

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

# Monitoramento Inteligente da Saúde: Aprendizado de Máquina na Prevenção de Doenças Relacionadas ao Estresse

Gabriel Fernandes Silva

JUIZ DE FORA  
DEZEMBRO, 2023

# Monitoramento Inteligente da Saúde: Aprendizado de Máquina na Prevenção de Doenças Relacionadas ao Estresse

GABRIEL FERNANDES SILVA

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Departamento de Ciências da Computação  
Bacharelado em Ciência da Computação

Orientador: Victor Stroele de Andrade Menezes

JUIZ DE FORA  
DEZEMBRO, 2023

MONITORAMENTO INTELIGENTE DA SAÚDE: APRENDIZADO  
DE MÁQUINA NA PREVENÇÃO DE DOENÇAS RELACIONADAS  
AO ESTRESSE

Gabriel Fernandes Silva

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS  
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-  
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE  
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Victor Stroele de Andrade Menezes  
Doutor em Engenharia de Sistemas e Computação pela UFRJ

Edelberto Franco Silva  
Doutor em Ciência da Computação pela UFF

Regina Maria Maciel Braga Villela  
Doutor em Engenharia de Sistemas e Computação pela UFRJ

Mário Antônio Ribeiro Dantas  
Doutor em Ciência da Computação pela University of Southampton (UK)

JUIZ DE FORA  
12 DE DEZEMBRO, 2023

*Aos meus amigos e irmãos.*

*Aos pais, pelo apoio e sustento.*

## Resumo

Doenças causadas por danos lentos e progressivos são a principal causa de óbitos. Os hábitos diários do indivíduo tem forte influência no funcionamento do organismo. O estresse vivenciado no decorrer do dia pode ser o causador de muitas enfermidades. Esse estresse é responsável por diminuir as defesas do organismo. Nesses casos, a prevenção é um componente fundamental, obtido pelo monitoramento dos indivíduos que ocorre, geralmente, através de um processo que dependente majoritariamente da intervenção humana. Em vista disso, torna-se necessário o desenvolvimento de soluções capazes de realizar um monitoramento automatizado para ajudar os indivíduos no dia a dia. O objetivo deste trabalho é promover a saúde dos indivíduos monitorando-os por meio de pulseiras inteligentes e apresentando notificações que os permitam adquirir mais conhecimento sobre si mesmos. Dessa forma, o trabalho foca no desenvolvimento de um ambiente computacional composto por uma pulseira e um aplicativo que integra a um modelo de aprendizado de máquina. Esse modelo é responsável por realizar a predição de dados de frequência cardíaca do usuário e, assim, gerar as notificações para o usuário. Os resultados mostram que é possível realizar o monitoramento em tempo real do usuário e identificar momentos de estresse usando um modelo de aprendizado de máquina e gerar uma notificação.

**Palavras-chave:** Internet das Coisas, Aprendizado de Máquina, Promoção da Saúde, Estresse

## Abstract

Diseases caused by slow and progressive damage are the leading cause of mortality. An individual's daily habits strongly influence the functioning of the body. The stress experienced throughout the day can be the cause of many illnesses, as it is responsible for diminishing the body's defenses. In such cases, prevention is a fundamental component achieved through monitoring individuals, usually through a process that heavily depends on human intervention. Therefore, the development of solutions capable of automated monitoring becomes necessary to assist individuals in their daily lives. The goal of this work is to promote individuals' health by monitoring them through smart bracelets and providing notifications that enable them to gain more knowledge about themselves. Thus, the work focuses on the development of a computational environment composed of a bracelet and an application integrated with a machine learning model. This model is responsible for predicting the user's heart rate data and generating notifications accordingly. The results show that real-time monitoring of the user is possible, and moments of stress can be identified using a machine learning model, leading to the generation of notifications.

**Keywords:** Internet of Things, Machine Learning, Health Promotion, Stress

## Agradecimentos

A toda minha família e amigos, pelo encorajamento e apoio.

Ao meu orientador Victor pela orientação, amizade e principalmente, pela paciência, que tornaram possível o desenvolvimento deste trabalho

Aos professores do Departamento de Ciência da Computação pelos seus ensinamentos e aos funcionários do curso, que durante esses anos, contribuíram de algum modo para minha formação pessoal e profissional.

*“Faça o teu melhor na condição que você tem, enquanto você não tem condições melhores para fazer melhor ainda!”*

*Mario Sergio Cortella*

# Conteúdo

<b>Lista de Figuras</b>	<b>7</b>
<b>Lista de Tabelas</b>	<b>8</b>
<b>Lista de Abreviações</b>	<b>9</b>
<b>1 Introdução</b>	<b>10</b>
<b>2 Fundamentação Teórica</b>	<b>15</b>
2.1 Internet das Coisas . . . . .	15
2.2 Data Lake . . . . .	17
2.3 Aprendizado de Máquina . . . . .	18
2.4 Arquitetura Lambda . . . . .	19
2.5 Considerações Finais do capítulo . . . . .	20
<b>3 Trabalhos Relacionados</b>	<b>21</b>
3.1 Machine Learning aplicado à Saúde . . . . .	21
3.2 Cuidados de saúde e detecção de anomalias . . . . .	22
3.3 Uma Arquitetura Fog-Cloud para o Monitoramento de Sinais Corporais . .	23
3.4 Revisão Sistemática sobre Algoritmos de Aprendizado de Máquina . . . .	24
3.5 Investigação sobre o uso de sensores corporais para diagnóstico de COVID-19	25
3.6 Uma proposta de arquitetura para apoiar notificações E-Health . . . . .	26
3.7 Considerações Finais do Capítulo . . . . .	27
<b>4 Materiais e Métodos</b>	<b>30</b>
<b>5 Resultados</b>	<b>36</b>
5.1 Olivia . . . . .	36
5.2 Pulseira Inteligente . . . . .	37
5.3 Modelo de Aprendizado de Máquina . . . . .	37
5.4 Notificações . . . . .	39
5.5 Limitações . . . . .	40
<b>6 Considerações Finais</b>	<b>42</b>
<b>Bibliografia</b>	<b>44</b>

## Lista de Figuras

2.1	Arquitetura básica IoT . . . . .	16
2.2	Crescimento global de dispositivos e conexões . . . . .	17
2.3	Três principais categorias de aprendizado de máquina: aprendizado não supervisionado, aprendizado supervisionado, aprendizado por reforço . . .	19
2.4	Arquitetura Lambda . . . . .	20
3.1	Ranking dos modelos preditivos para risco cardiovascular . . . . .	21
3.2	Classificação de anomalias com Random Forest em dados com 2,5% de anomalias . . . . .	23
3.3	Predição Gradient Boosting Regressor . . . . .	24
3.4	Sensores usados para monitoramento . . . . .	25
3.5	Rede Neural Recorrente . . . . .	26
3.6	Comparação entre frequência prevista e real . . . . .	28
4.1	Arquitetura Conceitual . . . . .	31
4.2	Boxplot do <i>dataset</i> de testes . . . . .	31
4.3	Histograma do <i>dataset</i> de testes . . . . .	31
4.4	Distribuição do <i>dataset</i> de testes em relação a localização . . . . .	32
4.5	Valores Preditos x Valores Reais no <i>dataset</i> de testes . . . . .	33
4.6	Implementação da Arquitetura . . . . .	35
5.1	Tela de Login e Monitoramento da Conexão . . . . .	37
5.2	Valores Preditos x Valores Reais . . . . .	38
5.3	Valores Preditos x Valores Reais . . . . .	39
5.4	Erro na leitura de informações da pulseira . . . . .	41

## Lista de Tabelas

3.1	Comparação entre os trabalhos relacionados . . . . .	29
-----	------------------------------------------------------	----

## Lista de Abreviações

DCC Departamento de Ciência da Computação

UFJF Universidade Federal de Juiz de Fora

# 1 Introdução

O desenvolvimento na medicina proporciona estratégias mais eficientes de saúde pública. Doenças responsáveis pela maior atenção dos profissionais da área séculos atrás não são as mesmas que promovem preocupação nos dias atuais. Doenças infecciosas ou causadas por falta de saneamento básico não são a principal causa de mortes (BAUER, 2002).

Na atualidade, a principal causa de mortes está relacionada a doenças causadas pelo acúmulo de danos lento e progressivo. Exemplos dessas doenças são: câncer, diabetes, doenças cardiovasculares e psicossomáticas (BAUER, 2002).

Segundo Bauer (2002), os hábitos diários do indivíduo tem forte influência no funcionamento do organismo. O estresse vivenciado no decorrer do dia pode ser o causador de muitas enfermidades. Esse estresse é responsável por diminuir as defesas do organismo levando ao desenvolvimento de doenças como o câncer e aumento da susceptibilidade a infecções (BAUER, 2002).

Com o avanço tecnológico a prevenção se torna menos invasiva no cotidiano das pessoas. Assim, a saúde pode ser monitorada de forma contínua sem a necessidade de demandar uma quantidade alta de esforço diário do indivíduo (MASSOLA; PINTO, 2018).

Dentro do escopo da medicina, a prevenção se mostra de suma importância para facilitar a promoção da saúde. Um diagnóstico precoce aumenta a chance de sucesso de diversos tratamentos. O setor primário da saúde é o principal responsável por esse acompanhamento (NORMAN; TESSER, 2009).

Esse monitoramento é feito de diversas formas. Um deles é através de um agente de saúde que frequenta periodicamente a casa dos indivíduos com o intuito de atender as necessidades médicas dos mesmos, como exames periódicos, vacinas, reforço, entre outras. Os agentes passam as demandas para os postos de saúde responsáveis pelas respectivas regiões e assim a demanda médica é repassada.

O acompanhamento permite ao sistema de saúde estabelecer medidas preventivas de forma mais assertiva a partir das informações coletadas. Esse procedimento também é benéfico para pessoas com alguma comorbidade de risco, como doenças cardiovasculares.

A periodicidade do processo permite a análise de desenvolvimento do quadro (MASSOLA; PINTO, 2018).

A motivação deste estudo é obter informações em tempo real dos usuários e informá-los da sua situação atual. A partir dessas notificações, permitir que as pessoas adquiram mais conhecimento sobre si mesmos.

Em virtude da prevenção ser uma medida fundamental na promoção da saúde, este trabalho propõe que o autoconhecimento absorvido pelo usuário irá auxiliar seu acompanhamento de saúde. Assim é possível que o usuário entenda o cenário no qual ele se encontra e, dessa forma, tomar as providências necessárias. Apesar do uso de sensores corporais, a proposta é que o sistema funcione no dia a dia do usuário sem interferência no seu cotidiano.

Essas informações obtidas através do aplicativo são complementares aos procedimentos preventivos adotados pelo usuário. Pessoas com comorbidade cardíaca podem ter um acompanhamento mais detalhado do seu estado ao longo do dia.

Em vista do processo citado, esse monitoramento se mostra muito dependente da ação humana. Essa relação gera uma janela de desinformação entre um ponto de contato e outro. Seja pela omissão de alguma informação importante por parte do paciente, ou por descuido do agente de saúde. Dessa forma, aprimoramos esse acompanhamento do indivíduos através do uso de sensores corporais.

Com o intuito de solucionar o problema mencionado anteriormente, o monitoramento é realizado nesse estudo através de pulseiras inteligentes que são responsáveis pela coleta dos sinais vitais do usuário. Com esse aparelho fixado no pulso do indivíduo é captada a frequência cardíaca.

Os dados captados pela pulseira são enviados para um aplicativo instalado no dispositivo móvel do usuário. O aplicativo é responsável por enviar os dados para a aplicação responsável pela tratativa desses dados. O aplicativo também fornece para a aplicação dados referentes a localização do usuário.

Ao receber os dados a aplicação utiliza um modelo de aprendizado de máquina cujo propósito é encontrar instabilidade nos dados coletados, ou seja, medições que fogem do padrão apresentado pelo indivíduo que está sendo monitorado. A partir das in-

formações extraídas no processo da execução do algoritmo, outra aplicação gera, quando necessário, uma notificação para o usuário no aplicativo em seu dispositivo.

