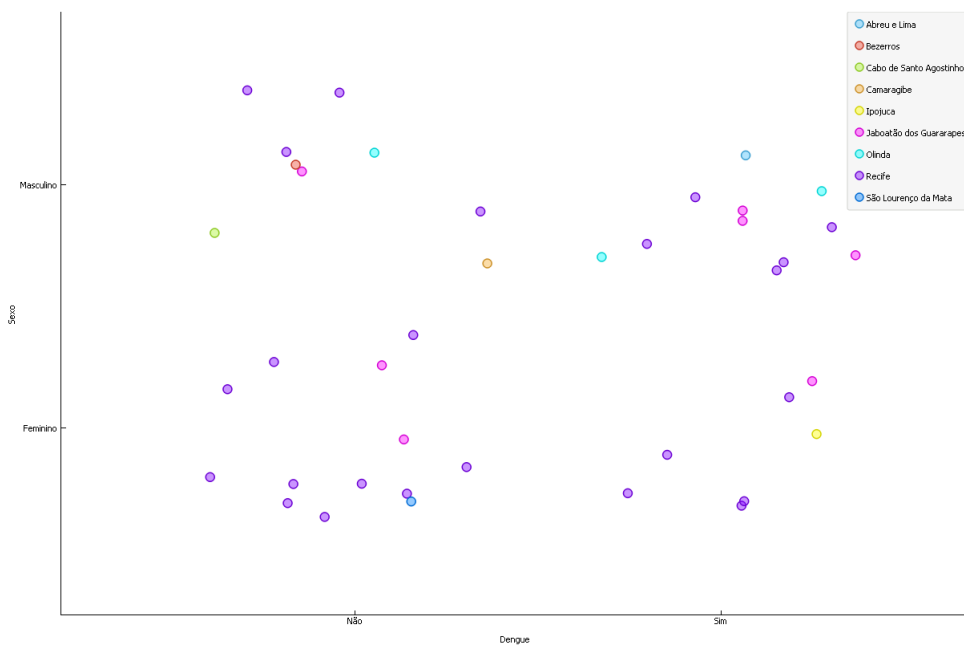
Ao todo, 41 pessoas responderam o formulário e essas informações foram modeladas e inseridas, de maneira manual, no banco de dados utilizado no sistema. Também foi realizado uma análise estatística de incidência de cada doença baseado no sexo e região onde residiam os voluntários. Os resultados podem ser observados nas Figuras 11, 12 e 13.

Figura 11 – Incidência de casos de chikungunya por região.



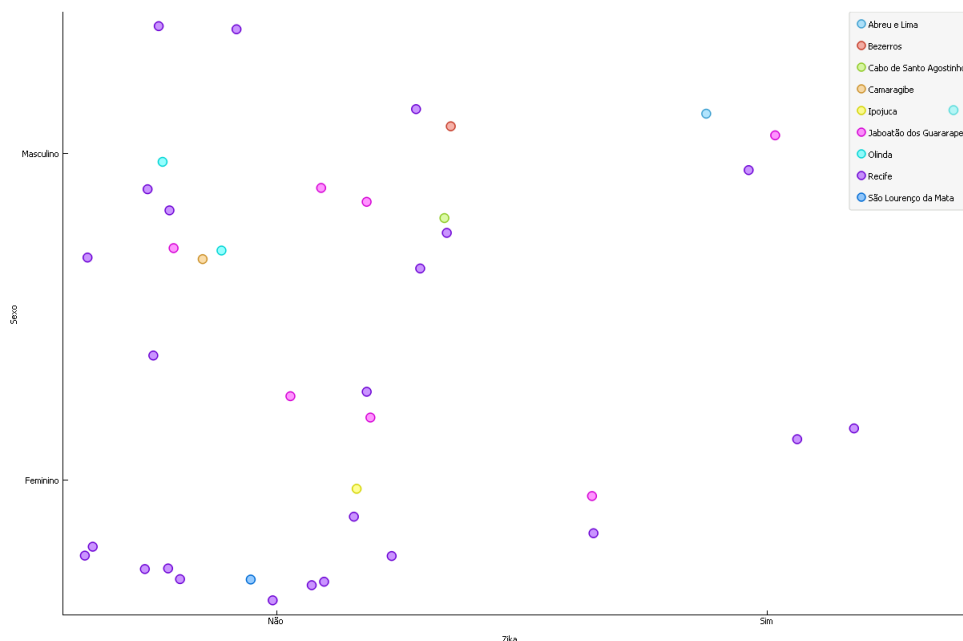
Fonte: Autoria própria (2018)

