INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Sistema de apoio à decisão preditivo baseado em dados de tele-saúde

Sara Emídio Nunes

Mestrado em Métodos Analíticos para a Gestão (Business Analytics)

Orientadores:

Doutor Luís Manuel Nobre de Brito Elvas, Investigador, Centro de Investigação em Tecnologias e Arquitetura (ISTAR-Iscte)

Doutor João Carlos Amaro Ferreira, Professor Auxiliar (com agregação), ISCTE-Instituto Universitário de Lisboa

Outubro, 2025

BUSINESS
SCHOOL

Sistema de apoio à decisão preditivo baseado em dados de tele-saúde

Departamento de Métodos Quantitativos para Gestão e Economia

Sara Emídio Nunes

Mestrado em Métodos Analíticos para a Gestão (Business Analytics)

Orientadores:

Doutor Luís Manuel Nobre de Brito Elvas, Investigador, Centro de Investigação em Tecnologias e Arquitetura (ISTAR-Iscte)

Doutor João Carlos Amaro Ferreira, Professor Auxiliar (com agregação), ISCTE-Instituto Universitário de Lisboa

Outubro, 2025

Dedico esta dissertação à minha família, cuja presença e apoio foram fundamentais em todas as etapas da minha vida.

Agradecimentos

A realização desta dissertação não teria sido possível sem o apoio, a paciência e o incentivo de várias pessoas que marcaram este percurso.

Agradeço à minha família, pelo amor e compreensão nos momentos mais exigentes e por acreditarem sempre nas minhas capacidades.

Aos amigos que estiveram presentes, pela amizade e força transmitida nas fases mais desafiantes.

Agradeço igualmente às empresas com as quais colaborei durante este período, pela disponibilidade e pela oportunidade de conciliar a atividade profissional com o desenvolvimento deste trabalho e, em particular, à Hope Care S.A., pela colaboração e pela disponibilização dos dados utilizados neste estudo.

Dirijo, por fim, um especial agradecimento aos dois professores que me orientaram e acompanharam neste processo, pela sua orientação científica, disponibilidade e contributos essenciais para a concretização desta dissertação.

A todos, obrigada!

Resumo

A transformação digital no setor da saúde tem impulsionado o desenvolvimento de sistemas de *telemonitorização* que permitem o acompanhamento contínuo e personalizado dos utentes fora do ambiente hospitalar. A presente dissertação propõe o desenvolvimento de um sistema de apoio à decisão preditivo baseado em dados de *tele-saúde*, com o objetivo de antecipar situações clínicas críticas através da análise e modelação de dados fisiológicos recolhidos remotamente.

Assente na metodologia *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*), o estudo utilizou dados fornecidos pela empresa Hope Care, S.A., integrando medições de frequência cardíaca, saturação de oxigénio, pressão arterial, parâmetros de sono e atividade física. Após a limpeza e preparação dos dados, realizou-se a análise estatística descritiva e exploratória que permitiu identificar padrões e correlações entre variáveis fisiológicas e estados de criticidade clínica.

Na fase de modelação preditiva, aplicaram-se técnicas de *machine learning* — *Logistic Regression*, *Support Vector Machine (SVM)*, *K-Nearest Neighbours (KNN)*, *Random Forest* (RF*)* e *Complement Naive Bayes* — para construir e comparar modelos capazes de prever a ocorrência de alertas clínicos. Os resultados obtidos demonstram o potencial destas abordagens na deteção precoce de anomalias e na otimização do processo de decisão médica.

Este trabalho contribui para o avanço da monitorização remota em saúde, promovendo uma abordagem preditiva, proativa e suportada por dados, que potencia a qualidade e a eficiência dos cuidados clínicos.

Palavras-chave: *tele-saúde*; *machine learning*; sistemas preditivos; apoio à decisão clínica; monitorização remota.

Classificação JEL: I10; C55.

Abstract

The digital transformation of healthcare has driven the development of *telemonitoring* systems that enable continuous and personalized patient follow-up outside hospital settings. This dissertation proposes the development of a predictive decision-support system based on *telehealth* data, aiming to anticipate critical clinical conditions through the analysis and modeling of remotely collected physiological data.

Following the *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) methodology, the study used data provided by Hope Care, S.A., which included heart rate, blood oxygen saturation, blood pressure, sleep parameters, and physical activity metrics. After data cleaning and preprocessing, a descriptive and exploratory statistical analysis was conducted to identify patterns and correlations between physiological variables and clinical criticality states.

During the predictive modeling phase, several *machine learning* techniques — *Logistic Regression*, *Support Vector Machine (SVM)*, *K-Nearest Neighbours (KNN)*, *Random Forest* e *Complement Naive Bayes* — were applied to build and compare models capable of predicting the occurrence of clinical alerts. The results demonstrated the potential of these approaches for early anomaly detection and optimization of medical decision-making processes.

This work contributes to the advancement of remote health monitoring by promoting a predictive, data-driven, and patient-centered approach that enhances the quality and efficiency of clinical care.

Keywords: *telehealth*; *machine learning*; predictive systems; clinical decision support; remote monitoring.

JEL classification: I10; C55.

Índice

Agradecimentos iii

Resumo v

Abstract vii

Lista de abreviaturas, acrónimos e siglas xi

1.Introdução 1

1.1. Tema e Delimitação 1

1.2. Problema e Questões de Investigação 2

1.3. Objetivos e Contributos 3

1.4. Abordagem Metodológica 4

2. Revisão da Literatura 7

2.1. Random forest 7

2.2. Redes neuronais 8

2.3. Seleção das *features* 9

2.4. Dispositivos de recolha de dados biométricos 9

2.5 Adesão dos pacientes 11

3. Metodologia 13

3.1. Compreensão do Negócio 13

3.2. Compreensão dos Dados 14

3.2.1 Diagnóstico Inicial da Qualidade dos Dados 15

3.2.2. Preparação dos Dados 16

3.3 Análise estatística descritiva 19

3.3.1. Análise de correlações entre variáveis 23

3.3.2. Correlação entre as variáveis criticity 24

3.3.2.1. A variável pulse_criticity 25

3.3.2.2. A variável SpO_2_criticity 26

3.3.2.3. A variável Systolic_criticity 28

3.4. Modelação dos dados 30
3.4.1. Modelos Preditivos 32
3.4.1.1 Dataframe Pulse Criticity.....33
3.4.1.2. Dataframe SpO_2 Criticity.....35
3.4.1.3. Dataframe Systolic Criticity 36
3.4.2. Conclusão Modelos.....38
4. Resultados e discussão 39
4.1. Dashboard 1 – Análise Individual do Paciente.....39
4.2. Dashboard 2 – Análise Transversal por Variável 41
5. Conclusões e trabalho futuro 43
Referências bibliográficas 46
Anexo.....51

Índice de tabelas

Tabela 1- Accuracy dos modelos preditivos da feature pulse criticity……....33
Tabela 2- Accuracy dos modelos preditivos da feature SpO_2 crriticity……35
Tabela 3- Accuracy dos modelos preditivos da feature Systolic Criticity……36

Índice de figuras

Figura 1- Dashboard - Standard.....33
Figura 2- Dashboard - Com seleção 40
Figura 3 - Dashboard - Paciente 41

Lista de abreviaturas, acrónimos e siglas

BPM – Batimentos por Minuto

CNNs – *Convolutional Neural Networks*

COVID-19 – *Coronavirus Disease 2019*

CRISP-DM – *Cross-Industry Standard Process for Data Mining*

EMA – Ecological Momentary Assessment

IA – Inteligência Artificial

IoMT – *Internet of Medical Things*

IoT – *Internet of Things*

KPIs – *Key Performance Indicators*

LSTM – *Long Short-Term Memory*

ML – *Machine Learning*

PRISMA – *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*

RF- *Random Forest*

RNA / RNAs – Rede(s) Neuronal(is) Artificial(is)

RPM – Monitorização Remota de Pacientes

SVM – *Support Vector Machine*

WoS – *Web of Science*

1.Introdução

1.1. Tema e Delimitação

A transformação digital no setor da saúde tem vindo a alterar o modo como os cuidados são prestados, migrando de práticas clínicas centradas na instituição para modelos cada vez mais focados no utente e apoiados em tecnologias emergentes (*Smith & Jones*, 2022). Neste contexto, a adoção de dispositivos de monitorização remota permite o acompanhamento contínuo de parâmetros fisiológicos fora do ambiente hospitalar, promovendo uma vigilância ativa e personalizada do estado de saúde dos utentes (*Brown et al.*, 2023). Paralelamente, a telemedicina tem-se consolidado como um pilar fundamental, facilitando a comunicação entre profissionais e pacientes e alargando o acesso a cuidados de forma remota (*Stoumpos, Kitsios & Talias*, 2023).

A *Hope Care, S.A.,* destaca-se neste panorama ao desenvolver soluções que integram dispositivos de recolha de dados biométricos e plataformas analíticas que suportam o processo de decisão clínica. A capacidade de reunir e processar grandes volumes de informação biométrica em tempo real potencia a deteção precoce de alterações fisiológicas, permitindo intervenções mais céleres e eficazes (*Miller & Davis*, 2021). Contudo, esta revolução não está isenta de desafios: a heterogeneidade dos dispositivos, a segurança e privacidade dos dados e a necessidade de compatibilidade entre sistemas constituem barreiras relevantes à implementação plena destas tecnologias (*Genezini,* 2022).

Com o crescimento exponencial das iniciativas de telemedicina, torna-se imperativo o desenvolvimento de ferramentas de análise avançada, que tirem partido de técnicas de ML e inteligência artificial (IA) para extrair valor dos dados recolhidos (*Oliveira*, 2024). Estas ferramentas são cruciais não apenas para apoiar a decisão clínica, mas também para gerar informações preditivas sobre o risco de eventos adversos, contribuindo para uma gestão proativa do acompanhamento dos utentes. Além disso, a inclusão dos próprios pacientes no processo — através de aplicações que lhes permitam visualizar e interpretar os seus próprios dados — reforça a adesão aos regimes terapêuticos e promove a literacia em saúde.

Esta dissertação centra-se na análise e previsão de criticidade clínica com base nos dados de monitorização remota fornecidos pela Hope Care, S.A. Serão explorados métodos de deteção de anomalias e modelos preditivos capazes de identificar padrões de risco em tempo útil. Pretende-se, assim, contribuir para um cuidado personalizado e eficiente alinhado com as tendências globais de transformação digital na saúde.

1.2. Problema e Questões de Investigação

O crescente volume, velocidade e heterogeneidade dos dados gerados pelos sistemas de monitorização remota impõem desafios de tratamento, qualidade e interpretação que vão para lá da simples recolha de informação (*Dupljak & Halili,* 2024). Estes dados, provenientes de diversos dispositivos - como *wearables*, sensores domiciliários e aplicações móveis - caracterizam-se pelas "4 V" do *Big Data* (volume, velocidade, variedade e veracidade), exigindo arquiteturas e métodos analíticos robustos capazes de extrair, em tempo real, sinais clínicos relevantes e, simultaneamente, garantir a segurança e a privacidade dos utentes (*Santos, Pereira & Costa*, 2022).

A transição de um paradigma reativo para um paradigma proativo na prestação de cuidados de saúde pretende a antecipação de eventos críticos através de modelos preditivos que sinalizem precocemente alterações fisiológicas anómalas (*Miller & Davis*, 2021; *Li & Wang*, 2023). Para além das dificuldades inerentes à modelação de séries temporais biométricas – geralmente com um grande volume de dados ausentes – há que ressalvar a necessidade de explicabilidade dos algoritmos, de forma que os profissionais de saúde possam confiar nas previsões e integrar as mesmas nos processos de decisão clínica.

Face a este cenário, o problema central deste estudo é: "De que forma é possível, a partir de dados biométricos recolhidos remotamente, identificar padrões temporais e selecionar modelos de aprendizagem automática que permitam antecipar, com elevado rigor, estados de alerta clínico em utentes?"

Para abordar esta problemática, a presente investigação centra-se em três questões de investigação:

1. Quais são as principais abordagens, modelos e tecnologias utilizadas na literatura para o desenvolvimento de sistemas preditivos e de apoio à decisão em tele-saúde?
2. Indicadores discriminantes – Como pode um sistema de apoio à decisão, baseado em dados de telemonitorização, integrar análise descritiva e modelação preditiva para antecipar alertas clínicos e apoiar a decisão médica em tempo real?
3. Desempenho de algoritmos de aprendizagem automática - Que algoritmos de *machine learning* (ML) apresentam o melhor equilíbrio entre sensibilidade e especificidade na previsão de alertas clínicos?

Ao centrar-se nestas três dimensões, esta investigação visa desenvolver e validar um protótipo integrado, capaz de conjugar as componentes num sistema de apoio à decisão

clínica que deverá permitir a sinalização precoce de situações de risco, promovendo intervenções mais atempadas, precisas e sustentadas por parte dos profissionais de saúde.

1.3. Objetivos e Contributos

A definição rigorosa dos objetivos de investigação é uma etapa fundamental no processo científico, pois orienta o desenho metodológico e assegura a coerência entre as questões formuladas e os resultados a alcançar. Segundo Creswell e Creswell (2018), a explicitação clara dos objetivos é determinante para garantir a consistência interna e a relevância científica de um estudo. Assim, os objetivos desta investigação decorrem da problemática identificada e das questões formuladas, articulando-se com o enquadramento teórico e a metodologia adotada.

O principal objetivo desta dissertação consiste em conceber e validar um modelo integrado de análise, visualização e previsão da criticidade clínica, sustentado na metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*) (*Wirth & Hipp*, 2000) e nos dados disponibilizados pela empresa Hope Care, S.A. Pretende-se, deste modo, contribuir para a otimização da monitorização remota de utentes e para o reforço da capacidade preditiva dos sistemas de apoio à decisão em contexto clínico.

Para concretizar este objetivo geral, definem-se os seguintes objetivos específicos:

1. Rever o estado da arte sobre sistemas de apoio à decisão e soluções preditivas aplicadas à telemonitorização, identificando as principais abordagens metodológicas, algoritmos e tecnologias descritas na literatura científica recente.
2. Preparar e caracterizar o conjunto de dados fornecido pela Hope Care, S.A., assegurando a sua qualidade e relevância na análise, através do tratamento de valores ausentes, normalização das variáveis e seleção dos indicadores clínicos e fisiológicos mais relevantes.
3. Realizar uma análise descritiva e exploratória dos dados, com o objetivo de identificar padrões, tendências e correlações entre variáveis, contribuindo para a compreensão do comportamento clínico dos utentes e da sua evolução ao longo do tempo.
4. Desenvolver e comparar modelos preditivos baseados em técnicas de ML de modo a antecipar alertas clínicos e eventos críticos, avaliando o desempenho de diferentes algoritmos quanto à sensibilidade, especificidade e equilíbrio entre ambas as métricas.
5. Validar e interpretar os resultados obtidos, avaliando a capacidade do modelo proposto para responder às questões de investigação delineadas e evidenciando o contributo

do sistema para a melhoria dos processos de decisão clínica e de monitorização remota.

A definição destes objetivos garante uma estrutura metodológica coerente, orientando todas as fases da investigação — desde a análise de dados até ao desenvolvimento e validação do modelo preditivo — e assegurando a aplicabilidade prática dos resultados no contexto da telemonitorização clínica.

Numa primeira fase, o estudo centra-se na caracterização do comportamento dos utentes no programa de monitorização, avaliando (1) a frequência de utilização dos dispositivos, (2) a taxa de adesão segmentada por género e faixa etária e (3) os padrões de medição registados ao longo do tempo. Paralelamente, procede-se à exploração da distribuição estatística dos parâmetros fisiológicos, com o objetivo de identificar variações significativas, valores atípicos (*outliers*) e, correlações entre variáveis. Uma componente essencial desta etapa é a análise das variáveis de criticidade clínica — *pulse_criticity, spo2_criticity* e *systolic_criticity* — com o objetivo de indentificar os principais fatores associados à ocorrência de estados críticos.

Após a compreensão dos perfis e padrões de utilização, o projeto avança para a fase de modelação preditiva. Nesta fase, são desenvolvidos e comparados vários algoritmos de ML, selecionados e avaliados de forma rigorosa, com o objetivo de estimar a probabilidade de ocorrência de situações clínicas críticas. Simultaneamente, são concebidos *dashbords* interativos que disponibilizam às equipas clínicas uma visualização clara, estruturada e intuitiva dos resultados analíticos e preditivos. Estas ferramentas visuais não só permitem a identificação imediata de padrões relevantes, como também possibilitam o acompanhamento individualizado do estado de cada utente. A base desta abordagem consiste na conversão de dados brutos em valor clínico tangível, permitindo antecipar riscos, personalizar cuidados e promover uma monitorização remota mais eficiente e centrada no utente.

1.4. Abordagem Metodológica

A metodologia adotada neste estudo é o *CRISP-DM*. Esta escolha justifica-se pelo facto de ser um modelo de processo amplamente reconhecido no âmbito de projetos de *data mining* e ciência de dados, devido à sua natureza estruturada, iterativa e flexível (*Chapman et al.*, 2000). A sua aplicação permite uma abordagem sistemática que abrange desde a formulação do problema até à implementação da solução.

O CRISP-DM organiza-se em seis fases principais: compreensão do negócio, compreensão dos dados, preparação dos dados, modelação, avaliação e implementação.

A natureza iterativa desta metodologia constitui um pilar fundamental, permitindo um movimento cíclico e flexível entre as diversas etapas. Esta característica possibilita o aprimoramento contínuo e a adaptação do projeto à medida que se adquire maior conhecimento sobre os dados e os objetivos. Esta flexibilidade é vital em projetos de análise de dados, onde a compreensão inicial pode evoluir significativamente ao longo do tempo, permitindo o retrocesso a fases anteriores conforme surjam novos conhecimentos ou necessidades durante o projeto (*Wirth & Hipp*, 2000).

Na primeira etapa, a compreensão do negócio, define-se o contexto clínico e operacional, traduzindo as necessidades em objetivos mensuráveis e critérios de sucesso. Este processo visa garantir a congruência entre os objetivos do projeto de DM e as necessidades, bem como as metas estratégicas, da organização, contribuindo para o desenvolvimento de soluções relevantes e executáveis (*Chapman et al.*, 2000).

A aplicabilidade transversal a diferentes setores de atividade constitui uma das principais vantagens do CRISP-DM. Esta característica possibilita a sua aplicação em diversos contextos, demonstrando ser particularmente apropriada para o estudo em questão, no âmbito da saúde (*Mariscal, Marbán & Fernández*, 2010).

A fase de preparação dos dados assegura a robustez e a pertinência do conjunto de dados para a modelação dos dados. A limpeza, transformação, integração e seleção de dados, realizadas nesta etapa, são fundamentais para minimizar o ruído e otimizar a qualidade, um fator crítico para a fiabilidade dos resultados futuros (*Han, Kamber & Pei*, 2011).

No âmbito da modelação, a metodologia CRISP-DM é caracterizada pela sua versatilidade na seleção e implementação de múltiplas técnicas de extração de dados. Esta fase não prescreve algoritmos específicos, mas permite a seleção das abordagens mais apropriadas para os objetivos definidos, otimizando o desempenho dos modelos (*Larose & Larose*, 2014).

A fase de avaliação, por sua vez, permite uma validação rigorosa dos modelos face aos critérios de negócio e aos dados. Este rigor assegura que as soluções desenvolvidas não são apenas estatisticamente sólidas, mas também práticas e capazes de gerar os *insights* ou predições esperadas (*Provost & Fawcett*, 2013).

A fase final, designada implementação, visa garantir que os resultados do projeto se traduzam em ações concretas e mensuráveis, com um impacto positivo para a organização. Independentemente da abordagem selecionada para a aplicação prática do conhecimento adquirido, seja a elaboração de relatórios, a integração em sistemas existentes ou a definição

de recomendações, esta fase conclui o processo, e assegura o retorno do investimento associado ao projeto (*Pyle & San Jose*, 2005).

2. Revisão da Literatura

Este capítulo apresenta a revisão da literatura existente, com o objetivo de contextualizar o tema em estudo e fundamentar teoricamente a investigação desenvolvida.

A metodologia adotada seguiu a norma PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) (Page *et al.*, 2021), tendo sido realizada uma pesquisa sistemática nas bases de dados *Scopus* e *Web of Science* (*WoS*). Utilizaram-se os seguintes termos de pesquisa, combinados com operadores booleanos: "*self-monitoring*" *OR "remote monitoring" AND "machine learning" OR "predictive models" AND "health" OR "clinical report" OR "healthcare" AND "prevent" OR "predict".* Foram definidos como critérios de inclusão os documentos do tipo *article* ou *review*, publicados em publicações científicas (*journals*) entre os anos de 2019 e 2024, exclusivamente em língua inglesa.

A pesquisa resultou inicialmente em 88 artigos na base *Scopus* e 76 na *WoS*, totalizando 138 registos. Após a remoção de duplicados, obtiveram-se 91 artigos únicos. Destes, 18 foram excluídos com base no título por não se enquadrarem na temática da revisão. Restaram assim 73 artigos, dos quais 25 foram eliminados após a leitura do *abstract* por apresentarem um enquadramento temático distinto do foco proposto. Por fim, 17 artigos foram descartados após leitura integral, por se revelarem insuficientemente alinhados com os objetivos da análise, culminando numa amostra final de 31 artigos considerados pertinentes para esta revisão.

2.1. Random forest

A técnica de aprendizagem automática supervisionada, designada por *Random forest* (RF), é amplamente reconhecida pela robustez, versatilidade e elevada capacidade preditiva que apresenta, especialmente no domínio da saúde digital (*Kanza et al.*, 2024; *Maqbool et al.*, 2023; *Mishra e Mohapatro*, 2019). A escolha desta metodologia deve-se à eficácia demonstrada em contextos de monitorização remota de pacientes (*RPM*), nos quais a variabilidade dos dados fisiológicos e a exigência de respostas em tempo real constituem fatores críticos (*Wang et al*., 2024). Para além de lidar eficazmente com dados heterogéneos e variáveis categóricas sem necessidade de pré-processamento complexo, o RF tem demonstrado elevadas taxas de acerto em tarefas de deteção precoce de anomalias e previsão de eventos clínicos, com resultados superiores a 80% em ambientes reais e de simulação (*Kanza et al*., 2024; *Hanumantharaju et al*., 2022).