É fundamental que o processo de captação do dado até a saída da notificação no dispositivo do usuário aconteça de forma rápida. O sistema proposto neste trabalho deve monitorar usuários em tempo real. Dessa forma, a arquitetura de execução da aplicação precisa ser escalável, de forma que seja sustentável com o aumento do número de usuários.

Considerando a complexidade computacional envolvida no grande fluxo de dados gerados continuamente, foi usada a cooperação *Fog-Cloud*. Essa proposta promove uma arquitetura descentralizada, que permite o processamentos de dados em maior escala. Assim, é possível evitar um congestionamento no serviço final (*Cloud*), mantendo parte do processamento computacional em ponto intermediário *Fog*.

Como o sistema foi desenvolvido para trabalhar em tempo real, é importante ressaltar que ele pode ser usado tanto como alertas em casos de instabilidade nos dados captados pelos sensores como na análise de dados históricos para repasse médico periódico. Completando a anamnese realizada pelos profissionais da saúde. Acredita-se que as notificações podem estimular o usuário em tempo real para a adoção de hábitos mais saudáveis, visto que ao identificar instabilidade o mesmo será avisado.

Durante o desenvolvimento do trabalho existem alguns desafios. Dentre eles está extrair os dados dos sensores da pulseira e localização de forma consistente para gerar informações relevantes para o usuário.

Outro desafio é explorar um modelo de aprendizado de máquina que consiga fornecer informações confiáveis a partir dos dados inseridos. Além de ser otimizado, visto que os dados gerados pelo modelo possuem um intervalo de validade, podendo se tornar obsoleto caso o processo crie um gargalo.

Dito isso, outro fator fundamental para o bom funcionamento do sistema é a arquitetura implementada. Ela é responsável por garantir que o fluxo contínuo de dados coletados passem por todas as camadas até a saída do dado para o usuário. É importante que a sua estrutura não seja obstruída durante um grande volume de dados gerados continuamente pelos sensores.

Dessa forma, a questão de pesquisa elaborada para esse trabalho é a seguinte:

- *A partir da extração de dados em tempo real de uma pulseira inteligente e processamento por um modelo de aprendizado de máquina, é possível gerar notificações relevantes para auxiliar o usuário na promoção da saúde pessoal?*

Nesse trabalho foi usado como inspiração o trabalho desenvolvido por Leão et al. (2023), que realiza o monitoramento do usuário em tempo real usando dados de frequência cardíaca. Este trabalho de conclusão de curso propõe um avanço em relação ao trabalho de inspiração em virtude do desenvolvimento de um aplicativo para coleta de dados de forma mais ubíqua e automatizada, além da ampliação da dimensionalidade dos dados (localização) com intuito de enriquecê-los para aprimorar a identificação de momentos de estresse no cotidiano do usuário.

Considerando a questão de pesquisa norteadora deste trabalho, o objetivo principal é o desenvolvimento de um ambiente computacional formado por um aplicativo para dispositivos móveis Android conectado a uma pulseira inteligente, que fornece dados para o processamento de um modelo de aprendizagem de máquina com o intuito de gerar informações para auxiliar o usuário na promoção da saúde.

Como objetivos específicos da presente pesquisa pretende-se:

1. Definição da arquitetura do ambiente computacional;
2. Desenvolvimento de um aplicativo para realizar a coleta dos dados de forma consistente;
3. Desenvolvimento de um modelo de aprendizado de máquina para processamento dos dados;
4. Desenvolvimento de um fluxo contínuo de treinamento do modelo para mantê-lo atualizado;
5. Desenvolvimento de serviços para que todos os dados coletados sejam armazenados de forma segura, garantindo a privacidade dos usuários.

Segundo Pereira et al. (2018) e Creswell (2003) esse estudo pode ser classificado como uma **pesquisa exploratória** em virtude da busca por informações na área da saúde e computação para desenvolvimento de uma arquitetura que utiliza sensores corporais para

---

extrair dados do usuário com o intuito de gerar informações relevante. Ademais, essa classificação pode ser complementada como uma **pesquisa bibliográfica e aplicada**, visto que foi necessária a busca de conhecimento em livros e artigos científicos, além de envolver o desenvolvimento de um ambiente computacional.

Este trabalho está organizado da seguintes forma: no Capítulo 2 é apresentada a fundamentação teórica; no Capítulo 3 são apresentados os trabalhos relacionados; no Capítulo 4 é detalhada a proposta desse trabalho; no Capítulo 5 são descritos os resultados obtidos; e no Capítulo 6 são apresentadas as conclusões e trabalhos futuros.

## 2 Fundamentação Teórica

Neste capítulo são explicados alguns conceitos fundamentais para o desenvolvimento deste trabalho. Ele é composto por quatro seções de fundamentos: A Seção 2.2 explica os fundamentos da Internet das Coisas e seu funcionamento; Na Seção 2.3 discorre-se sobre o conceito de *Data Lake*; Já na Seção 2.4 é discutido sobre o aprendizado de máquina e suas principais características; Na Seção 2.5 é apresentada a arquitetura Lambda.

Cada seção detalha os princípios de cada conceito e suas principais aplicações. Dessa forma, é possível entender todo o contexto que o trabalho usa para seu desenvolvimento. Por fim, a Seção 2.6 aborda as considerações finais do capítulo, fazendo a conexão de como cada fundamento se aplica no desenvolvimento desse trabalho.

### 2.1 Internet das Coisas

A Internet das Coisas, ou *IoT - Internet of Things* o termo em inglês, é uma tecnologia emergente que permite a conexão de dispositivos físicos, como sensores, à internet. Esses dispositivos são capazes de coletar e transmitir dados em tempo real, possibilitando o monitoramento e o controle remoto de processos e sistemas. A *IoT* tem o potencial de transformar diversos setores da sociedade, desde a indústria até a saúde, trazendo benefícios como aumento da eficiência, redução de custos e melhoria na qualidade de vida das pessoas (SANTOS et al., 2016).

A Internet das Coisas, de forma resumida, é uma extensão da internet, onde objetos com recursos computacionais e de comunicação se conectam a internet. Essa conexão permite que os objetos forneçam diferentes dados para a rede de computadores. De acordo com Tanenbaum (2002), "Rede de Computadores é um conjunto de computadores autônomos interconectados por uma tecnologia", essa tecnologia pode ser de diversos tipos, como fibra ótica, fios de cobre, ondas eletromagnéticas ou outras.

Os objetos com capacidade computacional e sensores são fundamentais para evolução desse conceito de rede. Pois essas novas características permitem que esses

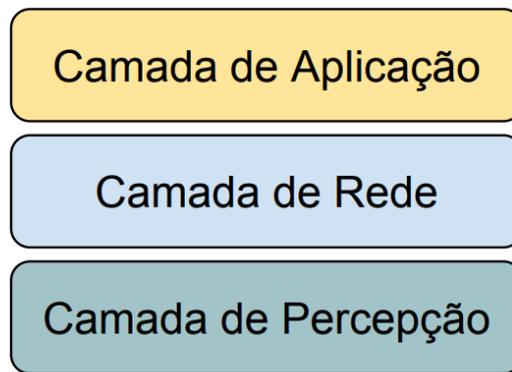


Figura 2.1: Arquitetura básica IoT (SANTOS et al., 2016)

objetos se conectem a rede, e não somente os computadores convencionais. Entre esses objetos estão pulseiras inteligentes, automóveis, câmeras, aparelhos domésticos, como geladeiras entre muitos outros. Essas características permitem expandir a utilidade desses objetos dentro da sociedade.

Segundo Santos et al. (2016) o funcionamento da *IoT* é baseado em três elementos principais como mostra a Figura 2.1: dispositivos conectados (Camada de Percepção), redes de comunicação (Camada de Rede) e plataformas de processamento (Camada de Aplicação) de dados. Os dispositivos conectados à internet são equipados com sensores que permitem coletar dados sobre o ambiente físico, processá-los e enviá-los para a rede. As redes de comunicação são responsáveis por conectar esses dispositivos à internet e podem utilizar diferentes tecnologias, como Wi-Fi, Bluetooth, ZigBee e 4G/5G. As plataformas de processamento de dados fornecem o processamento e armazenamento de grandes quantidades de dados gerados pelos dispositivos conectados, permitindo a extração de insights valiosos que podem ser usados para melhorar os processos, serviços e experiências do usuário.

As aplicações da *IoT* são amplas e diversificadas. Na saúde, a *IoT* pode ser utilizada para monitorar a saúde de pacientes em tempo real, coletando informações vitais e permitindo intervenções precoces em caso de emergência.

Em um relatório realizado pela Cisco (2018), mostra que até 2023, serão mais de 30 bilhões de dispositivos IoT em todo o mundo, o que representa um aumento significativo

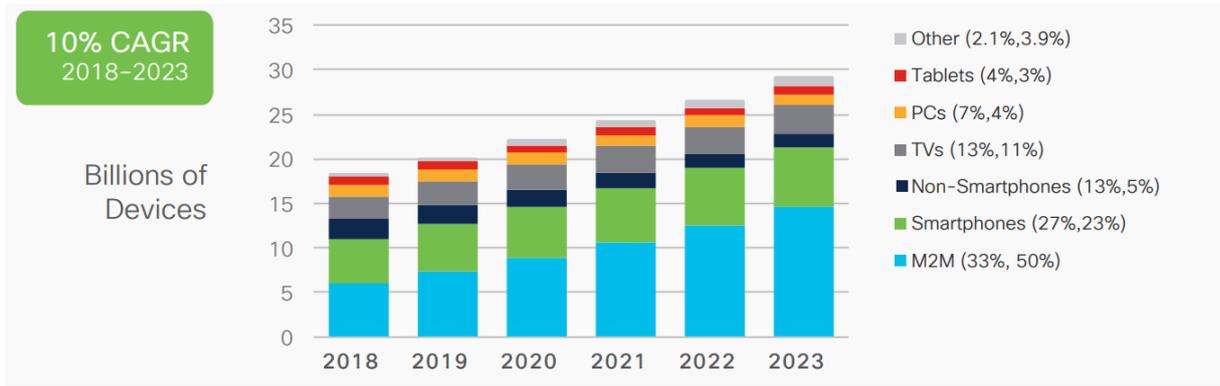


Figura 2.2: Crescimento global de dispositivos e conexões (CISCO, 2018)

em relação aos cerca de 13 bilhões em 2020. Esse desenvolvimento mostra uma perspectiva de crescimento acelerado para essa tecnologia.

## 2.2 Data Lake

O *Data Lake* é uma solução de armazenamento de dados que tem como objetivo permitir que empresas e organizações armazenem grandes volumes de dados estruturados e não estruturados em seu formato bruto, sem a necessidade de pré-processamento ou modelagem. O conceito de *Data Lake* foi proposto pela primeira vez por Dixon (2010), e tem se tornado cada vez mais popular como uma alternativa ao armazenamento de dados em *data warehouses* tradicionais (INMON, 2002).

Diferentemente dos *data warehouses*, que são projetados para armazenar dados estruturados em um esquema definido e pré-configurado, como definido por Inmon (2002), o *Data Lake* permite que os dados sejam armazenados em seu formato bruto, sem limitações em relação a sua estrutura, formato ou volume, segundo Dixon (2010). Isso significa que o *Data Lake* pode armazenar dados estruturados, semi-estruturados e não estruturados, provenientes de diversas fontes, tais como sensores, dispositivos *IoT*, sistemas transacionais, redes sociais, entre outros.

De acordo com Dixon (2010), as características do *Data Lake* incluem escalabilidade, flexibilidade, agilidade e baixo custo. Sua arquitetura permite que os dados sejam armazenados em grande escala, o que o torna uma solução ideal para empresas que precisam lidar com grandes volumes de dados. A flexibilidade do *Data Lake* permite que

os dados sejam armazenados em seu formato bruto, sem a necessidade de modelagem ou pré-processamento, o que aumenta a agilidade no acesso e análise dos dados. Além disso, o Data Lake apresenta baixo custo em comparação com os *data warehouses* tradicionais, uma vez que utiliza tecnologias de código aberto e armazenamento em nuvem.

