



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

DIJAY DE LIMA PEREIRA

**HRV-ANALYTICS: USO DE CIÊNCIAS DE DADOS E MICROSERVIÇOS PARA
ANÁLISE DA TAXA DE VARIABILIDADE DE FREQUÊNCIA CARDÍACA (VFC)
COM IDOSOS**

**CAMPINA GRANDE
2021**

**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA**

DIJAY DE LIMA PEREIRA

SÊNIOR SAÚDE MÓVEL

**HRV-ANALYTICS: USO DE CIÊNCIAS DE DADOS E MICROSSERVIÇOS PARA
ANÁLISE DA TAXA DE VARIABILIDADE DE FREQUÊNCIA CARDÍACA (VFC)
COM IDOSOS**

Relatório Técnico e Científico apresentado ao Curso de Ciência da Computação do Centro Ciências e Tecnologias da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Computação.

Orientador: Prof. Dr. Paulo Eduardo e Silva Barbosa.

**CAMPINA GRANDE
2021**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

P436h Pereira, Dijay de Lima.
HRV-Analytics [manuscrito] : uso de ciências de dados e microsserviços para análise da taxa de variabilidade de frequência cardíaca (VFC) com idosos / Dijay de Lima Pereira. - 2021.
41 p. : il. colorido.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2021.
"Orientação : Prof. Dr. Paulo Eduardo e Silva Barbosa , Departamento de Computação - CCT."
1. Ciência de Dados. 2. Arquitetura de microsserviços. 3. Frequência cardíaca. 4. Python. I. Título

21. ed. CDD 005.3

DIJAY DE LIMA PEREIRA

HRV-ANALYTICS: USO DE CIÊNCIAS DE DADOS E MICROSERVIÇOS PARA
ANÁLISE DA TAXA DE VARIABILIDADE DE FREQUÊNCIA CARDÍACA (VFC) COM
IDOSOS

Relatório Técnico e Científico apresentado ao
Curso de Ciência da Computação do Centro
Ciências e Tecnologias da Universidade
Estadual da Paraíba, como requisito parcial à
obtenção do título de bacharel em Computação.

Área de concentração: Ciência de Dados.

Aprovado em: 30 / 11 / 2021.


BANCA EXAMINADORA



Prof. Dr. Paulo Eduardo e Silva Barbosa (Orientador)
Universidade Estadual da Paraíba (UEPB)



Me. Eujessika Katielly Rodrigues Silva
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Fábio Luiz Leite Júnior
Universidade Estadual da Paraíba (UEPB)

Dedico esse Trabalho aos meus pais, pela dedicação,
companheirismo e esforço, sempre me incentivando
para busca dos meus objetivos.

AGRADECIMENTOS

Agradeço acima de tudo a Deus, por sempre me abençoar e permitir a plena realização na área pela qual tenho extrema felicidade de estar inserido. Além disso, sou grato por ter me guiado e me sustentado mesmo em meio aos momentos mais difíceis.

Aos meus pais Maria Lima e Daniel João, assim como aos meus irmãos que estiveram sempre ao meu lado durante todos esses anos me incentivando e motivando a persistir, assim como sou grato pela compreensão por minha ausência nas reuniões familiares.

Sou extremamente grato também a minha namorada Paloma Araújo, companheira que pude compartilhar todas as dificuldades e alegrias, e que sempre fez questão de estar presente e disponível para compartilhar todos os momentos, desde os mais difíceis aos mais especiais.

Aos professores do Curso de Ciência da Computação da Universidade Estadual da Paraíba por todo o empenho e conhecimento compartilhado, em especial ao meu orientador Paulo Eduardo e Silva Barbosa por ter me ajudado durante toda produção deste trabalho, como também ao longo de todo o curso ter me engajando em projetos de extensão. Agradeço ainda de maneira especial ao NUTES-UEPB, lugar onde adquiri muito conhecimento e principalmente experiência sendo fundamental para meu crescimento como aluno e como profissional.

Por último, agradeço a todos os meus colegas e amigos de curso, pessoas que foram fundamentais para conseguir finalizar cada etapa assim como dividir todos os momentos da graduação.

RESUMO

O presente trabalho descreve o processo de desenvolvimento de um microsserviço, criado com o objetivo de identificar os dados faltantes da série temporal vinda de dispositivos de captura de batimentos cardíacos. Além disso, buscou-se estratégias para o preenchimento dessas lacunas, através da utilização de Ciência de Dados, Interpolação Linear, Regressão Linear, Análise de Similaridade, bem como bibliotecas da linguagem Python, ressaltando a `heartpy`, que possibilitou a aquisição de métricas extremamente importantes. A aquisição dos dados é realizada por dispositivos que deixam uma lacuna na série temporal provocando, desse modo, uma grande perda de informações. A ausência deles impossibilita que conclusões relevantes sobre o estado de saúde dos pacientes sejam observadas e possíveis problemas sejam evitados. A partir do microsserviço foi possível disponibilizar na plataforma Sênior Saúde Móvel, métricas de Variabilidade de Frequência Cardíaca, que possibilitam uma análise mais profunda sobre os dados capturados, permitindo aos profissionais de saúde uma ferramenta fundamental para acompanhamentos dos dados de batimentos cardíacos dos pacientes.

Palavras-Chave: Ciência de Dados. Variabilidade de Frequência Cardíaca. Arquitetura de Microsserviço.

ABSTRACT

This paper describes the process of developing a microservice, created with the goal of identifying missing data from the time series coming from heartbeat capture devices. Furthermore, In addition, strategies for filling these gaps were sought through the use of Data Science, Linear Interpolation, Linear Regression, Similarity Analysis, as well as Python language libraries, highlighting heartpy, which enabled the acquisition of extremely important metrics. The data acquisition is performed by devices that leave a gap in the time series, thus causing a great loss of information. Their absence makes it impossible for relevant conclusions about the patients' health status to be observed and possible problems to be avoided. From the microservice it was possible to make available on the Senior Mobile Health platform, Heart Rate Variability metrics, which enable a deeper analysis on the captured data, allowing health professionals a fundamental tool for monitoring patients' heartbeat data.

Keywords: Data Science. Heart Rate Variability. Microservice Architecture.

LISTA DE ILUSTRAÇÕES

Figura 1 - Diagrama da arquitetura do projeto Sênior Saúde Móvel.....	22
Figura 2 - Processo para implantar microsserviço hrv-analytics.....	23
Figura 3 - Processo para desenvolvimento do microsserviço hrv-analytics.....	24
Figura 4 - Comparação entre valores de coleta entre Fitbit Inspire e Polar H10	28
Figura 5 - Comparação entre valores processados pela Interpolação Linear, entre Fitbit Inspire e Polar H10	31
Figura 6 - Exemplo de um intervalo RR.....	32
Figura 7 - Diagrama da Arquitetura após a integração do HRV-Analytics.....	35
Figura 8 - Exemplo de um registro do hrv-analytics no banco de dados	37

LISTA DE TABELAS

Tabela 1 - Comparação entre dispositivos para coleta de dados.	26
Tabela 2 - Métricas calculadas pelo algoritmo desenvolvido.....	32
Tabela 3 - Benchmarking sobre as técnicas abordadas	34

LISTA DE ABREVIATURAS E SIGLAS

AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
ECG	Eletrocardiograma
HTTP	Hypertext Transfer Protocol
IOS	Sistema Operacional Móvel da Apple
IoT	Internet of Things
JSON	JavaScript Object Notation
MAE	Mean Absolute Error
ML	Machine Learning
PA	Pressão Arterial
PaaS	Plataforma como Serviço
PPG	Photoplethysmography
RF	Requisitos Funcionais
RMSE	Root Mean Squared Error
RR	Ondas do tipo R
SOA	Arquitetura Orientada a Serviços
VFC	Variabilidade de Frequência Cardíaca

SUMÁRIO

1 INTRODUÇÃO	12
2 REVISÃO DE LITERATURA	14
2.1 Ciência de Dados	14
2.2 Série Temporal.....	15
2.3 Interpolação Linear	16
2.4 Regressão Linear.....	17
2.5 Microsserviços	17
2.6 Variabilidade de Frequência Cardíaca	18
2.7 Bibliotecas Python	19
2.7.1 Flask	20
2.7.2 Pandas	20
2.7.3 Numpy	21
2.7.4 HRV-Analysis	21
3 METODOLOGIA	22
3.1 Definição dos Requisitos	23
3.2 Coleta dos Dados	24
3.2.1 Fitbit Inspire	25
3.2.2 Polar H10	25
3.2.3 Polar OH1	26
3.2.4 Definição do dispositivo	26
3.3 Pré-Processamento dos Dados	28
3.3.1 Tratamento de dados faltantes	28
3.3.2 Remoção de dados faltantes irrelevantes	29
3.4 Abordagem das Técnicas	29
3.4.1 Regressão Linear	29
3.4.2 Interpolação Linear	30
3.4.3 Métricas – Hrv-Analysis	31
3.4.4 Benchmarking das técnicas (Avaliação comparativa)	33
3.5 Implantação do microsserviço na plataforma	35
3.5.1 Aquisição dos dados para cálculo das métricas	36
3.5.2 Aplicar o algoritmo de extração de métricas na arquitetura do RabbitMQ	36
3.5.3 Salvar as métricas e disponibilizar na arquitetura	36
3.5.4 Conclusão da implementação	37
4 CONSIDERAÇÕES FINAIS	39

REFERÊNCIAS.....	40
-------------------------	-----------

1 INTRODUÇÃO

Desde o início do século vinte e um, a população idosa cresceu consideravelmente no mundo, e com a chegada da velhice alguns problemas de saúde começam a afetar o organismo humano, entre eles as mudanças do sistema cardiovascular são destacadas (ANDRADE et al, 2017). Diante disso, é notável que com o avançar da idade o coração sofre diminuição das batidas, relaxamento retardado do ventrículo que, associado a degenerações arteriais e acúmulo de gordura, diminui o fluxo dos vasos e provoca um aumento no índice de idosos com pressão alta.

