isec
Engenharia

MESTRADO EM ENGENHARIA INFORMÁTICA

Interpretabilidade em Modelos de Avaliação de Risco Cardiovascular

Autor

Maria Inês Dias Figueiredo

Orientador

Professor Simão Paredes

INSTITUTO POLITÉCNICO
DE COIMBRA

INSTITUTO SUPERIOR
DE ENGENHARIA
DE COIMBRA

Coimbra, maio de 2022

isec
Engenharia

DEPARTAMENTO DE INFORMÁTICA E SISTEMAS

Interpretabilidade em Modelos de Avaliação do Risco Cardiovascular

Relatório de Trabalho de Projeto para a obtenção do grau de Mestre em Engenharia Informática

Especialização em Desenvolvimento de Software

Autor

Maria Inês Dias Figueiredo

Orientador

Professor Simão Paredes

Co-Orientador

Professor Jorge Henriques

INSTITUTO POLITÉCNICO
DE COIMBRA

INSTITUTO SUPERIOR
DE ENGENHARIA
DE COIMBRA

Coimbra, maio de 2022

AGRADECIMENTOS

A realização do presente trabalho não teria sido possível sem a colaboração de várias pessoas, ao longo de todos os meses em que foi desenvolvido. Manifesto aqui o meu agradecimento a todos aqueles que, direta ou indiretamente, contribuíram para a conclusão desta tese.

Aos Professores Simão Paredes e Jorge Henriques, meus orientadores, agradeço a orientação, empenho e paciência com que sempre me acompanharam ao longo do trabalho. Desde o início que sempre me tentaram mostrar o melhor caminho para alcançar os objetivos do trabalho, tendo estado sempre disponíveis para me ajudar e apoiar, sem nunca me desmotivar.

A todos os docentes e não docentes do Instituto Superior de Engenharia de Coimbra, que fizeram parte do meu percurso académico, tanto na licenciatura como no mestrado, agradeço por todos os ensinamentos e contributos para o meu crescimento pessoal.

A todos os meus colegas da NFON, agradeço pelo apoio e motivação que me deram ao longos dos últimos meses, incentivando-me sempre a continuar este trabalho.

Aos meus amigos, em especial à Rita, por serem sempre um refúgio e me ajudarem em todas as situações.

Agradeço à minha família, ao meu pai, irmão, avô e, em especial, à minha mãe que sempre me apoiou e deu força para nunca desistir e conseguir terminar este trabalho. Um agradecimento especial à minha avó Amélia que, mesmo não estando comigo fisicamente, me dá uma grande força para continuar e ser melhor a cada dia.

Ao David, por toda a paciência e compreensão que sempre demonstrou para comigo desde o primeiro dia da realização desta tese.

A todos, o meu sincero obrigada.

RESUMO

Segundo a *World Health Organization*, as doenças cardiovasculares representam uma das principais causas de morte a nível mundial. Para tentar contrariar esta tendência, as medidas passam bastante pela prevenção. Neste campo, a análise de dados clínicos tendo por base vários modelos *Machine learning* de previsão de risco tem vindo a ser implementada, com o objetivo de prever o risco de ocorrência de um determinado evento (doença, morte, etc.). No entanto, os modelos apresentam algumas fragilidades, entre as quais é possível destacar a falta de interpretabilidade, pois alguns são modelos *black-box*. Estes, apesar dos elevados desempenhos que apresentam, nem sempre conseguem explicar as suas previsões ao utilizador final. No entanto, em áreas tão críticas como a da saúde, o aumento da interpretabilidade é fundamental para criar a confiança necessária nos clínicos e permitir a efetiva incorporação deste tipo de ferramentas na prática clínica. Desta forma, têm tentado desenvolver-se modelos cada vez mais *white-box* que, ao fornecerem explicações sobre os resultados que apresentam, transmitem essa confiança aos utilizadores.

O objetivo principal do trabalho é analisar modelos *Machine learning*, de entre os selecionados, e o modelo GRACE, para perceber qual consegue transmitir o melhor compromisso entre interpretabilidade e performance. Escolheu-se o modelo GRACE por ser, atualmente, um modelo de previsão de risco bastante utilizado por profissionais da área clínica para avaliar pacientes com doenças cardíacas agudas. Os algoritmos escolhidos são Árvore de decisão, *Naïve Bayes* e *Clustering*. A Árvore de decisão é aplicada visto ser bastante interpretável e permitir facilmente a extração de regras. O modelo Bayesiano permite calcular relações probabilísticas entre classes, neste caso entre sintomas e possibilidade de morte ou sobrevivência. Por último o *Clustering*, dado que permite agrupar pacientes com caraterísticas semelhantes. Os modelos referidos são executados com um *dataset* composto por dados clínicos de pacientes dos Hospitais de Santa Cruz e de Leiria. Para cada um dos modelos são calculadas métricas demonstrativas da respetiva performance. Simultaneamente, são extraídas regras a partir de cada um deles e tenta-se verificar a veracidade das mesmas com recurso a dados reais.