Figura 12 – Incidência de casos de dengue por região.



Fonte: Autoria própria (2018)

Figura 13 – Incidência de casos de zika por região.



Fonte: Autoria própria (2018)

4.2 Testes e Resultados

Das 41 amostras coletadas, separou-se 32 para treinamento do algoritmo de classificação e 9 para fins de teste. Utilizando o algoritmo Naive Bayes, o sistema conseguiu acertar 6 das 9 amostras durante os testes realizados, ou seja, atingiu 66,6% de acurácia, o que pode ser considerado uma taxa de sucesso boa dada a pouca quantidade de amostras utilizadas para treinamento do algoritmo. Também foram realizados testes com outros algoritmos para fins de análise e comparação com a abordagem utilizada neste trabalho.

Utilizando o algoritmo de indução baseado em regras, o sistema acertou apenas 4 das 9 amostras utilizadas (44,4% de acurácia) para teste, enquanto que o algoritmo de árvore de decisão acertou 5 das 9 amostras (55,5% de acurácia). É importante ressaltar que o algoritmo de árvore de decisão utilizado no teste foi o da biblioteca *scikit-learn*, presente na linguagem de programação Python, o que demonstra que mesmo bibliotecas consolidadas não conseguem obter um resultado satisfatório, em comparação ao Naive Bayes, quando trabalha-se com poucos dados.

Na Tabela 3 é possível ver os resultados das amostras selecionadas para teste utilizando o algoritmo Naive Bayes, junto às suas respectivas classificações, enquanto nas Tabelas 4 e 5 observa-se sua matriz de confusão e matriz de análise, respectivamente. Nas análises das matrizes de confusão é possível ver os verdadeiros positivos (VP), verdadeiros (Negativos), falsos positivos (FP) e falsos negativos (FN) obtidos por cada algoritmo.

Tabela 3 – Resultados obtidos usando o algoritmo Naive Bayes.

Amostra	Chikungunya	Dengue	Zika
Chikungunya	44.84%	16.21%	38.94%
Chikungunya	48.07%	28.55%	23.37%
Chikungunya	21.79%	76.12%	2.07%
Dengue	27.43%	30.86%	41.69%
Dengue	40.33%	50.91%	8.75%
Dengue	4.19%	92.95%	2.85%
Dengue	27.75%	39.16%	33.07%
Zika	27.76%	45.23%	27.00%
Zika	31.05%	21.73%	47.20%

Tabela 4 – Matriz de confusão usando o algoritmo Naive Bayes.

Classe Real	Previsto		
	Chikungunya	Dengue	Zika
Chikungunya	2	1	0
Dengue	0	3	1
Zika	0	1	1

Tabela 5 – Análise da matriz de confusão usando o algoritmo Naive Bayes.

Chikungunya	
(VP) 2	(VN) 4
(FP) 0	(FN) 1
Dengue	
(VP) 3	(VN) 3
(FP) 2	(FN) 1
Zika	
(VP) 1	(VN) 5
(FP) 1	(FN) 1
Acurácia	0,66

Na Tabela 6 é possível ver os resultados das classificações usando o algoritmo de indução baseado em regras, enquanto nas Tabelas 7 e 8 observa-se sua matriz de confusão e matriz de análise, respectivamente.

Tabela 6 – Resultados obtidos usando o algoritmo de indução baseado em regras.