Muitas dessas doenças são causadas pela falta de uma alimentação saudável e prática de exercícios físicos, bem como a resistência de muitos idosos a um frequente acompanhamento médico para ter conhecimento sobre seu estado clínico, provocando um retardamento da descoberta de algum problema que, por consequência, compromete o seu estado de saúde.

É comum encontrarmos profissionais de saúde, assim como de outras áreas, procurando e desenvolvendo ferramentas correspondentes para resolução deste problema. Temos exemplos mais concretos onde Naeini (2019), propõe uma abordagem de avaliação de qualidade de sinais PPG¹ para sistemas de monitoramento de saúde baseados em IoT². Utilizando Redes Neurais Convolucionais, cria uma função de hipótese comparando a frequência cardíaca no PPG com os valores de frequência cardíaca correspondentes extraídos do sinal de ECG (Naeini et al, 2019). Temos ainda Prabhakar (2019), na qual mostra um trabalho que foca na classificação de doenças cardiovasculares utilizando os dados do PPG (Prabhakar et al, 2019). É importante destacarmos uma ferramenta amplamente utilizada para avaliar a função cardiovascular, a taxa de Variabilidade de Frequência Cardíaca (VFC), que é um mecanismo que permite a investigação de oscilações nos intervalos entre batimentos cardíacos consecutivos (intervalos RR).

Neste cenário pode-se afirmar que é de fundamental importância que estudos como os citados sejam cada vez mais frequentes, e a partir de soluções viáveis consigam ser efetivamente colocados em prática, para que os pacientes possam ser cada vez mais beneficiados com as análises realizadas a partir de pesquisas, bem como o desenvolvimento de tecnologias que agreguem benefícios.

¹ Photoplethysmography (fotopleletismografia) é uma técnica utilizada no monitoramento da saturação de oxigênio e frequência cardíaca.

² Internet of Things (Internet das coisas) descreve a rede de equipamentos incorporados a sensores, software e outras tecnologias com o objetivo de conectar e trocar dados com outros dispositivos e sistemas pela internet.

Diante do exposto, é possível fazer uso da tecnologia e de algumas das suas diversas técnicas, com o objetivo de criar e utilizar algoritmos que possam contribuir com todo esse processo e que torne viável e praticável todas essas melhorias. Com isso, o trabalho dos profissionais de saúde será favorecido e os pacientes envolvidos serão também beneficiados, tendo em vista que a tecnologia aplicada à saúde facilitará o diagnóstico dos problemas já citados.

Desta maneira, este trabalho tem como objetivo entender o contexto de trabalho e os desafios enfrentados de acordo com a ausência das séries temporais na coleta de dados, bem como analisar as técnicas de preenchimento das séries temporais com o uso da Ciência de Dados para que seja possível a construção de um microsserviço capaz de dispor as informações relevantes sobre a Variabilidade de Frequência Cardíaca. Assim, podemos realizar o cálculo de métricas sobre os indicadores de VFC e, por fim, conseguimos disponibilizar em uma arquitetura de microsserviço que serão disponibilizados na plataforma Sênior Saúde Móvel³, com o intuito de auxiliar profissionais de saúde na análise dos dados, apresentando informações e métricas relevantes. O uso de Ciência de Dados se mostra essencial na presente pesquisa visto que, através da validação e acurácia dos modelos desenvolvidos, possibilita a determinação da técnica que melhor se encaixa e produz os resultados mais satisfatórios.

³ Disponível em: < <https://github.com/senior-saude-movel> >. Acesso em: 02 de set. 2021

2 REVISÃO DE LITERATURA

Neste capítulo serão abordados os conceitos que tratam das áreas de conhecimento presentes no decorrer de toda pesquisa. Sendo assim, serão estudados os seguintes conceitos: Ciência de Dados, Séries Temporais, Interpolação Linear, Regressão Linear, microsserviços, Análise de Similaridade, microsserviços e Variabilidade de Frequência Cardíaca, além das bibliotecas Python utilizadas no desenvolvimento da aplicação, tendo como destaque `hrvanalysis`.

2.1 Ciência de Dados

Quando mencionamos o termo Ciência de Dados associamos esse conceito a novos métodos, no entanto é uma área de conhecimento antiga e consolidada. O que acontece é que ela envolve uma série de outras áreas em conjunto, como estatística, matemática, computação e qualquer outra que se deseja aplicar. Percebemos que esses temas são bem solidificados em pesquisas científicas, sendo assim temos uma ferramenta forte para a resolução de problemas complexos e maiores, como o processamento de grandes massas de dados (GRUS, 2016).

Ciência de Dados é uma ferramenta que possibilita que empresas, cientistas ou pesquisadores consigam a aquisição de informações importantes que estão na maioria das vezes omissas, possibilitando a extração de determinados conhecimentos da base de dados que se trabalha, para uma possível tomada de decisão em diferentes tipos de aplicações.

Atualmente, toda e qualquer atividade que realizamos está associada a disponibilização ou aquisição de dados, seja websites que rastreiam os cliques dos usuários no acesso diário, aplicativos de streaming para músicas, catálogos de filmes, séries e muitos outros conteúdos. Com base no perfil dos usuários, a Ciência de Dados consegue realizar um direcionamento e indicações de conteúdos específicos que condizem com as suas preferências em cada plataforma ou site utilizado, por exemplo. (GRUS, 2016).

Podemos perceber a grande influência dos dados no nosso cotidiano, e como eles nos possibilitam experiências inovadoras nas áreas profissionais, educacionais, entre outras. Nota-se que a Ciência de Dados está cada vez mais ganhando espaço nas organizações que objetivam o avanço tecnológico, pois ela consegue trabalhar de maneira eficiente com uma grande quantidade de dados e, ainda assim trazer resultados esperados e positivos para certas tomadas de decisões em qualquer área. Como um exemplo mais recente na saúde, a Secretaria de Saúde

de Florianópolis⁴ (Santa Catarina) juntamente com o Laboratório de Engenharia de Ciências da Informação da Universidade Federal de Santa Catarina, têm utilizado algoritmos de inteligência artificial para analisar o subdiagnóstico de casos da COVID-19 e assim, ajudar a mitigar a transmissão de SARS-CoV-2.

Nessa pesquisa a Ciência de Dados possibilitará a leitura e visualização dos dados capturados pelos dispositivos eletrônicos utilizados, permitindo que sejam feitas análises e manipulações nos dados obtidos, como o preenchimento de dados faltantes que acontecem no decorrer da captura e extração de informações relevantes sobre os batimentos cardíacos, com o intuito de prever possíveis problemas de saúde nos pacientes, bem como a visualizar informações relevantes que alertem os profissionais de saúde responsáveis.

2.2 Série Temporal

Podemos dividir qualquer tipo de dado em duas categorias: atemporais e temporais. A primeira categoria, por sua vez, é bastante comum nos datasets e mais utilizada em Ciência de Dados, as características envolvidas independem de uma linha cronológica. Essa pode ter dados temporais, mas consiste em uma característica secundária, não sua essência (DHAMALA et al, 2018).

Os dados temporais, por outro lado, foram criados com um objetivo específico, o registro de informações é salvo a partir de uma linha cronológica. Exemplos clássicos são variações ao longo de um dia ou o tráfego de usuários por hora. Quando tratamos um dado atemporal que tem informações ausentes, normalmente, vemos explicações de que o mais simples é remover essas linhas. Isso porque muitos algoritmos não conseguem fazer análises com dados ausentes e, na maioria dos casos, eles são abundantes o suficiente para que a remoção de algumas linhas tenha pouca influência (DHAMALA et al, 2018).

No entanto, quando falamos de dados temporais, a remoção de um dado pode ser bastante problemática. Eles precisam de cadência, padrão, especialmente se formos fazer análises de frequência. Nossa série temporal, se trata exatamente de uma frequência de batimentos cardíacos e, por esse motivo, não podemos concluir a existência de um padrão, porém podemos concluir que esse padrão varia de acordo com o tipo de atividade que o usuário exerça (PRATAMA et al, 2016).

⁴ Disponível em: <<https://medilablog.com/2020/12/29/ciencia-de-dados-e-a-area-da-saude/>>. Acesso em: 02 de set. 2021

Sendo assim, na presente pesquisa a análise das séries temporais será realizada de maneira detalhada e minuciosa, visualizando os possíveis padrões existentes dentre as diferentes categorias de atividades físicas que os pacientes exerçam, para que dessa maneira tenhamos um nível de confiança sobre o tipo de preenchimento dos dados faltantes nas séries temporais.

2.3 Interpolação Linear

Para contribuir com a solução desenvolvida, foi necessário abordar uma série de soluções, com o intuito de encontrar a mais viável e que melhor se adequasse ao desafio proposto. Dessa maneira uma das primeiras técnicas encontradas e observadas foi a Interpolação Linear. Segundo Araújo (2004, p. 1) “A interpolação consiste em determinar, a partir de um conjunto de dados discretos, uma função ou um conjunto de funções analíticas que possam servir para a determinação de qualquer valor no domínio de definição”.

Tendo conhecimento dessas informações, podemos observar que a Interpolação Linear se adequa a uma possível solução, por ter como objetivo encontrar pontos intermediários entre dois outros, já que precisamos de um preenchimento na série temporal, ocasionada pelos dispositivos que não conseguem realizar a captura de algumas informações (ARAÚJO, 2004).

A interpolação pode ser aplicada em diversos campos da ciência, como por exemplo, na computação gráfica, no processamento de sinais e imagens. Pelo bom desempenho ela vem ganhando cada vez mais espaço quando se trata de resolução de problema para preenchimento de dados faltantes. Apesar de ser uma ferramenta básica na integração numérica e nos rigorosos métodos de solução de equações diferenciais, os níveis de complexidade na solução são cada vez mais elevados.