As aplicações do *Data Lake* são amplas e diversificadas. Na área de negócios, o Data Lake pode ser utilizado para análise de dados, *business intelligence*, inteligência artificial e aprendizado de máquina. Na área de ciência de dados, o Data Lake pode ser usado para análise de dados complexos, *Big Data* e modelagem estatística. Na área de tecnologia da informação, o *Data Lake* pode ser aplicado para armazenamento e gerenciamento de dados, backup e recuperação de desastres

## 2.3 Aprendizado de Máquina

O aprendizado de máquina é uma área da ciência da computação que atualmente tem um papel fundamental na tomada de decisões automatizadas em muitos domínios, como finanças, medicina, transporte, entre outros. De acordo com Jordan. e Mitchell (2015), o aprendizado de máquina é uma área da inteligência artificial que se concentra em dar às máquinas a capacidade de aprender com dados, sem a necessidade de programação explícita.

Os algoritmos de aprendizado de máquina podem ser divididos em três categorias principais como mostra a figura: 2.3: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. Segundo Murphy (2012), o aprendizado supervisionado é um processo em que o algoritmo aprende a partir de dados rotulados, ou seja, dados que já foram classificados previamente por um ser humano. Já o aprendizado não supervisionado é um processo em que o algoritmo aprende a partir de dados não rotulados, encontrando padrões e estruturas ocultas nos dados. O aprendizado por reforço é um processo em que o algoritmo aprende a partir de tentativa e erro, recebendo recompensas ou penalidades em função das ações que toma em determinado ambiente.

Os algoritmos de aprendizado de máquina têm sido cada vez mais utilizados em muitas aplicações práticas. Por exemplo, o reconhecimento de voz, a detecção de fraudes em cartões de crédito e a previsão de demanda em empresas são algumas das aplicações

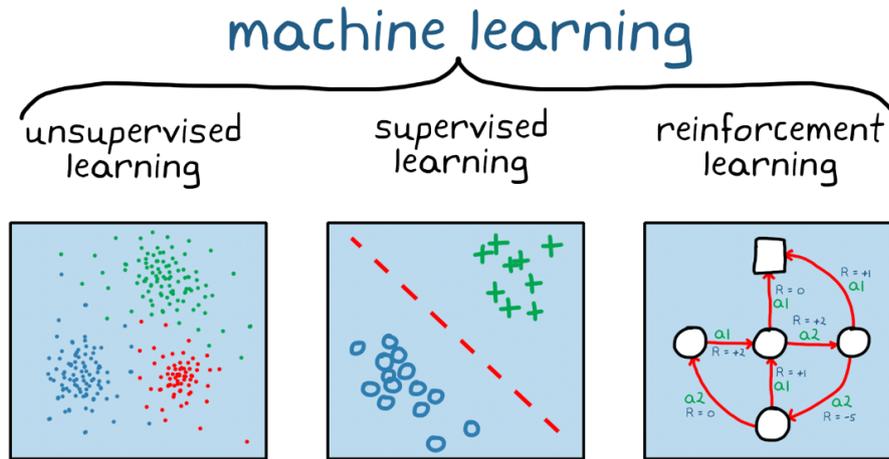


Figura 2.3: Três principais categorias de aprendizado de máquina: aprendizado não supervisionado, aprendizado supervisionado, aprendizado por reforço

que fazem uso desses algoritmos. Segundo Domingos (2015), o aprendizado de máquina tem o potencial de revolucionar muitas áreas, tornando possível o desenvolvimento de sistemas mais inteligentes e autônomos.

## 2.4 Arquitetura Lambda

A arquitetura lambda é um modelo de processamento de dados em tempo real que tem sido amplamente utilizado em sistemas de grande escala, permitindo que grandes volumes de dados sejam processados e analisados em tempo hábil. Segundo Kleppmann (2015), a arquitetura lambda é um modelo de processamento de dados em várias camadas que combina processamento em tempo real e em lote.

A arquitetura lambda é composta por quatro componentes principais: camada de ingestão de dados, camada de processamento em tempo real, camada de processamento em lote e armazenamento de dados, segundo Kleppmann (2015). A camada de ingestão de dados é responsável por receber os dados de entrada, enquanto a camada de processamento em tempo real processa os dados em tempo real. A camada de processamento em lote processa os dados de forma diferida, utilizando ferramentas como o *Apache Hadoop*.

A arquitetura lambda é altamente escalável e pode ser usada para lidar com grandes volumes de dados, permitindo que os dados sejam processados em tempo real, bem como em lotes, permitindo que a arquitetura possa ser usada em uma ampla variedade

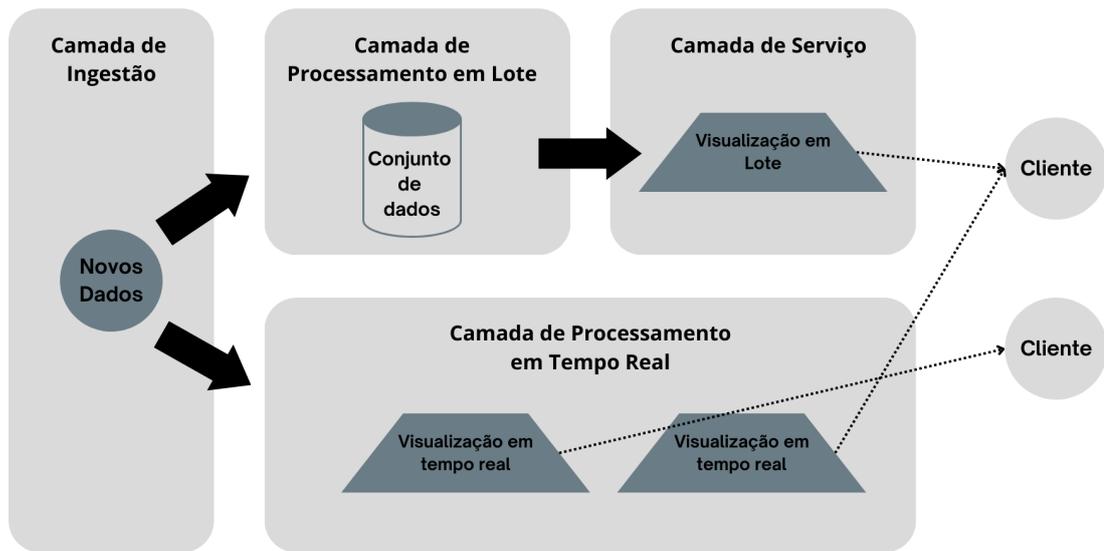


Figura 2.4: Arquitetura Lambda

de aplicações, como análise de dados em tempo real, processamento de dados em grande escala, detecção de anomalias e detecção de fraudes (KLEPPMANN, 2015).

## 2.5 Considerações Finais do capítulo

Nesse capítulo, foram introduzidos alguns conceitos essenciais para o entendimento do trabalho. Primeiro, foi descrito o funcionamento da Internet das Coisas, junto a sua arquitetura básica, em seguida foi apresentado o conceito de *Data Lake*, que será fundamental para a implementação do trabalho proposto, em vista da sua versatilidade.

Além disso, foi apresentado também o conceito de aprendizado de máquina, e suas principais características, pois serão fundamentais para aplicação do modelo proposto que possui um sistema de treinamento supervisionado. E por fim, foi introduzido a arquitetura Lambda que fornece o suporte necessário para o processamento dos dados em tempo real e em lote.

Todos os conceitos apresentados nesse capítulo contribuem de forma complementar para o desenvolvimento desse trabalho.

## 3 Trabalhos Relacionados

Com o intuito de recuperar trabalhos relacionados a pesquisa deste trabalho de conclusão de curso foram realizadas buscas nas bases de dados do *Google Scholar* e *IEEE* utilizando palavras-chave como *Machine Learning E-Health*, aprendizado de máquina na saúde, *Machine Learning Heart Rate*, *IoT* na saúde, monitoramento em tempo real.

Dessa forma, este capítulo apresenta os trabalhos selecionados que auxiliaram no desenvolvimento deste trabalho. Cada seção explica um trabalho relacionado e ao final do capítulo uma seção de considerações finais apresenta as contribuições deste trabalho em relação aos trabalhos apresentados.

### 3.1 Machine Learning aplicado à Saúde

O trabalho desenvolvido por Batista e Filho (2019) apresenta o panorama de uso das técnicas de *Machine Learning* aplicadas na área de saúde. Ele foi desenvolvido com o intuito de utilizar algoritmos de classificação para identificar a presença de doenças cardiovasculares a partir de dados clínicos do paciente, com dados contínuos e categóricos, como idade e se o paciente é fumante.

Para o desenvolvimento dos modelos preditivos foram utilizados os dados sintéticos criados por Laderas et al. (2018). Esses dados foram preparados e os autores optaram por filtrar o conjunto de dados para registros de indivíduos com idade superior a 55 anos. A preparação consistiu na padronização das variáveis contínuas e decomposição das variáveis categóricas em um conjunto de variáveis indicadoras.

Modelo	AUC	Precisão	Sensibilidade	Brier
Gradient Boosted Trees	0,780	0,647	0,344	0,162
Redes Neurais Artificiais	0,780	0,645	0,341	0,161
Random Forest	0,780	0,643	0,342	0,162
K-vizinhos mais próximos	0,770	0,633	0,349	0,164
Regressão Logística	0,770	0,629	0,363	0,163
Árvore de Decisão	0,750	0,617	0,311	0,169

Figura 3.1: Ranking dos modelos preditivos para risco cardiovascular (BATISTA; FILHO, 2019)

A Figura 3.1 mostra o desempenho dos algoritmos avaliados, em ordem decrescente por AUC(*area under curve*) e precisão. Como visto na figura, o algoritmo *Gradient Boosted Trees* apresentou maior eficiência entre os modelos elaborados no estudo.

Dessa forma, o estudo em questão mostrou que os algoritmos possibilitaram o ajuste de modelos preditivos em um conjunto de dados sintéticos sobre o risco cardiovascular, apresentando resultados próximos. Como ponto negativo vale ressaltar a ausência de dados reais na consolidação da eficiência dos algoritmos.

## 3.2 Cuidados de saúde e detecção de anomalias

Sabic et al. (2020) realizaram um estudo para detectar anomalias nos dados de frequência cardíaca através de modelos de aprendizado de máquina. Foram analisados cinco algoritmos nessa pesquisa utilizando duas bases de dados simuladas, uma com 0,5% de anomalias e outra com 2,5% de anomalias.

Segundo Sabic et al. (2020), o processo de detecção de anomalias necessita de uma definição operacional que assume duas verdades, (1) as anomalias são raras na ocorrência e (2) as anomalias são diferentes de dados normais. O trabalho desenvolvido por Sabic et al. (2020) avaliou os algoritmos *Local Outlier Factor (LOF)*, *K-Nearest Neighbors (KNN)*, *Support Vector Machines (SVM)*, *Isolation Forest (IF)* e *Random Forest (RF)* na abordagem de aprendizado supervisionado e não supervisionado.

Os dados foram gerados por um simulador e não foram inicialmente rotulados. No entanto, com o intuito de aprimorar o aprendizado dos modelos foi estipulado que dados de frequência cardíaca fora da faixa de 60 a 100 bpm (batimentos por minuto) sejam considerados como uma anomalia.

Após realizar o experimento nos modelos, foi possível observar que os modelos se adaptaram de formas diferentes em relação a base de dados, onde *KNN* e *SVM* apresentaram melhor resultado no conjunto de dados de 0,5% de anomalias, enquanto o *Random Forest* apresentou o melhor desempenho no conjunto de 2,5%, como mostra a Figura 3.2 com marcações em vermelho onde foi detectada uma anomalia. Nesse estudo, Sabic et al. (2020) mostrou que o uso de dados simulados podem ajudar a aprimorar algoritmos quando dados reais rotulados não estão disponíveis.