A metodologia assenta na construção de múltiplas árvores de decisão treinadas com subconjuntos aleatórios de dados, permitindo uma agregação final que reduz o risco de sobreajuste (*overfitting*) e melhora a capacidade de generalização do modelo. Em estudos anteriores, como o de *Mishra e Mohapatro* (2019), esta técnica alcançou taxas de acerto superiores a 90% na previsão de anomalias cardíacas, utilizando dados do repositório *UCI Heart Disease*. Adicionalmente, investigações orientadas para ambientes domiciliários e também em contextos de saúde demonstraram a eficácia desta abordagem na recolha passiva de dados fisiológicos, mantendo elevados níveis de fiabilidade (*Wang et al*., 2024).

2.2. Redes neuronais

As Redes Neuronais Artificiais (RNAs) têm sido amplamente exploradas em investigações recentes no domínio da saúde digital, evidenciando uma capacidade notável para identificar padrões não lineares e complexos em conjuntos de dados biométricos. Inspiradas na arquitetura do cérebro humano, estas estruturas têm demonstrado especial eficácia em contextos onde a tomada de decisão clínica rápida e precisa é essencial, nomeadamente na previsão de eventos críticos e na monitorização remota de pacientes (*Pappada et al*., 2020; *Singh, Hashmi e Sharma,* 2020).

A literatura aponta para a existência de diferentes variantes de RNAs*,* entre as quais se destacam as *Convolutional Neural Networks* (*CNNs*) e as *Long Short-Term Memory* (*LSTM*). Estas têm sido aplicadas com sucesso à análise de sinais fisiológicos recolhidos por dispositivos móveis e sensores vestíveis, especialmente em ambientes onde se pretende garantir continuidade de cuidados ou resposta imediata a situações de emergência. No estudo de *Pappada et al*. (2020), a aplicação de RNA evidenciou uma elevada precisão na previsão dos níveis de glicose em pacientes, contribuindo para a tomada de decisões terapêuticas em tempo real. De forma complementar, *Singh, Hashmi e Sharma* (2020) aplicaram RNAs à vigilância remota de pacientes com paralisia cerebral, recorrendo à análise de eletroencefalograma para deteção precoce de crises epiléticas.

Outras abordagens, como a proposta por *Lavanya et al.* (2024), integraram RNAs em dispositivos de Internet das Coisas Médicas (*I*oMT), viabilizando a monitorização remota em contextos com infraestruturas tecnológicas limitadas. Na avaliação de diversos algoritmos, as RNAs demonstraram um desempenho superior, mesmo em condições de amostragem reduzida. Adicionalmente, as CNNs, conforme demonstrado por *Khanwalkar* (2024) e *Basavaiah et al.* (2024), têm demonstrado um potencial substancial em atividades de previsão em tempo real, tanto em modelos centrados na gestão preventiva de doenças crónicas como em sistemas de alerta para crises epiléticas iminentes, com taxas de acerto superiores a 95%.

Apesar da elevada eficácia, nem todos os estudos apontam para a superioridade inequívoca destas redes. *Zaitin et al*. (2024) demonstraram que, em cenários com séries temporais curtas, modelos híbridos baseados em regressão estatística combinados com classificadores supervisionados (SVM) podem alcançar desempenhos superiores às RNAs. Este dado reforça a importância de uma seleção criteriosa do modelo, tendo em consideração as características específicas do problema em análise.

2.3. Seleção das *features*

A eficácia dos modelos preditivos em saúde digital não é determinada unicamente pela seleção do algoritmo, mas também pela qualidade das *features* utilizadas. O processo de seleção e transformação de variáveis relevantes tem-se revelado um elemento central no desenvolvimento de soluções de monitorização inteligente.

Conforme demonstrado por *Dave et al*. (2020), através de uma análise comparativa entre modelos com e sem integração de variáveis contextuais, a seleção criteriosa de variáveis relevantes revela-se um fator determinante para a melhoria da precisão preditiva. Este estudo mostra que o volume de dados, por si só, não garante a eficácia do modelo.

2.4. Dispositivos de recolha de dados biométricos

O desenvolvimento das tecnologias digitais no âmbito da saúde, impulsionado pelos desafios impostos pela pandemia de COVID-19 *(COronaVIrus Disease – 2019*), tem acelerado a implementação de sistemas de RPM. Neste novo paradigma, a integração de dispositivos vestíveis e sensores inteligentes com plataformas IoMT possibilita a recolha e análise contínua de dados fisiológicos, promovendo cuidados médicos mais preventivos, personalizados e acessíveis.

Um conjunto crescente de estudos tem demonstrado a eficácia desta abordagem em contextos diversos. *Asghari* (2021) propôs um sistema de RPM para populações idosas, com sensores biométricos capazes de recolher dados como a frequência cardíaca e a temperatura corporal. Estes dados são utilizados por algoritmos de ML para antecipar riscos associados ao cancro. De forma semelhante, *Omar et al.* (2023) desenvolveram uma plataforma de IA conversacional baseada em sensores Internet das Coisas (*IoT*), que monitoriza sinais vitais e fornece orientações médicas automatizadas em tempo real, especialmente útil em situações de isolamento ou emergência. A metodologia adotada é centrada no utilizador permitindo a adaptação do sistema às necessidades do paciente. Na mesma perspetiva, *Veyron et al.* (2024) demonstraram que a aplicação de sensores a registos clínicos preenchidos por cuidadores permite antecipar hospitalizações.

Thacharodi et al. (2024), por sua vez, apresentam uma revisão sistemática que confirma os benefícios da convergência entre IA, IoT, dispositivos vestíveis e *big data* na transformação dos cuidados de saúde, com benefícios claros na precisão diagnóstica, personalização de tratamentos e autonomia do paciente.

A saúde mental tem igualmente auferido benefícios provenientes destas tecnologias. *Adler et al.* (2020) desenvolveram uma metodologia inovadora, na qual sensores integrados em *smartphones* são utilizados para analisar comportamentos passivos e prever recaídas psicóticas em indivíduos com perturbações do espectro da esquizofrenia. A sensibilidade do modelo foi moderada (0,25), contudo, a especificidade foi elevada (0,88), demonstrando que os dados contextuais captados pelos sensores podem gerar *insights* clínicos relevantes.

No acompanhamento de condições crónicas, os sensores vestíveis têm revelado uma utilidade significativa. Conforme demonstrado por *Kallam et al.* (2023), foi proposta uma infraestrutura IoT de múltiplas camadas para a monitorização de doenças crónicas, tais como a diabetes e patologias cardiovasculares. Este estudo recorreu a sensores fisiológicos e a *smartphones*, com o objetivo de criar um ambiente inteligente de vida assistida. Já *Bus et al*. (2024) centraram-se na prevenção de úlceras em pacientes com pé diabético, recorrendo a palmilhas inteligentes e a uma bota inteligente que monitorizam a distribuição da pressão plantar e fornecem *feedback* em tempo real, promovendo uma aderência mais efetiva ao tratamento.

A aplicação de sensores é uma área em que se verificou um sucesso considerável, particularmente no domínio da reabilitação física. *Fellger et al.* (2020) desenvolveram um sistema destinado à previsão de padrões de sono e atividade física em pacientes hospitalizados. O sistema recorreu a dados recolhidos através de diversos dispositivos vestíveis, apresentando uma taxa de erro média inferior a 12%. Este resultado reforça o potencial da padronização de dados de origens múltiplas na melhoria da continuidade dos cuidados.

No domínio da monitorização cardíaca, *Chen, Valehi e Razi* (2019) introduziram uma abordagem preditiva em duas fases, fundamentada em registos da atividade elétrica do cérebro — sinais de eletroencefalograma — obtidos por sensores. A primeira etapa atua como um sistema de alarme para anomalias graves, enquanto a segunda realiza uma análise personalizada para detetar alterações subtis nos sinais individuais, atingindo uma accuracy global de 96,6%. Em populações mais vulneráveis, como pessoas com demência, *Fletcher-Lloyd et al.* (2021) demonstraram que sensores ambientais e fisiológicos podem prever episódios de agitação com precisão considerável (*F1-score* de até 0,80), mesmo com dados

simples. Já *Subba Reddy et al.* (2024) desenvolveram um sistema de deteção de quedas em tempo real baseado em sensores com capacidade para medir a temperatura e os batimentos cardíacos, obtendo resultados notáveis em termos de precisão (accuracy de 99,864%).

Os sensores vestíveis e inteligentes, integrados em plataformas de monitorização remota, constituem um pilar fundamental da saúde digital contemporânea (*Thacharodi et al.,* 2024; *Kallam et al*., 2023). Contudo, o seu verdadeiro impacto não reside apenas na sofisticação dos algoritmos que os suportam, mas na qualidade e continuidade da recolha de dados, na sua contextualização clínica e na integração efetiva com os fluxos reais de cuidado (*Veyron et al*., 2024; *Omar et al.,* 2023). A articulação sinérgica entre tecnologia, contexto e prática possibilita à medicina digital evoluir para um modelo mais preventivo, proativo e centrado no paciente, fundamentado em dados, mas com um carácter marcadamente humano.

2.5 Adesão dos pacientes

A transformação digital no setor da saúde tem promovido uma mudança paradigmática na gestão das doenças crónicas, destacando dois eixos temáticos centrais: por um lado, a adesão e o envolvimento sustentado dos pacientes em plataformas digitais de monitorização; por outro, a crescente utilização de dispositivos vestíveis e sensores inteligentes para a recolha contínua e não invasiva de dados clínicos (*Vouzis & Maglogiannis*, 2023; *Guthrie et al.,* 2019). Estes elementos não apenas redefinem a relação entre pacientes e prestadores de cuidados, como também suscitam desafios metodológicos, técnicos e éticos que estão a ser progressivamente abordados na literatura científica (*Davergne et al*., 2020; *Zhang et al.*, 2022; *Kallam et al.,* 2023; *Bus et al.,* 2024).

A questão da adesão digital tem sido objeto de estudo aprofundado no âmbito da monitorização remota. Segundo a investigação de *Viziris e Maglogiannis* (2023), a questão da desistência antecipada em programas digitais de saúde constitui um desafio significativo, que não só compromete a eficácia clínica, mas também a sustentabilidade económica dessas intervenções. Os autores propõem um modelo de *stacked generalization*, uma técnica de aprendizagem automática baseada na combinação de diferentes algoritmos, para prever comportamentos de abandono, atingindo uma accuracy de 97,6%. Esta abordagem destaca o potencial dos sistemas digitais preditivos na aplicação de estratégias motivacionais adaptadas, procurando reforçar a adesão e a continuidade da monitorização destes sistemas.

Numa perspetiva complementar, *Guthrie et al*. (2019) exploram o impacto das terapias digitais na gestão da hipertensão arterial, associando intervenções comportamentais com modelos de previsão baseados em *Random Forest*. O estudo demonstrou não só a relevância da

inteligência artificial como ferramenta de apoio ao diagnóstico, mas também como mecanismo facilitador da continuidade terapêutica e do autocuidado sustentado.

Em contraste, *Zhang et al.* (2022) apresentam uma metodologia centrada na análise contextual e ecológica da adesão ao tratamento, especificamente em adolescentes com diabetes tipo 1, recorrendo à metodologia *Ecological Momentary Assessment* (EMA). Os autores procederam à combinação de dados contextuais com medições glicémicas em tempo real, construindo modelos preditivos com a capacidade de identificar episódios de não adesão com elevada precisão. Este estudo destaca a importância de considerar os fatores emocionais, sociais e situacionais no design de intervenções digitais eficazes, sobretudo em populações vulneráveis e de difícil monitorização.

O segundo eixo relevante diz respeito à monitorização passiva e contínua através de tecnologias vestíveis. *Davergne et al.* (2020) procedem ao exame da utilização de dispositivos de monitorização da atividade física em pacientes com doenças reumáticas, demonstrando a capacidade destas tecnologias para a previsão de surtos inflamatórios com elevada sensibilidade. Contudo, o estudo identifica limitações substanciais, incluindo a heterogeneidade entre dispositivos, a adesão inconsistente dos utilizadores e a escassez de estudos robustos sobre a eficácia destas intervenções a longo prazo.

A evolução deste paradigma é também aprofundada nos estudos de *Kallam et al*. (2023) e *Bus et al.* (2024). Os dois estudos demonstram que as tecnologias vestíveis e a análise automatizada desempenham um papel crucial na promoção de intervenções personalizadas e reativas, com um impacto direto na qualidade de vida dos pacientes.

A saúde digital visa ampliar e qualificar o cuidado humano, promovendo uma abordagem centrada na personalização e na prevenção. Neste sentido, a performance técnica dos algoritmos, embora essencial, revela-se insuficiente se dissociada da realidade prática dos cuidados de saúde e da experiência subjetiva dos pacientes. A integração eficaz destas soluções pressupõe uma abordagem multidisciplinar capaz de articular as dimensões técnica, clínica e humana que sustentam a inovação em saúde (*Thacharodi et al.,* 2024).

3. Metodologia

O presente capítulo detalha a metodologia aplicada para desenvolver o modelo de análise, visualização e previsão da criticidade clínica a partir dos dados da *Hope Care, S.A.*, recorrendo à metodologia CRISP-DM, reconhecida pela sua estrutura sistemática e flexibilidade.

Neste capítulo apresentam-se as etapas que orientaram o desenvolvimento do estudo, abrangendo as fases da investigação, os métodos de recolha e análise de dados, bem como os critérios adotados para assegurar a validade e a fiabilidade. Esta abordagem tem como objetivo garantir a transparência do processo e possibilitar a replicação do estudo.

3.1. Compreensão do Negócio

O ponto de partida da metodologia CRISP-DM, é a compreensão do negócio, etapa crucial para estabelecer uma base sólida em projetos de data mining. O principal objetivo consiste em definir uma relação clara entre os objetivos do projeto de dados e as metas estratégicas da organização , assegurando que a solução desenvolvida gere valor real e mensurável. No presente estudo, a atenção centra-se na *Hope Care, S.A.*, entidade de referência no domínio da saúde digital em Portugal, cuja necessidade principal é otimizar a monitorização de pacientes por meio da previsão de situações de alerta clínico.

A Hope Care, S.A., posiciona-se como uma organização inovadora no setor da saúde digital, com a ambição de transformar o paradigma dos cuidados, promovendo o conceito de "Hospital em Casa". Com o intuito de reduzir internamentos hospitalares e otimizar a utilização dos recursos de saúde, a empresa oeferece a monitorização remota contínua e proativa de pacientes, em epsecial daqueles que vivem com doenças crónicas.

Este projeto visa reforçar a capacidade preditiva da Hope Care,S.A., na identificação de situações de alerta clínico iminente em pacientes monitorizados remotamente.

Entende-se por alerta clínico uma alteração significativa e desfavorável no estado de saúde do paciente, detetada a partir da análise de dados biométricos, que indica uma potencial necessidade de intervenção médica ou de enfermagem. Embora a Hope Care S.A., já realize uma monitorização robusta, o desafio reside em passar de uma abordagem reativa ou baseada em regras fixas para uma abordagem preditiva, capaz de antecipar a deterioração do estado de saúde de um paciente antes que esta se manifeste de forma crítica.

Embora a Hope Care, S.A. já disponha de um sistema de monitorização robusto, o desafio atual consiste em evoluir de uma abordagem reativa, baseada em regras fixas, para uma abordagem preditiva, capaz de antecipar a deterioração do estado de saúde antes que esta se manifeste de forma crítica. A antecipação das intervenções pelas equipas médicas permite o contacto precoce com os pacientes ou cuidadores, evitando a progressão do quadro clínico, reduzindo episódios de emergência e internamentos hospitalares, e promovendo, assim, uma maior eficiência operacional.

Deste modo, as equipas de monitorização passam a concentrar-se nos pacientes com maior risco efetivo, diminuindo o número de falsos positivos. Esta evolução reforça a proposta de valor e o compromisso da Hope Care, S.A. com a qualidade de vida e o bem-estar dos seus pacientes.

Para a *Hope Care, S.A.,* o sucesso da solução preditiva de alertas clínicos assenta em três pilares: (1) funcionalidade e utilidade para as equipas médicas, com interfaces e painéis intuitivos; (2) accuracy e fiabilidade técnica assegurando a deteção de alertas reais e a minimização de falsos positivos; (3) impacto positivo nos indicadores de desempenho (KPIs) traduzido na redução dos tempos de resposta, diminuição de internamentos e otimização da carga de trabalho.

3.2. Compreensão dos Dados

De acordo com a metodologia CRISP-DM*,* a fase de compreensão dos dados pretende entender a origem, a estrutura e as características intrínsecas do conjunto de dados, bem como avaliar a sua qualidade inicial. No presente caso, os dados têm origem numa base de dados fornecida pela Hope Care, S.A. Estes dados foram disponibilizados em formato *Excel*, contendo dois *worksheets*: o primeiro com informações demográficas básicas de cada paciente (identificador único do paciente, género e idade) e o segundo com os registos longitudinais de medições de saúde. No total, o *dataset* abrange 30 pacientes, com 56.963 registos de medições recolhidos entre junho de 2022 e março de 2025, englobando 23 indivíduos do sexo masculino e 7 do sexo feminino, com idades entre 40 e 89 anos. Cada registo temporal está identificado pelo código do paciente (*clientID*) , por carimbos de data/hora (quando o dado foi recebido pelo sistema – *recivedOn* – e quando foi gerado pelo dispositivo – *createdOn*), inclui o tipo de medida realizada (*ProviderMNameStandard*) juntamente com o valor numérico observado (*ValueStandard*) e a respetiva unidade de medida (*UnitsStandard*). Adicionalmente, são fornecidos metadados sobre a origem da medição, como o método de aquisição (*SourceType*) e, a identificação do dispositivo utilizado (*DeviceManufacturer* e *DeviceModel*).

No conjunto de dados, que abrange uma variedade de indicadores de saúde, estão presentes sinais vitais, como a frequência cardíaca (*pulse*), a saturação de oxigénio no sangue (SpO_2), a pressão arterial sistólica e diastólica (*systolic* e *diastolic*), a temperatura corporal e a glicemia capilar (*blood_glucose*). Incluem-se também medidas antropométricas e relacionadas com composição corporal, como o peso corporal (*weight*). Além disso, são registadas métricas de atividade física como o número de passos dados, distância percorrida e energia ativa despendida, e parâmetros de sono obtidos através de dispositivos *wearable* ou aplicações de saúde — dispositivos eletrónicos concebidos para serem usados sobre o corpo ou integrados em peças de roupa, que incorporam sensores capazes de recolher e transmitir dados fisiológicos em tempo real (*Iqbal et al.*, 2021).

A Hope Care, S.*A.,* forneceu um *dataset* adicional que inclui variáveis de criticidade clínica, concebidas para classificar o nível de risco associado a cada medição biométrica. Segundo os critérios definidos pela empresa, cada registo de um sinal vital é automaticamente categorizado num de cinco níveis — *normal, light, important, severe* ou *urgent* — com base em limiares clínicos pré-estabelecidos que correspondem a intervalos de valores considerados seguros ou esperados para cada métrica. Assim, as variáveis de criticidade, integram conhecimento clínico diretamente no *dataset*, sinalizando automaticamente medições potencialmente preocupantes conforme regras médicas estabelecidas e permitindo uma triagem imediata dos casos que exigem maior atenção.

A heterogeneidade na frequência de registo das variáveis constitui uma característica distintiva deste conjunto de dados, decorrente do facto de todas as medições serem efetuadas manualmente com diferentes periodicidades. Algumas variáveis apresentam uma maior frequência de registo, como a contagem diária de passos ou a medição da frequência cardíaca, enquanto outras são registadas de forma menos regular, como a pressão arterial, a glicemia ou o peso corporal, geralmente anotadas apenas uma vez por dia ou por semana. A este conjunto de dados acrescem ainda os parâmetros de sono, que aparecem tipicamente uma única vez por dia, de manhã, como resumo da noite anterior. Consequentemente, os dados apresentam uma granularidade temporal irregular, com períodos em que determinados sinais vitais registam múltiplas medições por hora, intercalados com longos intervalos sem qualquer registo em variáveis medidas apenas de forma ocasional.

3.2.1 Diagnóstico Inicial da Qualidade dos Dados

A qualidade dos dados brutos foi analisada na fase de compreensão com o objetivo de identificar problemas de completude e coerência que pudessem comprometer etapas subsequentes do processo. Tal como esperado, considerando os padrões de recolha

previamente descritos, foi detetada uma elevada percentagem de *missing values* (valores em falta) na maioria das variáveis fisiológicas e comportamentais. O conjunto de dados original, composto por 56.963 registos temporais, revelou, através da contagem de valores não nulos, uma ausência considerável de informação em diversos atributos essenciais.

No que respeita à completude dos dados, os campos de identificação e demografia — *clientID*, género e idade — encontram-se integralmente preenchidos, não se registando missing values. A variável *weight* exibe igualmente um nível de cobertura quase total, com apenas 0,04 % de falhas.

Em contraste, observam-se lacunas substanciais nas variáveis clínicas e comportamentais. As métricas de sono revelam-se particularmente críticas: a duração total do sono (*sleep*) está ausente em 94,8 % dos registos, e a contagem de despertares noturnos (*awake_count*) em 95,1 %, uma vez que, por norma, estas medições resultam de um único registo diário que sintetiza o período noturno. A frequência cardíaca apresenta uma percentagem de valores em falta que oscila entre 60% e 70%, refletindo períodos em que o dispositivo não foi utilizado ou não transmitiu medições. Situações análogas ocorrem na pressão arterial sistólica/diastólica e na saturação de oxigénio. As métricas de atividade — energia ativa gasta, passos (*steps*), deteção de fibrilhação auricular (*afib*) ou registos de humor (*mood*) — evidenciam igualmente uma marcada ausência de dados, ultrapassando 75% de lacunas. Para mitigar o impacto destes constrangimentos evitar enviesamentos nas análises posteriores, torna-se necessário implementar um *pipeline* de pré-processamento baseado em: (i) harmonização temporal — através de reamostragem ou agregação em janelas uniformes — e (ii) imputação dirigida ao tipo de variável, de modo a transformar o *corpus* original num *dataset* denso, consistente e apto a suportar uma modelação preditiva robusta.