Podemos constatar ainda que o método de Interpolação Linear possui um nível de precisão elevado, por suas técnicas de validação Mean Absolute Error (MAE) e Root Mean Squared Error (RMSE). A primeira consiste em uma medida de erros entre observações pareadas que expressam o mesmo fenômeno, a segunda por sua vez é a medida que calcula "a raiz quadrática média" dos erros entre valores observados (reais) e previsões (hipóteses) (NOOR et al, 2007). Para criação do modelo, foi necessário fazer uso de diversas tecnologias e bibliotecas que possibilitam e facilitam todo o processo de desenvolvimento. Nas seções a seguir elas serão abordadas para melhor entendimento de seus papéis no contexto deste trabalho.

2.4 Regressão Linear

Outra tecnologia com capacidade de contribuir no desenvolvimento da solução é a Regressão Linear, pois busca encontrar uma correlação de forma linear sobre os dados de cada série temporal. Essa técnica é considerada uma subárea da Ciência de Dados, sendo muito utilizada em muitos algoritmos de Machine Learning (ML).

A regressão linear pode ser aplicada em diversos contextos e com os mais variados objetivos, na prática consiste em uma reta traçada a partir de uma relação em um diagrama de dispersão. Dessa maneira encontramos exemplos mais concretos quando precisamos realizar a predição de uma variável, a partir de uma correlação predefinida por outros dados presentes no modelo, quando por exemplo, queremos prever altura dos filhos a partir da altura de seus pais, variação dos salários e taxa de desemprego (GRUS, 2016).

Para o andamento do projeto, a regressão será aplicada com o objetivo de obter as predições de possíveis valores de batimentos cardíacos entre as séries de tempo não padronizadas, para que dessa forma os dados faltantes possam ser preenchidos, possibilitando assim uma perda mínima das informações da variabilidade da frequência cardíaca dos pacientes.

Para criação do modelo, foi necessário fazer uso de diversas tecnologias e bibliotecas que possibilitam e facilitam todo o processo de desenvolvimento. Nas seções a seguir elas serão abordadas para melhor entendimento de seus papéis no contexto deste trabalho.

2.5 Microserviços

Em um sistema monolítico, à medida que aparece necessidade de inserção de novas funcionalidades a um determinado sistema, a quantidade de código aumenta até que, em determinado momento a sua manutenção se torna indispensável, visto que princípios da Engenharia de Software - como coesão e acoplamento - não foram respeitados durante a implementação, o que tornou a manutenção da aplicação baixa.

Baseado no Princípio da Responsabilidade Única de Robert C. Martin, Newman (2015) propôs uma arquitetura de microserviços, onde serviços pequenos e autônomos trabalham de forma cooperativa. Sendo assim cada serviço possui uma única responsabilidade, de acordo com a regra de negócio que ele implementa, conseqüentemente torna-se fácil evitar que o código aumente de forma a prejudicar a sua manutenibilidade. Por ser uma entidade dissociada, a implantação de cada um dos microserviços devem ser realizadas de forma independente,

como um serviço isolado em uma plataforma como serviço (PaaS), e toda a sua comunicação deverá ser realizada via chamadas de rede (NEWMAN, 2015).

Pode-se dizer que microsserviços nada mais são do que serviços bem definidos, pouco acoplados, seguindo os princípios da boa e velha arquitetura orientada a serviços (SOA), com um grau de maturidade maior, justamente por usa independência e capacidade de trabalhada em conjunto com outros microsserviços ou de forma autônoma ao mesmo tempo (MOREIRA; BEDER, 2015).

Cada um dos microsserviços expõe uma interface de programação de aplicações, cujo acrônimo API vem do inglês Application Programming Interface, que será utilizada por outros serviços para comunicação. Visando possibilitar que cada microsserviço seja escrito em diferentes linguagens de programação, Moreira e Beder (2015) recomendam que sejam utilizados protocolos de comunicação padronizados, como o HTTP (Hypertext Transfer Protocol), e formatos de dados multiplataforma, como o JSON (JavaScript Object Notation).

A utilização de microsserviços oferece um conjunto de vantagens que podem ser identificadas como heterogeneidade tecnológica, resiliência, escalabilidade e facilidade de implantação (NEWMAN, 2015; MOREIRA; BEDER, 2015). Em um sistema com múltiplos serviços colaborativos, é possível definir visando a performance, qual a tecnologia mais adequada para implementar uma determinada funcionalidade. Em uma rede social, por exemplo, seria possível elaborar um microsserviço na linguagem de programação Java que armazenasse as interações entre os usuários em uma base de dados orientada a grafos. Por outro lado, o serviço responsável pela gestão dos conteúdos postados seria em Ruby e utilizaria um banco orientado a documentos.

Tendo em vista que a Plataforma Sênior Saúde Móvel é uma SOA, fez-se necessário a criação de um microsserviço. Ele nos possibilitará a execução das tarefas definidas, considerando a integração entre os Serviços já disponíveis. O RabbitMQ⁵ é a arquitetura de SOA utilizada para realização desse gerenciamento de comunicação via microsserviços, diante disso, tornou-se imprescindível o uso dela para a troca de mensagens entre as aplicações.

2.6 Variabilidade de Frequência Cardíaca

A frequência cardíaca é o número de batimentos cardíacos por minuto. Dessa maneira, a variabilidade da frequência cardíaca (VFC) consistirá na flutuação nos intervalos de tempo entre os batimentos cardíacos adjacentes (MCCRATY et al, 2015). A partir de tal característica

⁵ Disponível em: <<https://www.rabbitmq.com/>>. Acesso em: 20 de set. 2021

perceberemos que VFC é uma propriedade emergente de sistemas regulatórios interdependentes que operam em diferentes escalas de tempo para ajudar a nos adaptar aos desafios ambientais e psicológicos. VFC, de maneira direta, reflete na regulação do equilíbrio autonômico, pressão arterial (PA), trocas gasosas, intestino, coração e tônus vascular, que se refere ao diâmetro dos vasos sanguíneos que regulam a PA e, possivelmente, os músculos faciais (SHAFFER et al, 2017).

Um coração saudável não é caracterizado por possuir um padrão de pulsos ou batimentos constantes, muito pelo contrário as oscilações de um coração saudável são complexas e não lineares. A variabilidade dos sistemas não lineares fornece a flexibilidade para lidar rapidamente com um ambiente incerto e em mudança. Enquanto os sistemas biológicos saudáveis exibem complexidade espacial e temporal, a doença pode envolver uma perda ou aumento de complexidade (SHAFFER et al, 2017).

Uma VFC mais alta nem sempre é melhor, pois as condições patológicas podem produzir VFC. Quando as anormalidades da condução cardíaca elevam as medidas de VFC, isso está fortemente relacionado ao aumento do risco de mortalidade (particularmente entre os idosos). Um nível ideal de VFC está associado à saúde e capacidade de autorregulação e adaptabilidade ou resiliência. Níveis mais altos de VFC em repouso vagamente mediado estão ligados ao desempenho de funções executivas, como atenção e processamento emocional pelo córtex pré-frontal (SHAFFER et al, 2017).

Dentro do escopo deste trabalho, as VFC serão utilizadas de modo que as informações mencionadas anteriormente permitirão encontrar padrões de variabilidade que não sejam anormais, além disso, funcionarão como direcionamento para possíveis tomadas de decisões, através de códigos desenvolvidos que possibilitará entender mais a fundo métricas e informações relevantes sobre a VFC dos pacientes.

2.7 Bibliotecas Python

A linguagem de programação escolhida para o desenvolvimento do projeto foi Python⁶, uma linguagem de programação interpretada, de alto nível e que dá suporte a vários paradigmas. Consiste em uma linguagem de código aberto, utilizada para diversos fins, como em aplicações com Ciência de Dados, ML, desenvolvimento web, desenvolvimento de aplicativos, automação de aplicações etc. Na pesquisa apresentada, utilizaremos Python com dois principais objetivos,

⁶ Disponível em: < <https://www.python.org/> >. Acesso em: 20 de set. 2021

resolução do problema em ciências de dados, bem como o desenvolvimento de um microsserviço para disponibilizar dos dados na plataforma Sênior Saúde Móvel.

Python é a principal tecnologia utilizada nesta pesquisa, visto que está presente em toda parte de seu desenvolvimento, desde o trabalho com o processamento dos dados, até a criação do modelo. Desse modo, podemos destacar ainda o micro framework Flask⁷, utilizado para o desenvolvimento do serviço. Apesar de sua classificação como “micro”, em algumas extensões o Flask é poderoso e ideal para quem busca simplicidade, rapidez, soluções para projetos pequenos e aplicações robustas. Sendo assim, as seções posteriores abordarão de forma mais específica bibliotecas da linguagem Python que foram utilizadas no desenvolvimento da solução.

2.7.1 Flask

O Flask é framework Web pequeno e leve, que fornece ferramentas e recursos úteis que facilitam a criação de aplicativos Web em Python. Ele proporciona flexibilidade aos desenvolvedores especialmente aos iniciantes, pois viabiliza a elaboração de aplicativos Web rapidamente usando um único arquivo Python. Flask é extensível e não força uma estrutura de diretório em específico.

Desse modo, Flask se torna importante para o desenvolvimento, pois se trata de um microsserviço para processamento das séries temporais com Ciência de Dados, por conseguinte o framework permite a adequação aos algoritmos ML que realizam os devidos processamentos.