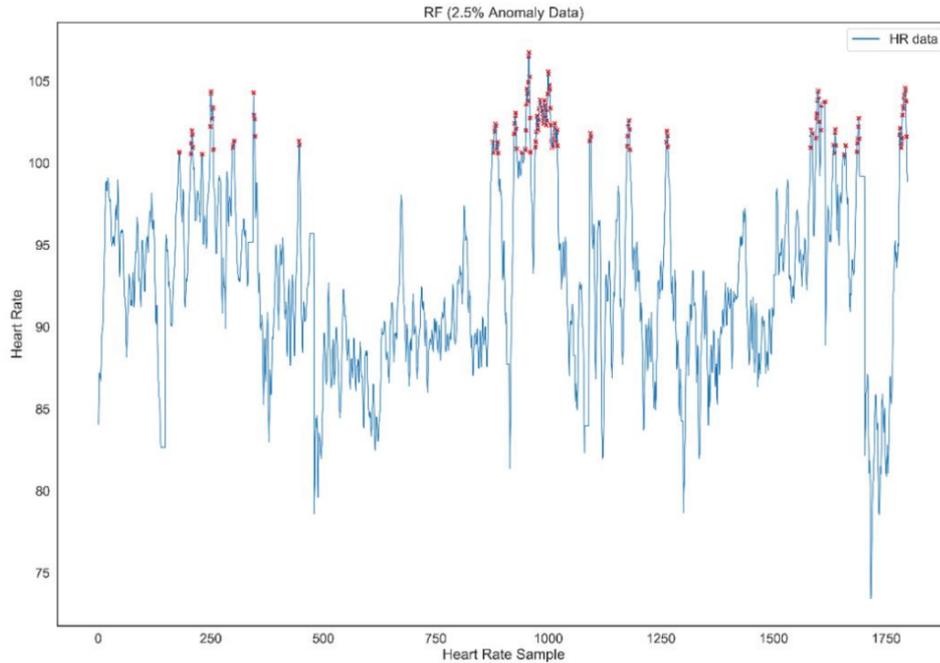


Figura 3.2: Classificação de anomalias com Random Forest em dados com 2,5% de anomalias

(SABIC et al., 2020)

Como ponto negativo do trabalho é importante ressaltar que, apesar dos dados simulados atingirem uma relevância significativa no desenvolvimento do modelo, a ausência de dados reais podem tornar os modelos limitados.

### 3.3 Uma Arquitetura Fog-Cloud para o Monitoramento de Sinais Corporais

Nesse trabalho, Leão et al. (2021) realizaram um estudo sobre uma proposta de arquitetura Fog-Cloud que coleta dados de sensores corporais do usuário, usando *Machine Learning* para determinar momentos de estresse do usuário.

Durante o estudo foram utilizados 4 algoritmos de aprendizado de máquina, *Random Forest Regressor*, *Decision Trees*, *K-Nearest Neighbors Regressor* e *Gradient Boosting Regressor*, para analisar os dados coletados e comparar os resultados obtidos. A Figura 3.3 mostra os resultados obtidos pelo algoritmo *Gradient Boosting Regressor* onde o eixo X representa a frequência real coletada pela pulseira e o eixo Y representa a frequência cardíaca prevista pelo algoritmo. Os pontos localizados abaixo da diagonal representam

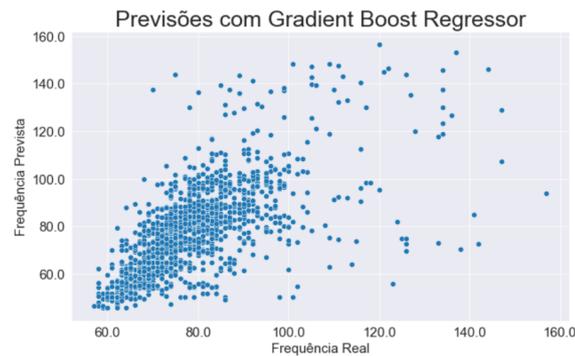


Figura 3.3: Predição Gradient Boosting Regressor

as medições que o algoritmo previu uma frequência cardíaca mais baixa que a real, dessa forma, esses pontos caracterizam momentos de estresse.

O trabalho mostrou a viabilidade de identificar momentos de estresse do usuário com base na frequência cardíaca e localização do usuário. Como trabalho futuro foi apontado a ampliação dos parâmetros para aprimorar as previsões feitas pelos algoritmos.

## 3.4 Revisão Sistemática sobre Algoritmos de Aprendizado de Máquina

Nesse trabalho, Site, Nurmi e Lohan (2021) reuniram 67 estudos sobre monitoramento da saúde com uso de algoritmos de aprendizado de máquina. Nesse estudo foram analisados diferentes tipos de sensores para obter os dados de saúde dos indivíduos (Figura 3.4). Esses sensores foram utilizados para monitorar os sinais vitais de pessoas com doenças crônicas, como doenças cardíacas, diabetes, distúrbios sanguíneos, entre outras.

Esse trabalho analisou as diferentes técnicas de processamento e análise nos dados da saúde. O resultado obtido identificou diferentes tipos de algoritmos e seus respectivos desempenhos nos cenários de cada estudo analisado, assim como os principais sensores usados. A maioria dos dados foram obtidos através de sensores como acelerômetros, monitores EEG (Eletroencefalogramas), ECG (Eletrocardiogramas) e dispositivos vestíveis/*smartwatches*.

Como resultado desse trabalho, foram encontradas diferentes características que podem ser extraídas dos dados coletados. Os estudos analisados utilizaram aprendizado

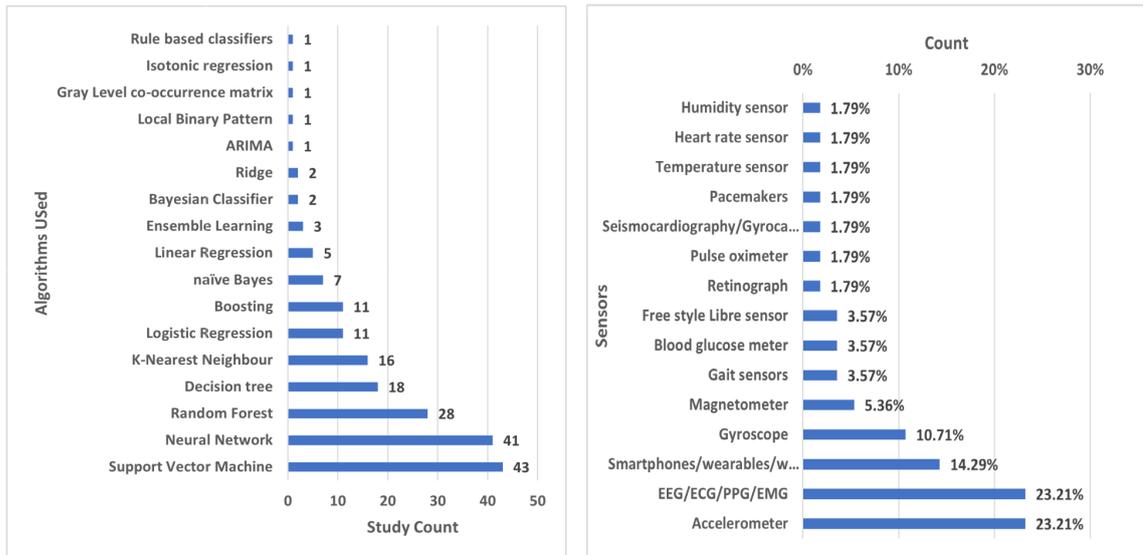


Figura 3.4: Sensores usados para monitoramento (SITE; NURMI; LOHAN, 2021)

supervisionado para treinar os respectivos modelos, outra informação relevante é que alguns estudos utilizaram métodos de seleção de características para selecionar as características mais relevantes para realizar a predição/detecção/classificação.

### 3.5 Investigação sobre o uso de sensores corporais para diagnóstico de COVID-19

A pesquisa desenvolvida por Risch et al. (2022), durante a pandemia do COVID-19, teve como objetivo a investigação dos dados obtidos de usuários para concluir se o indivíduo estava infectado com o vírus do COVID ou saudável. Segundo o estudo, os indivíduos infectados apresentavam alterações nos dados coletados pelos sensores corporais dias antes de apresentarem sintomas da gripe. Assim, seria possível detectar essa oscilação através de um algoritmo desenvolvido no trabalho.

Durante esse estudo, os autores utilizam uma Rede Neural Recorrente (RNN, na sigla em inglês) com células longas de memória de curto prazo (LSTM, na sigla em inglês) para realizar uma classificação binária de um indivíduo em um determinado dia como saudável ou infectado (positivo para COVID-19). A estrutura dessas redes podem memorizar e facilmente buscar ou ativar esses estados, mesmo que não tenham sido criados

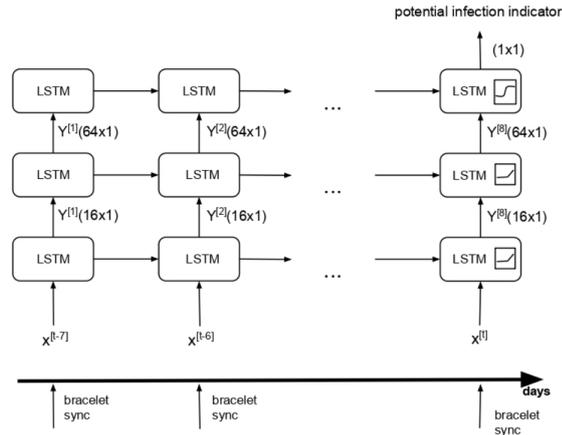


Figura 3.5: Rede Neural Recorrente

recentemente. A implementação usada nesse projeto consiste em duas camadas ocultas de 16 e 64 células, como mostra a Figura 3.5. A entrada dessa rede foram oito valores de sinais fisiológico originados de oito noites consecutivas de dados e a saída foi a indicação sobre a potencial infecção do indivíduo por COVID-19.

O estudo foi realizado com 1163 pessoas, onde 127 pessoas contraíram o vírus. No conjunto de testes, o algoritmo identificou 68% dos casos 2 dias antes do início dos sintomas. Dessa forma, o resultado do estudo mostrou a eficiência do uso de aprendizado de máquina na saúde, prevendo informações relevantes para tomada de decisão mais assertiva em benefício do indivíduo e da sociedade.

Esse trabalho mostrou que o uso de *Machine Learning* em dados extraídos de sensores corporais pode gerar uma informação relevante para apoiar a tomada de decisão pelos profissionais da saúde. Dessa forma, essas novas informações ajudaram a tomada de medidas preventivas para evitar a proliferação do vírus COVID-19 na sociedade.

### 3.6 Uma proposta de arquitetura para apoiar notificações E-Health

Nesse estudo, Leão et al. (2023) propõem uma arquitetura que permita que dados coletados a partir de sensores corporais sejam processados por um modelo de aprendizado de máquina com o objetivo de gerar notificações para o usuário quando são identificados momentos de estresse.

O trabalho em questão utilizou o conceito *Fog-Cloud* aliado com a implementação da arquitetura *Lambda* para superar os desafios de desenvolver um sistema de monitoramento em tempo real. Dessa forma, foi utilizado um aplicativo móvel para coletar dados de um *smartwatch*. Ao coletar os dados de frequência cardíaca do usuário a cada 5 min eles são enviados para um serviço de *Data Lake* através de uma requisição HTTP que foi implementado em um formato de API WEB. O serviço de armazenamento disponibiliza esses dados via API WEB para outro serviço onde os dados são processados para realizar as predições. Dessa forma, após realizar a predição esse resultado é comparado com o valor real coletado, visto que o resultado da comparação apresenta uma discrepância significativa, uma requisição é feita a um serviço de agendamento para agendar uma notificação para aquele usuário específico.

Durante o desenvolvimento do trabalho foram utilizados vários modelos de aprendizado de máquina, mas para fins de validação foi utilizado o *K-Nearest Neighbor Regressor (KNN)*. A Figura 3.6 mostra uma diferença entre o valor previsto e o de fato coletado pelos sensores no intervalo de 18 horas às 20 horas, onde a frequência prevista possui um valor menor do que a frequência cardíaca do dado coletado, o que indica que o usuário estava em um momento de estresse.

Os pontos positivos encontrados no estudo foi a validação da viabilidade da arquitetura para monitoramento do usuário em tempo real. Assim foi possível identificar momentos de estresse e notificá-lo com o objetivo de promover o autoconhecimento do usuário juntamente com o histórico de dados que podem ser usados por profissionais da saúde para compreender o comportamento do paciente. Como ponto negativo, citado pelos autores, está o restrito número de usuário no qual o experimento foi submetido. Ademais, foi sugerido pelos autores, como trabalhos futuros, a ampliação da dimensionalidade dos dados coletados com intuito de enriquecer o modelo.