3.2.2. Preparação dos Dados

No modelo *CRISP-DM,* a fase de preparação de dados engloba todas as atividades necessárias para construir o conjunto de dados final que servirá de base às técnicas de modelação, a partir dos dados brutos iniciais" (*Chapman et al.,* 2000). De acordo com a bibliografia especializada, esta definição deve ser interpretada como um alerta operacional: as atividades de limpeza, integração e transformação podem representar cerca de 70 a 80% do tempo total dedicado a um projeto de ciência de dados, uperando frequentemente a própria fase de modelação (*Press,* 2016). Consequentemente, a qualidade do produto analítico é determinada pelo princípio "*garbage in, garbage out*", ou seja, modelos treinados com dados incompletos ou ruídosos tendem a produzir inferências imprecisas, independentemente da sua sofisticação algorítmica (*Redman,* 2001).

Neste estudo, a preparação de dados foi implementada em Python *(Pandas 2.2, NumPy 1.26)* e organizada num processo em três etapas: (1) Harmonização temporal — reamostragem/agrupamento em janelas regulares de 30 minutos; (2) Imputação de sinais vitais contínuos — preenchimento *forward/backward* específico por paciente; (3) Tratamento das variáveis de sono e atividade — propagação intradiária e interpolação interdiária com sinalização dos valores imputados.

Este encadeamento de transformações visa converter um conjunto de dados originalmente disperso e irregular num dataset denso, coerente e estatisticamente consistente, apto para a modelação preditiva.

(1) Harmonização temporal - Com o objetivo de tornar comparáveis séries temporais cuja cadência de medição varia de segundos a dias, procedeu-se à uniformização do eixo temporal por meio de reamostragem em janelas fixas de 30 minutos. A cada registo foi atribuída uma nova marca temporal correspondente ao início do bloco de meia hora em que ocorreu. Deste modo, medições registadas às 10:05:23 e às 10:29:59 foram ambas arredondadas para 10:00:00. Posteriormente, o conjunto de dados foi agrupado por identificador de paciente e marca temporal arredondada, retendo-se apenas os valores não nulos em cada grupo. Esta opção preserva medições efetivamente observadas — em vez de médias agregadas — conferindo maior fiabilidade aos valores analisados, e garante que o ponto selecionado corresponde sempre ao momento mais próximo do início do intervalo.

Por fim, a coluna auxiliar de tempo foi redefinida como marca temporal oficial, resultando em, no máximo, um registo por variável e por paciente em cada intervalo de 30 minutos. Testes preliminares, realizados com janelas de 15 minutos, mostraram-se insuficientes, uma vez que se verificou a existência de um número excessivo de blocos vazios; por isso, a janela de meia hora foi considerada a opção mais adequada, pois preserva variações clínicas relevantes, garantindo que a maioria dos intervalos inclua pelo menos uma medição válida. Esta etapa reduziu o número total de linhas de 56 963 para 13 645, produzindo um conjunto de dados substancialmente mais compacto e com base temporal estável e adequado às análises modelação preditiva subsequentes.

(2) Imputação de sinais vitais contínuos - Para colmatar as lacunas remanescentes nos sinais vitais após a reamostragem de 30 e 30 minutos, foi aplicada em Python uma imputação bidirecional por paciente. A sequência temporal de cada individuo foi ordenada e percorrida em duas passagens: (i) o método *forward fill (ffill*), que copia para a frente o último valor válido; (ii) o método *backward fill (bfill)*, que percorre a série na direção inversa, preenchendo eventuais lacunas iniciais. Por exemplo, às 08:30 não existia

medição, foi atribuído o primeiro valor posterior disponível, 70 bpm. A imputação foi executada individualmente por *clientID*, garantindo que os dados de diferentes indivíduos não se misturam e, assumindo uma evolução gradual dos sinais vitais. Após a aplicação do procedimento, todos os blocos de 30 minutos passaram a conter medições de temperatura corporal, pressão arterial sistólica e diastólica, frequência cardíaca e saturação de oxigénio no sangue. A variável peso corporal atingiu 99,97% de completude. Estas séries contínuas e sem *missing values* constituem agora uma base robusta para a extração de *features* e para a subsequente modelação preditiva.

(3) Tratamento das variáveis de sono - As métricas de sono distinguem-se dos sinais vitais contínuos por apresentarem, na melhor das hipóteses, periodicidade diária. A replicação de um valor noturno em janelas de 30 minutos fora do seu contexto fisiológico poderia gerar artefactos analíticos. Assim, foi definido um protocolo de imputação em dois planos, ajustado à escala temporal do fenómeno:

a) Propagação intradiária: a partir de cada *timestamp*, foi extraída a componente de data (AAAA-MM-DD). Sempre que registada uma observação válida de sono, para determinado dia e paciente, o respetivo valor foi difundido ao longo das 48 janelas de 30 minutos desse mesmo dia. Por exemplo, um registo de 7 horas de sono obtido às 08:00 foi aplicado a todos os intervalos entre 00:00 e 23:30 do mesmo dia, suprimindo lacunas intradiárias (NaN) sem propagação indevida para dias adjacentes.

b) Interpolação interdiária — Falhas no registo noturno resultaram em lacunas de um ou mais dias completos. Para cada hiato, foi imputado o valor $\frac{X+Y}{2}$, resultante da média aritmética dos dias imediatamente anterior (X) e seguinte (Y) do mesmo paciente, assumindo-se a continuidade do padrão de sono. Cada imputação foi marcada por uma *flag* binária, assegurando a rastreabilidade dos dados estimados. A aplicação deste processo resultou num aumento expressivo da cobertura das variáveis de sono:

- Duração total do sono *(sleep): 5,2 % → 29,6 % de preenchimento;*
- Número de despertares *(awake_count):* 4,9 % → 28,2 % de preenchimento.

Em termos absolutos, cerca de um terço dos registos diários anteriormente vazios passaram a conter valores plausíveis, originando séries mais densas e acompanhadas de indicadores de incerteza que podem ser considerados, ponderados ou descartados em modelação subsequente.

3.3 Análise estatística descritiva

Após o tratamento e a preparação dos dados, procede-se à análise estatística descritiva, etapa destinada a caracterizar o conjunto de dados final e a compreender a sua estrutura e consistência interna. Esta fase exploratória tem como propósito avaliar a distribuição e a variabilidade das variáveis fisiológicas e comportamentais recolhidas, permitindo identificar tendências gerais, padrões de comportamento e potenciais anomalias que possam influenciar as etapas subsequentes de modelação preditiva.

A análise incide sobre o perfil global dos utentes e dos sinais vitais monitorizados, com o objetivo de estabelecer uma base interpretativa sólida que sustente a construção de modelos de previsão de criticidade clínica.

Uma componente desta etapa consiste na análise das variáveis associadas à criticidade clínica — *pulse_criticity, spo2_criticity* e *systolic_criticity* —, de modo a compreender os fatores que influenciam o surgimento de estados críticos e a validar a coerência dos dados utilizados nos modelos preditivos.

O conjunto de dados compreende mais de 13.000 registos e inclui 32 *features* relacionadas com parâmetros fisiológicos e métricas de atividade, tais como a frequência cardíaca (*pulse*), a qualidade e duração do sono (*deep sleep* e *sleep*), a saturação de oxigénio no sangue (SpO_2) e o índice de massa corporal (*IMC*). Foram definidas três *features target* — *pulse_criticity*, *spo2_criticity* e *systolic_criticity* — para as quais se calcularam estatísticas descritivas que indicam se uma determinada leitura se enquadra num intervalo crítico de saúde, nomeadamente média, mediana, desvio padrão, e valores mínimo e máximo. Estas variáveis foram analisadas em relação a outras características registadas, de modo a identificar padrões associados a um risco fisiológico elevado.

A variável *pulse*, que corresponde à frequência cardíaca, apresenta uma média de 68 bpm e uma mediana próxima, com valores a variar entre 33 e 140 bpm, sugerindo uma distribuição aproximadamente normal. Observa-se ainda que os valores mais elevados de frequência cardíaca tendem a coincidir com níveis superiores da variável *pulse_criticity*, o que pode indicar a presença de um problema de saúde associado a um aumento anómalo da frequência cardíaca.

A variável SpO_2, que corresponde à saturação de oxigénio, apresenta uma média de 95,9% e uma dispersão reduzida, o que reflete uma estabilidade geral deste parâmetro vital entre os utilizadores. No entanto, os casos em que os níveis de SpO_2 descem até aos 31,7%

indicam situações de stress fisiológico grave ou a presença de condições respiratórias críticas, em consonância com classificações elevadas na variável *spo2_criticity*.

No caso da variável *systolic*, que corresponde à pressão arterial sistólica, observou-se um intervalo de valores entre 65 mmHg e 209 mmHg, com uma média de 112 mmHg. Embora este valor médio se situe dentro dos limites clínicos considerados normais, a amplitude do intervalo poderá refletir uma ligeira tendência para hipertensão em determinados utentes.

As estatísticas descritivas obtidas permitiram a caracterização do perfil geral dos utentes monitorizados e a identificação de *outliers,* em articulação com visualizações gráficas (*boxplots* e histogramas) que proporcionaram uma compreensão clara da dispersão e distribuição dos dados, a identificação de assimetrias e de eventuais padrões extremos. Através dessas visualizações, verificou-se que, apesar da existência de alguns valores nos limites superiores das distribuições, estes se mantêm compatíveis com as variações fisiológicas expectáveis no contexto clínico e na população em estudo. Consequentemente, não se identificou a necessidade de eliminação ou correção de *outliers*, uma vez que os dados se mantiveram dentro dos intervalos previamente definidos pela Hope Care, S.A., não comprometendo a qualidade da análise.

Para caracterizar o perfil demográfico da população estudada, procedeu-se à análise da distribuição etária dos participantes, cujos resultados se encontram representados no Anexo (Figura 1). Observa-se uma predominância clara do sexo masculino, que corresponde a 71,7% da amostra total, enquanto o sexo feminino representa 28,3%.

No grupo masculino, a distribuição etária revela maior heterogeneidade, com múltiplos picos de frequência, predominantemente entre os 70 e 85 anos, conforme ilustrado pela curva de densidade na Figura 1. Já no grupo feminino, a concentração etária é mais acentuada entre os 65 e 70 anos, com um ligeiro pico aos 67 anos. Esta distribuição apresenta uma simetria relativa, ainda que com tendência ligeira para a esquerda, o que sugere uma maior concentração de idades mais jovens.

Para avaliar o envolvimento dos utilizadores e a fiabilidade dos dados, analisou-se a taxa de adesão ao programa de monitorização remota, apresentada no Anexo (Figura 2). A taxa de adesão global é de aproximadamente 33,3%, valor que representa a proporção de participantes que efetuaram as medições em conformidade com o protocolo clínico estabelecido. Embora em número inferior relativamente aos participantes do sexo masculino, as mulheres demonstram um maior nível de envolvimento com o processo de monitorização, refletido numa taxa de adesão de 57,3%, em contraste com o grupo masculino, cuja adesão é de 42,7%. A análise da taxa de adesão revela que esta se mantém relativamente baixa e

estável na maioria dos grupos etários, embora se verifiquem picos distintos em idades específicas — nomeadamente aos 59, 63, 66, 70 e 78 anos, sendo o mais proeminente aos 59 anos. Apesar da variabilidade geral, as taxas de adesão concentram-se maioritariamente entre os valores 1 e 2, registando-se, pontualmente, valores atípicos que ultrapassam este intervalo. Esta distribuição, marcada por flutuações irregulares, sugere que a idade pode influenciar o comportamento de adesão, possivelmente em resultado de fatores como maior sensibilização para a saúde ou interesse acrescido na monitorização autónoma de parâmetros clínicos (Anexo, Figura 3).

A análise da distribuição das medições ao longo dos diferentes dias do mês, revela uma frequência relativamente uniforme, sem indícios de irregularidades sistemáticas. Verifica-se uma variação entre 410 e 480 medições diárias, o que indica uma consistência significativa no comportamento de registo. As pequenas flutuações detetadas poderão ser atribuídas à variabilidade natural na atividade dos participantes, e não a qualquer tipo de enviesamento sistémico. Esta estabilidade reforça a fiabilidade da componente temporal do conjunto de dados (Anexo, Figura 4).

A distribuição do número de medições por dia da semana, revela-se equilibrada. No entanto, exite uma ligeira redução na frequência de registos durante os fins de semana, em particular aos sábados, que poderá refletir padrões comportamentais ou contextuais, nomeadamente alterações na rotina diária dos utilizadores que afetam a regularidade das medições (Anexo , Figura 5).

A frequência de medições entre participantes do sexo masculino e feminino revela padrões distintos. Enquanto o grupo masculino apresenta uma distribuição dispersa, com vários picos localizados em torno das 250, 500, 1000 e 1250 medições - o que evidencia uma maior heterogeneidade na forma como os homens realizam os registos - as participantes do grupo feminino revelam um padrão mais polarizado, com um número reduzido de mulheres nas faixas de medição mais baixas e uma concentração acentuada no extremo superior da distribuição, com cerca de 2000 medições, indicando um maior grau de envolvimento por parte deste grupo (Anexo, Figura 6).

A distribuição proporcional dos alertas de criticidade teve por base a categorização visual dos alertas, sendo atribuída a cor laranja às categorias *“Important”* e “*Severe”*, e a cor vermelha à categoria “*Urgent”*. A análise revela que os alertas laranja representam a maioria dos casos (69,1%), enquanto os alertas vermelhos correspondem aos restantes 30,9%. Esta distribuição sugere que, embora uma proporção relevante dos alertas requeira atenção imediata, a maior parte das ocorrências reflete situações de elevada gravidade, ainda sem

atingir um nível crítico de urgência. Este padrão evidencia a eficácia do modelo na deteção precoce de condições potencialmente críticas, favorecendo intervenções atempadas e promovendo uma abordagem preventiva no acompanhamento dos utilizadores (Anexo, Figura 7).

A distribuição dos alertas de criticidade ao longo do mês mantém-se relativamente estável, sem que se observem desvios extremos em dias específicos. A maioria dos eventos sinalizados está enquadrada nas categorias "*Important*" e "*Severe*", em contraste com a menor proporção de alertas classificados como "*Urgent*". Esta estabilidade temporal nas proporções de alerta reforça a fiabilidade dos mecanismos de monitorização e não evidencia padrões sistemáticos de concentração de eventos de alto risco em determinados períodos. (Anexo, Figura 8).

A distribuição dos alertas de criticidade por parâmetro de saúde, revela uma assimetria acentuada. Existe uma concentração significativa de alertas vermelhos associados ao parâmetro *steps_criticity*, que, isoladamente, representa mais de metade de todos os alertas urgentes e, por esse motivo, identificada como a variável com maior relevância em termos de criticidade no sistema.

Em contraste, os restantes parâmetros fisiológicos — como a pressão arterial sistólica – *systolic* - e a pressão arterial diastólica – *diastolic* -, a saturação de oxigénio - SpO_2-, a frequência cardíaca – pulse -, a temperatura corporal – *body_temperature* – e, o número de passos - geram um número substancialmente inferior de alertas. Entre estes, destaca-se a *systolic,* que apresenta a maior proporção de alertas laranja. As variáveis *pulse* e SpO_2 , embora associadas a alguns alertas laranja, registam volumes muito reduzidos (Anexo, Figura 9).

O parâmetro *systolic* regista uma incidência expressiva de alertas nos níveis *"Important*" e "*Severe*". Já os parâmetros SpO_2, *pulse* e *body_temperature* apresentam um número mais moderado de alertas em níveis críticos, refletindo uma menor frequência de desvios extremos.

A variável *diastolic*, por sua vez, evidencia uma distribuição relativamente equilibrada entre os níveis *"Light", "Important"* e "*Severe*", com uma incidência residual de alertas classificados como "*Urgent*" (Anexo, Figura 10).

A distribuição dos valores de *systolic* em diferentes faixas etárias evidencia uma maior representatividade de dados provenientes de participantes do sexo masculino. As medições

realizadas por homens distribuem-se de forma consistente ao longo de um intervalo etário alargado, entre os 40 e os 90 anos. Em contraste, os dados relativos às participantes do sexo feminino são menos numerosos e abrangem um conjunto mais restrito de faixas etárias. Não se observa uma relação linear evidente entre a idade e o aumento dos valores de *systolic* na população estudada (Anexo, Figura 11).

A distribuição dos valores de *diastolic* em função da idade apresenta-se dentro de um intervalo relativamente constante, não se verificando uma tendência clara de variação dos seus valores em função da idade. No grupo masculino não se observam grandes oscilações deste parâmetro já no grupo feminino, registam-se valores mais elevados por volta dos 50 anos, aproximando-se dos 78 mmHg (Anexo, Figura 12).

3.3.1. Análise de correlações entre variáveis

Após a exploração inicial do conjunto de dados, procede-se à análise das correlações entre variáveis, com o objetivo de identificar associações estatisticamente relevantes e compreender as interações entre diferentes parâmetros de monitorização. A partir da matriz de correlação (Anexo, Figura 13), foram selecionadas para análise detalhada as variáveis que apresentaram os valores mais elevados de correlação, por evidenciarem relações potencialmente significativas entre si.

A relação entre *steps* e *sleep_duration* revela uma concentração de dados na faixa inferior da variável relativa à duração do sono. A linha de tendência apresenta-se praticamente horizontal, indicando uma correlação muito fraca ou inexistente entre o número de passos diários e o tempo total de sono (Anexo, Figura 14).

A associação entre *weight* e *pulse* mostra uma concentração dos dados entre 50 e 100 kg e 50 a 100bpm, respetivamente. A linha de tendência apresenta uma ligeira inclinação ascendente, sugerindo uma correlação positiva fraca entre as duas variáveis. Este padrão indica que indivíduos com maior peso corporal tendem a registar frequências cardíacas ligeiramente superiores, o que é coerente com o comportamento fisiológico esperado, dada a maior exigência cardiovascular associada a maior massa corporal (Anexo, Figura 15).

A relação entre *sleep_duration* e *energy* evidencia uma ligeira correlação positiva, com os dados concentrados principalmente entre 20.000 e 60.000 *kcal.* Observam-se ainda valores anómalos que refletem dispêndios energéticos excecionalmente elevados. Este padrão sugere que indivíduos com maior gasto energético tendem a apresentar durações de sono ligeiramente superiores, uma resposta fisiológica plausível à exigência física aumentada (Anexo, Figura 16).

Entre as variáveis *total_sleep* e *wake_time*, verifica-se uma correlação positiva de intensidade moderada, revelando que períodos mais longos de vigília tendem a estar associados a durações maiores de sono total, possivelmente como mecanismo de compensação fisiológica destinado a equilibrar o esforço e a recuperação (Anexo, Figura 17).

Na relação entre SpO_2 e *pulse_criticity*, as leituras de saturação de oxigénio concentram-se entre 90% e 100%, enquanto os valores de criticidade do pulso se distribuem de forma mais irregular. A linha de tendência apresenta uma inclinação ligeiramente descendente, exibindo uma correlação negativa fraca, na qual níveis mais baixos de SpO_2 tendem a associar-se a valores ligeiramente mais elevados de criticidade da frequência cardíaca (Anexo, Figura 18).

A relação entre *active_steps* e *sleep_duration* mostra uma dispersão concentrada nos intervalos inferiores de ambas as variáveis, com uma linha de tendência praticamente plana. Este comportamento indica ausência de uma relação linear significativa, sugerindo que o número de passos ativos não exerce influência substancial sobre a duração do sono, nem esta parece impactar diretamente o nível de atividade física (Anexo, Figura 19).

Por fim, a associação entre *weight* e *atrial_fibrillation*, representada por uma variável categórica que indica a presença ou ausência da condição, apresenta uma tendência descendente, sugerindo uma correlação negativa entre o peso corporal e a ocorrência de fibrilação auricular. Este padrão indica que indivíduos com menor peso corporal tendem a apresentar uma prevalência ligeiramente superior de fibrilação auricular dentro da amostra analisada (Anexo, Figura 20).

Os resultados obtidos não justificam a criação de novas *features* combinadas, dado que as correlações identificadas entre variáveis não se traduzem em ganhos relevantes de desempenho preditivo, nem em reforço significativo das associações com as *features-target*.

3.3.2. Correlação entre as variáveis criticity

Na sequência da análise das relações entre as variáveis originais, procede-se à avaliação da criticidade das variáveis pulse, SpO_2 e *systolic*, em relação às restantes variáveis do conjunto de dados. Este passo tem como objetivo estudar a forma como a criticidade das variáveis se associam a alterações noutras dimensões fisiológicas ou comportamentais, permitindo detetar possíveis interdependências em contextos de maior risco clínico e reforçar a compreensão das dinâmicas subjacentes aos episódios de criticidade.

As variáveis *"_criticity"* assumem valores discretos entre 0 e 4, sendo que os valores mais elevados indicam maior gravidade clínica.

3.3.2.1. A variável pulse_criticity

A variável *pulse_criticity*, assume uma distribuição extremamente assimétrica, com um predomínio claro do valor 0, que representa a quase totalidade dos registos — mais de 11.000 casos. Este valor sugere que, na maioria da amostra, a frequência cardíaca se encontra dentro dos limites normais ou não críticos. Os restantes níveis de criticidade ocorrem com frequência muito reduzida, sendo praticamente residuais quando comparados com o valor 0. A presença de um número reduzido de casos com criticidade elevada, associada ao predomínio dos valores normais, reforça a existência de um forte desbalanceamento nesta variável. Este aspeto deverá ser tido em consideração na fase de modelação, particularmente na aplicação de estratégias de balanceamento de classes e na avaliação de métricas de desempenho, para compensar a sub-representação das categorias minoritárias e garantir a imparcialidade e robustez dos resultados obtidos (Anexo, Figura 21).

Para examinar a relação desta variável - *pulse_criticity* - com diversas variáveis fisiológicas e demográficas, os dados foram desagregados por género.

Não há evidencia de uma relação forte entre as variáveis *pulse_criticity* e *age* (Anexo, Figura 22).

Na análise conjunta das variáveis *pulse_criticity* e body_temperature, observa-se uma ligeira tendência negativa, mais pronunciada no grupo feminino. Este padrão sugere que temperaturas corporais mais elevadas estão associadas a níveis ligeiramente inferiores de criticidade do pulso, indicando um possível efeito compensatório fisiológico (Anexo, Figura 23).

A relação entre a *pulse_criticity* e a *diastolic* revela uma correlação negativa mais acentuada nos participantes do sexo masculino enquanto nas mulheres a relação é ligeiramente positiva, refletindo diferenças fisiológicas ou comportamentais entre os grupos. Este comportamento indica que valores mais elevados de pressão diastólica tendem a coincidir com níveis inferiores de criticidade do pulso, sugerindo uma possível regulação hemodinâmica protetora (Anexo, Figura 24).