2.7.2 Pandas

Para contribuir com o processamento das séries temporais será utilizada a biblioteca Pandas⁸, uma biblioteca de código aberto, utilizada para manipulação e análise de dados. Ela foi desenvolvida na linguagem de programação Python e é caracterizada por ser rápida, potente, flexível e fácil de usar. É uma biblioteca fundamental para realização de análises exploratórias de dados, já que possibilita a realização de diversas operações em bases de dados, como leitura, manipulação e agregação de dados de forma simplificada.

⁷ Disponível em: < <https://flask.palletsprojects.com/en/2.0.x/> >. Acesso em: 20 de set. 2021

⁸ Disponível em: < <https://pandas.pydata.org/> >. Acesso em: 20 de set. 2021

A biblioteca Pandas é de extrema importância para o desenvolvimento deste trabalho, pois por meio dela são realizadas as leituras, escritas e manipulações das bases de dados trabalhadas.

2.7.3 Numpy

Outra biblioteca fundamental para contribuição neste trabalho é a Numpy⁹, ela também consiste em um módulo de código aberto. Desenvolvida na Linguagem Python, ela possui uma particularidade por conter cálculos internos que são processados na Linguagem C, uma das mais rápidas. Por esse motivo ela se torna uma ferramenta significativa, sendo utilizada principalmente para os cálculos matemáticos e métricas na pesquisa. Desse modo é importante destacarmos esse módulo do Python pelo seu grande potencial de processamento, bem como sua velocidade e desempenho em uma grande massa de dados.

2.7.4 HRV-Analysis

O pacote hrv-analysis¹⁰ consiste em um módulo do Python voltado para a análise de Variabilidade de Frequência Cardíaca sobretudo no processamento de intervalos RR. Corresponde a um pacote de código aberto para o ambiente de computação estatística Python, que oferece suporte a uma ampla variedade de métodos de análise VFC, como tempo, frequência e linearidade. Ele também inclui várias funções para pré-processamento de sinais, como outliers e remoção de batimento ectópico.

Este pacote é adequado para pesquisadores, profissionais que trabalham na área de saúde, bem como para desenvolvedores e, mais amplamente, para qualquer pessoa disposta a realizar análises detalhadas sobre variabilidade do batimento cardíaco. Essa biblioteca possibilita uma maior facilidade no trabalho de manipulação e aquisição de métricas sobre os dados e sua utilização é importante neste trabalho, tendo em vista que os picos dos intervalos RR serão avaliados, bem como a aquisição de métricas que se tornam fundamentais para os profissionais podendo realizar análises das VFC dos pacientes.

⁹ Disponível em: < <https://numpy.org/> >. Acesso em: 20 de set. 2021

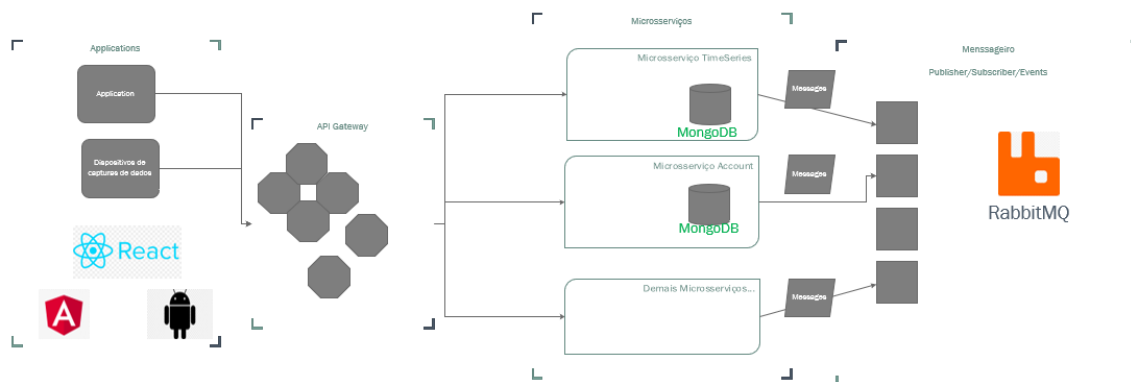
¹⁰ Disponível em: < <https://aura-healthcare.github.io/hrv-analysis/readme.html#hrv-analysis> >. Acesso em: 23 de set. 2021

3 METODOLOGIA

De acordo com as constantes interações dentre as áreas de conhecimento abordadas, foram apresentados e analisados os principais problemas enfrentados pelas lacunas presentes nas séries temporais, provocadas pelos dispositivos que coletam as frequências cardíacas dos pacientes, conseqüentemente a inviabilidade de extração de informações das séries temporais diante da coleta.

O projeto Sênior Saúde Móvel possui uma arquitetura de microsserviços que utiliza o RabbitMQ em sua implementação. Ele possibilita que as aplicações, com funcionalidades objetivas, realizem a troca de mensagens de forma rápida e confiável, além da compatibilidade com diversas linguagens de programação, contendo ainda uma interface de administração de fácil manuseio. A seguir podemos visualizar a Figura 1 que consiste no diagrama de sua arquitetura antes da inserção do microsserviço hrv-analytics, que serão apresentados em seções posteriores.

Figura 1 - Diagrama da arquitetura do projeto Sênior Saúde Móvel



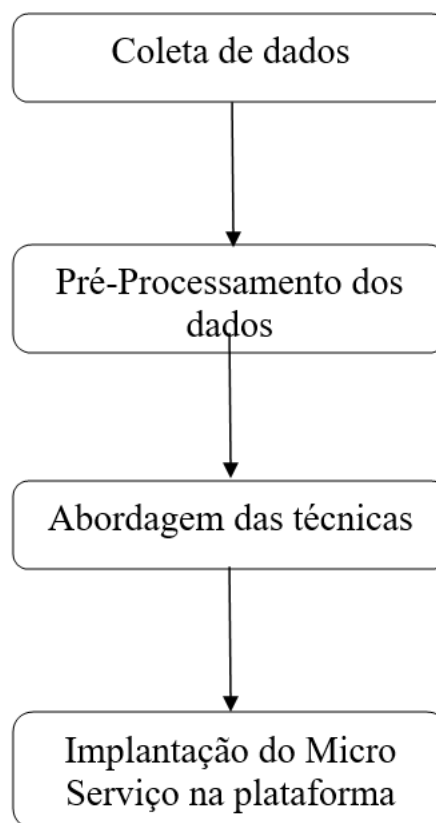
A partir disso, foram avaliadas e levantadas soluções com o objetivo de aperfeiçoar o processo utilizado atualmente no desempenho das coletas de séries temporais. Visando agregar qualidade sobre a extração de informações e preenchimento dos dados ausentes, foram utilizados algoritmos de Ciência de Dados com a possibilidade de oferecer resultados satisfatórios, com o intuito final de disponibilizar aos profissionais de saúde, uma aplicação que colaborasse no diagnóstico dos pacientes.

Sendo assim, foram definidas quatro grandes etapas a serem aplicadas para tornar possível os objetivos estabelecidos. A Figura 2 apresenta os principais processos para a criação do microsserviço que utiliza Ciência de Dados para disponibilização das métricas dos pacientes. Inicialmente serão validados e definidos os dispositivos usados para coleta dos dados, avaliando vantagens e desvantagens dos equipamentos. Em seguida será efetuado o pré-processamento nas

séries temporais, tendo como principais atividades a limpeza dos dados observando características importantes, como a padronização de batimentos de acordo com cada atividade física.

A partir disso tornou-se necessária uma abordagem sobre técnicas de preenchimento verificando e analisando técnicas viáveis para complementar informações faltantes, contribuindo com maior eficiência. Por fim temos a etapa de implantação da plataforma como um microsserviço, permitindo a disponibilização dos resultados encontrados. Mediante ao citado os temas expostos serão abordados com mais detalhamento, para que esclareça todo o processo realizado até a conclusão da aplicação final.

Figura 2 - Processo para implantar microsserviço hrv-analytics



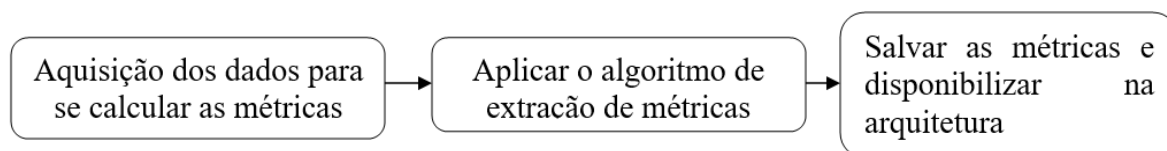
3.1 Definição dos Requisitos

Antes de iniciarmos o processo de desenvolvimento, assim como de análise de algoritmos de Ciência de Dados para resolução do problema, se faz necessário a definição dos requisitos funcionais de toda a aplicação. Sendo assim, precisamos definir como se dará a implementação do serviço que realizará todo o processo de disponibilização das métricas na arquitetura.

Com o uso da arquitetura que o projeto Sênior Saúde Móvel possui, podemos assim resgatar os dados que são coletados pela plataforma, e dessa maneira disponibilizar através do

protocolo de troca de mensagens entre os microsserviços, Advanced Message Queuing Protocol (AMQP), a possibilidade de envio de informações para as devidas necessidades dos demais serviços. A Figura 3 possui o fluxograma que nos mostra as atividades a serem realizadas para conseguir se adequar a arquitetura.

Figura 3 - Processo para desenvolvimento do microsserviço hrv-analytics



Diante do exposto, foram estabelecidas as etapas a serem implementadas, definindo os requisitos (Requisitos Funcionais - RF) que a aplicação necessita, como estão dispostos a seguir:

[RF001]: Deve ser definido o dispositivo a ser usado na coleta.

[RF002]: Deve ser definido o período de avaliação das séries temporais para extrações de informações relevantes.

[RF003]: Deve ser definida a técnica de Ciência de Dados que tenha o melhor desempenho nas séries temporais.

[RF004]: O microsserviço deve adquirir informações sobre os pacientes, que está presente em um serviço disponível do Sênior Saúde Móvel, o Account.