### 3.7 Considerações Finais do Capítulo

A Tabela 3.1 traz a análise dos trabalhos relacionados detalhados anteriormente mostrando as características consideradas relevantes para o desenvolvimento deste trabalho. Uso de *Machine Learning*, identificando se o estudo faz uso de algoritmos de aprendizado

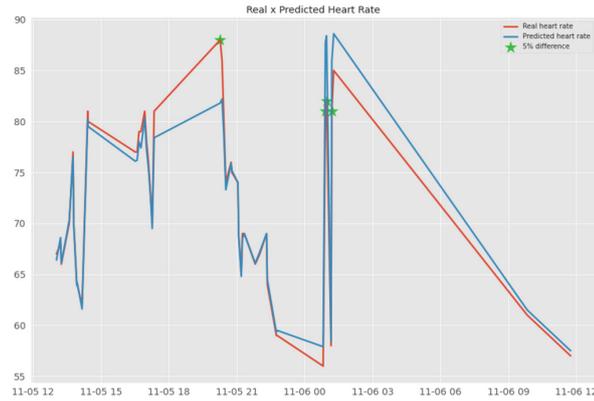


Figura 3.6: Comparação entre frequência prevista e real (LEão et al., 2023)

de máquina em seu desenvolvimento, como acontece na maior parte. Tempo-Real, analisando se o estudo usa como premissa o monitoramento em tempo real, como priorizado no trabalho de Leão et al. (2023). Comparativo, verificando se o trabalho realiza uma comparação entre diferentes algoritmos de aprendizado de máquina, analisando as métricas e o contexto aplicado, como visto nos trabalhos de Site, Nurmi e Lohan (2021), Sabic et al. (2020) e Batista e Filho (2019). Por fim, é avaliado se os trabalhos utilizaram dados reais para desenvolvimento do estudo ou foi usada uma base de dados simulada.

Este trabalho propõe uma expansão da dimensionalidade dos dados em relação ao trabalho do Leão et al. (2023) considerando dados de localização do usuário, além de aprimorar a coleta dos dados em ambiente que não proporciona uma conexão com internet. A escolha do modelo de aprendizado de máquina também foi baseado no trabalho de Leão et al. (2023) e Site, Nurmi e Lohan (2021).

Dessa forma, os trabalhos apresentados neste capítulo são responsáveis por formar a base para o desenvolvimento deste trabalho de conclusão de curso. Esses estudos foram fundamentais para complementar o atual trabalho em aspectos técnicos e bibliográficos, o que permitiu aprimorar o experimento para atingir os objetivos definidos no início desse artigo.

Tabela 3.1: Comparação entre os trabalhos relacionados

Referências	Ano	Uso de <i>Machine Learning</i>	Tempo-Real	Comparativo	Dados Reais
Batista e Filho (2019)	2019	✓	-	✓	-
Sabic et al. (2020)	2020	✓	-	✓	-
Site, Nurmi e Lohan (2021)	2021	-	-	✓	✓
Leão et al. (2021)	2021	✓	-	-	✓
Risch et al. (2022)	2022	✓	-	-	✓
Leão et al. (2023)	2023	✓	✓	-	✓

## 4 Materiais e Métodos

A arquitetura proposta nesse trabalho está representada na Figura 4.1. Essa arquitetura é fundamentada no paradigma *Edge-Fog-Cloud*. Onde os dispositivos vestíveis, como *smartbands*, responsáveis por coletar os valores de frequência cardíaca e enviá-la para o *smartphone* do usuário representam a camada de borda ou *edge*. Já o *smartphone* do usuário representa a camada de névoa ou *fog*, onde a frequência cardíaca coletada pela pulseira persiste até que o ambiente permita que uma sincronização com a nuvem seja realizada. O mesmo vale para os dados de localização, que são coletados a partir do GPS do dispositivo e ficam armazenados até que uma conexão de sincronização com o banco de dados principal aconteça.

O banco de dados principal, o treinamento e armazenamento dos modelos treinados acontece na nuvem ou *cloud*. O *Data Lake* é responsável por armazenar todos os dados coletados pelos sensores de frequência cardíaca e GPS. A partir desses dados armazenados no *Data Lake*, os modelos são treinados semanalmente com uma janela fluante dos últimos 30 dias de coleta. Dessa forma, o fluxo de treinamento do modelo é responsável por buscar esses dados, realizar o pré-processamento, treinar e, em seguida, armazenar esses modelos junto a suas respectivas métricas de desempenho.

Após esse processo, os modelos treinados e armazenados na nuvem ficam disponíveis para realizar predições da frequência cardíaca do usuário. A informação de localização foi adicionada com o intuito de expandir a dimensionalidade dos dados para que o modelo se ajuste melhor aos hábitos do usuário. Assim, é possível agrupar os dados de um usuário por localização e horário, obtendo a frequência cardíaca média, máxima e mínima em cada minuto do dia em todos os locais onde foram coletadas.

Durante o desenvolvimento do modelo foi montado um *dataset* de testes com 225.003 registros <sup>1</sup> de medidas de frequência cardíaca, localização e horário da coleta dos dados. Os dados de localização foram obtidos a partir da exportação dos dados coletados pelo *Google Maps* do dispositivo do autor. Já os dados de frequência cardíaca foram

---

<sup>1</sup><https://github.com/gabrielfsil/dataset-heart-rate-location>

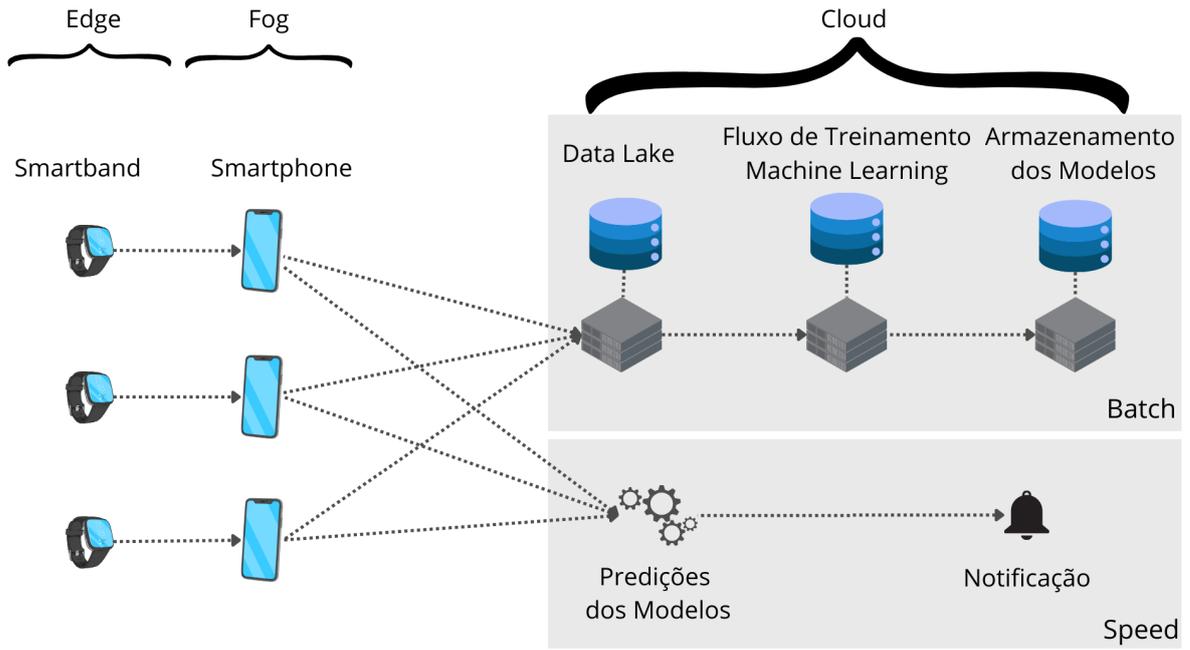


Figura 4.1: Arquitetura Conceitual

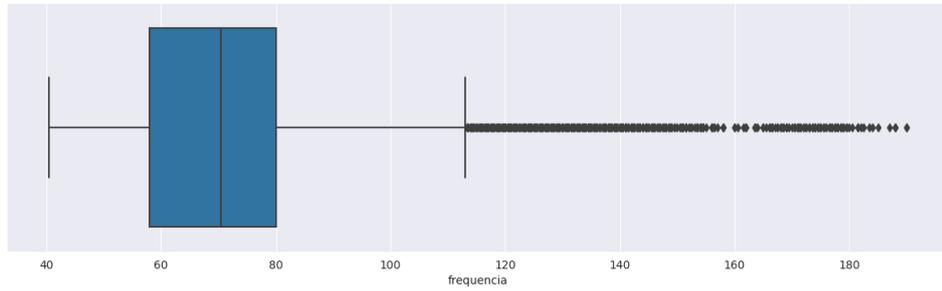


Figura 4.2: Boxplot do *dataset* de testes

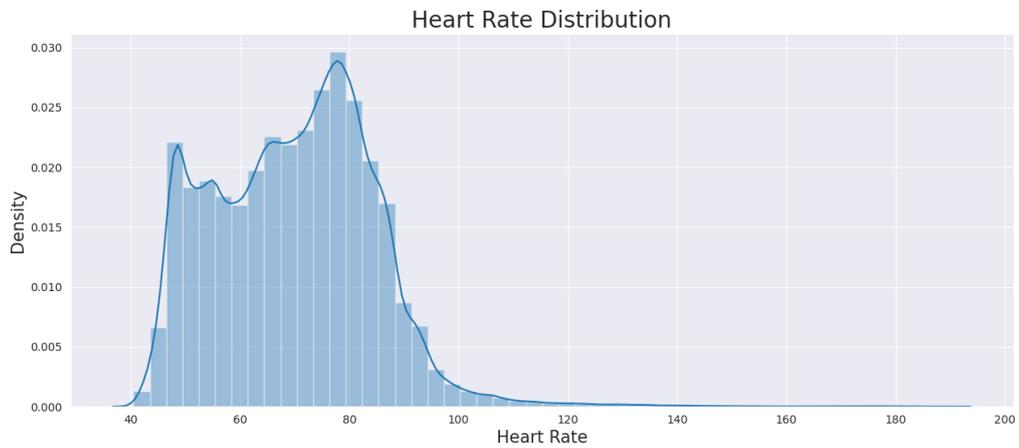


Figura 4.3: Histograma do *dataset* de testes

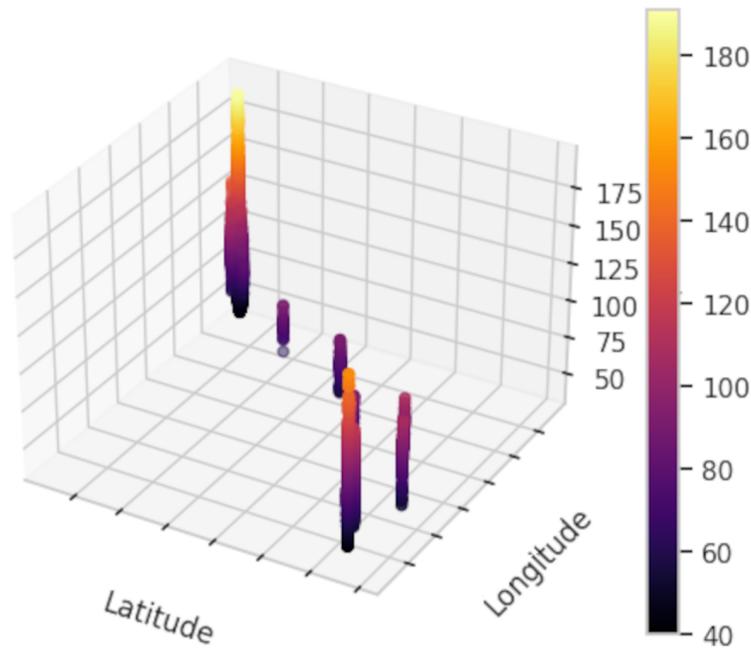


Figura 4.4: Distribuição do *dataset* de testes em relação a localização

obtidos pela exportação dos dados do *Zepp Life*, que é o aplicativo disponibilizado pela *Xiaomi* para se conectar com a pulseira inteligente *Mi Band 5*, também do próprio autor. Esses dados foram utilizados para testar o modelo, assim como entender a distribuição de dados reais. Após a exportação desses dados, eles foram unificados pelo horário para criar um *dataset* com a dimensionalidade desejada. As figuras 4.2, 4.3 e 4.4 mostram o comportamento dos dados do *dataset*.