Evidencia-se uma tendência positiva entre a SpO_2 e a *pulse_criticity* – no grupo masculino, indicando que níveis mais elevados de oxigenação estão associados a uma maior criticidade

do pulso. No grupo feminino, a linha de tendência mantém-se praticamente horizontal, não evidenciando uma relação significativa entre as duas variáveis (Anexo, Figura 25).

A *systolic* evidencia uma elevada dispersão dos dados e uma relação fraca com a *pulse.* Apenas no grupo feminino se observa uma ligeira inclinação positiva na linha de tendência, sugerindo uma correlação marginal entre estas variáveis (Anexo, Figura 26).

Quando relacionadas, as variáveis *pulse_criticity* com *systolic_criticity* e diastolic*_criticity* revelam padrões distintos. No primeiro caso, observa-se uma ligeira correlação negativa para ambos os géneros, indicando que valores mais elevados de criticidade sistólica tendem a ocorrer quando a criticidade do pulso é menor (Anexo, Figura 27). No segundo, verifica-se uma correlação positiva no grupo masculino, sugerindo uma possível interação entre estas duas dimensões de criticidade fisiológica (Anexo, Figura 28).

Embora algumas variáveis — como idade, peso e pressão arterial sistólica — apresentem correlações fracas com a criticidade do pulso, outras, como a pressão arterial diastólica, a saturação de oxigénio e a criticidade diastólica, evidenciam padrões mais consistentes e diferenciados entre géneros, assumindo maior relevância na análise exploratória.

3.3.2.2. A variável SpO_2_criticity.

A distribuição da variável *spo2_criticity*, evidencia um desequilíbrio significativo entre as classes. A maioria dos registos encontra-se concentrada no nível 0, correspondente à ausência de criticidade, enquanto os níveis mais elevados - nomeadamente os valores 3 e 4 - apresentam uma frequência substancialmente inferior (Anexo, Figura 29).

Ao analisar os valores da varíavel *SpO_2_criticity* em função da idade dos utentes, observa-se uma tendência para o aumento da criticidade em faixas etárias mais avançadas, sendo esta mais acentuada no sexo masculino. Este padrão poderá indicar um potencial impacto da idade nos níveis de saturação de oxigénio, embora se mantenha uma elevada variabilidade individual entre os casos (Anexo, Figura 30).

No que diz respeito à temperatura corporal, constata-se que os dados se concentram maioritariamente no intervalo entre 35ºC e 40ºC. No grupo masculino, observa-se uma ligeira correlação negativa entre as variáveis *body_temperature* e a *SpO_2_criticity,* o que pode indicar que temperaturas corporais mais elevadas coincidem com uma melhor oxigenação. Contudo, no grupo feminino, não se deteta um padrão tão evidente, o que pode refletir uma maior variabilidade individual ou a influência de outros fatores fisiológicos (Anexo, Figura 31).

A relação entre a *SpO_2_criticity* e a *diastolic* não evidencia uma correlação forte à primeira vista. No entanto, a linha de tendência ascendente observada no grupo masculino sugere uma possível associação entre valores mais elevados de pressão diastólica e um aumento da *SpO_2_criticity*. Por outro lado, no grupo feminino, os dados apresentam-se mais neutros, não revelando um padrão claro que permita estabelecer uma relação consistente entre estas variáveis (Anexo, Figura 32).

A análise conjunta da *SpO_2_criticity*, com a *pulse*, revela uma tendência ligeiramente positiva, embora fraca, em ambos os géneros. Este padrão sugere que valores mais elevados de frequência cardíaca poderão estar associados a níveis superiores de criticidade da *SpO_2*, ainda que a correlação observada não seja suficientemente robusta para permitir conclusões definitivas (Anexo, Figura 33).

Entre as variáveis *SpO_2* e *SpO_2_criticity*, observa-se uma correlação inversa - à medida que os valores de oxigenação diminuem, os níveis de criticidade aumentam. Esta associação, que se alinha com as expectativas clínicas, é confirmada pelas linhas de tendência descendentes em ambos os géneros e reforça a validade da variável *spo2_criticity* como um reflexo clínico da condição respiratória (Anexo, Figura 34).

Observa-se uma possível correlação positiva entre as variáveis *SpO_2_criticity* e *systolic*, especialmente no grupo masculino, sugerindo que valores mais elevados de pressão sistólica podem associar-se a níveis superiores de criticidade da oxigenação. Este padrão poderá indicar uma relação entre hipertensão sistólica e maior comprometimento da saturação de oxigénio, embora a variabilidade observada recomende uma interpretação prudente (Anexo, Figura 35).

Não se identifica uma relação estatisticamente relevante com entre *SpO_2_criticity* e *weight.* A ausência de inclinação nas linhas de tendência sugere que esta variável, considerada de forma isolada, não exerce um impacto direto sobre a saturação de oxigénio. Estes resultados indicam que o peso corporal, por si só, poderá não constituir um preditor significativo da condição respiratória avaliada através da *SpO_2_criticity* (Anexo, Figura 36).

As relações entre *SpO_2_criticity* e as outras variáveis de criticidade - *systolic_criticity, e diastolic_criticity* - evidenciam padrões distintos que poderão refletir associações fisiológicas de relevância clínica.

No caso de *systolic_criticity*, observa-se uma relação positiva entre níveis mais elevados desta variável e aumento da *SpO_2_criticity*, particularmente entre os indivíduos do sexo

masculino, como indicado pela inclinação ascendente da tendência geral dos dados. Este comportamento poderá refletir situações clínicas em que tanto a pressão arterial como a oxigenação periférica se encontram comprometidas, sugerindo uma possível coocorrência de disfunções cardiovasculares e respiratórias (Anexo, Figura 37).

Relativamente a *diastolic_criticity*, verifica-se uma tendência semelhante, na qual, especialmente entre os homens, os níveis desta variável aumentam paralelamente aos de *SpO_2_criticity,* ainda que com alguma dispersão nos dados. Esta correlação poderá estar relacionada com a interdependência entre a função vascular e a capacidade de oxigenação, sendo particularmente relevante em faixas etárias mais avançadas, em que ambas as funções tendem a deteriorar-se de forma paralela (Anexo, Figura 38).

3.3.2.3. A variável Systolic_criticity

A distribuição da variável *systolic_criticity*, representa diferentes níveis de criticidade associados à pressão arterial sistólica, assumindo valores entre 0 a 4. Observa-se que a maioria dos registos se concentra no valor 0, com mais de 8.000 ocorrências, o que sugere que grande parte da população analisada apresenta valores de pressão sistólica dentro dos parâmetros considerados não críticos. Os níveis 1 e 2, que correspondem a níveis ligeiros a moderados de criticidade, registam uma frequência significativamente inferior, ambos com menos de 1.000 casos. O valor 3 destaca-se com uma quantidade expressiva de registos, aproximadamente 4.000, indicando uma proporção relevante de indivíduos com níveis elevados de pressão sistólica. Por sua vez, o valor 4, que se admite representar o grau mais severo de criticidade, é o menos frequente, com menos de 500 registos. Em termos gerais, verifica-se uma distribuição assimétrica e desbalanceada, fortemente dominada pelos casos não críticos, com uma menor, mas ainda significativa, representação de casos com maior criticidade (Anexo, Figura 39).

A relação entre systolic_criticity e body_temperature evidencia uma tendência geral negativa, indicando que, à medida que a temperatura corporal aumenta, os níveis de criticidade sistólica tendem a diminuir. Este comportamento sugere que indivíduos com temperaturas corporais mais baixas apresentam maior probabilidade de registar níveis sistólicos mais críticos, refletindo uma possível associação entre hipotermia relativa e agravamento da função cardiovascular. Esta relação é mais pronunciada no grupo masculino, enquanto no grupo feminino a correlação negativa também se verifica, embora de forma mais subtil (Anexo, Figura 40).

A variável *systolic_criticity*, em função da *diastolic_blood_pressure*, revela uma relação fraca, com uma tendência ligeiramente negativa para ambos os géneros. A dispersão considerável dos pontos sugere a ausência de uma associação clara entre a pressão diastólica e o grau de criticidade sistólica. No contexto deste conjunto de dados, estas duas variáveis não apresentam uma correlação significativa, podendo atuar de forma relativamente independente no perfil clínico dos utentes analisados (Anexo, Figura 41).

A relação entre *systolic_criticity* e *heart_rate* revela padrões distintos entre géneros. Nos indivíduos do sexo masculino, observa-se uma correlação ligeiramente positiva, sugerindo que valores mais elevados de frequência cardíaca estão associados a níveis superiores de criticidade sistólica. Em contraste, no grupo feminino, a relação entre estas variáveis é praticamente nula, não se evidenciando qualquer padrão consistente. A elevada dispersão dos dados denota uma variabilidade individual significativa, o que poderá indicar diferenças fisiológicas ou comportamentais que atenuam a força da associação global (Anexo, Figura 42).

A variável SpO_2, que mede a saturação de oxigénio no sangue, apresenta uma correlação negativa evidente com *systolic_criticity*, sendo esta relação mais pronunciada no sexo masculino. De forma geral, observa-se que, à medida que os valores de oximetria diminuem, o risco de ocorrência de níveis mais elevados de criticidade sistólica tende a aumentar. Este padrão está em consonância com a evidência clínica, segundo a qual uma oxigenação insuficiente pode estar associada a disfunções cardiovasculares, nomeadamente ao agravamento do estado da pressão arterial sistólica (Anexo, Figura 43).

A relação entre *systolic_criticity* e *systolic_blood_pressure* demonstra uma correlação positiva entre as duas variáveis, mais evidente nos indivíduos do sexo masculino. Observa-se que valores mais elevados de pressão arterial sistólica estão associados a níveis superiores de criticidade, o que poderá refletir um agravamento do estado hemodinâmico em situações de pressão arterial persistentemente elevada (Anexo, Figura 44).

A relação entre *systolic_criticity* e *weight* revela uma tendência praticamente nula em ambos os géneros, com linhas de tendência quase horizontais. A dispersão acentuada dos dados, particularmente em determinados intervalos de peso, evidencia uma variabilidade individual significativa. Estes resultados sugerem que o peso corporal, por si só, não apresenta uma associação clara com os níveis de criticidade da pressão arterial sistólica no contexto da amostra analisada (Anexo, Figura 45).

A variável *diastolic_criticity*, em relação à *systolic_criticity*, apresenta uma ligeira tendência crescente no sexo masculino, indicando que níveis mais elevados de criticidade diastólica estão associados a valores superiores de criticidade sistólica. No grupo feminino, contudo, a relação é substancialmente mais fraca ou mesmo inexistente, evidenciando diferenças nos padrões de associação entre géneros (Anexo, Figura 46).

A variável *systolic_criticity* apresenta uma correlação positiva moderada com SpO_2*_criticity* em ambos os géneros, sendo esta ligeiramente mais acentuada no sexo masculino. Esta relação sugere que níveis mais elevados de criticidade na saturação de oxigénio tendem a estar associados a uma maior criticidade sistólica, reforçando a interligação entre estas duas dimensões fisiológicas. Tal associação poderá indicar a existência de um perfil clínico comum de compromisso simultâneo das funções cardiovascular e respiratória (Anexo, Figura 47).

Verifica-se uma correlação ligeiramente positiva *systolic_criticity* e *body_temperature_criticity* no grupo feminino, indicando que, à medida que a criticidade da temperatura corporal aumenta, também tende a aumentar a criticidade sistólica. No grupo masculino, a tendência é praticamente nula, com uma linha de tendência essencialmente plana, o que sugere a ausência de uma associação clara entre estas variáveis neste conjunto de participantes (Anexo, Figura 48).

A variável *pulse_criticity* apresenta uma correlação positiva mais pronunciada com *systolic_criticity* no grupo masculino, sugerindo que níveis superiores de criticidade do pulso estão associados a valores mais elevados de criticidade sistólica. No grupo feminino, a tendência é menos acentuada. Esta relação reforça a hipótese de que determinadas variáveis fisiológicas exercem influência direta sobre a criticidade da pressão arterial sistólica, podendo ser relevantes para a caracterização global do estado clínico dos utentes (Anexo, Figura 49).

3.4. Modelação dos dados

A fase de modelação preditiva corresponde ao momento em que o conhecimento obtido nas etapas anteriores é transformado em modelos analíticos capazes de prever a ocorrência de situações clínicas críticas, com base nos padrões identificados e nos dados previamente preparados. Nesta etapa, aplicam-se técnicas de ML orientadas para a construção de modelos capazes de detetar relações complexas entre variáveis fisiológicas e indicadores de criticidade clínica. O processo envolve o estudo e seleção dos algoritmos mais adequados, a definição de estratégias de validação e a otimização dos respetivos parâmetros, de modo a garantir equilíbrio entre desempenho preditivo e capacidade de generalização. Durante esta etapa, é igualmente essencial prevenir fenómenos de *overfitting* e *underfitting*, assegurando

que os modelos mantêm a capacidade de generalizar os resultados para novos conjuntos de dados.

A avaliação dos modelos é realizada com recurso a métricas de desempenho adequadas, e selecionadas de acordo com as características do problema, o tipo de variável-*target* e eventuais desequilíbrios nas classes. Para garantir a qualidade e consistência do *dataset* utilizado, processo de preparação iniciou-se com a eliminação das colunas que apresentavam uma proporção igual ou superior a 20% de valores nulos. Concluída esta filtragem inicial, foi realizada uma limpeza manual dos dados, uma vez que a elevada proporção de valores ausentes comprometia a sua utilidade no processo de modelação. Adicionalmente, foram removidas *features* consideradas irrelevantes ou redundantes, por não contribuírem diretamente para os objetivos do projeto nem para o desempenho dos modelos a desenvolver.

No caso do *dataframe df_pulsec*, foram removidas as seguintes colunas: *active_energy_burned, active_energy_burned_steps, activity_type, afib, awake_count, awake_duration, bmi, deep_sleep, distance, distance_steps, elevation_steps, energy_burned_steps, height, light_sleep, mood, sleep,* e *time_to_fall_asleep.* Estas colunas foram eliminadas com o comando *df_pulsec.drop (columns=remover, inplace=True).* No *dataframe df_spo₂c*, a limpeza seguiu os mesmos critérios. As colunas removidas foram: *active_energy_burned, active_energy_burned_steps, activity_type, afib, awake_count, awake_duration, bmi, deep_sleep, distance, distance_steps, elevation_steps, energy_burned_steps, height, light_sleep, mood, sleep,* e *time_to_fall_asleep*, utilizando também o método *drop().* No *dataframe df_systc*, foram retiradas as mesmas *features* que as do *dataframe* anterior.

Concluída esta filtragem, foram tratadas as lacunas existentes nas *features* preservadas. Tendo em conta a elevada variabilidade individual dos dados clínicos, optou-se por não utilizar métodos estatísticos convencionais, como a imputação pela média ou mediana, pois estes poderiam comprometer a integridade das observações.

Em substituição, foi seguida uma estratégia de imputação baseada na dimensão temporal, tirando partido da sequência cronológica dos registos de cada paciente. Esta abordagem mostrou-se particularmente adequada à variável *weight*, que está presente em todos os *dataframes*. Para esta variável, os valores em falta foram preenchidos utilizando, de forma sequencial, os métodos *backfill()* e *ffill(),* assegurando que a imputação fosse realizada com base em medições anteriores ou posteriores do mesmo indivíduo. Esta técnica respeita a lógica temporal inerente aos dados clínicos e minimiza o risco de enviesamento introduzido por imputações genéricas.

Como resultado, foi possível estruturar conjuntos de dados consistentes, compostos apenas por colunas relevantes e sem lacunas críticas, o que contribuiu para garantir a fiabilidade e a coerência necessárias na fase subsequente de modelação dos dados.

3.4.1. Modelos Preditivos

Com o intuito de avaliar a eficácia dos modelos preditivos desenvolvidos, foram adotadas diferentes estratégias de seleção de *features* com abordagens baseadas na correlação centre variáveis independentes e a variável-*target*. O objetivo consistiu em identificar a combinação de variáveis e algoritmos de ML que permitisse obter a melhor *performance* na classificação dos níveis de criticidade dos sinais vitais.

Foram avaliados cinco algoritmos de *machine learning*: *Logistic Regression*, *Support Vector Machine (SVM)*, *K-Nearest Neighbours (KNN)*, *Random Forest* (RF)e *Complement Naive Bayes*.

A métrica utilizada para aferir o desempenho dos modelos foi a *accuracy*, dada a sua natureza intuitiva e a capacidade de expressar, de forma simples, a proporção de classificações corretamente realizadas (Han, Kamber & Pei, 2012).

Para garantir a imparcialidade e replicabilidade da avaliação, o conjunto de dados foi dividido em dois subconjuntos: 80% para treino e 20% para teste. Esta divisão foi realizada utilizando a função *train_test_split*, assegurando que os modelos fossem treinados com uma amostra representativa e validados com um subconjunto independente. Neste processo, foi aplicada estratificação da variável-*target*, mantendo a preservação da distribuição das classes nos dois subconjuntos. Este passo revelou-se particularmente face ao desequilíbrio de classes observado no *dataset*, evitando enviesamentos e favorecendo uma melhor capacidade de generalização dos modelos.

Durante o treino, os hiperparâmetros dos modelos foram ajustados com recurso à técnica de validação cruzada, *GridSearchCV*. Esta abordagem permite testar automaticamente um conjunto de combinações de parâmetros pré-definidos, avaliando o seu desempenho de forma sistemática, com o objetivo de encontrar a configuração ideal para cada algoritmo (Géron, 2019). A utilização deste método foi especialmente eficaz em modelos mais sensíveis à parametrização, como o *SVM*, o RF e o *KNN*, contribuindo para uma otimização robusta do desempenho final.

As previsões foram realizadas com base em quatro abordagens distintas de seleção de *features*. A primeira abordagem consistiu na utilização integral das variáveis disponíveis no

dataset, sem qualquer processo de seleção prévio (Sem *Feature Selection*). Na segunda, foi aplicada uma seleção estatística, tendo sido consideradas apenas as *features* com valores de correlação absoluta superiores a ±0,1 (Com *Feature Selection*). A terceira abordagem baseou-se na inclusão exclusiva de variáveis com correlação positiva significativa com a variável-target, caracterizando-se como uma estratégia de correlação direta. Por fim, a quarta abordagem considerou unicamente as *features* com correlação fraca ou negativa em relação à variável-*target*, permitindo explorar o impacto de preditores menos evidentes, numa lógica de correlação indireta.

A aplicação destas diferentes estratégias permitiu comparar o desempenho relativo dos algoritmos e analisar o efeito da seleção de variáveis na capacidade de generalização dos modelos. Esta abordagem revelou-se particularmente útil na avaliação da robustez de cada classificador e na sua sensibilidade ao conjunto de *features* utilizado, aspetos com implicações diretas na fiabilidade e na aplicabilidade clínica dos modelos desenvolvidos.

3.4.1.1 Dataframe Pulse Criticity

Abordagem / Modelo	Logistic Regression	SVM	KNN	Random Forest	Complement Naive Bayes
Sem FS	0.9667	0.9227	0.9736	0.9749	0.4288
Com FS (±0.1)	0.9212	0.9572	0.9285	0.9636	0.3814
Correlação Direta	0.9352	0.9572	0.9785	0.9789	0.5159
Correlação Indireta	0.9227	0.9227	0.9164	0.9759	0.5057

Tabela 2- Accuracy dos modelos preditivos da feature pulse criticit

A Tabela 1 apresenta os resultados da avaliação dos cinco modelos de ML aplicados à tarefa de previsão da variável *pulse_criticity*, considerando as quatro abordagens distintas de seleção de *features.* Esta análise permite comparar o desempenho individual de cada algoritmo e avaliar o impacto das diferentes estratégias de seleção de variáveis na capacidade preditiva dos modelos desenvolvidos.

Os modelos RF e KNN apresentaram o melhor desempenho, com taxa de acerto elevada. em todas as abordagens. O modelo RF destacou-se pela estabilidade da sua performance, com valores de accuracy entre 0.9636 e 0.9789, enquanto o modelo *KNN* obteve a mais elevada entre todos os testes realizados (0.9785), na abordagem baseada em correlação direta, revelando-se particularmente eficaz na identificação dos padrões associados à criticidade do pulso.

O modelo SMV também apresentou resultados satisfatórios, especialmente nas abordagens com seleção de *features*, e atingindo uma taxa de acerto próximas de 0.9572. Em contraste, a *Logistic Regression* demonstrou maior sensibilidade à redução de variáveis, registando uma diminuição de desempenho de 0.9667 (sem seleção de *feature*) para 0.9212 (com seleção baseada na correlação ±0.1), o que evidencia dependência em relação ao conjunto completo de variáveis.

O modelo *Complement Naive Bayes,* por sua vez, apresentou desempenho significativamente inferior em todas as abordagens testadas, com accuracy entre 0.3814 e 0.5159. Estes resultados indicam que o modelo não se adequa ao problema em estudo, possivelmente devido à sua limitação em capturar relações complexas entre variáveis, uma vez que pressupõe independência entre *features*.

Relativamente às estratégias de seleção de *features*, a abordagem baseada em correlação direta revelou-se a mais eficaz, proporcionando melhores resultados para a maioria dos modelos. Esta estratégia, que considera apenas as variáveis com correlação positiva significativa com a variável-*target*, permitiu maximizar a capacidade preditiva dos algoritmos sem comprometer a integridade do modelo. Por outro lado, a seleção com limiar absoluto de correlação (±0.1) teve impacto negativo no desempenho da maioria dos modelos, sugerindo que este método poderá eliminar variáveis com valor informativo relevante. A abordagem de correlação indireta, embora menos eficaz que a correlação direta, superou o desempenho obtido com o limiar absoluto em vários casos — em especial no *Naive Bayes* — sugerindo que variáveis com correlação fraca ou negativa podem, em determinados contextos, contribuir positivamente para a previsão.

Os resultados demonstram a superioridade dos modelos RF e KNN na previsão da criticidade do pulso, tanto pela robustez como pela consistência perante diferentes estratégias de seleção de variáveis. Além disso, verifica-se que uma escolha criteriosa das *features* — em particular através da análise da correlação direta — pode melhorar significativamente o desempenho preditivo, reforçando a importância desta etapa na construção de modelos fiáveis e aplicáveis ao contexto clínico.