[RF005]: A aplicação deve buscar as informações sobre os sinais de batimentos cardíacos no serviço presente na arquitetura, TimeSeries.

[RF006]: A aplicação deverá extrair as métricas dos pacientes de acordo com o período determinado.

[RF007]: O microsserviço deve disponibilizar na arquitetura, as métricas contidas no banco de dados, mantendo um histórico dos pacientes.

3.2 Coleta dos Dados

O projeto Sênior Saúde Móvel consiste em uma arquitetura de microsserviços, onde exercem uma série de funcionalidades que atuam em conjunto. Contudo o projeto hrv-analytics¹¹, é um estudo voltado especificamente para a análise da variabilidade da frequência cardíaca que utilizará o TimeSeries, microsserviço responsável pela coleta e armazenamento

¹¹ Disponível em: < <https://github.com/senior-saude-movel/hrv-analytics> >. Acesso em: 9 de out. 2021

das informações referentes aos batimentos cardíacos dos pacientes, contendo também dados relacionados aos tipos de atividades que o paciente exerça como corrida, pedalada, descanso etc. Dessa maneira, como a pesquisa está voltada para pacientes idosos a primeira etapa consiste na definição do dispositivo utilizado para realizar a coleta.

O Sênior Saúde Móvel possui três tipos de dispositivos a serem escolhidos, o Fitbit Inspire, Polar H10 e Polar OH1. Eles foram avaliados considerando seus benefícios e malefícios que interfiram diretamente no projeto. Por consequência o primeiro fator de análise foi a viabilidade dos pacientes idosos utilizarem o dispositivo no maior tempo possível, que levou a percepção da necessidade de um equipamento confortável. Foram consideradas ainda outros dois critérios de análise, eficiência do dispositivo e intervalo de tempo entre os sinais coletados.

3.2.1 Fitbit Inspire

Fitbit Inspire¹² consiste em uma pulseira que é composta por um rastreador de análise de atividades físicas, com uma interface amigável para análise diária sobre gastos calóricos, bem como informações sobre batimentos cardíacos que ajuda o usuário a construir hábitos saudáveis, de forma dinâmica e prática. O dispositivo ainda utiliza a tecnologia Bluetooth 4.0 se conectando a dispositivos Android e Sistema Operacional Móvel da Apple Inc (IOS), facilitando a conexão com o serviço do Sênior Saúde Móvel.

3.2.2 Polar H10

O Polar H10¹³ consiste em uma cinta peitoral, que possui um sensor de frequência cardíaca, capturando informações sobre os batimentos do usuário. O grande diferencial desse dispositivo seria a localização da captura dos dados. Por estar posicionado muito próximo do coração suas informações são bastante precisas, além disso o intervalo de tempo de captura é a cada segundo e assim como o Fitbit Inspire possui facilidade de conexão com dispositivos que possuam a tecnologia Bluetooth 4.0. Mediante a tais informações, o equipamento se tornou relevante na presente pesquisa para a análise dos dados, sendo considerado pelos profissionais de saúde como um padrão ouro em relação aos dispositivos dispostos.

¹² Disponível em: < https://staticcs.fitbit.com/content/assets/help/manuals/manual_inspire_en_US.pdf >. Acesso em: 10 de ago. 2021

¹³ Disponível em: < https://support.polar.com/e_manuals/H10_HR_sensor/Polar_H10_user_manual_Portugues/manual.pdf >. Acesso em: 10 de ago. 2021

3.2.3 Polar OH1

O Polar OH1¹⁴ se trata de um sensor ótico de frequência cardíaca compacto, diferentemente dos anteriores, esse dispositivo mede a frequência cardíaca no braço ou na têmpora. Por essa característica ele se torna versátil e oferece ótimas opções de dispositivos de frequência cardíaca localizado no braço. Com o Polar OH1, pode-se realizar transmissão da frequência cardíaca em tempo real para o seu relógio desportivo, smartwatch ou outra aplicação de fitness via Bluetooth e, simultaneamente.

3.2.4 Definição do dispositivo

De acordo com o exposto podemos observar que todos os dispositivos avaliados possuem total capacidade para serem utilizados na coleta de dados, possuindo facilidade de conexão, bem como resultados satisfatórios sobre a coleta de informações de frequência cardíacas. Diante disso, o primeiro ponto observado foi a localização que são dispostos para utilização. Por se tratar de idosos o dispositivo usado deveria possuir uma eficiência em sua coleta de dados permitindo o mínimo possível de incômodo, para que os idosos se sintam mais à vontade e usem na maior parte do dia.

Com a análise o Fitbit Inspire se destacou, visto que é minimamente invasivo, viabilizando a correlação com outros sensores do dispositivo, havendo ainda a possibilidade da coleta ser realizada a longo prazo, já que os idosos podem efetuar de sua própria casa, sem necessidade do seu deslocamento para o consultório ou hospital. Portanto a pulseira incomoda minimamente o usuário no desempenho de qualquer atividade física, bem como quando o usuário esteja descansando ou dormindo. Mediante a tais conhecimentos, foi realizado um levantamento de pontos positivos e negativos sobre os três dispositivos como mostra a Tabela 1.

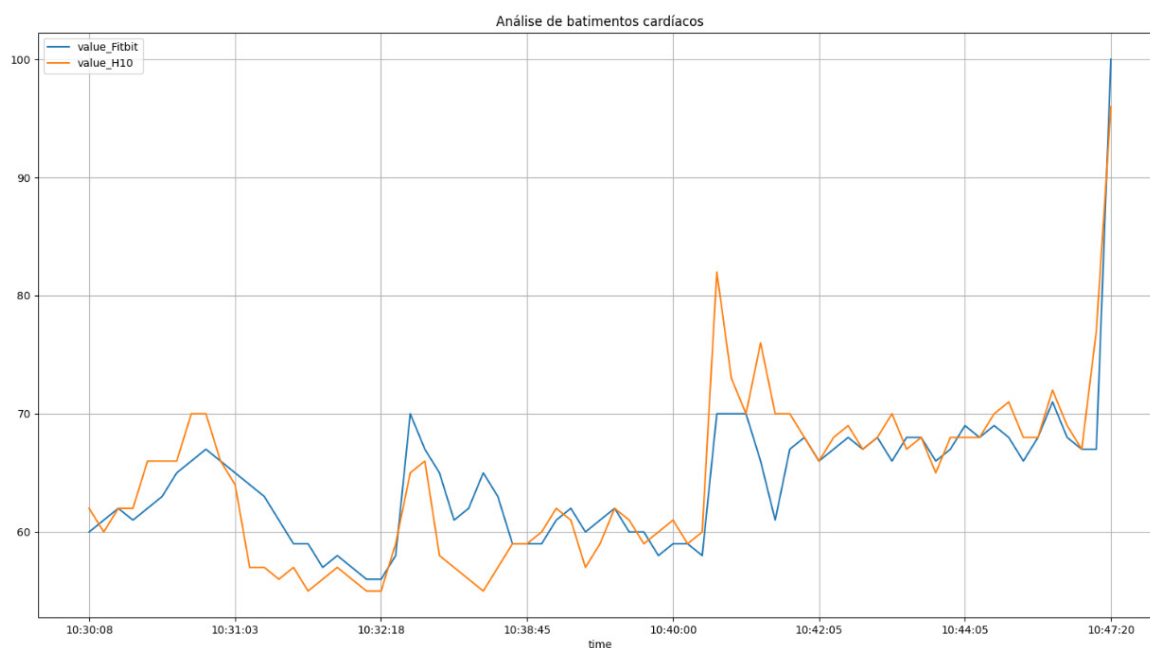
Tabela 1 - Comparação entre dispositivos para coleta de dados.

Fitbit Inspire		Polar H10		Polar OH1	
Favor	Contra	Favor	Contra	Favor	Contra
Leve e confortável de usar	Captura de frequência	Captura de frequência cardíaca a cada segundo	Incomoda em algumas atividades	Captura de frequência	Incomoda em algumas atividades

¹⁴ Disponível em: < https://support.polar.com/e-manuals/OH1/Polar_OH1_user_manual_Portugues/manual.pdf >. Acesso em: 10 de ago. 2021

	cardíaca a cada 5 segundos		ou no descanso/dormir	cardíaca a cada segundo	ou no descanso/dormir
Rastreamento de atividades básicas decente		Monitoramento de frequência cardíaca confiável	Não é possível editar dados no modo de automonitoramento	Monitoramento de frequência cardíaca confiável	Não é possível editar dados no modo de automonitoramento
Notificações por smartphone		Conexão com smartphones	Coleta realizada com o acompanhamento de um profissional de saúde ou no hospital	Conexão com smartphones	Coleta realizada com o acompanhamento de um profissional de saúde ou no hospital
Conexão com smartphones		Padrão Ouro, dispositivo que a coleta é mais precisa			
Correlação com outros sensores					
Coleta a longo prazo e realizada em domicílio					

Assim podemos observar que mesmo sendo o mais indicado a se utilizar, o Fitbit possui limitações. Por ser usada no braço a coleta da frequência cardíaca pode não ser tão precisa quanto os outros dispositivos, percebemos ainda que o intervalo de captura a cada cinco segundos limita as informações. Portanto foram analisadas capturas realizadas em um paciente com o Fitbit e o Polar H10, como na Figura 4, percebendo que mesmo com as limitações o dispositivo consegue um desempenho satisfatório, não havendo divergência suficiente para interferir o intuito principal da pesquisa.

Figura 4 - Comparação entre valores de coleta entre Fitbit Inspire e Polar H10

Dessa maneira foi escolhido o dispositivo Fitbit Inspire, para coleta das informações pela comodidade oferecida ao usuário com o objetivo de permitir aos pacientes o maior conforto, como maneira de incentivá-los a usar o maior tempo possível realizando qualquer tipo de atividade.