Amostra	Resultado obtido
Chikungunya	Chikungunya
Chikungunya	Dengue
Chikungunya	Dengue
Dengue	Chikungunya
Dengue	Dengue
Dengue	Dengue
Dengue	Chikungunya
Zika	Dengue
Zika	Zika

Tabela 7 – Matriz de confusão usando o algoritmo de indução baseado em regras.

Classe Real	Previsto		
	Chikungunya	Dengue	Zika
Chikungunya	1	2	0
Dengue	2	2	0
Zika	0	1	1

Tabela 8 – Análise da matriz de confusão usando o algoritmo de indução baseado em regras.

Chikungunya	
(VP) 1	(VN) 3
(FP) 2	(FN) 2
Dengue	
(VP) 2	(VN) 2
(FP) 3	(FN) 2
Zika	
(VP) 1	(VN) 3
(FP) 0	(FN) 1
Acurácia	0,44

Na Tabela 9 é possível ver os resultados das classificações usando o algoritmo de árvore de decisão, enquanto nas Tabelas 10 e 11 observa-se sua matriz de confusão e matriz de análise, respectivamente.

Tabela 9 – Resultados obtidos usando o algoritmo de árvore de decisão

Amostra	Resultado obtido
Chikungunya	Dengue
Chikungunya	Chikungunya
Chikungunya	Chikungunya
Dengue	Dengue
Dengue	Dengue
Dengue	Chikungunya
Dengue	Chikungunya
Zika	Zika
Zika	Chikungunya

Tabela 10 – Matriz de confusão usando o algoritmo de árvore de decisão.

Classe Real	Previsto		
	Chikungunya	Dengue	Zika
Chikungunya	2	1	0
Dengue	2	2	0
Zika	1	0	1

Tabela 11 – Análise da matriz de confusão usando o algoritmo de árvore de decisão.

Chikungunya	
(VP) 2	(VN) 3
(FP) 3	(FN) 1
Dengue	
(VP) 2	(VN) 3
(FP) 1	(FN) 2
Zika	
(VP) 1	(VN) 4
(FP) 0	(FN) 1
Acurácia	0,55

Finalmente na Tabela 12 é possível ver uma comparação entre os resultados obtidos pelos algoritmos utilizados.

Tabela 12 – Comparação de resultados entre os algoritmos utilizados.

Algoritmo	Amostras classificadas corretamente	Acurácia obtida
Árvore de decisão	5	55,5%
Indução baseado em regras	4	44,4%
Naive Bayes	6	66,6%

É possível perceber observando a Tabela 3 que algumas das amostras classificadas incorretamente pelo sistema possuem probabilidades com valores próximos uns dos outros,

o que dá margem para o usuário decidir se quer aceitar a resposta fornecida pelo sistema ou não, no fim a decisão final sempre será a do usuário.

Ao realizar os testes em um dispositivo móvel com as especificações descritas na seção 3.2, o sistema obteve um tempo médio de 4 à 6 segundos durante a fase de treinamento e 2 à 3 segundos durante a fase de classificação. Considerando que a fase de treinamento será executada apenas 1 vez por dia, sendo ela feita antes do primeiro paciente ser consultado neste dia, chega-se a conclusão que o tempo é considerado satisfatório.

5 Conclusão

Este trabalho de conclusão de curso apresentou uma nova abordagem no que se refere ao uso de sistemas especialistas, trazendo esse recurso de uma forma mais conveniente e acessível para os padrões atuais. Um recurso poderoso que estará disponível na palma da mão de um profissional recém-graduado ou mesmo os que estão atuando em seu ofício por anos.

Também foi demonstrado que um sistema desse tipo pode ajudar, na maioria dos casos, a atender pacientes de forma eficiente mesmo possuindo uma base de conhecimentos pequena e que seu poder de classificação pode aumentar progressivamente a medida que mais amostras são fornecidas.

De maneira geral, o sistema apresentou resultados satisfatórios dada a base de conhecimentos que ele possuía, e que suas aplicações podem estender-se por outras áreas, sendo elas áreas da saúde com outro escopo de atuação ou áreas completamente diferentes, como análise de crédito em empresas bancárias. Como o sistema depende unicamente de uma base de conhecimentos para seu funcionamento, ao mudar o domínio da base, muda-se também o escopo de atuação, sendo feitas modificações mínimas no produto final (como interface com o usuário).