3.4.1.2. Dataframe SpO_2 Criticity

Abordagem / Modelo	Logistic Regression	SVM	KNN	Random Forest	Complement Naive Bayes
Sem FS	0.9663	0.9359	0.9454	0.9934	0.5907
Com FS (±0.1)	0.9359	0.9359	0.9355	0.9418	0.3756
Correlação Direta	0.9414	0.9377	0.9652	0.9933	0.7721
Correlação Indireta	0.9238	0.9359	0.9337	1.0000	0.8923

Tabela 2- Accuracy dos modelos preditivos da feature SpO_2 Criticity

A Tabela 2 apresenta os resultados da avaliação dos modelos de ML aplicados à previsão da variável *SpO_2_criticity*, considerando as quatro abordagens de seleção de *features* anteriormente descritas.

À semelhança do observado anteriormente, o modelo RF destacou-se pelo seu desempenho superior e consistente. Em todas as abordagens, obteve valores de *accuracy* elevados, com destaque para 0.9934 (sem seleção de *features*) e 1.000 (com correlação indireta), este último representando um desempenho perfeito no conjunto de teste. Estes resultados evidenciam a robustez e a capacidade de generalização do modelo, independentemente da abordagem de seleção de variáveis aplicada.

O modelo KNN apresentou igualmente um desempenho competitivo, com uma taxa de acerto próxima de 0.945 na maioria das abordagens, atingindo 0.9652 na estratégia de correlação direta. De forma semelhante, o *modelo Logistic Regression* obteve resultados sólidos, com *accuracy* de 0.9663 na abordagem sem seleção de *features* e desempenho estável nas restantes, embora ligeiramente inferior.

O modelo SVM demonstrou uma performance consistente em todas as abordagens, com *accuracy* de 0.9359, sugerindo que a seleção de *features* teve impacto limitado neste algoritmo em particular. Em contrapartida, o modelo *Complement Naive Bayes* registou uma melhoria expressiva em relação à tarefa de previsão de *pulse_criticity*. Embora a sua *accuracy* inicial (sem seleção de *features*) fosse de apenas 0.5907, os valores aumentaram significativamente com as abordagens de correlação direta (0.7721) e, sobretudo, de correlação indireta (0.8923), revelando-se mais eficaz neste segundo cenário de previsão.

No que respeita às estratégias de seleção de *features*, a abordagem baseada em correlação direta voltou a apresentar resultados consistentes e positivos, especialmente nos modelos KNN, RF e *Naive Bayes*. Contudo, neste caso, a estratégia de correlação indireta demonstrou

um desempenho particularmente notável, sobretudo para o RF que atingiu uma taxa de acerto perfeita (1.000), e para o *Naive Bayes*, que obteve a melhor performance global entre todas as abordagens aplicadas. Esta tendência sugere que, ao contrário do esperado, *features* com correlação fraca ou negativa com a variável-*target* podem conter informação complementar valiosa, capaz de melhorar o desempenho preditivo em determinados modelos.

A seleção baseada num limiar de correlação (±0.1) revelou, mais uma vez, impacto negativo na maioria dos modelos, resultando em quedas acentuadas de desempenho, em especial no *Naive Bayes* (0.3756). Este comportamento reforça as limitações desta abordagem simplista, que poderá eliminar variáveis informativamente relevantes e reduzir a capacidade dos modelos em identificar padrões significativos.

Em conclusão, os resultados obtidos na previsão de SpO_2*_criticity* confirmam a superioridade do modelo RF, não apenas pela precisão alcançada, mas também pela sua estabilidade face às diferentes abordagens de seleção de *features*. Os modelos *KNN* e *Logistic Regression* também demonstraram desempenho sólido, enquanto o *Naive Bayes* se revelou substancialmente mais eficaz neste cenário do que na tarefa anterior. De forma global, as estratégias de correlação direta e indireta mostraram-se mais eficazes do que a abordagem baseada em limiar absoluto, reforçando a importância de uma seleção criteriosa das variáveis na construção de modelos preditivos fiáveis e clinicamente relevantes.

3.4.1.3. Dataframe Systolic Criticity

Abordagem / Modelo	Logistic Regression	SVM	KNN	Random Forest	Complement Naive Bayes
Sem FS	0.9238	0.9359	0.9337	0.9252	0.8223
Com FS (±0.1)	0.7401	0.7401	0.7599	0.9035	0.5433
Correlação Direta (CD)	0.9455	0.7401	0.9876	0.9975	0.5866
Correlação Indireta (CI)	0.9678	0.7401	0.9950	1.0000	0.8923

Tabela 3- accuracy dos modelos preditivos da feature Systolic Criticity

A Tabela 3 apresenta os resultados obtidos na tarefa de previsão da variável *systolic_criticity*, com base na aplicação dos cinco modelos de ML anteriormente descritos, testados sob quatro abordagens distintas de seleção de *features*.

O modelo RF voltou a destacar-se como o algoritmo com melhor desempenho, atingindo *accuracy* perfeita (1.000) na abordagem de correlação indireta e valores muito próximos de

1.000 nas restantes abordagens — 0.9975 (correlação direta) e 0.9252 (sem seleção de *features*). Estes resultados confirmam a robustez e a adaptabilidade deste modelo, independentemente da técnica de seleção aplicada.

O modelo KNN apresentou igualmente desempenho elevado, com destaque para as abordagens de correlação direta (0.9876) e correlação indireta (0.9950), evidenciando a capacidade do algoritmo em captar padrões relevantes mesmo quando as *features* utilizadas não apresentam, à partida, correlação positiva com a variável-*target*. Nas abordagens sem seleção e com limiar absoluto (±0.1), o desempenho foi inferior, embora ainda satisfatório (0.9337 e 0.7599, respetivamente).

O modelo *Logistic Regression* obteve resultados intermédios, com melhor desempenho na correlação indireta (0.9678), seguida da correlação direta (0.9455). Verificou-se, contudo, uma quebra acentuada de desempenho na abordagem com limiar absoluto (±0.1), em que a taxa de acerto desceu para 0.7401, reforçando a limitação desta técnica na retenção de variáveis relevantes.

O modelo SMV demonstrou uma estabilidade notável, apresentando valores de *accuracy* constantes (0.9359) em todas as abordagens. Esta insensibilidade à seleção de *features* poderá estar associada à forma como o modelo constrói as suas fronteiras de decisão, tornando-o menos dependente da filtragem de variáveis.

Por sua vez, o *Complement Naive Bayes* voltou a apresentar desempenhos inferiores em comparação com os restantes algoritmos. Embora o valor da *accuracy* sem seleção (0.8223) tenha melhorado nas abordagens de correlação indireta (0.8923), os resultados mantiveram-se baixos para FS ±0.1 (0.5433) e correlação direta (0.5866). Este comportamento, consistente com as análises anteriores, reflete as limitações estruturais do modelo, que pressupõe independência entre as variáveis e não se adequa a relações complexas.

No que respeita às estratégias de seleção de *features*, destaca-se novamente o fraco desempenho da abordagem baseada num limiar absoluto (±0.1), que comprometeu significativamente a performance da maioria dos modelos. Em contraste, a correlação indireta revelou-se particularmente eficaz, sobretudo nos modelos FR, KNN, *Logistic Regression* e *Naive Bayes*, desafiando a ideia tradicional de que apenas correlações positivas com a variável-*target* são úteis.

Os resultados da previsão da *systolic_criticity* confirmam a superioridade dos modelos RF e KNN, tanto pela precisão como pela consistência. A abordagem de correlação indireta,

destacou-se como a mais eficaz para esta variável, superando mesmo a correlação direta em vários modelos. Estes resultados reforçam a importância de explorar diferentes estratégias de seleção de variáveis e de adequar o modelo preditivo às características específicas da variável-target.

3.4.2. Conclusão Modelos

Os resultados obtidos demonstram que os algoritmos baseados em árvores *(Random Forest)* e em distância *(KNN)* se revelaram os mais eficazes na previsão da criticidade dos sinais vitais, mantendo os níveis da taxa de acerto elevados mesmo sob diferentes abordagens de seleção de *features*.

O modelo *Random Forest*, em particular, destacou-se pela obtenção de *accuracy* perfeita (100%) em duas variáveis, evidenciando robustez face a diferentes subconjuntos de dados e elevada capacidade de generalização. Por outro lado, o *Complement Naive Bayes*, apesar da sua simplicidade e rapidez computacional, mostrou-se inadequado para este tipo de problema, apresentando desempenho consideravelmente inferior, sobretudo quando as *features* disponíveis não se encontravam bem ajustadas.

De forma geral, esta análise reforça o potencial da aprendizagem automática no domínio da saúde, evidenciando que modelos adequadamente treinados e validados podem ser integrados em sistemas de apoio à decisão clínica, contribuindo para a deteção precoce de situações de risco e para a personalização da monitorização de pacientes.

4. Resultados e discussão

Após o desenvolvimento e a avaliação dos modelos, de modo a verificar se os resultados obtidos correspondiam aos definidos inicialmente, foi possível avançar para a etapa de implementação .

Com base nas necessidades identificadas pela clínica, foram desenvolvidos dois *dashboards* interativos no âmbito de um sistema de apoio à decisão orientado para a análise de dados clínicos. O principal objetivo destes elementos consiste em disponibilizar uma visualização integrada, dinâmica e intuitiva da informação, permitindo aos utilizadores monitorizar, de forma eficiente, os principais indicadores e níveis de criticidade associados aos pacientes.

Os *dashboards* foram desenvolvidos na ferramenta *Microsoft Power BI*, selecionada pela sua capacidade de integração com múltiplas fontes de dados, flexibilidade na criação de visualizações interativas e possibilidade de atualização automática da informação. Esta solução de *business intelligence* permite transformar dados brutos em informação estruturada e visualmente acessível, facilitando a análise exploratória e a tomada de decisão baseada em evidência (Talakola, 2022).

4.1. Dashboard 1 – Análise Individual do Paciente

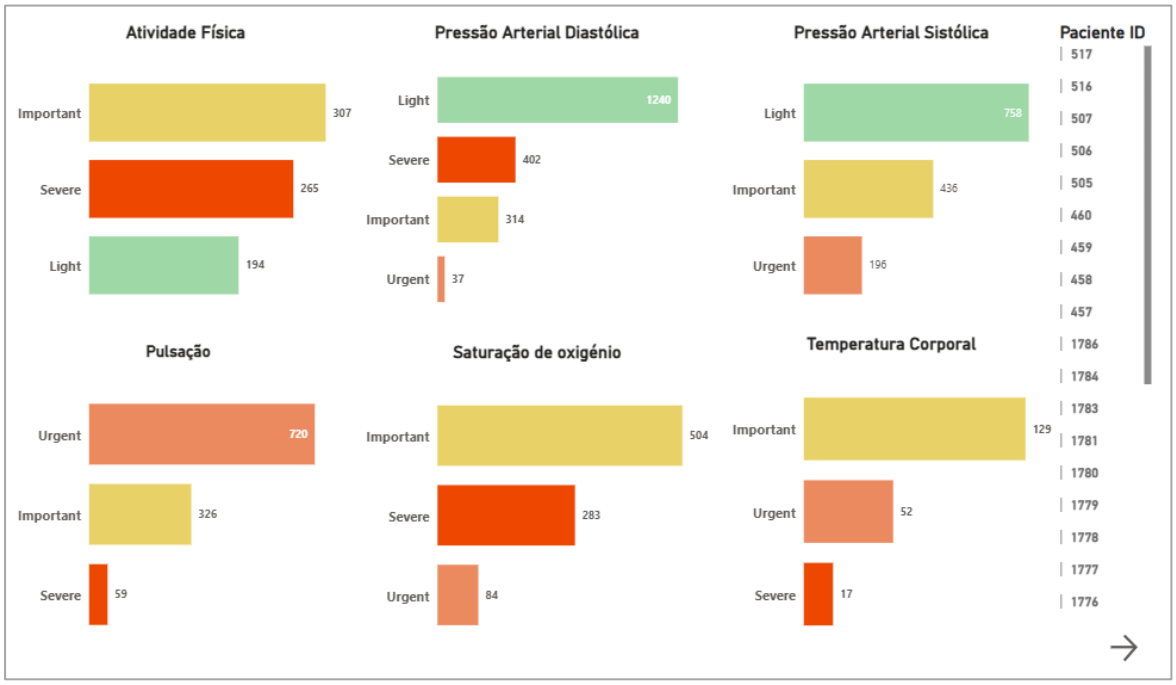


Figura 1- Dashboard - Standard

A primeira página do *dashboard* (figura 1) desenvolvida no *Power BI*, tem como finalidade apresentar, de forma clara e intuitiva, a distribuição e a gravidade de diferentes indicadores fisiológicos recolhidos de um conjunto de pacientes. Trata-se de um painel interativo que permite observar, em simultâneo, a variação de seis parâmetros vitais distintos, possibilitando uma compreensão imediata do estado de saúde geral e da existência de situações de risco que mereçam atenção clínica.

O painel é composto por seis gráficos de barras horizontais, cada um dedicado a uma variável fisiológica específica: atividade física, pressão arterial diastólica, pressão arterial sistólica, pulsação, saturação de oxigénio e temperatura corporal. Cada gráfico mantém uma estrutura coerente, o que facilita a leitura e a comparação entre os diferentes indicadores. No eixo vertical de cada visual encontram-se as categorias de severidade atribuídas a cada registo — nomeadamente *Light, Important, Urgent* e *Severe* —, enquanto no eixo horizontal se representa a quantidade de observações ou o valor correspondente a cada nível de gravidade.

As categorias de severidade estão associadas a uma escala cromática progressiva que facilita a interpretação visual e reforça a leitura imediata da informação. A cor verde representa a categoria *Light*, correspondente a uma condição normal e estável. O amarelo identifica a categoria *Important*, que indica a necessidade de atenção ou monitorização, ainda sem risco elevado. O laranja, associado à categoria *Urgent*, sinaliza situações que requerem uma resposta rápida, e o vermelho, correspondente à categoria *Severe*, representa estados críticos que exigem intervenção imediata. Esta sequência de cores — do verde ao vermelho segue uma lógica universal de gravidade crescente, garantindo uma leitura intuitiva e imediata dos dados apresentados.

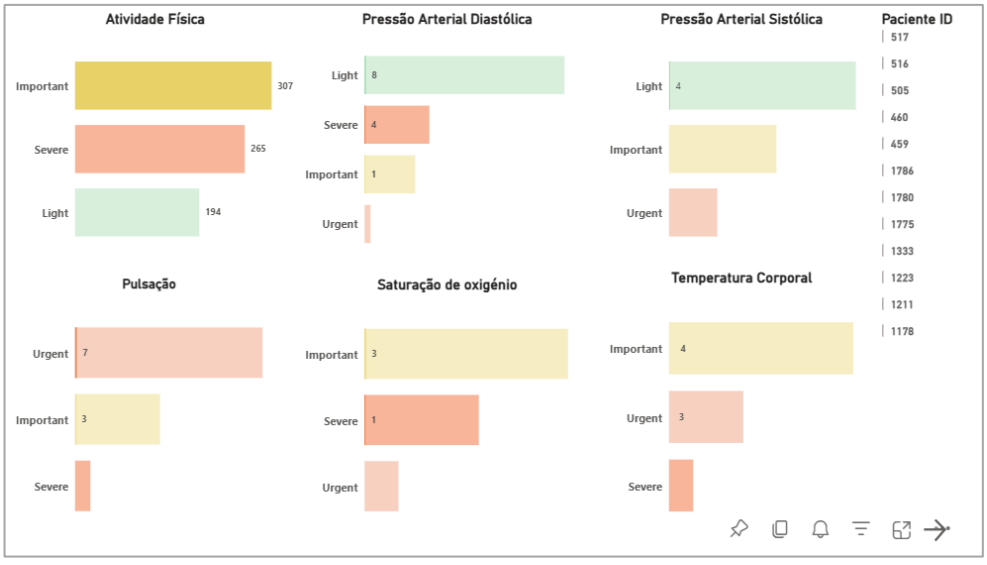


Figura 2- Dashboard - Com seleção

No lado direito do *dashboard (*figura 2*),* encontra-se um filtro que contém o identificador de cada paciente, designado por “Paciente *ID*”. Através deste filtro é possível selecionar um indivíduo específico e visualizar exclusivamente os dados que lhe dizem respeito, passando de uma visão global para uma análise individualizada. O painel é totalmente interativo: a seleção de uma categoria num dos gráficos faz com que os restantes se ajustem automaticamente, refletindo apenas os dados relacionados e promovendo uma exploração dinâmica e integrada da informação. Além disso, permite que o utilizador selecione uma variável específica e o respetivo nível de criticidade, visualizando de imediato todos os pacientes que apresentam essa condição. Esta funcionalidade facilita a identificação de padrões de risco e contribui para a priorização de casos que requerem acompanhamento mais próximo, reforçando o carácter analítico e de apoio à decisão clínica do painel.

4.2. Dashboard 2 – Análise Transversal por Variável

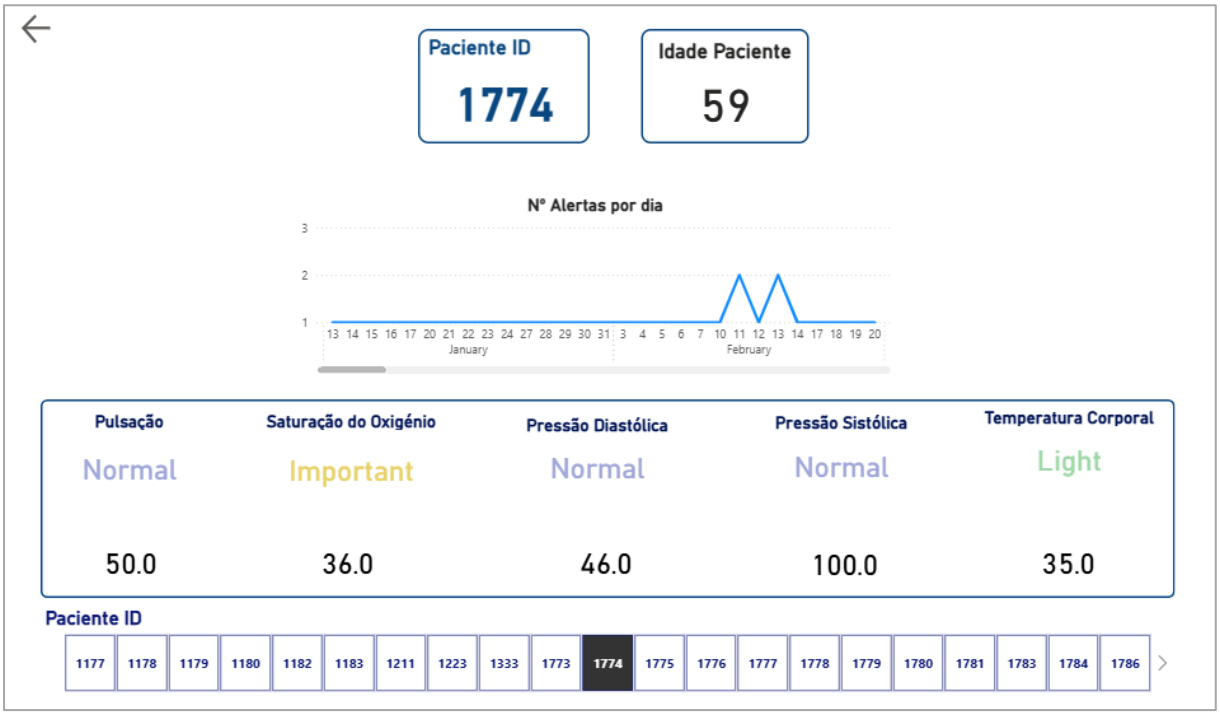


Figura 3- Dashboard - Standard

A segunda página do *dashboard* (figura3), foi concebida para apresentar uma análise detalhada e individual de cada paciente, permitindo visualizar de forma clara o seu estado clínico atual e a evolução de alertas registados ao longo do tempo. Esta página é acedida a partir da visão geral da primeira página e centra-se na observação personalizada dos indicadores fisiológicos de um único paciente.

Na parte superior do painel encontram-se dois cartões informativos que identificam o paciente selecionado, exibindo o respetivo *ID* e a idade do indivíduo em análise. Logo abaixo,

apresenta-se um gráfico de linhas que mostra o número de alertas por dia, permitindo acompanhar a evolução temporal dos eventos registados. Esta representação facilita a deteção de picos de alerta e possíveis períodos críticos, contribuindo para a identificação de padrões de comportamento clínico e de momentos em que foi registada maior instabilidade nos parâmetros fisiológicos. Na secção inferior do *dashboard* são apresentados, de forma organizada, os principais indicadores vitais do paciente: pulsação, saturação de oxigénio, pressão arterial diastólica, pressão arterial sistólica e temperatura corporal. Cada um destes parâmetros é acompanhado pela respetiva classificação de severidade (por exemplo, *Normal, Important* ou *Light)* e pelo valor numérico correspondente. As classificações mantêm a lógica cromática utilizada na primeira página, permitindo uma leitura coerente e imediata do estado clínico: tons de verde representam condições normais, o amarelo indica situações que exigem atenção e o laranja ou vermelho assinalam estados mais críticos.

Na parte inferior, encontra-se uma barra de navegação com os identificadores de todos os pacientes disponíveis. Esta funcionalidade permite alternar rapidamente entre perfis individuais, tornando o painel altamente dinâmico e interativo. Assim, o utilizador pode selecionar qualquer paciente e visualizar instantaneamente os seus dados específicos, sem necessidade de regressar à página principal.

De forma geral, esta segunda página funciona como um painel de monitorização personalizada, oferecendo uma visão centrada no paciente e complementando a análise global apresentada na primeira página. A sua estrutura facilita a observação imediata dos principais sinais vitais, a deteção de anomalias e o acompanhamento da evolução clínica ao longo do tempo, constituindo uma ferramenta de apoio essencial à tomada de decisão médica e à gestão de cuidados individualizados.

O segundo painel complementa o anterior, ao fornecer uma visão global e comparativa das variáveis monitorizadas, apoiando a tomada de decisão baseada em dados (*data-driven decision making).*

O conjunto de *dashboards* desenvolvidos em *Power BI* constitui uma ferramenta integrada de monitorização e análise clínica, que permite uma leitura eficiente e interativa dos dados fisiológicos recolhidos. A primeira página apresenta uma visão global e comparativa dos principais indicadores de saúde, possibilitando identificar rapidamente padrões de gravidade e áreas que requerem maior atenção. Já a segunda página complementa essa análise com uma abordagem individualizada, centrada em cada paciente, permitindo acompanhar em detalhe os seus valores vitais e a evolução temporal dos alertas registados.