A Figura 4.2 mostra a distribuição da frequência cardíaca coletada no *dataset*, onde é possível observar *outliers* definidos como frequências acima de 117 bpm (batimentos por minuto). Na Figura 4.3 é possível entender a distribuição, com poucos dados acima do limite superior definido na figura 4.2. Já a Figura 4.4 mostra o comportamento dos dados em relação a latitude e longitude coletada, representadas no eixo X e Y, e eixo Z mostra o valor da frequência cardíaca, assim é possível observar onde alguns *outliers* estão localizados.

Esse *dataset* de testes foi fundamental para entender como a localização pode fornecer um enriquecimento do dado e aprimorar o funcionamento do modelo. A Figura 4.5 mostra o comportamento do *KNeighborsRegressor (KNN)* nesse *dataset* entre valores preditos e valores reais, onde uma linha diagonal representaria o modelo perfeito. Na

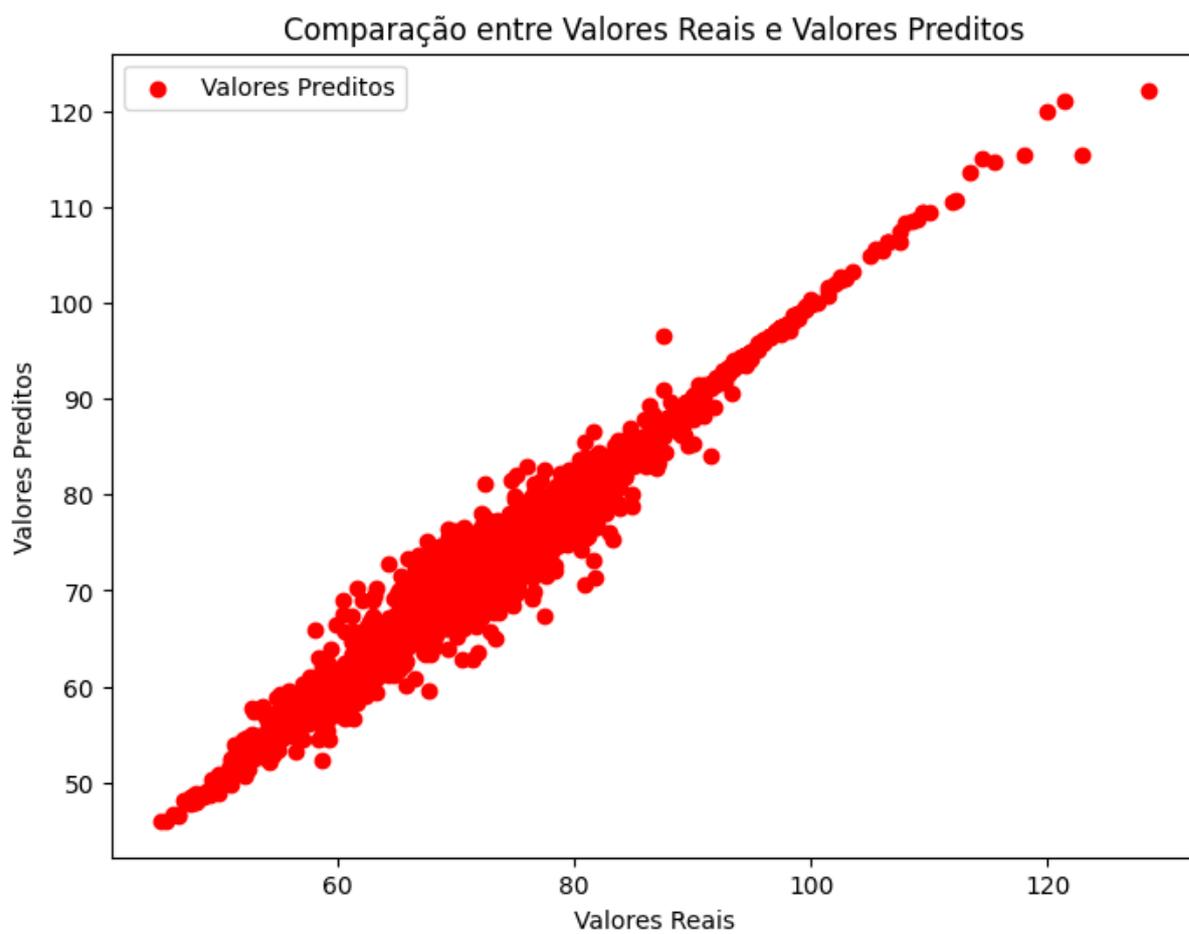


Figura 4.5: Valores Preditos x Valores Reais no *dataset* de testes

Figura 4.5 é possível observar uma distribuição dos pontos próximos a uma diagonal, com alguns valores acima ou abaixo dessa linha. Durante esse experimento foi observado um Coeficiente de Determinação ( $R^2$ ) e Erro Quadrado Médio ( $MSE$ ) de 0,977 e 2,129, respectivamente, para o modelo.

Nesse contexto é importante ressaltar o funcionamento do modelo em relação a identificação de momentos de estresse do usuário. O treinamento do modelo com os dados do usuário identifica o padrão de comportamento, que é específico daquele usuário. Após treinado, o modelo é disponibilizado para realizar as previsões de frequência cardíaca. Quando esse valor predito se distancia de forma significativa do valor real coletado pelo sensor é caracterizado um momento de possível estresse e uma notificação é enviada para o usuário. Assim, quando o modelo realiza uma previsão mais baixa que o valor real coletado é um indicador que a frequência do usuário está mais alta que o padrão de comportamento identificado pelo modelo.

A camada de névoa ou *fog* foi implementada por um aplicativo para coletar os dados dos usuário em tempo real. Os dados foram coletados de uma pulseira inteligente, que compõe a borda ou *edge*, conectada ao aplicativo através de uma conexão sem fio BLE (*Bluetooth Low Energy*). O aplicativo foi desenvolvido em *React Native* com o princípio de *offline first*, que permite que o usuário fique *offline* e continue o monitoramento dos dados coletados pela pulseira. Esses dados são armazenados no banco de dados do dispositivo, que foi implementando usando *Realm*. Essa ferramenta permite que os dados fiquem armazenados no *smartphone* do usuário até que uma conexão com a internet seja estabelecida para que uma sincronização se inicie automaticamente. Outra funcionalidade implementada no aplicativo foi uma tarefa em segundo plano, aproximadamente a cada 10 minutos, ela busca pela pulseira configurada para reestabelecer a conexão caso ela tenha sido perdida. Essa tarefa também é responsável por atualizar o *access token*, que expira em 30 minutos, a partir de um *refresh token*, que expira em 60 dias, para permitir que a sincronização continue sem que o usuário precise autenticar novamente nesse intervalo de 60 dias.

A sincronização dos dados acontece com um banco de dados *MongoDB* de forma flexível. O *Realm* inicia uma sessão para sincronizar os novos dados coletados e a faz de

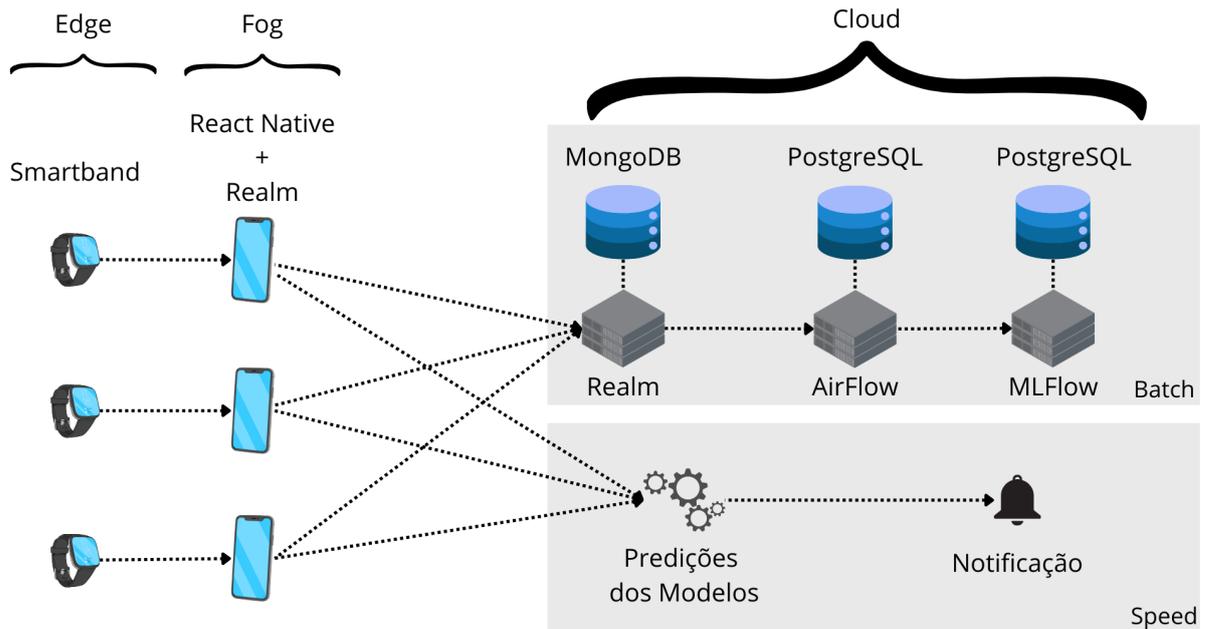


Figura 4.6: Implementação da Arquitetura

acordo com as especificações da conexão e quantidade de registros a serem inseridos. Essa estratégia permite que o processo de sincronização não seja sobrecarregado pela grande quantidade de registros ou instabilidade da rede.

Após toda a etapa de coleta e sincronização dos dados acontece o fluxo de treinamento do modelo. Esse serviço foi implementado utilizando o *AirFlow*. Semanalmente uma tarefa é executada no *AirFlow*, essa tarefa busca os dados coletados nos últimos 30 dias, realiza o pré-processamento e treina os modelos. O modelo é treinado com auxílio do *MLFlow*, para realizar o armazenamento do modelo, junto as métricas de desempenho observadas. O *AirFlow* e o *MLFlow* usam um banco de dados *PostgreSQL* para registrar os resultados obtidos nas tarefas, as métricas de desempenho e o modelo treinado.

O modelo treinado é disponibilizado para o usuário. Assim, de acordo com a saída gerada pelo modelo, é feita uma comparação com a frequência real, caso a frequência seja significativamente inferior em relação a coletada pelos sensores, a aplicativo gera uma notificação para o usuário. A Figura 4.6 ilustra o ambiente computacional proposto nesse trabalho.

## 5 Resultados

Os resultados do presente trabalho foram divididos nas seções a seguir: A Seção 5.1 explica o funcionamento do aplicativo desenvolvido. A Seção 5.2 discorre sobre o dispositivo inteligente usado para realizar as medição no usuário. Na Seção 5.3 mostra o comportamento do modelo de aprendizado de máquina usado para realizar as predições. A Seção 5.4 exemplifica uma notificação ao usuário. Por fim, a Seção 5.5 expõe algumas limitações encontradas durante o desenvolvimento da ambiente computacional.

### 5.1 Olivia

O aplicativo, denominado Olivia, foi desenvolvido usando *React Native*, sendo responsável por autenticar os usuários e realizar a conexão com a pulseira inteligente, onde os dados de frequência cardíaca são coletados. O aplicativo solicita algumas permissões do usuário para funcionar corretamente, como acesso a localização em segundo plano, permissão de execução em segundo plano e acesso ao *Bluetooth*.

A Figura 5.1 mostra a tela de login onde o usuário informa o email e a senha para acessar o aplicativo. Após essa etapa ele é direcionado para a tela de monitoramento da conexão, Figura 5.1, onde ele irá conectar a pulseira que fará as medições de frequência cardíaca. Nessa tela ele irá clicar em "Conectar Dispositivo" e o aplicativo listará os dispositivos próximos. Ao selecionar o dispositivo desejado o aplicativo realiza a conexão via *BLE* e já inicia o monitoramento do sensor de frequência cardíaco do aparelho conectado. O monitoramento da localização acontece a partir do GPS do *smartphone* do usuário, que informa o aplicativo sempre que a localização do usuário é alterada.