3.3 Pré-Processamento dos Dados

A Ciência de Dados é uma tecnologia fundamental quando há a necessidade de visualizar, organizar, limpar e extrair informações de dados em grandes quantidades. Dessa forma, em posse dos dados coletados e com o auxílio da biblioteca Pandas, foi realizada a leitura dos dados presente no microsserviço TimeSeries, para visualização e manipulação.

Posteriormente, foi efetuada uma análise exploratória das informações, buscando examinar de maneira geral possíveis processamentos, alterações e tratamentos a serem aplicados, além de potenciais comportamentos anômalos e problemas que pudessem surgir.

3.3.1 Tratamento de dados faltantes

Durante o desenvolvimento de aplicações que realizam manipulações em bases de dados, é comum a presença de informações faltantes, seja por falta de preenchimento ou problemas no armazenamento. Como citado anteriormente o dispositivo permite lacunas nas séries temporais já que a coleta é realizada a cada cinco segundos. Essas informações ausentes

são fundamentais para análises sobre a saúde dos pacientes, portanto fez-se necessário o tratamento com prudência para que o preenchimento não interfira negativamente sobre o modelo.

O preenchimento precisa nos garantir que as informações sobre os batimentos cardíacos sejam preservadas, necessitam que o algoritmo de preenchimento se adeque com as categorias de atividades que o paciente exerça, permitindo uma análise realista sobre o status de saúde dele.

3.3.2 Remoção de dados faltantes irrelevantes

Existem diversas formas diferentes de lidar com dados faltantes, seja aplicando média entre os dados, fazendo comparações ou até mesmo removendo-os. Desta forma, se fez necessário uma busca sobre técnicas de preenchimento sobre os dados, que mais se assemelhe a variações de batimentos cardíacos reais, possibilitando que a perda de informações fosse mínima e que se torne possível uma confiabilidade sobre as técnicas aplicadas.

3.4 Abordagem das Técnicas

Diante do exposto, percebe-se a necessidade de preenchimento dos dados sobre frequência cardíaca. Portanto fez-se necessário a busca e análise sobre as técnicas em Ciência de Dados que pudessem disponibilizar resultados satisfatórios sobre complementação das séries temporais faltantes. Além disso foi observado os dados preenchidos após a aplicação das técnicas com os capturados pelo Polar H10 dispositivo com maior precisão, para que assim fosse determinado qual a técnica mais viável e mais se assemelha aos dados reais.

3.4.1 Regressão Linear

Regressão Linear é um algoritmo de ML muito conhecido e usado por diversos sistemas atualmente, se trata de um modelo estatístico e um tipo de algoritmo supervisionado, que consiste em um aprendizado aplicado quando se pretende encontrar a relação entre as variáveis encontrando um certo padrão de correlação. Neste trabalho, essa técnica foi abordada tendo o objetivo de visualizar os padrões sobre os batimentos cardíacos, de acordo com a atividade que o paciente realiza, dessa forma percebeu-se um padrão na frequência cardíaca de acordo com a

intensidade dos movimentos dos usuários, sendo assim tal característica foi usada para preenchimento.

Com a abordagem da Regressão Linear, no preenchimento das séries temporais faltantes, se observou que modelo apesar de ter comportamentos pertinentes em situações de complementação de dados ausentes, não conseguiu retornar resultados satisfatórios. O principal motivo para o ocorrido é a variância que os batimentos cardíacos apresentam em situações extremas, como excesso de esforço físico em um curto espaço de tempo.

O algoritmo não consegue adequar-se a uma variação tão significativa e considera como um outlier, desconsiderando dados como os propostos previamente em seu treinamento do modelo. Dessa forma foi constatado que mesmo sendo uma técnica que teria capacidade de prever resultados, sua aplicação não permite a solução esperada.

3.4.2 Interpolação Linear

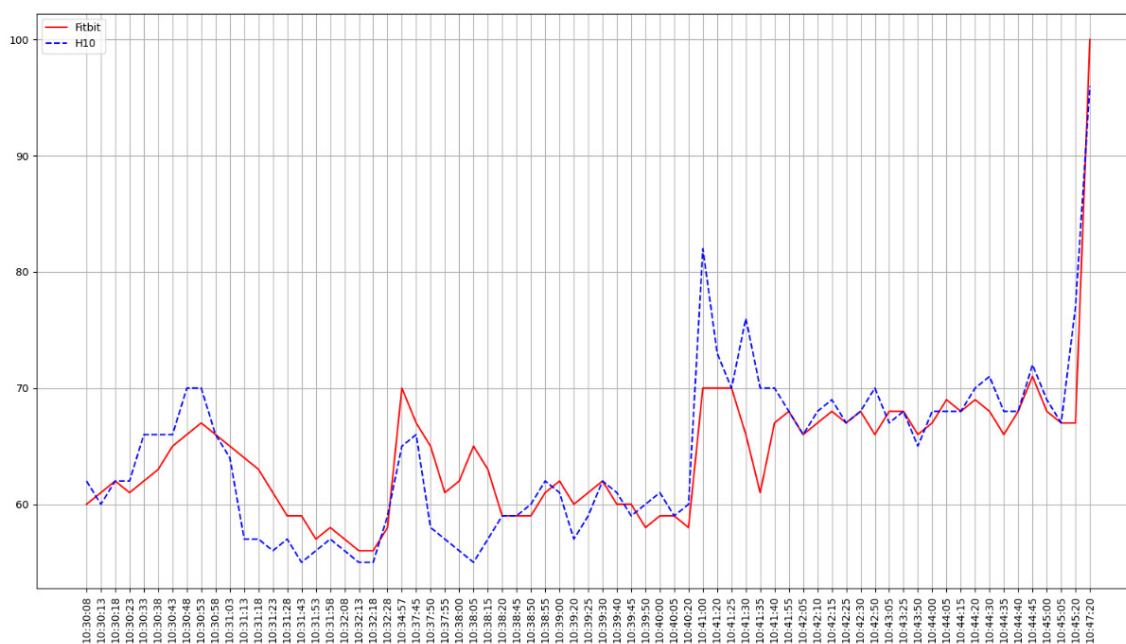
Outra técnica abordada no decorrer do trabalho foi a análise da interpolação linear e qual seu comportamento nas séries ausentes. A interpolação linear é um método matemático que ajusta uma função ao seu dado e usa essa função para extrapolar os dados faltantes, ela calcula a média entre os valores anterior e posterior do dado ausente.

É inviável constatar que todas as séries temporais possuam um padrão. Na presente pesquisa os dados apesar de possuir uma pertinência sobre o comportamento em determinadas atividades realizadas pelos pacientes, não é possível constatar um padrão bem definido, pois é frequente e normal a existência de alterações sobre os batimentos cardíacos em qualquer paciente.

Podemos constatar ainda a dificuldade enfrentada pela técnica de interpolação nos dados, pois como temos uma lacuna de pelo menos cinco segundos na coleta, o algoritmo tende a preencher os dados faltantes do intervalo com o mesmo valor, inviabilizando a análise de picos que pode representar um problema de saúde. Assim não conseguimos prever dados ou informações que foram perdidas pelo preenchimento, não sendo possível afirmar existência de problemas ou alterações significativas. Portanto, a interpolação linear não conseguiu um desempenho satisfatório, pois o padrão entre dados coletados não é linear.

A Figura 5 nos mostra o comportamento da aplicação da interpolação linear comparadas com os dados do Polar H10. Podemos observar que mesmo tendo um comportamento aceitável em certas situações, por outro lado em outras situações a discrepância é considerável, tornando-se um modelo que não nos permite um nível de confiança em sua implementação.

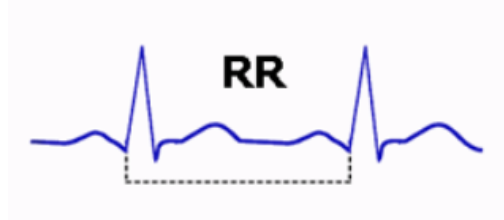
Figura 5 - Comparação entre valores processados pela Interpolação Linear, entre Fitbit Inspire e Polar H10



3.4.3 Métricas – Hrv-Analysis

Diante das técnicas apresentadas, se percebeu que o preenchimento de dados faltantes nas séries temporais seria uma tarefa que ao ser implantada se correria o risco. O algoritmo ou modelo pode perder informações importantes, que são essenciais quando estamos trabalhando com um serviço que pretende disponibilizar aos profissionais de saúde, uma análise minuciosa sobre o estado de saúde dos pacientes.

Partindo desse princípio, quando trabalhamos com séries temporais conseguimos visualizar características como os sinais de um eletrocardiograma, e a partir desses sinais, como vimos nas figuras anteriores, podemos encontrar ondas, intervalos e variações que aparecem nos seus resultados, podendo indicar uma série de doenças e eventos cardíacos, como o infarto. Com isso, se observou essencialmente que mesmo com a ausência de alguns dados essas análises seriam possíveis para compreensão e avaliação dos profissionais de saúde a partir do conceito Intervalo RR. O intervalo R-R, é a distância entre duas ondas R sucessivas, e é medido desde o início de uma onda R até o início da próxima R e sua duração depende da frequência cardíaca, como mostra a Figura 6.

Figura 6 - Exemplo de um intervalo RR

Tendo conhecimento sobre o exposto, foi encontrada uma biblioteca da linguagem Python, *hrv-analysis*, que permite a extração dessas informações, como de outras métricas fundamentais para uma análise mais criteriosa sobre as séries temporais que representam as VFC's. A Tabela 2 nos mostra as métricas que foram extraídas dos dados e a descrição de cada uma delas.