Um problema que pode surgir durante o uso do sistema é a avaliação equivocada do usuário que pode acabar avaliando erroneamente a classificação fornecida, dessa forma comprometendo a base de conhecimentos e conseqüentemente futuros diagnósticos realizados pelo sistema. Uma maneira na qual podemos tratar isso é realizando uma validação por meio de *feedback* material, ou seja, as avaliações só seriam incluídas na base de conhecimentos após validação por meio de resultados obtidos por exames laboratoriais.

Por fim, é importante ressaltar que apesar da autonomia do sistema, recomenda-se sempre a existência de um mediador utilizando-o, pois o uso incorreto por parte de um usuário inexperiente pode causar diversos problemas. Para este trabalho, o sistema desenvolvido foi criado visando a utilização por meio de um profissional certificado na área de saúde, caso um usuário não certificado utilizasse o sistema por conta própria, o mesmo poderia realizar ações problemáticas, como automedicação.

6 Trabalhos Futuros

Como já descrito anteriormente, o trabalho apresentado possui inúmeras possibilidades de aplicação, tendo apenas como requisito para sua replicação uma base de conhecimentos que possa ser utilizada pelo sistema e um especialista na área de atuação que possa auxiliar em seu desenvolvimento.

Para os trabalhos futuros, pode-se realizar as seguintes ações:

- Aperfeiçoar o classificador do sistema: Apesar de satisfatório, o classificador utilizado é bastante simples e pode ser substituído por classificadores mais complexos que forneçam melhor acurácia e desempenho;
- Estender as áreas de aplicação do sistema: Dados os resultados promissores obtidos neste trabalho, as aplicações para esse sistema podem ser estendidas para outras áreas de atuação, como diagnóstico de outros tipos de doenças.
- Integração com prontuários médicos: É possível a integração com sistemas digitais de prontuários médicos para obtenção de melhores resultados nas classificações, partindo da premissa que alguns resultados podem variar de acordo com o histórico médico do paciente;
- Integração com outros sistemas de classificação: É possível integrar o presente sistema em outros sistemas de classificação que utilizam *chatbots* para identificação de termos informais de sintomas.

Referências

- [1] LIMA, M. A. d. C. S. L. *Doenças*. Acesso em: 12/08/2018. Disponível em: <<https://mundoeducacao.bol.uol.com.br/doencas/>>. Citado na página 15.
- [2] BYNUM, W. História da medicina. *Souto Maior F, tradutora. Porto Alegre: L&PM*, 2011. Citado na página 15.
- [3] VEJA. *Erro médico mata mais que câncer no Brasil*. Acesso em: 05/06/2018. Disponível em: <<https://veja.abril.com.br/saude/erro-medico-mata-mais-que-cancer-no-brasil/>>. Citado na página 15.
- [4] WIKIPÉDIA. *Falso positivo*. Acesso em: 05/06/2018. Disponível em: <https://pt.wikipedia.org/wiki/Falso_positivo>. Citado na página 15.
- [5] BARONE, D. *Sociedades Artificiais - a Nova Fronteira da Inteligência Nas Máquinas*. Porto Alegre: Bookman, 2003. Citado 2 vezes nas páginas 16 e 24.
- [6] DINO. *A nova geração de aplicativos móveis*. Acesso em: 05/06/2018. Disponível em: <<https://exame.abril.com.br/negocios/dino/a-nova-geracao-de-aplicativos-moveis-apps-dino89089167131>>. Citado na página 16.
- [7] OLHARDIGITAL. *Mais de 70 milhões de brasileiros acessam a internet pelo celular*. Acesso em: 05/06/2018. Disponível em: <<https://olhardigital.com.br/noticia/mais-de-70-milhoes-de-brasileiros-acessam-a-internet-pelo-celular/51463/>>. Citado na página 16.
- [8] GIGERENZER, G.; GAISSMAIER, W. Heuristic decision making. *Annual Review of Psychology*, 2011. Citado na página 18.
- [9] GIGERENZER, G. *Simple heuristics that make us smart*. New York: Oxford University Press, 1999. Citado 2 vezes nas páginas 19 e 20.
- [10] LONGO, D. L. *Medicina Interna de Harrison.18. ed.* [S.l.]: Mc Graw Hill, 2012. Citado 2 vezes nas páginas 20 e 21.
- [11] SAKAMOTO, R. *Técnicas básicas de Exame Físico*. Acesso em: 05/06/2018. Disponível em: <<http://www.enfermeiroaprendiz.com.br/tecnicas-basicas-de-exame-fisico/>>. Citado na página 21.
- [12] KASSIRER, J. *Learning clinical reasoning*. Baltimore, MD: Lippincott Williams & Wilkins Health, 2010. Citado na página 22.