A combinação destas duas perspetivas — global e individual — oferece uma visão abrangente e equilibrada do estado clínico dos pacientes, promovendo uma análise mais informada e precisa. A utilização de uma linguagem visual coerente, aliada à interatividade e à possibilidade de filtragem dinâmica, transforma o *dashboard* num instrumento eficaz de apoio à decisão médica, contribuindo para uma gestão mais eficiente dos cuidados de saúde e para uma resposta mais célere a situações de risco.

5. Conclusões e trabalho futuro

O presente trabalho teve como objetivo central conceber e validar um modelo integrado de análise, visualização e previsão da criticidade clínica, sustentado na metodologia CRISP-DM e nos dados disponibilizados pela empresa Hope Care, S.A. Este modelo pretende apoiar a decisão médica em tempo real, antecipando situações clínicas críticas através da análise de dados de telemonitorização. A investigação desenvolveu-se de forma estruturada, articulando a revisão da literatura, a compreensão e preparação dos dados, a análise estatística descritiva, a modelação preditiva e visualização dos resultados.

A revisão do estado da arte (capítulo 2) permitiu identificar as principais abordagens, modelos e tecnologias aplicadas à telemonitorização e à previsão de eventos clínicos. Verificou-se uma predominância de técnicas de machine learning supervisionadas — como *Random Forest, Support Vector Machines* e Redes Neuronais Artificiais —, bem como a adoção crescente de metodologias estruturadas como a CRISP-DM para a implementação de sistemas de apoio à decisão clínica. Esta análise evidenciou ainda a importância da qualidade dos dados, da interpretabilidade dos modelos e da integração entre módulos analíticos e interfaces de visualização. Assim, o primeiro objetivo foi alcançado, tendo a revisão bibliográfica fornecido o enquadramento teórico e metodológico que sustentou o desenho da investigação e a seleção das técnicas utilizadas nas fases posteriores.

A preparação e caracterização do conjunto de dados da Hope Care S.A. (capítulos 3.1 e 3.2) permitiram assegurar a sua consistência e qualidade analítica. O tratamento de *missing* values, a normalização e a seleção das variáveis mais relevantes garantiram a robustez necessária para as etapas de modelação. A análise estatística descritiva (capítulo 3.3) possibilitou caracterizar o comportamento dos utentes no programa de monitorização, avaliando a frequência de utilização dos dispositivos, a taxa de adesão segmentada por género e faixa etária e os padrões de medição registados ao longo do tempo. Foram igualmente exploradas as distribuições dos principais parâmetros fisiológicos — *pulse*, SpO_2 e *systolic* —, identificando tendências e variações relevantes, bem como *outliers*. A análise das categorias de criticidade clínica destas três variáveis permitiu compreender os

fatores associados à ocorrência de estados críticos, contribuindo para a definição das variáveis-*target* utilizadas nos modelos preditivos. Deste modo, os objetivos relacionados com a compreensão, preparação e análise exploratória dos dados foram integralmente cumpridos.

A fase de modelação preditiva (capítulo 4) constituiu o núcleo técnico do estudo. Foram implementados e comparados cinco algoritmos de ML, tendo sido avaliados com base em métricas de desempenho como a precisão, a sensibilidade e a especificidade. Os resultados obtidos demonstraram que os modelos baseados em RF e *RNA* apresentaram o melhor equilíbrio entre sensibilidade e especificidade, assegurando uma boa capacidade de generalização e um desempenho robusto na previsão de alertas clínicos. Com base nestes resultados, foram concebidos *dashboards* interativos que disponibilizam às equipas clínicas uma visualização clara e intuitiva dos resultados analíticos e preditivos, permitindo a identificação imediata de padrões relevantes e o acompanhamento individualizado dos utentes. Assim, os objetivos relativos ao desenvolvimento e validação de modelos preditivos foram atingidos, confirmando a viabilidade técnica e científica da abordagem proposta e o seu potencial contributo para a melhoria dos processos de apoio à decisão em saúde digital.

Como trabalho futuro, recomenda-se a expansão e automatização da base de dados, de modo a garantir uma recolha contínua, uniforme e fiável das medições provenientes dos dispositivos de monitorização. A integração direta entre estes dispositivos e a base de dados central da Hope Care, S.A. permitirá eliminar a necessidade de recolha manual de dados, reduzindo significativamente falhas humanas, atrasos e perdas de informação. Adicionalmente, recomenda-se que a Hope Care, S.A. integre esta solução nos seus sistemas internos, assegurando que os *dashboards* clínicos estejam conectados diretamente à base de dados e atualizados em tempo real. Esta integração permitirá fornecer aos profissionais de saúde indicadores dinâmicos e fiáveis, reforçando a capacidade de resposta imediata perante situações de risco e promovendo uma monitorização mais proativa e eficiente.

Paralelamente, a validação dos modelos preditivos em populações mais diversificadas constitui um passo essencial para assegurar a sua generalização e robustez. A aplicação do sistema a diferentes faixas etárias, patologias e contextos clínicos permitirá avaliar a estabilidade dos algoritmos e ajustar os parâmetros às especificidades de cada grupo de utentes. Esta abordagem contribuirá para o desenvolvimento de modelos mais

personalizados e adaptativos, capazes de refletir a variabilidade individual e de sustentar uma prática clínica mais preventiva e centrada no paciente.

Referências bibliográficas

Adler, D. A., Ben-Zeev, D., Tseng, V. W.-S., Kane, J. M., Brian, R., Campbell, A. T., Hauser, M., Scherer, E. A., & Choudhury, T. (2020). Predicting early warning signs of psychotic relapse from passive sensing data: An approach using encoder-decoder neural networks. JMIR mHealth and uHealth, 8(8). https://doi.org/10.2196/19962

Asghari, P. (2021). A diagnostic prediction model for colorectal cancer in elderlies via internet of medical things. International Journal of Information Technology (Singapore), 13(4), 1423–1429. https://doi.org/10.1007/s41870-021-00663-5

Basavaiah, J., Anthony, A. A., Mahadevaswamy, S., & Naveen Kumar, H. N. (2024). An efficient approach of epilepsy seizure alert system using IoT and machine learning. Journal of Reliable Intelligent Environments, 10(4), 449–461. https://doi.org/10.1007/s40860-024-00228-w

Brown, A., White, C., & Green, D. (2023). Remote patient monitoring: Trends and outcomes. Journal of Telemedicine and Telecare, 29(4), 210–220.

Bus, S. A., Reeves, N. D., Armstrong, D. G., & Najafi, B. (2024). Offloading and adherence through technological advancements: Modern approaches for better foot care in diabetes. Diabetes/Metabolism Research and Reviews, 40(2). https://doi.org/10.1002/dmrr.3769

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.

Chen, J., Valehi, A., & Razi, A. (2019). Smart heart monitoring: Early prediction of heart problems through predictive analysis of ECG signals. IEEE Access, 7, 120831–120839. https://doi.org/10.1109/ACCESS.2019.2937875

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage Publications.

Dave, D., DeSalvo, D. J., Haridas, B., McKay, S., Shenoy, A., Koh, C. J., Lawley, M., & Erraguntla, M. (2020). Feature-based machine learning model for real-time hypoglycemia prediction. Journal of Diabetes Science and Technology, 15(4), 842–855. https://doi.org/10.1177/1932296820922622

Davergne, T., Rakotozafiarison, A., Servy, H., & Gossec, L. (2020). Wearable activity trackers in the management of rheumatic diseases: Where are we in 2020? Sensors, 20(17). https://doi.org/10.3390/s20174797

Dupljak, J., & Halili, S. (2024). Big data challenges in remote patient monitoring. Journal of Medical Systems, 48(1), 15.

Fellger, A., Sprint, G., Weeks, D., Crooks, E., & Cook, D. J. (2020). Wearable device-independent next day activity and next night sleep prediction for rehabilitation populations. IEEE Journal of Translational Engineering in Health and Medicine, 8. https://doi.org/10.1109/JTEHM.2020.3014564

Fletcher-Lloyd, N., Soreq, E., Wilson, D., Nilforooshan, R., Sharp, D. J., & Barnaghi, P. (2021). Home monitoring of daily living activities and prediction of agitation risk in people living with dementia. Alzheimer's & Dementia, 17, e058614. https://doi.org/10.1002/alz.058614

Genezini, M. (2022). Challenges in digital health: Security, privacy, and interoperability. Health Informatics Journal, 28(2), 145–158.

Guthrie, N. L., Berman, M. A., Edwards, K. L., Appelbaum, K. J., Dey, S., Carpenter, J., Eisenberg, D. M., & Katz, D. L. (2019). Achieving rapid blood pressure control with digital therapeutics: Retrospective cohort and machine learning study. JMIR Cardio, 3(1). https://doi.org/10.2196/13030

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras & TensorFlow (2nd ed.). O'Reilly Media.

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.

Hanumantharaju, R., Shreenath, K. N., Sowmya, B. J., & Srinivasa, K. G. (2022). Leveraging the fog-based machine learning model for ECG-based coronary disease prediction. International Journal of Business Intelligence and Data Mining, 21(3), 328–353. https://doi.org/10.1504/ijbidm.2022.125214

Iqbal, S. M. A., Mackay, A. L., & Aitchison, J. S. (2021). Advances in healthcare wearable devices. npj Flexible Electronics, 5, Article 49. https://doi.org/10.1038/s41528-021-00107-x

Kallam, S., Chakrabarti, P., Hung, B. T., & Siva Shankar, S. (2023). Remote health monitoring IoT framework using machine learning prediction and advanced artificial intelligence (AI) model. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5), 42–47. https://doi.org/10.17762/ijritcc.v11i5.6523

Kanza, R. A., Al Rasyid, M. U. H., Sukaridhoto, S., Utomo, B., Fauziah, S., & Primajaya, G. A. (2024). A cloud-centric application for elderly heart disease detection with machine learning and confusion matrix. In A. A. Yunanto et al. (Eds.), Proceedings of the IEEE International Conference on Industrial Electronics and Systems (pp. 709–714). https://doi.org/10.1109/IES63037.2024.10665789

Khanwalkar, P. (2024). Patient model based personalized remote health care for chronic disease. Proceedings on Engineering Sciences, 6(3), 897–902. https://doi.org/10.24874/PES06.03.001

Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: An introduction to data mining. Wiley.

Lavanya, P., Subba Reddy, I. V., Selvakumar, V., & Deshpande, S. V. (2024). An intelligent health surveillance system: Predictive modeling of cardiovascular parameters through machine learning algorithms using LoRa communication and Internet of Medical Things (IoMT). Journal of Internet Services and Information Security, 14(1), 165–179. https://doi.org/10.58346/JISIS.2024.I1.011

Li, X., & Wang, Y. (2023). Attention-based recurrent models for physiological signal prediction. IEEE Transactions on Neural Networks and Learning Systems, 34(4), 1123–1134.

Maqbool, S., Bajwa, I. S., Maqbool, S., Ramzan, S., & Chishty, M. J. (2023). A smart sensing technologies-based intelligent healthcare system for diabetes patients. Sensors, 23(23). https://doi.org/10.3390/s23239558

Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining applications in health care. Journal of Healthcare Engineering, 1(2), 147–161.

Miller, R., & Davis, S. (2021). Leveraging big data analytics for proactive healthcare. Health Informatics Journal, 27(3), 1–15.

Miller, T., & Davis, R. (2021). Analytical tools for telemedicine data: Machine learning approaches. Computers in Biology and Medicine, 135, 104582.

Mishra, A., & Mohapatro, M. (2019). An IoT framework for bio-medical sensor data acquisition and machine learning for early detection. International Journal of Advanced Technology and Engineering Exploration, 6(54), 112–125. https://doi.org/10.19101/IJATEE.2019.650027

Oliveira, L. (2024). Predictive analytics in remote health monitoring: Empowering patients through data visualization. International Journal of Medical Informatics, 170, 104439.

Omar, A. A., Theofrida, M., Samuel, R., Frederic, N., & Jimmy, N. (2023). Developing an IoT-based conversational AI recommender assistant for vital sign predicted anomalies. In Proceedings of the ACM International Conference on Computing and Communication (pp. 23–27). https://doi.org/10.1145/3589883.3589887

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Pappada, S. M., Cameron, B. D., Rosman, P. M., Bourey, R. E., Papadimos, T. J., Olorunto, W., & Borst, M. J. (2011). Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. Diabetes Technology & Therapeutics, 13(2), 135–141. https://doi.org/10.1089/dia.2010.0104

Press, G. (2016, March 23). Cleaning big data: Most time-consuming, least enjoyable data science task, survey says. Forbes. https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.

Pyle, D., & San Jose, P. (2005). Business modeling and data mining. Morgan Kaufmann Publishers.

Redman, T. C. (2001). Data quality: The field guide. Digital Press.

Santos, M., Pereira, F., & Costa, L. (2022). Handling heterogeneity in remote health data. Journal of Healthcare Engineering, 2022, 5483921.

Singh, D. N., Hashmi, M. F., & Sharma, S. K. (2020). Predictive analytics & modeling for modern health care system for cerebral palsy patients. Multimedia Tools and Applications, 79, 10285–10307. https://doi.org/10.1007/s11042-019-07834-4

Smith, J., & Jones, P. (2022). Paradigm shift in clinical practice: Digital transformation in healthcare. Health Policy and Technology, 11(1), 100–110.

Stoumpos, S., Kitsios, F., & Talias, M. (2023). Telemedicine and patient engagement: Current trends and future directions. BMC Health Services Research, 23(1), 456.

Talakola, S. (2022). Leverage Microsoft Power BI reports to generate insights and integrate with the application. International Journal of AI, Big Data, Computational and Management Studies, 3(2), 31–40. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V3I2P104

Thacharodi, A., Singh, P., Meenatchi, R., Ahmed, Z. H. T., Kumar, R. R. S., V, N., et al. (2024). Revolutionizing healthcare and medicine: The impact of modern technologies for a healthier future—A comprehensive review. Health Care Science, 3(5), 329–349. https://doi.org/10.1002/hcs2.115

Veyron, J.-H., Deparis, F., Al Zayat, M. N., Belmin, J., & Havreng-Théry, C. (2024). Postimplementation evaluation in assisted living facilities of an eHealth medical device developed to predict and avoid unplanned hospitalizations: Pragmatic trial. Journal of Medical Internet Research, 26. https://doi.org/10.2196/55460

Vouzis, E., & Maglogiannis, I. (2023). Prediction of early dropouts in patient remote monitoring programs. SN Computer Science, 4(5). https://doi.org/10.1007/s42979-023-01843-9

Wang, K., Cao, S., Kaur, J., & Morita, P. P. (2024). Remote heart rate monitoring with contactless ambient technology using machine learning for aging population. Gerontechnology, 23, 1–1. https://doi.org/10.4017/GT.2024.23.S.886.OPP

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining (pp. 29–39). https://www.the-modeling-agency.com/crisp-dm.pdf

Zaitin, Y., Mansurova, M., Kunelbayev, M., Tyulepberdinova, G., Sarsembayeva, T., & Shomanov, A. (2024). Development of a patient health monitoring system based on the internet of things with a module for predicting vital signs. Indonesian Journal of Electrical Engineering and Computer Science, 33(1), 518–529. https://doi.org/10.11591/ijeecs.v33.i1.pp518-529

Zhang, P., Fonnesbeck, C., Schmidt, D. C., White, J., Kleinberg, S., & Mulvaney, S. A. (2022). Using momentary assessment and machine learning to identify barriers to self-

management in type 1 diabetes: Observational study. JMIR mHealth and uHealth, 10(3). https://doi.org/10.2196/21959