Todas as informações coletadas pelo aplicativo são armazenadas em um banco de dados no próprio dispositivo e periodicamente, de acordo com a disponibilidade de acesso a internet, é realizada uma sincronização com o banco de dados principal, *MongoDB*. Caso a conexão com a pulseira seja finalizada por algum motivo externo, o aplicativo tem uma tarefa que é executada em segundo plano para buscar a pulseira cadastrada pelo usuário

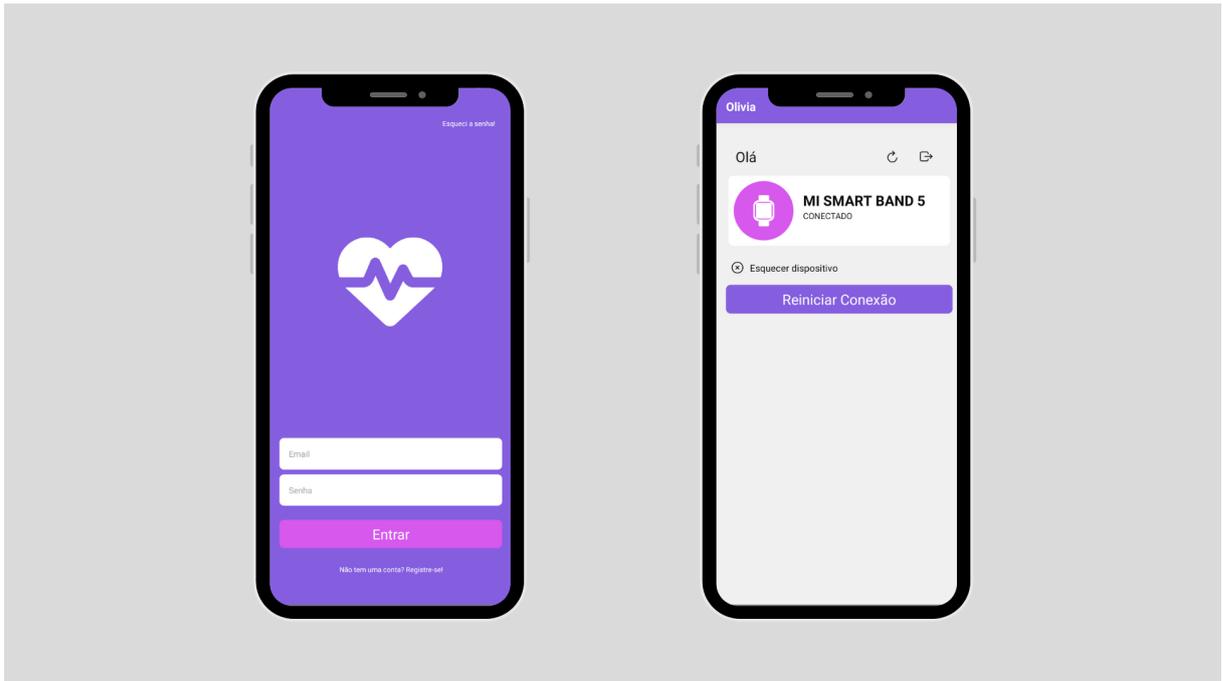


Figura 5.1: Tela de Login e Monitoramento da Conexão

e reestabelecer a conexão.

## 5.2 Pulseira Inteligente

Nesse trabalho foi tomado como objetivo o monitoramento do usuário a partir de uma pulseira inteligente. Durante o desenvolvimento foi utilizada uma *smartband* do modelo Mi Band 5 para realizar as medições. Para que o aplicativo desenvolvido no trabalho estabeleça a conexão com esse dispositivo é preciso acessar o aplicativo disponibilizado pelo fabricante, a permitir que a pulseira seja detectada por aplicativos de terceiros e desativar as restrições de emparelhamento. Após realizar essas configurações na pulseira será possível estabelecer a conexão via *BLE*.

## 5.3 Modelo de Aprendizado de Máquina

O modelo de aprendizado de máquina utilizado nesse trabalho foi o *KNeighborsRegressor* (*KNN*). Para realizar o treinamento do modelo os dados coletados pelo aplicativo e armazenados no banco de dados principal foram preprocessados.

O preprocessing dos dados consiste em buscar os dados de frequência cardíaca

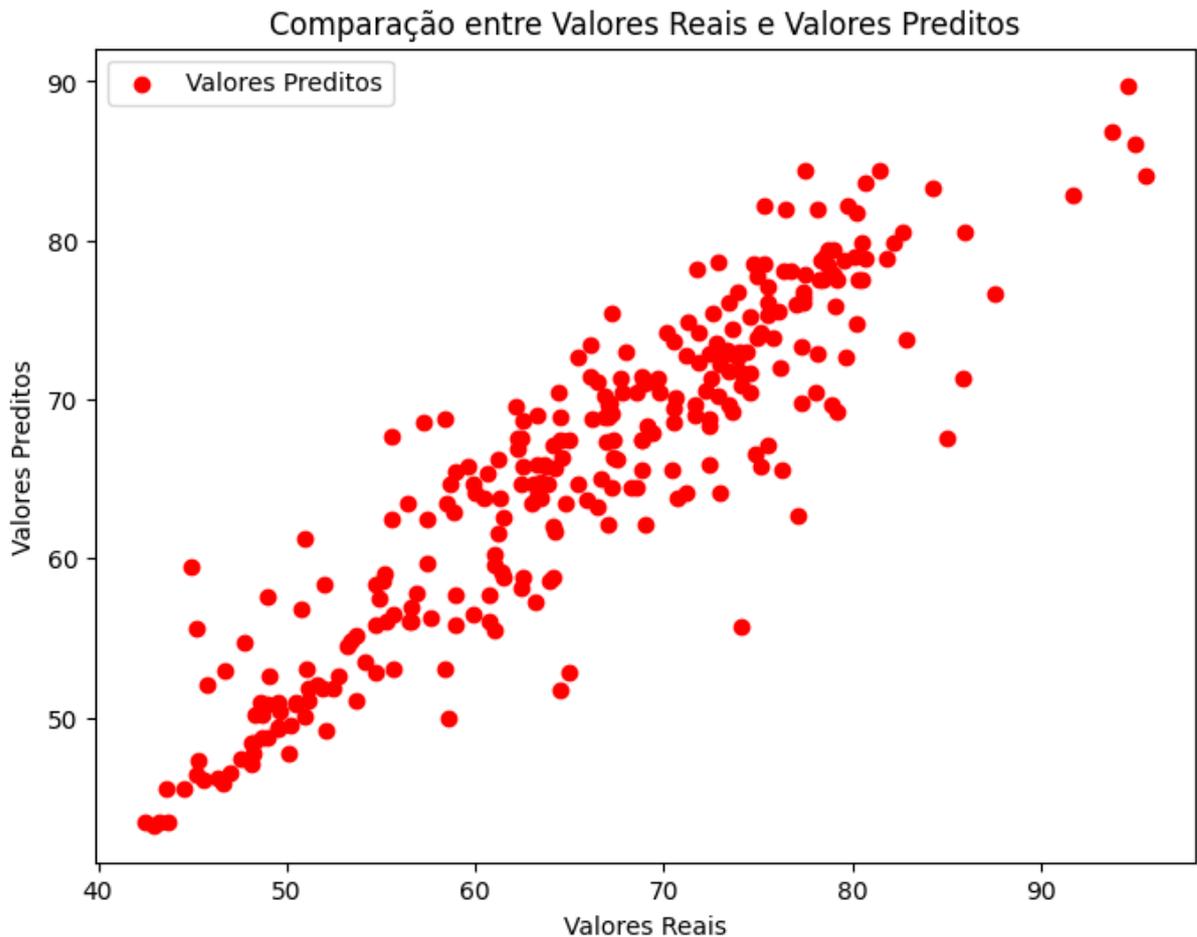


Figura 5.2: Valores Preditos x Valores Reais

e localização coletados nos últimos 30 dias e agrupá-los. Esse agrupamento dos dados foi feito de forma que a frequência cardíaca média, máxima e mínima de cada usuário fosse segmentada por latitude, longitude e horário.

Após o pré-processamento os dados são usados para treinar um modelo para cada usuário, tendo como valor objetivo a frequência cardíaca média do usuário. Durante o experimento o modelo foi treinado com dados coletados do próprio autor, que correspondem a 34126 frequências cardíacas coletadas, após agrupá-las foram obtidos 1440 registros, o que corresponde a pelo menos uma frequência cardíaca por minuto do dia em uma latitude e longitude.

Com relação ao desempenho do modelo implementado foi usada as métricas de Coeficiente de Determinação ( $R^2$ ) e Erro Quadrado Médio ( $MSE$ ). O modelo apresentou, respectivamente, 0,828 e 21,913 nas métricas de desempenho  $R^2$  e  $MSE$ . A Figura 5.2 mostra a distribuição dos comportamento do modelo relacionando valores previstos com

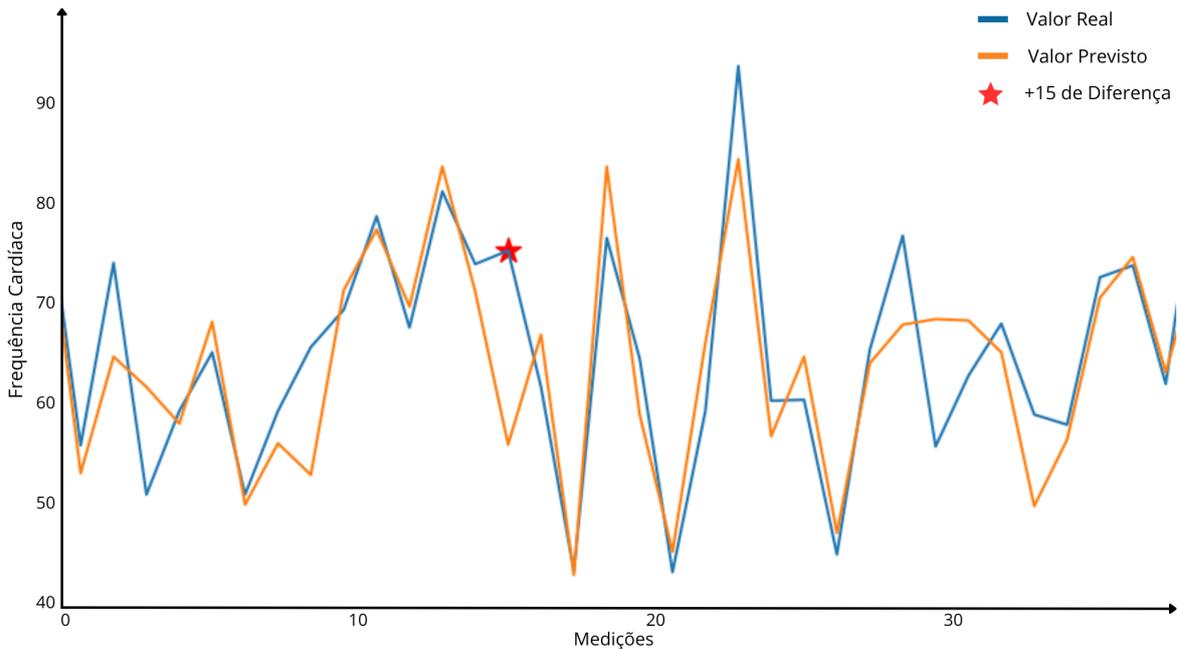


Figura 5.3: Valores Preditos x Valores Reais

valores reais. O gráfico de um modelo perfeito seria uma linha na diagonal, o modelo atual apresenta uma distribuição nessa linha com alguns pontos mais dispersos dessa diagonal.

Foram realizadas entrevistas com profissionais da área médica com o objetivo de identificar um valor para que uma frequência cardíaca seja considerada significativamente alterada. Durante essas entrevistas os valores de uma frequência alterada variam de 15 a 30 batimentos por minuto acima da frequência base do indivíduo. Nesse trabalho decidimos considerar 15 batimentos por minuto acima da frequência prevista como um momento de estresse do usuário.