Tabela 2 - Métricas calculadas pelo algoritmo desenvolvido

Métrica	Descrição	
mean	Mean	Média dos batimentos cardíacos
std	standard deviation	Desvio Padrão
min	Min	Valor Mínimo
max	Max	Valor Máximo
sddn	Standard deviation of NN intervals	Desvio padrão de intervalos NN
sdsd	The related standard deviation of successive RR interval differences	O desvio padrão relacionado de diferenças sucessivas de intervalo RR
nni_50	The number of adjacent NN intervals that differ from each other by more than 50 ms	O número de intervalos NN adjacentes que diferem uns dos outros por mais de 50 ms
pnni_50	Percentage of successive RR intervals that differ by more than 50 ms	Porcentagem de intervalos RR sucessivos que diferem por mais de 50 ms
nni_20	The number of adjacent NN intervals that differ from	O número de intervalos NN adjacentes que diferem uns

	each other by more than 20 ms	dos outros por mais de 20 ms
rmssd	Root mean square of successive RR interval differences	Raiz quadrada média de diferenças sucessivas de intervalo RR
mean_hr	Mean heart rate	média da frequência cardíaca
max_hr	Average difference between the highest and lowest heart rates during each respiratory cycle	Diferença média entre as frequências cardíacas mais altas e mais baixas durante cada ciclo respiratório
min_hr	Average difference between the highest and lowest heart rates during each respiratory cycle	Diferença média entre as frequências cardíacas mais altas e mais baixas durante cada ciclo respiratório
std_hr	standard deviation heart rate	desvio padrão da frequência cardíaca
triangular_index	Integral of the density of the RR interval histogram divided by its height	Integral da densidade do histograma do intervalo RR dividido por sua altura
sdnni	Mean of the standard deviations of all the NN intervals for each 5 min segment of a 24 h HRV recording	Média dos desvios padrão de todos os intervalos NN para cada segmento de 5 min de uma gravação de 24 h HRV
sdann	Standard deviation of the average NN intervals for each 5 min segment of a 24 h HRV recording	Desvio padrão dos intervalos NN médios para cada segmento de 5 min de uma gravação HRV de 24 h

3.4.4 Benchmarking das técnicas (Avaliação comparativa)

Após a aplicação das três técnicas apresentadas, se torna inevitável termos uma avaliação comparativa sobre as abordagens referente a busca de cada solução. Para isso foi

utilizada uma tabela para apresentar os devidos resultados referente a cada técnica, como mostra a tabela 3.

Tabela 3 - Benchmarking sobre as técnicas abordadas

Técnica	Resultados:	Problemas:
Interpolação Linear	<ul style="list-style-type: none"> • Conseguiu encontrar padrões nas atividades físicas; • Conseguiu preencher os dados faltantes; 	<ul style="list-style-type: none"> • Quando possui uma variância instantânea dos dados, o modelo não possui um bom comportamento, ou seja, problemas com presença de outliers; • Tende a preencher os dados faltantes do intervalo com o mesmo valor, inviabilizando a análise de picos que pode representar um problema de saúde;
Regressão Linear	<ul style="list-style-type: none"> • Conseguiu encontrar padrões nas atividades físicas; • Conseguiu preencher os dados faltantes; 	<ul style="list-style-type: none"> • Quando possui uma variância instantânea dos dados, o modelo não possui um bom comportamento, ou seja, problemas com presença de outliers; • O algoritmo não consegue adequar-se a uma variação tão significativa;
HRV-Analysis	<ul style="list-style-type: none"> • Mudança na técnica de extração de informações relevantes; • Consegue extrair informações relevantes em qualquer período de tempo determinado; • Através da extração de métricas relevantes, é possível constatar problemas de saúde sobre os pacientes; 	<ul style="list-style-type: none"> • Realiza a extração de métricas apenas os dados capturados.

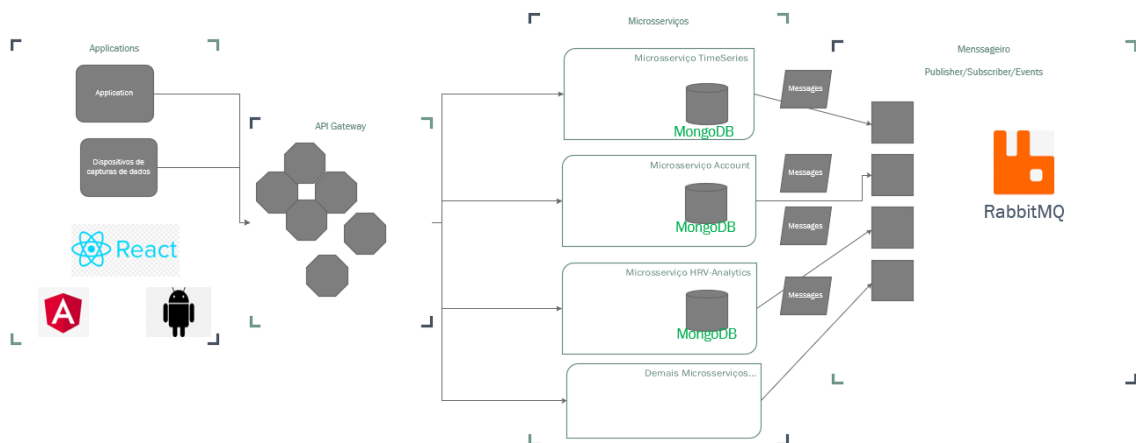
Dessa forma, podemos constatar que apesar da mudança na estratégia inicial do problema, percebeu o ótimo desempenho da técnica HRV-Analisys sobre a extração das métricas relevantes. Portanto, foi possível encontrar resultados muito satisfatórios, se estabelecendo o período para extração dessas informações a cada dia, ou seja, essas métricas serão extraídas sobre as séries temporais diariamente.

3.5 Implantação do microserviço na plataforma

O projeto Sênior Saúde Móvel consiste em uma aplicação formada por uma arquitetura de microserviços que utiliza o RabbitMQ em sua implementação. O RabbitMQ é um servidor de mensageria de código aberto (open source) desenvolvido em Erlang¹⁵, implementado para suportar mensagens em um protocolo denominado AMQP. Ele possibilita lidar com o tráfego de mensagens de forma rápida e confiável, além de ser compatível com diversas linguagens de programação, possui interface de administração nativa e ser multiplataforma. Dentre as aplicabilidades do RabbitMQ estão possibilitar a garantia de assincronicidade entre aplicações, diminuir o acoplamento entre elas, distribuir alertas, controlar fila de trabalhos em background.

Diante do exposto, a utilização do cálculo das métricas só poderia ser efetivada com a integração de um microserviço através da arquitetura já montada. Com a conclusão do serviço seria possível disponibilizar as métricas para que outros consumam através da plataforma da Sênior Saúde Móvel. A seguir na Figura 7, podemos visualizar um Diagrama contendo a arquitetura com a integração do microserviço desenvolvido.

Figura 7 - Diagrama da Arquitetura após a integração do HRV-Analytics



¹⁵ Disponível em: < <https://www.erlang.org/> >. Acesso em: 21 de set. 2021

3.5.1 Aquisição dos dados para cálculo das métricas

A partir da arquitetura montada no projeto Sênior Saúde Móvel, o primeiro passo foi a criação de uma comunicação entre o microsserviço de cálculo das métricas hrv-analytics, com os serviços da plataforma. Assim fez-se necessário a aquisição das informações dos usuários e as séries temporais que foram coletadas, nessa ordem subsequentemente foram acessados os serviços do account e em seguida no TimeSeries.

Com isso definido, foi montada uma estrutura de background, que no serviço seria uma maneira de validar a conexão entre os microsserviços certificando que não existem problemas na comunicação, bem como garantindo a segurança da troca de informações durante o envio das mensagens.

Dessa maneira, a primeira etapa foi realizada com sucesso e a partir de então se tornou possível a execução do serviço hrv-analytics. É importante ressaltar que a cada usuário do account, seria buscado as séries temporais referentes no TimeSeries, com isso a ordem de execução ficou definida, busca pelos id's dos usuários e em seguida busca da coleta dos batimentos cardíacos feita para o dia em execução.

3.5.2 Aplicar o algoritmo de extração de métricas na arquitetura do RabbitMQ

Como citado anteriormente, a preferência do Framework para desenvolvimento do serviço foi o Flask, que utiliza a linguagem de programação Python, e isso se tornou o principal fator de escolha, já que os algoritmos de Ciência de Dados para cálculo das métricas é realizado pela mesma.

A partir disso, as séries temporais coletadas foram buscadas nos microsserviços correspondentes, sendo preciso aplicar do algoritmo desenvolvido para cálculo de métricas. Portanto o código desenvolvido foi adequado no hrv-analytics, diante da arquitetura montada. Isso se tornou possível com o padrão de projeto Repository, que tem como objetivo dar apoio ao modelo de domínio fornecendo persistência na implementação.

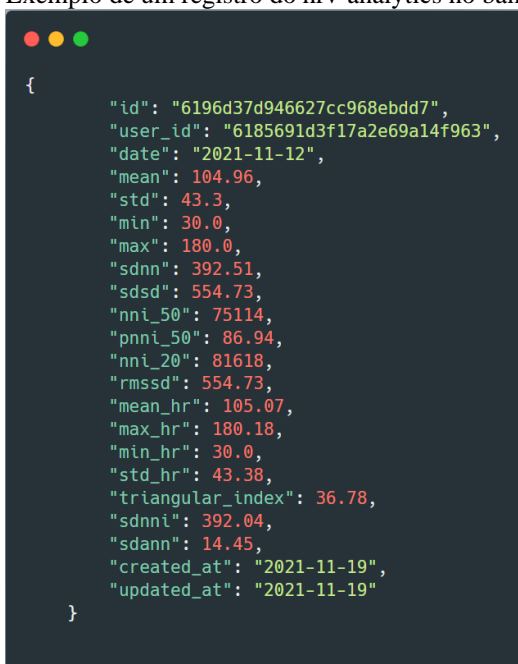
3.5.3 Salvar as métricas e disponibilizar na arquitetura

Com a execução do cálculo das métricas, a etapa seguinte seria manter as informações guardadas para acessos futuros, e para isso se faz necessário o uso de um banco de dados. Como estávamos trabalhando com mais de dezessete métricas a manipulação dos dados na linguagem

Python se tornou mais viável e convencional com a estrutura de dicionários, que vincula um valor a um identificador. O MongoDB¹⁶, foi o banco de dados escolhido para armazenamento das informações processadas, um banco de dados não relacional que permite uma integração de escala simples com Python, a partir da estrutura de dicionários utilizadas.