Anexo

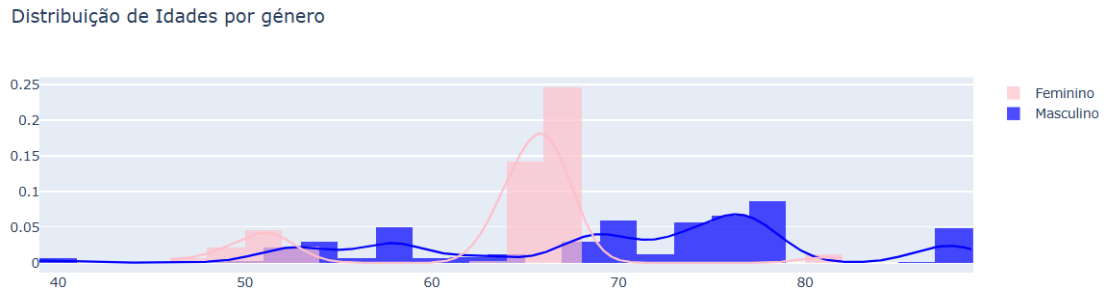


Figura 1- Distribuição de Idades por género

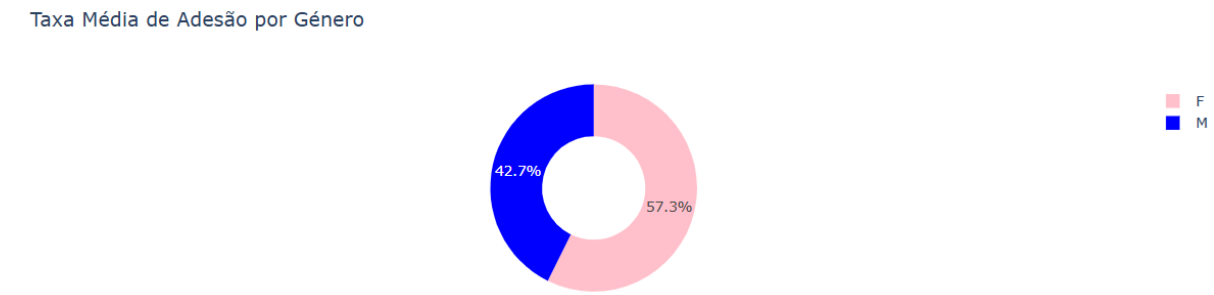


Figura 2 - Taxa média de adesão por género

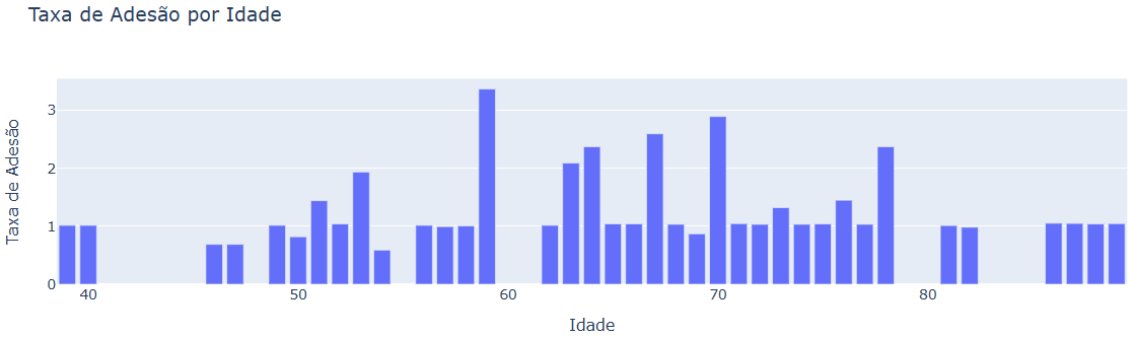


Figura 3 - Taxa média de adesão por idade



Figura 4 – Distribuição de medições por dia do mês

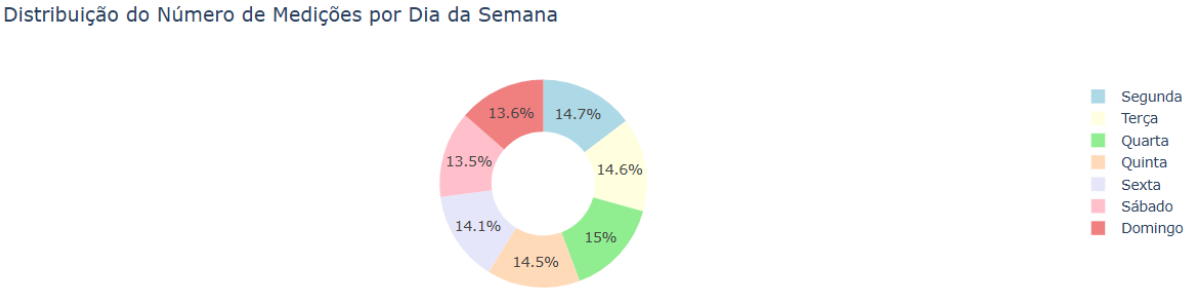


Figura 5 – Distribuição de medições por dia da semana

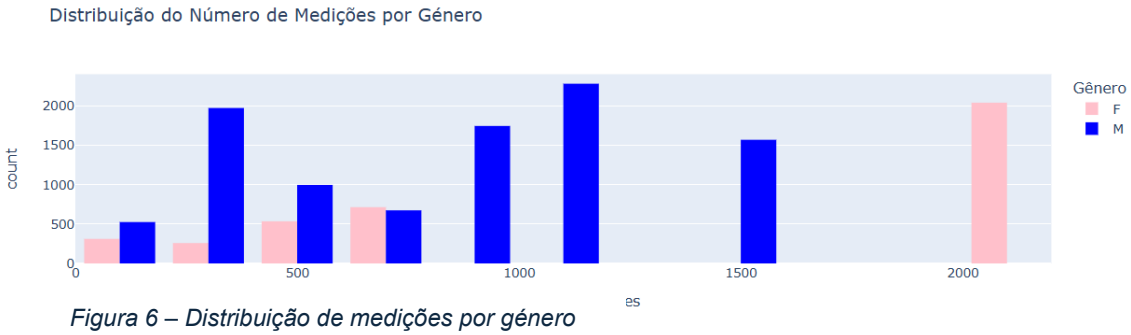


Figura 6 – Distribuição de medições por género

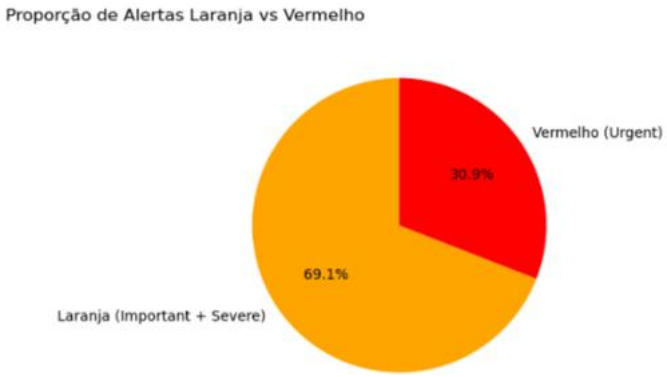


Figura 7 – Distribuição de Alertas Laranja vs Vermelho

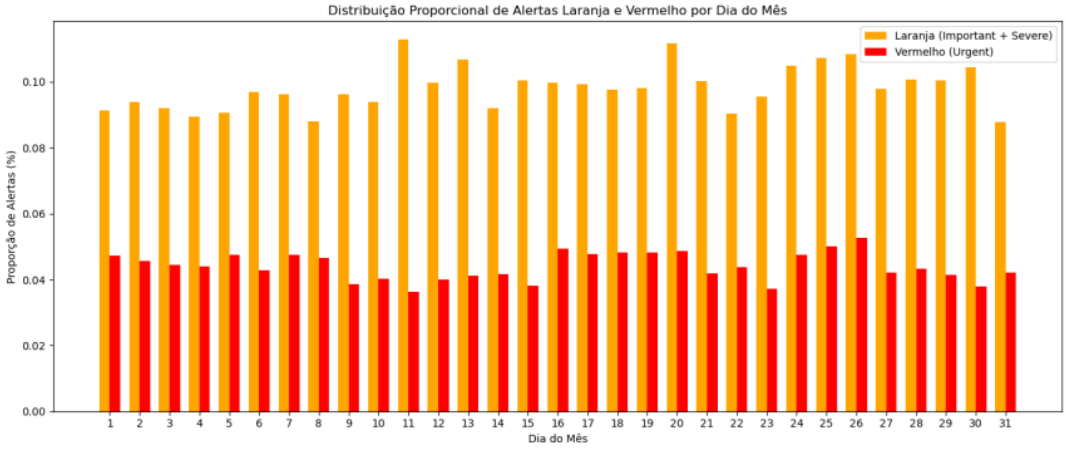


Figura 8 – Distribuição de Alertas Laranja vs Vermelho por dia do mês

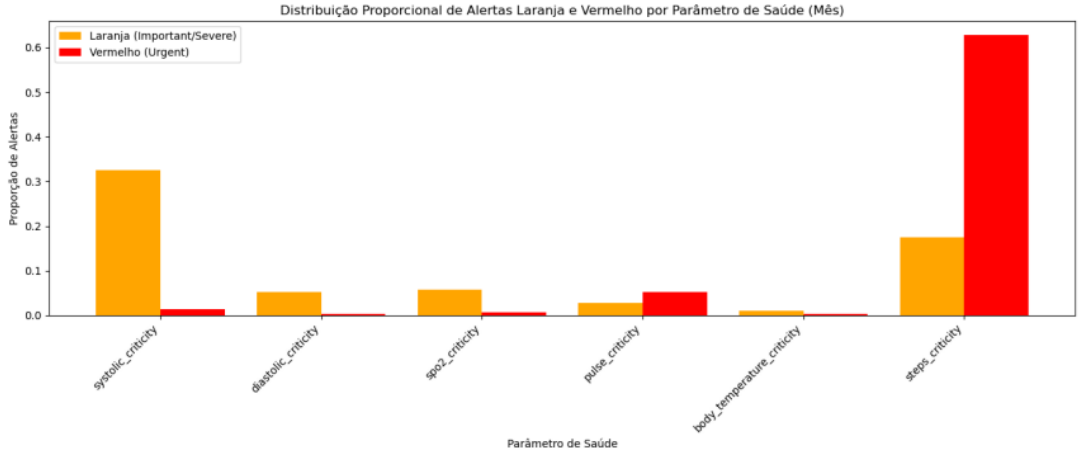


Figura 9 – Distribuição de Alertas Laranja vs Vermelho por parâmetro de saúde

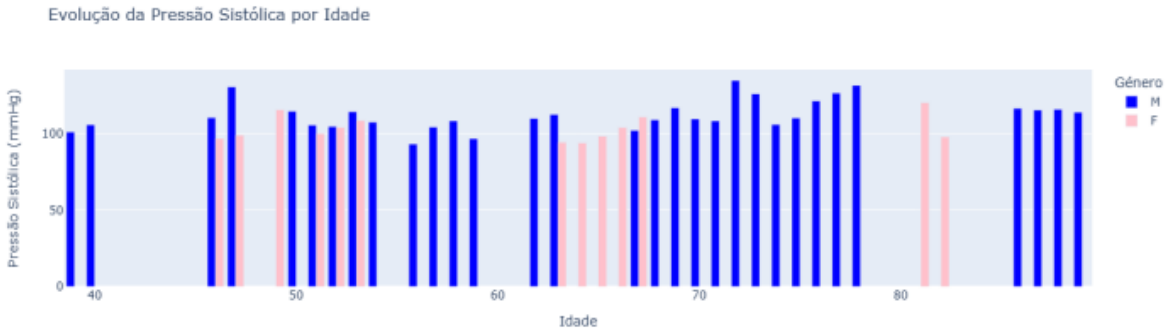


Figura 10 – Distribuição de Alertas por parâmetro de saúde e por nível de criticidade