- [13] SANTOS, M. A. d. S. *Inteligência Artificial*. Acesso em: 01/08/2018. Disponível em: <<https://brasilecola.uol.com.br/informatica/inteligencia-artificial.htm/>>. Citado na página 23.
- [14] TURING, A. M. Computing machinery and intelligence. *New York Times*, 1950. Citado na página 23.
- [15] COELHO, H. *Inteligência artificial em 25 lições*. Lisboa: Fundação Calouste Gulbenkian, 1995. Citado na página 23.
- [16] TURING, A. M. *The essential Turing : seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma*. Oxford New York: Clarendon Press Oxford University Press, 2004. Citado na página 23.
- [17] CÂMARA, G. et al. Spring: Integrating remote sensing and gis by object-oriented data modelling. *Computers & Graphics*, 1996. Citado na página 23.
- [18] GHALLAB, M. *Automated planning: theory and practice*. Amsterdam Boston: Elsevier/Morgan Kaufmann, 2004. Citado na página 24.
- [19] WANGENHEIM, C. *Raciocínio baseado em casos*. Barueri: Manole, 2003. Citado na página 24.
- [20] DOMBAL, F. T. d. et al. Human and computer-aided diagnosis of abdominal pain: Further report with emphasis on performance of clinicians. *British Medical Journal*, 1974. Citado na página 24.
- [21] SILVA, C. *Montadoras investem em robotização, mesmo com crise e fábricas ociosas*. Acesso em: 01/08/2018. Disponível em: <<https://economia.estadao.com.br/noticias/geral,montadoras-investem-em-robotizacao-mesmo-com-crise-e-fabricas-ociosas,70001935089/>>. Citado na página 24.
- [22] THEODORIDIS, S. *Pattern recognition*. Amsterdam Boston: Elsevier/Academic Press, 2006. Citado na página 24.
- [23] RAY, S. *6 Easy Steps to Learn Naive Bayes Algorithm*. Acesso em: 18/07/2018. Disponível em: <<https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>>. Citado na página 27.
- [24] CAELUM. *Apostila java e orientação a objetos*. Acesso em: 12/08/2018. Disponível em: <<https://www.caelum.com.br/apostila-java-orientacao-objetos/o-que-e-java/#java/>>. Citado na página 32.
- [25] MYSQL. Acesso em: 01/08/2018. Disponível em: <<https://www.mysql.com/>>. Citado na página 33.

-
- [26] SQLITE. Acesso em: 01/08/2018. Disponível em: <<https://www.sqlite.org/index.html>>. Citado na página 33.
- [27] SCIKITLEARN. Acesso em: 01/08/2018. Disponível em: <<http://scikit-learn.org/>>. Citado na página 33.
- [28] DANTAS, C.; LENHARO, M. *Febre amarela, dengue, zika e chikungunya: entenda as doenças do Aedes que afetam o Brasil*. Acesso em: 03/06/2018. Disponível em: <<https://g1.globo.com/bemestar/noticia/febre-amarela-dengue-zika-e-chikungunya-entenda-as-doencas-do-aedes-que-afetam-o-brasil.ghtml>>. Citado na página 42.