A Figura 8 nos permite visualizar um exemplo de um registro no banco, trazendo as informações das métricas citadas acima, como também outros registros, user_id que representa o identificador do usuário coletado, bem como a data referente ao registro das métricas e ainda data de criação como uma possível atualização, subseqüentemente date, created_at e updated_at.

Figura 8 - Exemplo de um registro do hrv-analytics no banco de dados

A screenshot of a terminal window showing a JSON document. The document contains various metrics for heart rate variability, including id, user_id, date, mean, std, min, max, sdnn, sdsd, nni_50, pnni_50, nni_20, rmssd, mean_hr, max_hr, min_hr, std_hr, triangular_index, sdnni, and sdann. It also includes timestamps for created_at and updated_at.

```
{
  "id": "6196d37d946627cc968ebdd7",
  "user_id": "6185691d3f17a2e69a14f963",
  "date": "2021-11-12",
  "mean": 104.96,
  "std": 43.3,
  "min": 30.0,
  "max": 180.0,
  "sdnn": 392.51,
  "sdsd": 554.73,
  "nni_50": 75114,
  "pnni_50": 86.94,
  "nni_20": 81618,
  "rmssd": 554.73,
  "mean_hr": 105.07,
  "max_hr": 180.18,
  "min_hr": 30.0,
  "std_hr": 43.38,
  "triangular_index": 36.78,
  "sdnni": 392.04,
  "sdann": 14.45,
  "created_at": "2021-11-19",
  "updated_at": "2021-11-19"
}
```

3.5.4 Conclusão da implementação

Sendo assim foi finalizado toda a implementação do microsserviço hrv-analytics, tornando-o responsável por realizar a variabilidade da frequência cardíaca, onde suas principais funcionalidades, consiste em:

- Geração de métricas do usuário;
- Calcula e salva métricas do banco de dados;
- Integração com RabbitMQ, serviços utilizados TimeSeries e Account.

¹⁶ Disponível em: < <https://www.mongodb.com/> >. Acesso em: 31 de set. 2021

Foi desenvolvido ainda a documentação da aplicação, com a plataforma Swagger, possibilitando aos desenvolvedores futuros, ou que necessitem do serviço, a análise de rotas utilizadas bem como a sua funcionalidade. Isso facilita o teste unitário de tal maneira que o desenvolvedor de forma simples faça uma análise do comportamento do serviço, assim como seu retorno. É importante destacar ainda que o entendimento da aplicação fica mais claro, com tal ferramenta, pois em cada rota disponível do serviço é mostrado uma documentação descrevendo sobre funcionalidades, entradas e saídas.

Podemos concluir dessa forma que apesar do desafio proposto o hrv-analytics, se tornou o produto final, atingindo os objetivos levantados no início do projeto. Destacamos ainda a capacidade de integração entre algoritmos de Ciência de Dados com uma arquitetura de microsserviço, capaz de se comunicar entre vários tipos de sistemas e composto por uma série de funcionalidades que tem um poder significativo para qualquer porte de projeto sugerido.

4 CONSIDERAÇÕES FINAIS

Quando se deu início ao trabalho de pesquisa, foi possível observar a dificuldade enfrentada para preenchimento dos dados de batimentos cardíacos, e ainda que sua ausência influencia de maneira significativa nas possíveis inferências, complicando ainda mais a análise dos profissionais de saúde sobre os pacientes observados. Desse modo, houve a necessidade de análise e desenvolvimento de uma ferramenta que contribuísse com todo o processo de preenchimento das séries temporais ausentes, sendo ainda visualizada a necessidade da criação de um microsserviço para disponibilização dos dados processados.

A partir disso, a pesquisa teve por objetivo desenvolver uma aplicação que fosse capaz de contribuir com todo o processo de preenchimento das séries temporais, de maneira a facilitar o trabalho dos profissionais de saúde e continuar realizando análise sobre as informações dos batimentos cardíacos. Isto posto, é possível constatar que o objetivo definido foi alcançado, pois o projeto conseguiu demonstrar todo o processo necessário para estruturação e desenvolvimento da ferramenta desejada.

Considerando as lacunas apresentadas na coleta de dados sobre os batimentos cardíacos de pacientes inseridos no projeto Sênior Saúde Móvel e que essa ausência impossibilitava um avanço nas análises sobre o estado de saúde dos mesmos, observou-se como uma oportunidade o desenvolvimento de um projeto que sanasse esse problema.

Para isso, foi realizado um estudo sobre as possíveis soluções viáveis, apresentadas nas seções 3.3 Abordagem das técnicas e seguintes, objetivando o preenchimento dos dados, de acordo com as análises feitas, percebeu-se a possibilidade de cálculo de métricas fundamentais sobre os batimentos cardíacos dos pacientes. Tendo essa aplicação implementada, foi desenvolvido ainda um microsserviço para disponibilização na plataforma Sênior Saúde Móvel, tendo uma facilidade na visualização dos resultados adquiridos bem como uma maior praticidade na avaliação sobre o diagnóstico do paciente de acordo com as informações dispostas na aplicação.

Por conseguinte, é cabível que posteriormente sejam desenvolvidos e incentivados projetos que busquem atuar nas barreiras enfrentadas pelo hrv-analysys, buscando soluções para ampliar as métricas calculadas que a ferramenta adquire, de modo que essas disponham da maior quantidade de informações necessárias para facilitação e alerta aos profissionais de saúde que utilizam a plataforma Sênior Saúde Móvel.

REFERÊNCIAS

ANDRADE, Paulo Evaristo de. et al. **Reduction of heart rate variability in hypertensive elderly**. Blood Pressure, 2017. Disponível em: < <http://dx.doi.org/10.1080/08037051.2017.1354285> >. Acesso em: 02 jun. 2021.

ARAÚJO, Fábio Meneghetti Ugulino De. **Interpolação**. Departamento de Engenharia da Computação e Automação - UFRN, 2011. Disponível em: < <https://www.dca.ufrn.br/~meneghet/FTP/MCEC/Transp05.pdf> >. Acesso em: 02 jun. 2021.

DHAMALA, Jwala. et al. **Multivariate Time-series Similarity Assessment via Unsupervised Representation Learning and Stratified Locality Sensitive Hashing: Application to Early Acute Hypotensive Episode Detection**. IEEE Xplore, 2018. Disponível em: < <https://ieeexplore.ieee.org/document/8506445> >. Acesso em: 02 jun. 2021.

GRUS, Joel. **Data Science From Scratch: First Principles with Python**. 1.ed. Sebastopol: O'Reilly Media, 30 abr. 2015.

MCCRATY, Rollin; SHAFFER, Fred. **Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health Risk**. Sage Journals, 2015. Disponível em: < <https://journals.sagepub.com/doi/abs/10.7453/gahmj.2014.073> >. Acesso em: 02 jun. 2021.

MOREIRA, Pedro Felipe Marques; BEDER, Delano Medeiros. **Desenvolvimento de Aplicações e Micro Serviços: Um estudo de caso**. Revista TIS, v. 4, n. 3, 2015. Disponível em: < <http://revistatis.dc.ufscar.br/index.php/revista/article/viewFile/364/127> >. Acesso em: 04 jun. 2021.

NAEINI, Emad Kasaeyan. et al. **A Real-time PPG Quality Assessment Approach for Healthcare Internet-of-Things**. Scopus, 2019. Disponível em: < <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85071917915&doi=10.1016%2fj.procs.2019.04.074&partnerID=40&md5=f01d61d350b8cd6fe4a9ee4648f47bde> >. Acesso em: 02 jun. 2021.

NEWMAN, Sam. **Building microservices**. Sebastopol: O'reilly Media, 2015. 280 p. 1 v. Disponível em: < https://www.nginx.com/wp-content/uploads/2015/01/Building_Microservices_Nginx.pdf >. Acesso em: 04 jun. 2021.

NOOR, Norazian Mohamed. et al. **Comparison of linear interpolation method and mean method to replace the missing values in environmental data set**. Scientific.net, 2007. Disponível em: < <https://www.scientific.net/MSF.803.278> >. Acesso em: 02 jun. 2021.

PRABHAKAR, Sunil Kumar; RAJAGURU, Harikumar; LEE, Seong-Whan. **Metaheuristic-Based Dimensionality Reduction and Classification Analysis of PPG Signals for Interpreting Cardiovascular Disease**. Scopus, 2019. Disponível em: < <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077627459&doi=10.1109%2fACCESS.2019.2950220&partnerID=40&md5=cf701e5c53d54af6102ac62065b87362> >. Acesso em: 02 jun. 2021.

PRATAMA, Irfan. et al. **A Review of Missing Values Handling Methods on Time-Series Data**. IEEE Xplore, 2016. Disponível em: < <https://ieeexplore.ieee.org/document/7858189> >. Acesso em: 02 jun. 2021.

SHAFFER, Fred; GINSBERG, Jean Paul. **An Overview of Heart Rate variability Metrics and Norms**. PubMed.gov, 2017. Disponível em: < <https://pubmed.ncbi.nlm.nih.gov/29034226/> >. Acesso em: 02 jun. 2021.