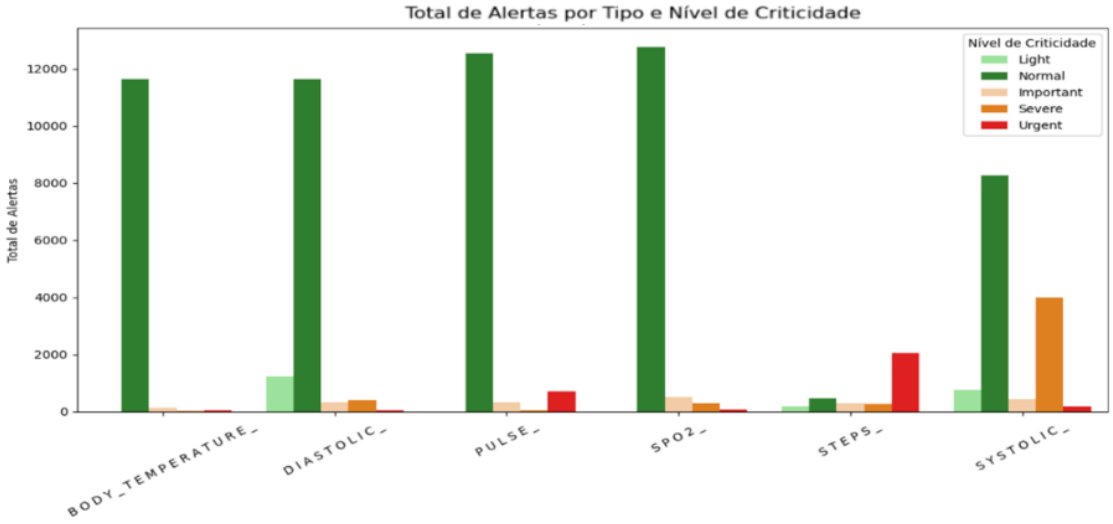


Figura 11 – Distribuição da Pressão sistólica por faixa etária e género

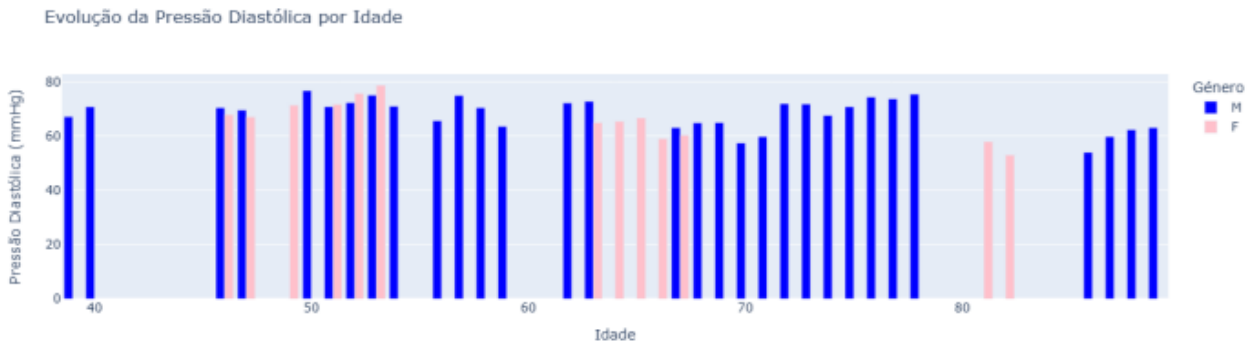


Figura 12 – Distribuição da Pressão distólica por faixa etária e género

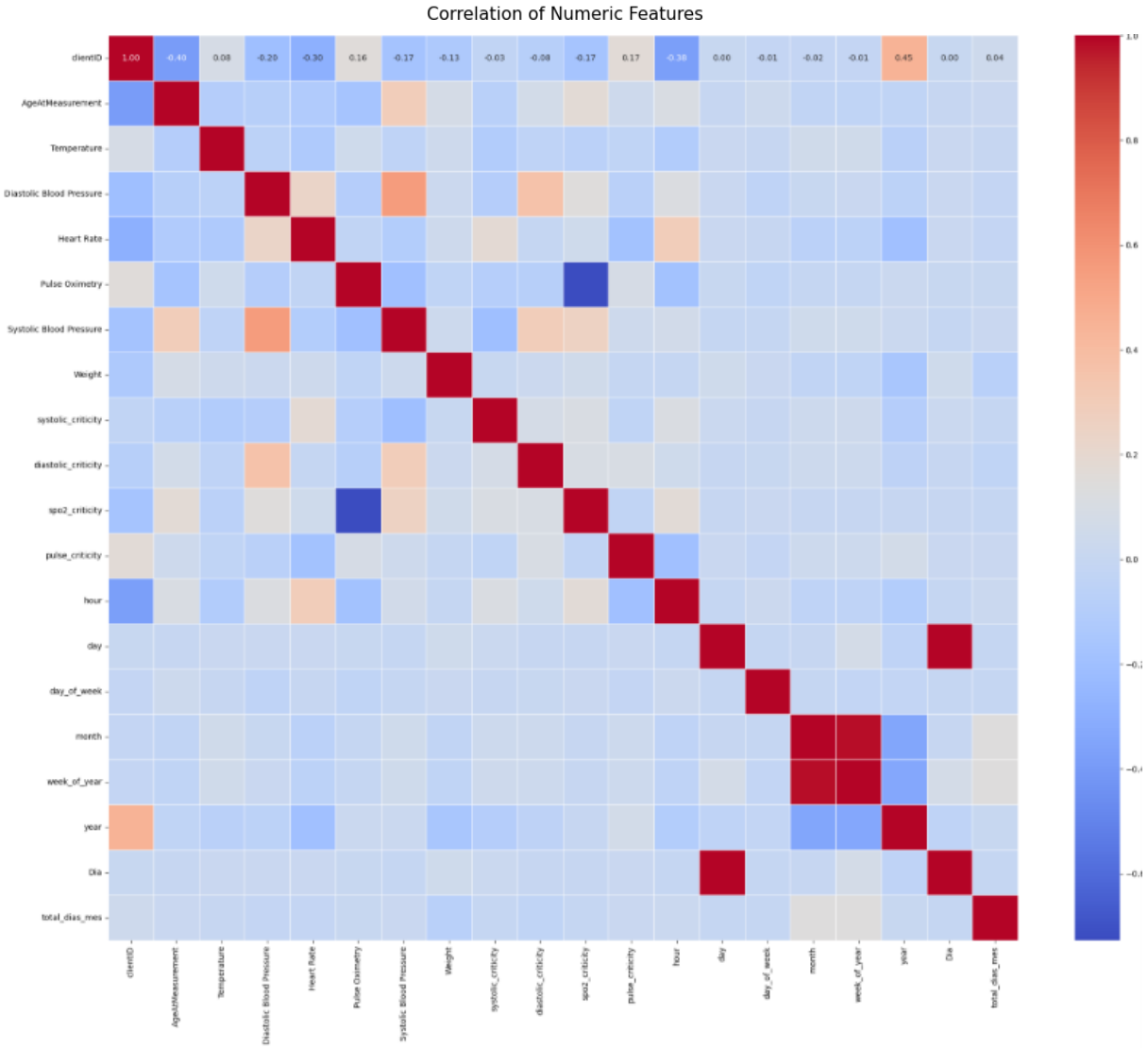


Figura 13 – Matriz de Correlações

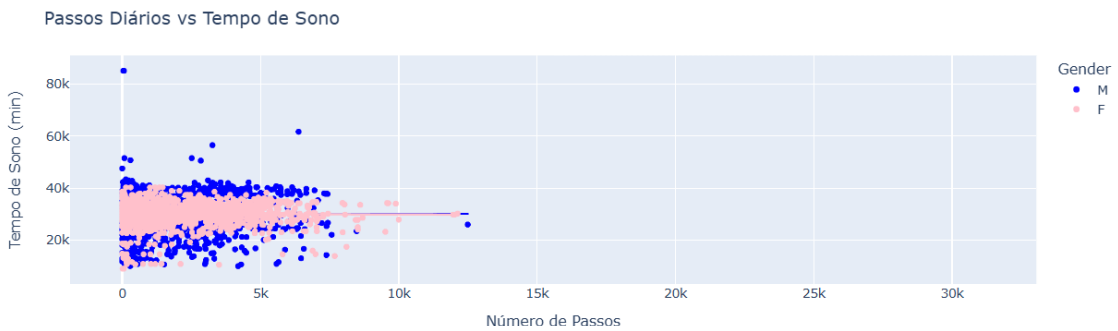


Figura 14 –Correlação entre steps e sleep_duration

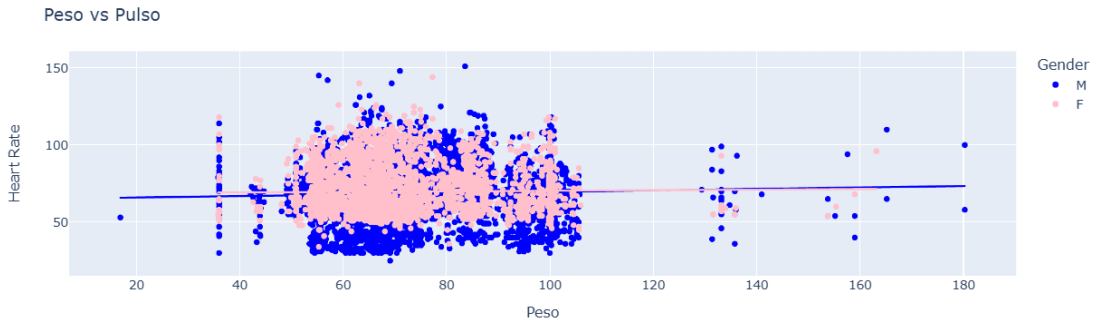


Figura 15 –Correlação entre weight e pulse

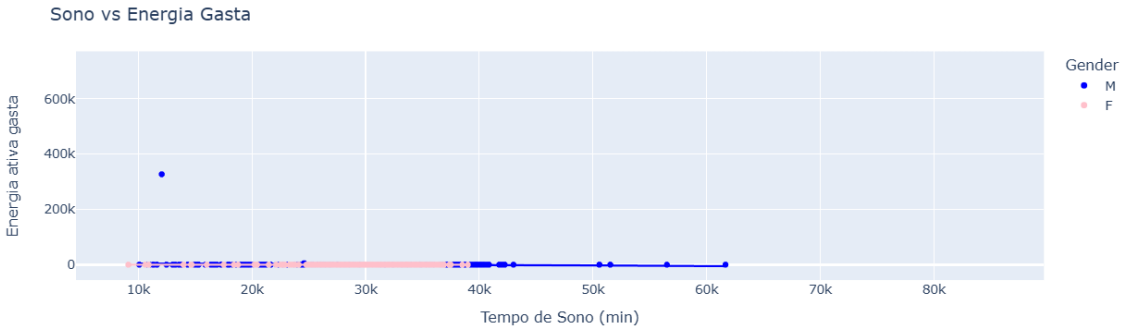


Figura 16 –Correlação entre sleep_duration e energy

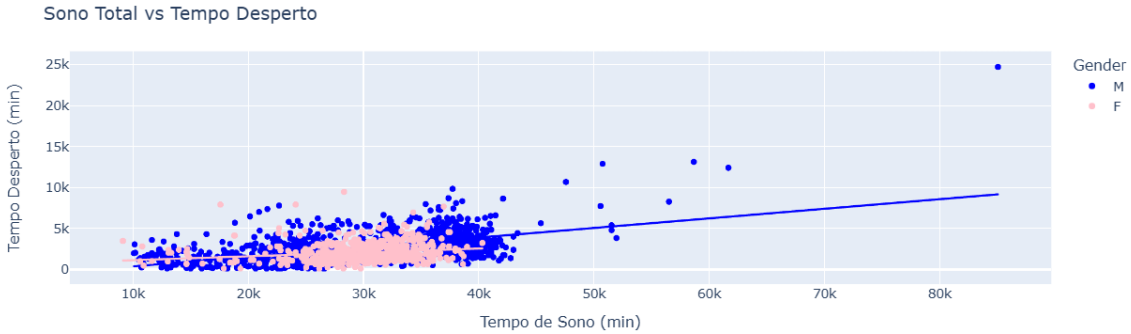


Figura 17–Correlação entre total_sleep e wake_time

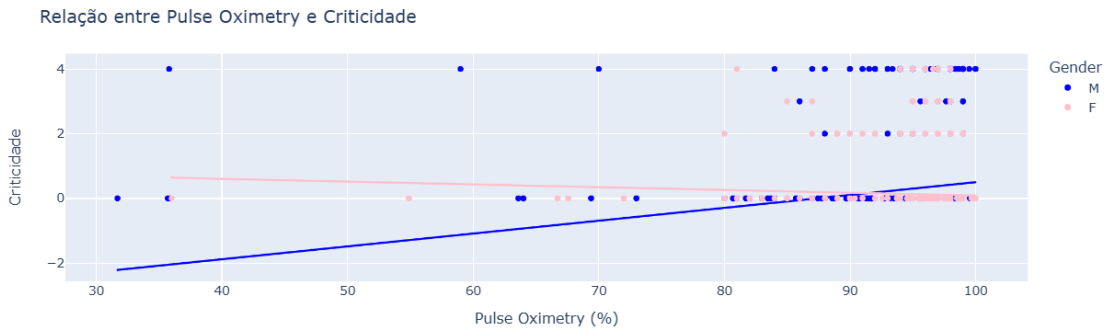


Figura 18–Correlação entre SpO_2 e pulse_criticity

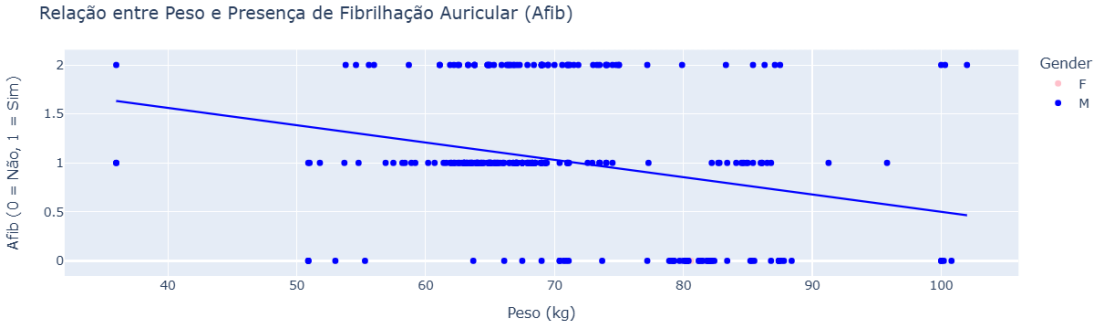


Figura 19–Correlação entre active_steps e sleep_duration

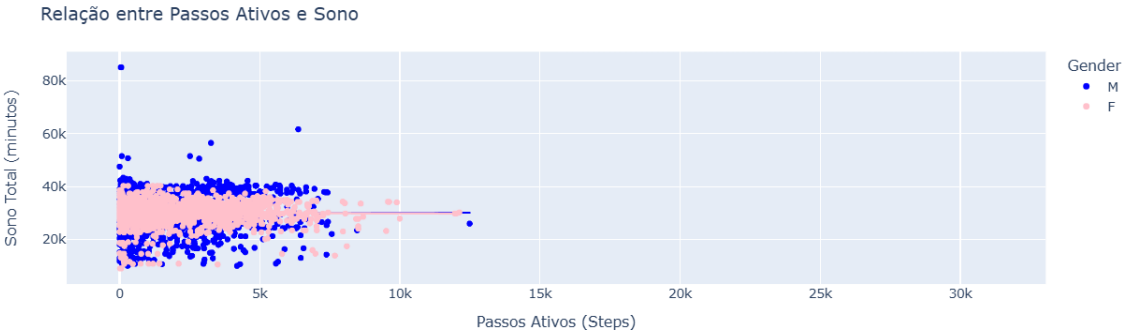


Figura 20–Correlação entre weight e atrial_fibrillation



Figura 21– Distribuição da variável pulse_criticity



Figura 22– Correlação entre as variáveis pulse_criticity e age

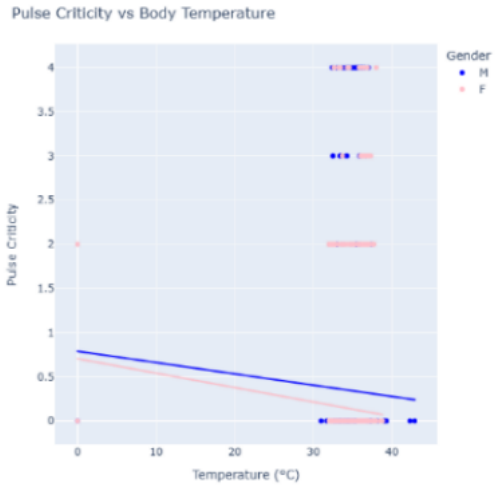


Figura 23– Correlação entre as variáveis pulse_criticity e body_temperature

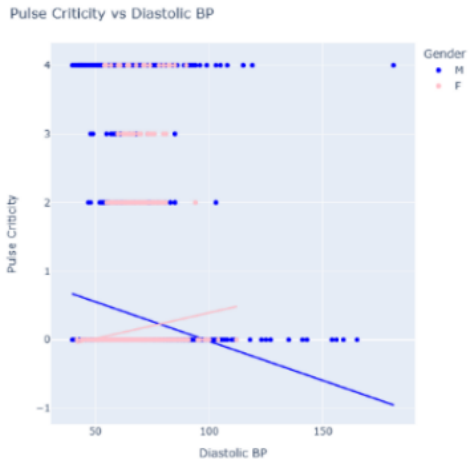


Figura 24– Correlação entre as variáveis pulse_criticity e diastolic

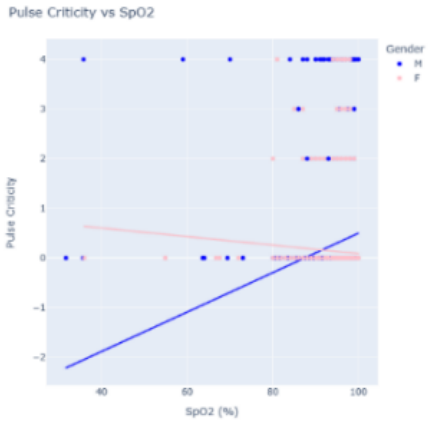


Figura 25– Correlação entre as variáveis pulse_criticity e SpO_2

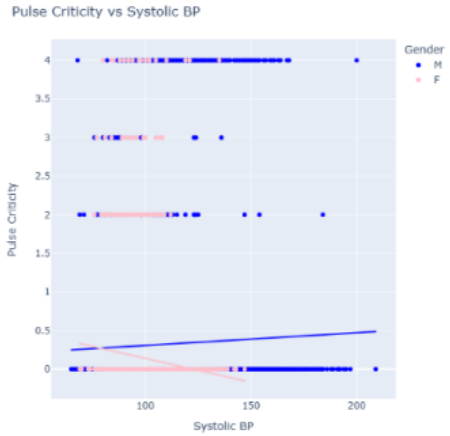


Figura 26– Correlação entre as variáveis pulse_criticity e systolic

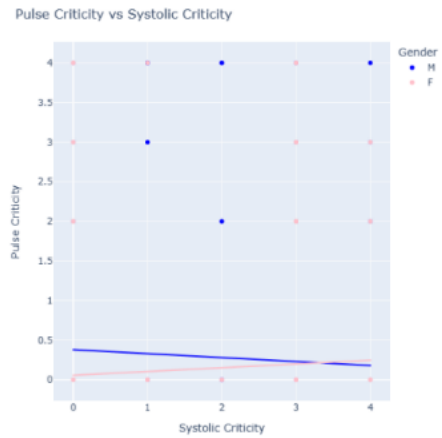


Figura 27– Correlação entre as variáveis pulse_criticity e systolic_criticity

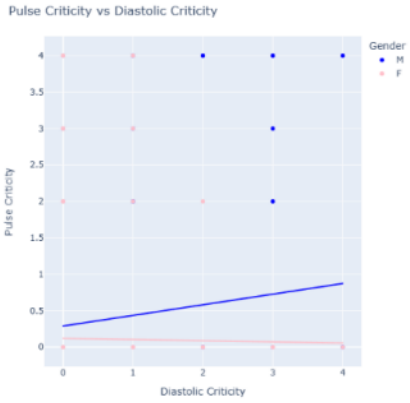


Figura 28– Correlação entre as variáveis pulse_criticity e diastolic_criticity

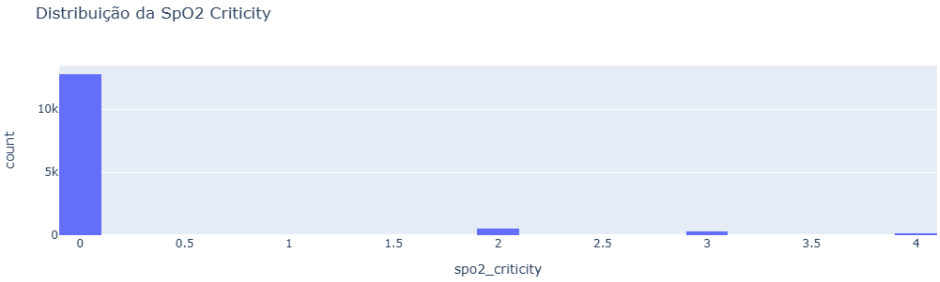


Figura 29– Distribuição da variável SpO_2_criticity

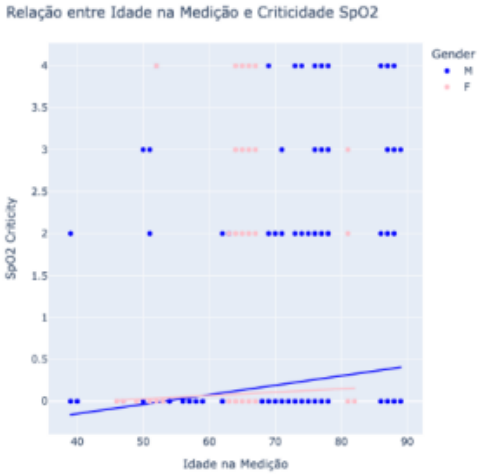


Figura 30– Correlação entre as variáveis SpO_2_criticity e age

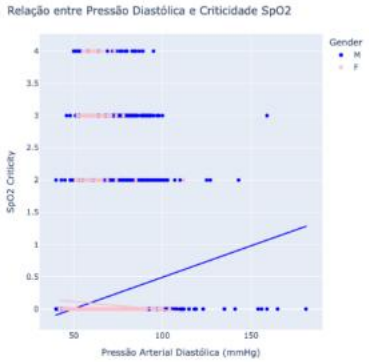


Figura 31– Correlação entre as variáveis SpO_2_criticity e body_temperature

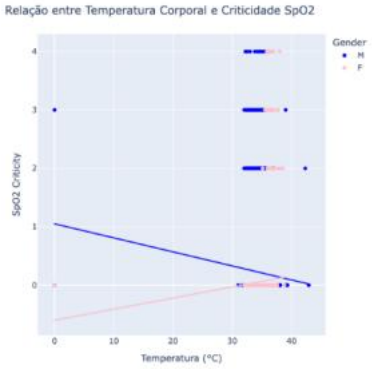


Figura 32– Correlação entre as variáveis SpO_2_criticity e diastolic

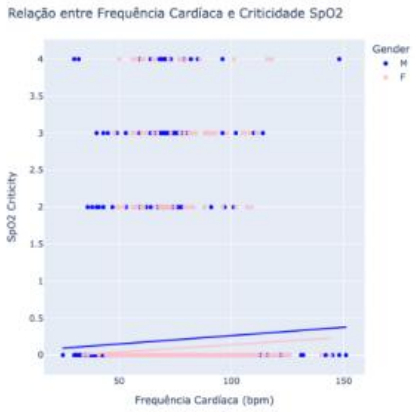


Figura 33– Correlação entre as variáveis SpO_2_criticity e pulse

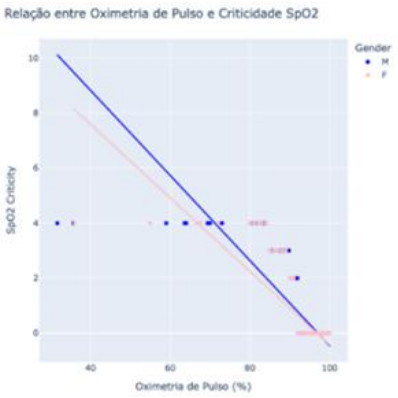


Figura 34– Correlação entre as variáveis SpO_2_criticity e SpO_2

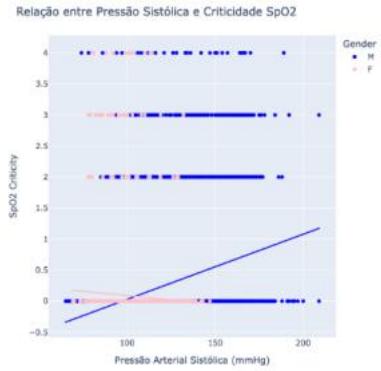


Figura 35– Correlação entre as variáveis SpO_2_criticity e systolic

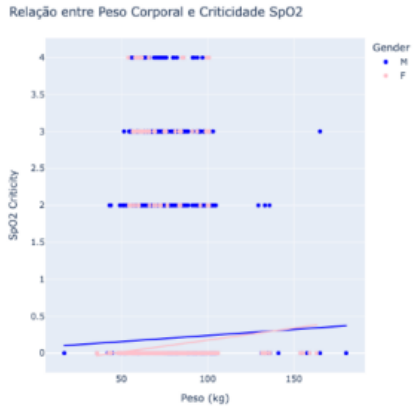


Figura 36– Correlação entre as variáveis SpO_2_criticity e weight

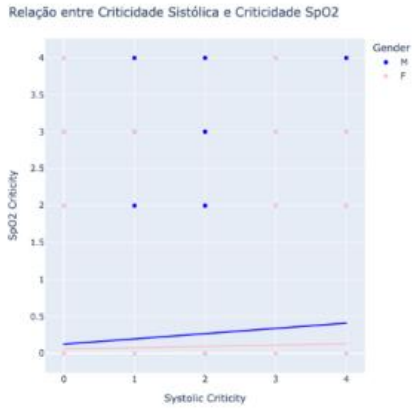


Figura 37– Correlação entre as variáveis SpO_2_criticity e systolic_criticity

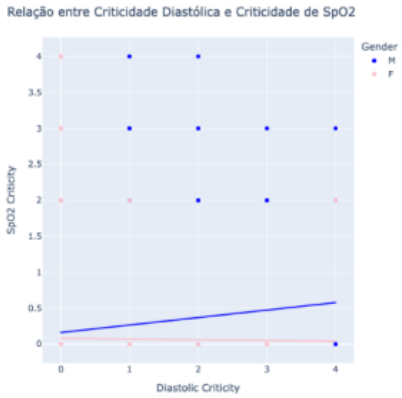


Figura 38– Correlação entre as variáveis SpO_2_criticity e diastolic_criticity

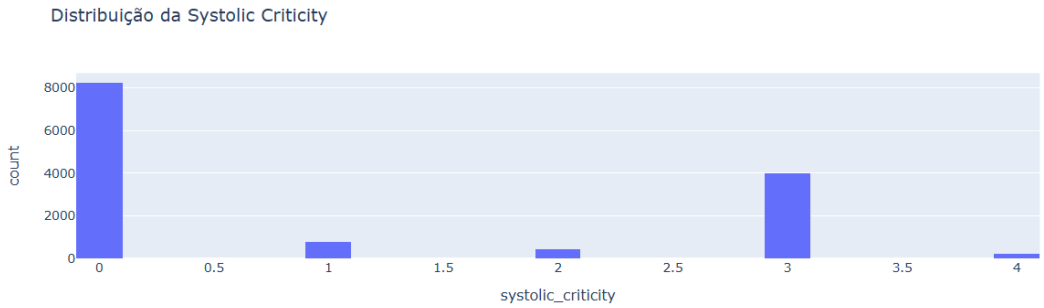


Figura 39– Distribuição da systolic_criticity

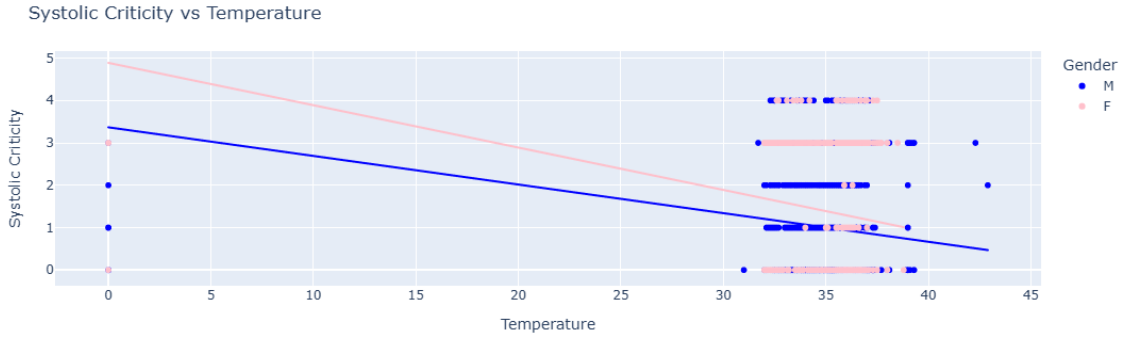


Figura 40– Correlação entre a systolic_criticity e body_temperature

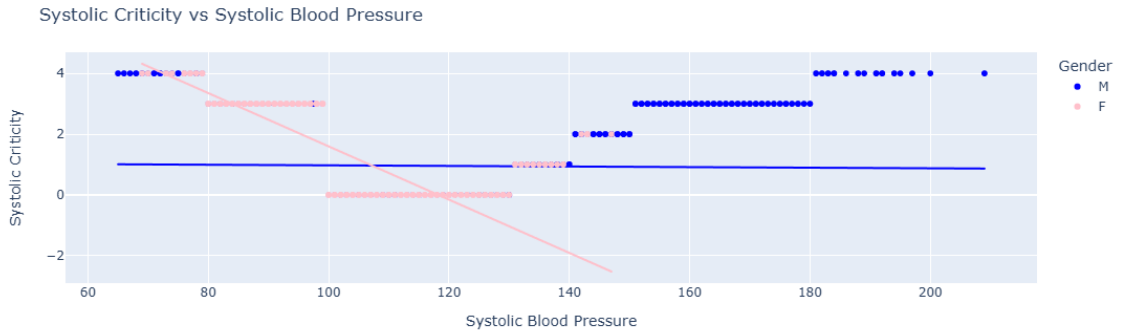


Figura 41– Correlação entre a systolic_criticity e blood_pressure

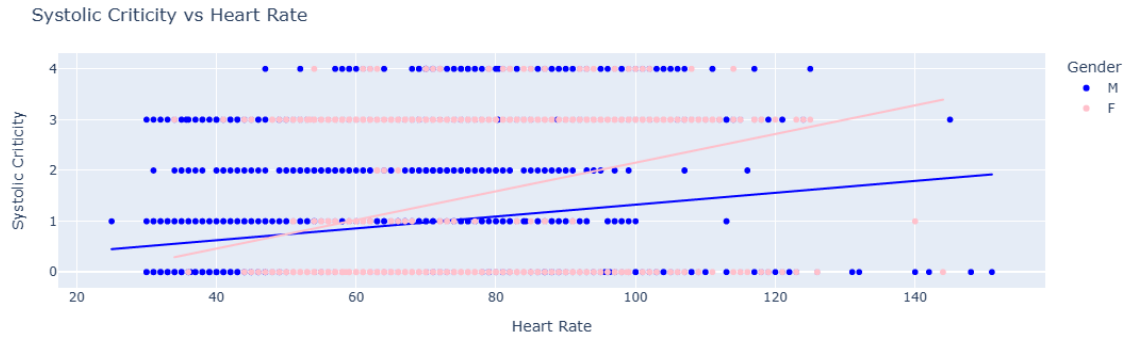


Figura 42– Correlação entre a systolic_criticity e heart_rate

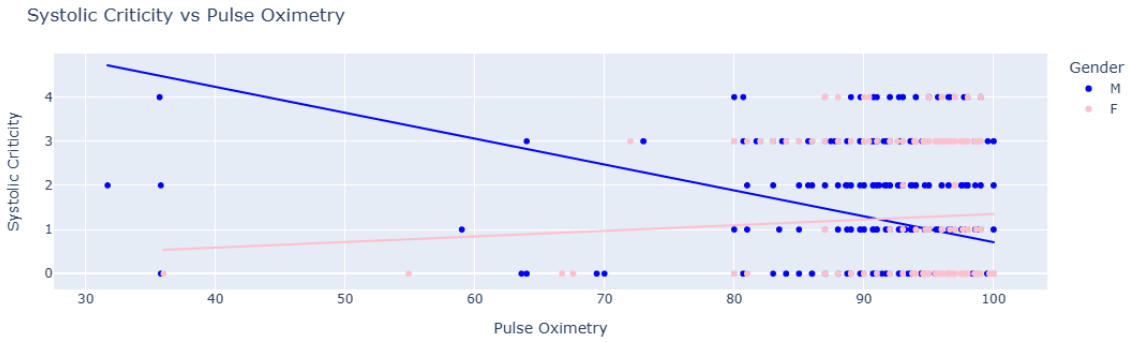


Figura 43– Correlação entre a systolic_criticity e SpO_2

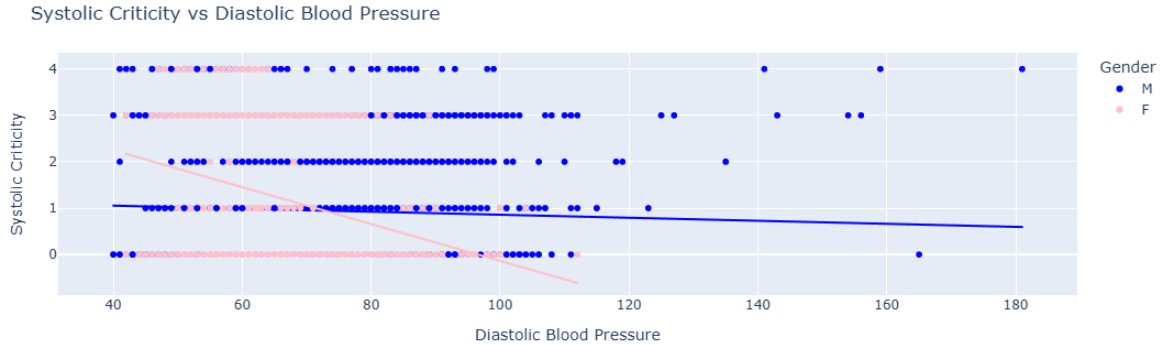


Figura 44– Correlação entre a systolic_criticity e blood_pressure

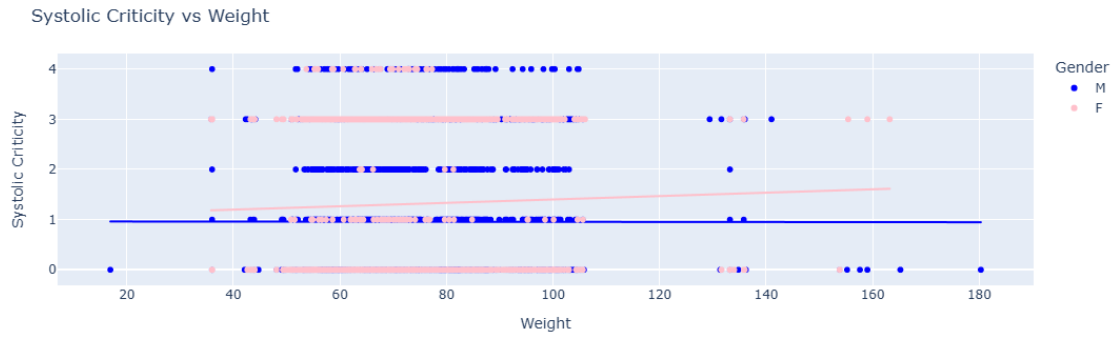


Figura 45– Correlação entre a systolic_criticity e weight

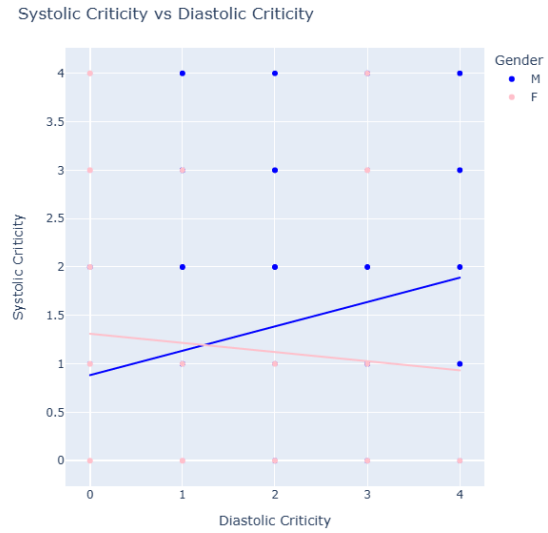


Figura 46– Correlação entre a systolic_criticity e diastolic_criticity



Figura 47– Correlação entre a systolic_criticity e SpO_2_criticity

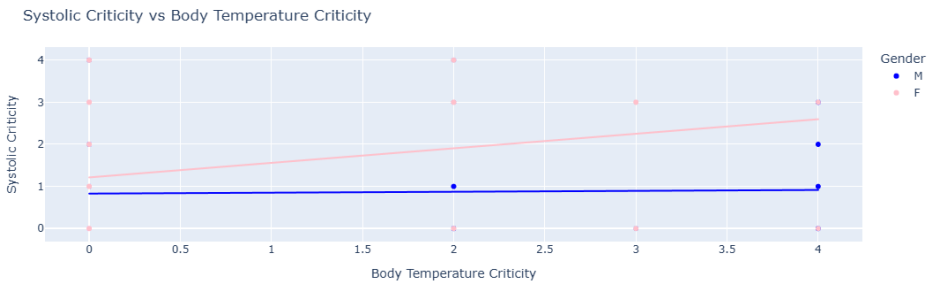


Figura 48– Correlação entre a systolic_criticity e body_temperature_criticity

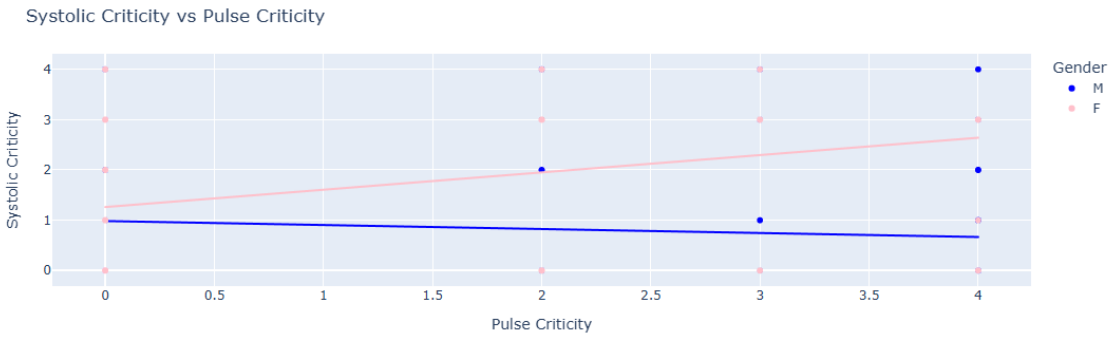


Figura 49– Correlação entre a systolic_criticity e pulse_criticity