A Figura 5.3 mostra o comportamento dos valores preditos e valores coletados pela pulseira, em locais onde a diferença entre o valor real e o valor previsto é maior que 15 batimentos por minuto (bpm) aparece marcado por uma estrela. Esse ponto marcado seria um momento de estresse em que a frequência cardíaca do usuário está mais alta que o esperado e uma notificação seria gerada para o usuário.

## 5.4 Notificações

Após identificar momentos de estresse do usuário é importante notificá-lo. Essas notificações são fundamentais para permitir que o usuário entenda os eventos que estão

provocando esse estresse em seu dia a dia. Dessa forma, a notificação de um evento de estresse detectado pelo modelo seria a seguinte:

- *Notei uma variação significativa na sua frequência cardíaca no dia [data] às [horário], enquanto você estava na [localização]. Isso pode indicar um momento de estresse. Recomendamos que você tire um tempo para relaxar e praticar técnicas de redução de estresse.*

## 5.5 Limitações

Durante o desenvolvimento do trabalho foram observadas algumas limitações relacionadas a implementação do ambiente computacional. Uma dessas limitações está relacionada a leitura dos dados coletados pela pulseira. Em alguns momentos, ao iniciar a conexão do aplicativo ao dispositivo inteligente, o aplicativo inicia o monitoramento dos dados coletados pelo sensor, entretanto acontecia um erro de leitura do *BLE* informando que o serviço solicitado não estava disponível na pulseira, como mostra a Figura 5.4. Para solucionar essa limitação é necessário abrir o aplicativo do fabricante e conectar a pulseira, após isso, o usuário precisa acessar o aplicativo desenvolvido no trabalho e clicar no botão "Reiniciar Conexão". Após realizar esse procedimento a conexão e o monitoramento é executado com sucesso.

Outra limitação observada é a periodicidade que uma tarefa em segundo plano pode ser executada nos dispositivos *Android*. Na implementação das tarefas é possível definir um intervalo mínimo para que a tarefa seja executada, entretanto a execução em si fica a critério da disponibilidade de recursos e outras tarefas presentes no sistema operacional. Segundo a documentação do *Android*, tarefas em segundo plano são executadas entre 10 e 15 min, dependendo da disponibilidade do sistema e características do aplicativo/serviço.

Esse experimento foi realizado com uma quantidade pequena de usuários, o que não permite avaliar tão bem o desempenho da arquitetura em uma escala maior de dados. Questões de escalabilidade seria melhor observada com um fluxo maior no processamento dos dados. Assim, essa também se mostrou uma limitação nesse trabalho. Nesse viés

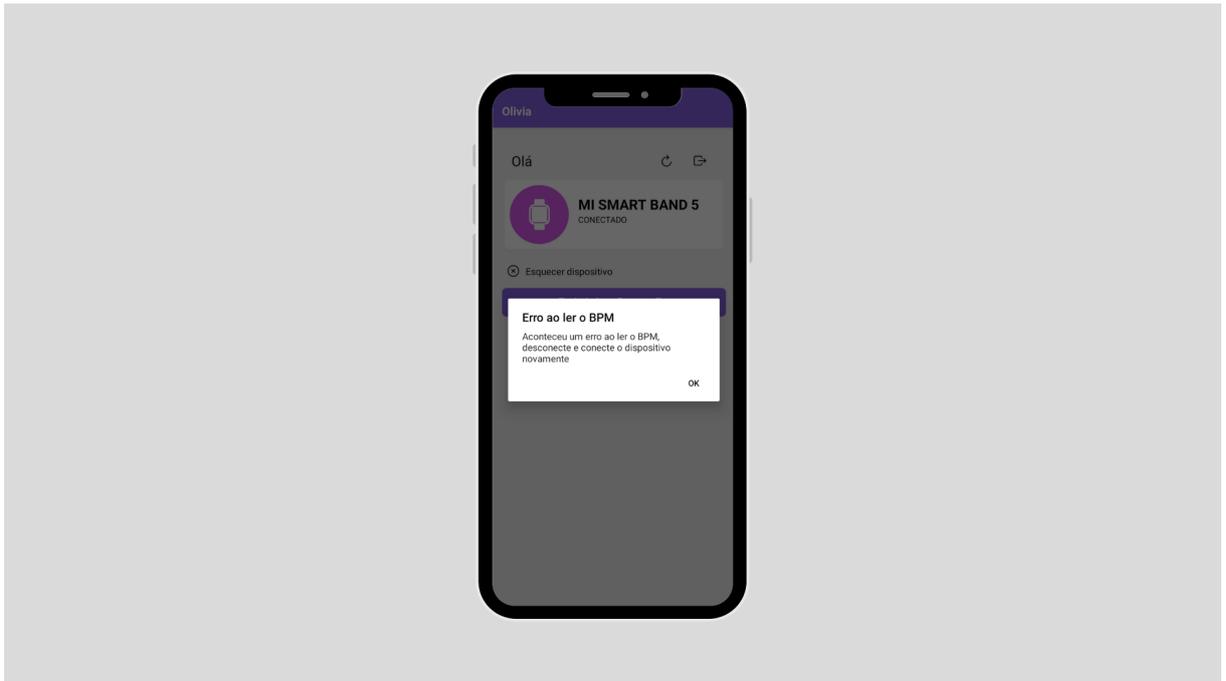


Figura 5.4: Erro na leitura de informações da pulseira

de maior volume de dados alguns gargalos podem surgir, como o treinamento do modelo que atualmente acontece na nuvem. Dado um volume maior de dados coletados pelos dispositivos o treinamento seria sobrecarregado. Dessa forma é importante ressaltar esse ponto como uma limitação do projeto.

## 6 Considerações Finais

Nesse trabalho foi desenvolvido um ambiente computacional responsável por monitorar usuários, utilizando dados coletados através de sensores corporais e localização para notificar o usuário sobre momentos de estresse detectados pelo modelo de aprendizado de máquina também desenvolvido no estudo.

Através dos experimentos realizados foi possível observar que a arquitetura proposta consegue realizar o monitoramento em tempo real do usuário e identificar momentos de estresse para gerar uma notificação. O monitoramento se mostrou bastante eficiente, suportando momentos de instabilidade nas conexões, como permanecer monitorando mesmo sem uma conexão com a internet, além de restabelecer a conexão com a pulseira caso ela fosse encerrada por algum fator externo, como a distância entre os dispositivos. Mesmo com as limitações encontradas no funcionamento do aplicativo ele apresentou um resultado satisfatório.

A partir dos dados coletados pelos sensores e sincronizados no banco de dados principal da arquitetura foi possível validar o *pipeline* de treinamento dos modelos para cada usuário com métricas de desempenho também satisfatórias. Esses resultados mostraram que a arquitetura proposta consegue identificar momentos de estresse do usuário a partir da frequência cardíaca e localização.

Como trabalhos futuros, são previstas novas abordagens que permitirão o aumento da eficiência e impacto da solução. Dentre elas está a implantação do modelo preditivo no dispositivo do usuário para permitir notificações mesmo quando estiver *offline*, junto a uma atualização periódica do modelo usado em seu dispositivo. Outra abordagem que pode contribuir para a evolução desse trabalho é o monitoramento de outros indicadores referentes a saúde do usuário, como temperatura, número de passos, saturação de oxigênio, entre outros. Como forma de aprimorar o trabalho atual é interessante analisar também o comportamento de outros modelos de aprendizado de máquina para realizar as predições do usuário. Para expandir um universo de perfis de usuário do sistema seria interessante também utilizar técnicas de aprendizado federado a partir dos

---

modelos treinados para cada usuário com o intuito de evitar que o modelo se ajuste a um comportamento de estresse.

## Bibliografia

- BATISTA, A. F. de M.; FILHO, A. D. P. C. Machine learning aplicado à saúde. Laboratório de Big Data e Análise Preditiva em Saúde (LABDAPS) e Faculdade de Saúde Pública da Universidade de São Paulo, 2019.
- BAUER, M. E. Estresse. In: *Como ele abala as defesas do corpo?* [S.l.]: Pontifícia Universidade Católica do Rio Grande do Sul, 2002.
- CISCO. Cisco annual internet report(2018-2023). In: .  
<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>: [s.n.], 2018.
- CRESWELL, J. W. Research design. In: . [S.l.]: Sage Publications, 2003.
- DIXON, J. Data lake. In: . [S.l.: s.n.], 2010.
- DOMINGOS, P. The master algorithm. In: *How the Quest for the Ultimate Learning Machine Will Remake Our World*. [S.l.: s.n.], 2015.
- INMON, W. H. Building the data warehouse. In: *The most comprehensive introduction to the core concepts and methods of data warehousing*. Canada: John Wiley and Sons, Inc, 2002.
- JORDAN., M. I.; MITCHELL, T. M. Machine learning. In: *Trends, perspectives, and prospects. Science*. [S.l.: s.n.], 2015.
- KLEPPMANN, M. Designing data-intensive applications. In: *The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. [S.l.: s.n.], 2015.
- LADERAS, T.; VASILEVSKY, N.; PEDERSON, B.; HAENDEL, M.; MCWEENEY, S.; DORR, D. A. Teaching data science fundamentals through realistic synthetic clinical cardiovascular data. bioRxiv, 2018.
- LEÃO, W. S.; SILVA, G. D. iorio; STRÖELE, V.; DANTAS, M. Uma arquitetura fog-cloud para o monitoramento de sinais corporais. Universidade Federal de Juiz de Fora, 2021.
- LEÃO, W. S.; SILVA, G. D. iorio; STRÖELE, V.; DANTAS, M. An architecture proposal to support e-healthcare notifications. Universidade Federal de Juiz de Fora, 2023.
- MASSOLA, S. C.; PINTO, G. S. O uso da internet das coisas (iot) a favor da saúde. In: . [S.l.]: Revista Interface Tecnológica, 2018.
- MURPHY, K. P. Machine learning. In: *A Probabilistic Perspective*. [S.l.: s.n.], 2012.
- NORMAN, A. H.; TESSER, C. D. Prevenção quaternária na atenção primária à saúde: uma necessidade do sistema Único de saúde. Centro de Saúde Tapera, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina, 2009.
- PEREIRA, A. S.; SHITSUKA, D. M.; PARREIRA, F. J.; SHITSUKA, R. Metodologia da pesquisa científica. In: . [S.l.]: UNIVERSIDADE FEDERAL DE SANTA MARIA, 2018.

RISCH, M.; GROSSMANN, K.; AESCHBACHER, S.; WEIDELI, O. C.; KOVAC, M.; PEREIRA, F.; WOHLWEND, N.; RISCH, C.; HILLMANN, D.; LUNG, T.; RENZ, H.; TWERENBOLD, R.; ROTHENBÜHLER, M.; LEIBOVITZ, D.; KOVACEVIC, V.; MARKOVIC, A.; KLAVER, P.; BRAKENHOFF, T. B.; FRANKS, B.; MITRATZA, M.; DOWNWARD, G. S.; DOWLING, A.; MONTES, S.; GROBBEE, D. E.; CRONIN, M.; CONEN, D.; GOODALE, B. M.; RISCH, L. Investigation of the use of a sensor bracelet for the presymptomatic detection of changes in physiological parameters related to covid-19: an interim analysis of a prospective cohort study (covi-gapp). In: . [S.l.]: BMJ open, British Medical Journal Publishing Group, 2022.

SABIC, E.; KEELEY, D.; HENDERSON, B.; NANNEMANN, S. Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data. Springer Nature, 2020.

SANTOS, B. P.; SILVA, L. A. M.; CELES, C. S. F. S.; NETO, J. B. B.; PERES, B. S.; VIEIRA, M. A. M.; VIEIRA, L. F. M.; GOUSSEVSKAIA, O. N.; LOUREIRO, A. A. F. Internet das coisas: da teoria à prática. In: . [S.l.]: Universidade Federal de Minas Gerais, 2016.

SITE, A.; NURMI, J.; LOHAN, E. S. Systematic review on machine-learning algorithms used in wearable-based ehealth data analysis. IEEE Access, 2021.

TANENBAUM, A. S. Computer networks. In: *Prentice Hall Professional Technical Reference*. [S.l.: s.n.], 2002.