Por fim, todos os modelos são comparados em termos das performances obtidas, tendo-se concluído que o GRACE é o que consegue obter o melhor compromisso entre a performance e a interpretabilidade que fornece. Isto acaba por revelar um equilíbrio com a realidade, visto que este é um modelo bastante aplicado na prática clínica.

Palavras-Chave: Doenças Cardiovasculares, *Machine learning*, GRACE, Árvore de Decisão, Modelo Bayesiano, *Clustering*, Interpretabilidade, Performance

ABSTRACT

According to the World Health Organization, cardiovascular diseases represent one of the leading causes of death worldwide. To try to counteract this situation, the measures are largely based on prevention. In this field, clinical data analysis based on several Machine learning models for risk prediction have been implemented, with the objective of predicting the risk of occurrence of a given event (disease, death, etc.). However, the models have some weaknesses, among which is the lack of interpretability, since some of them are black-box models. These models, despite their high performance, are not always able to explain their predictions to the user. But, in critical areas such as health, it is not acceptable to present clinicians with a model that is not interpretable. Increased interpretability is essential to create the necessary confidence in clinicians and thus allow the effective incorporation of this type of tool in clinical practice. Thus, attempts have been made to develop increasingly white-box models that, by providing explanations of the results they present, convey confidence to users.

The main objective of this work is to analyze machine learning models, among those selected, and the GRACE model, to understand which one can convey the best compromise between interpretability and performance. The GRACE model was chosen because it is currently a risk prediction model widely used by clinical professionals to evaluate patients with acute heart disease. The algorithms chosen are Decision Tree, Naïve Bayes, and Clustering. Decision Tree is applied because it is very interpretable and allows easy extraction of rules. The Bayesian model allows to calculate probabilistic relationships between classes, in this case between symptoms and possibility of death or survival. Finally, Clustering, since it allows grouping patients with similar characteristics. The models mentioned above are run with a dataset composed of clinical data from patients at the Santa Cruz and Leiria Hospitals. For each of the models, metrics demonstrating their respective performance are calculated. Simultaneously, rules are extracted from each of them, and their veracity is verified using real data.

Finally, all models are compared in terms of their performance, and it is concluded that GRACE is the one that achieves the best compromise between performance and the interpretability it provides. This ultimately reveals a balance with reality, since this is a model that is widely applied in clinical practice.

Keywords: Cardiovascular Diseases, Machine Learning, GRACE, Decision Tree, Bayesian Model, Clustering, Interpretability, Performance

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

SIMBOLOGIA E ABREVIATURAS

IA	Inteligência Artificial
AVC	Acidente Vascular Cerebral
CHF	*Congestive heart failure*
CVD	Doença Cardiovascular
DAC	Doença Arterial Coronariana
AD	Árvore de decisão
ECG	Eletrocardiograma
FP	*False Positive*
FN	*False Negative*
GRACE	*Global Registry of Acute Coronary Events*
MCI	*Mild Cognitive Impairment*
MI	*Myocardial Infarction*
ML	*Machine Learning*
NB	*Naïve Bayes*
NSTEMI	*Non-ST-elevation myocardial infarction*
SE	Sensibilidade
SMO	*Sequential Minimal Optimization*
SMOTE	*Synthetic Minority Oversampling Technique*
SP	Especificidade
STEMI	*ST-elevation myocardial infarction*
SVM	*Support vector machine*
PURSUIT	*Platelet glycoprotein IIb/IIIa in Unstable Angina*
TIMI	*The thrombolysis In Myocardial Infarction*
TP	*True Positive*
TN	*True Negative*
UA	*Unstable Angina*

CAPÍTULO 1

INTRODUÇÃO

Atualmente, uma das principais causas de morte a nível mundial são as doenças cardiovasculares (CVD). Estas doenças caraterizam-se como um grupo de patologias no coração e nos vasos sanguíneos. Incluem algumas anomalias como ataques cardíacos, que ocorrem quando o fluxo de sangue para uma parte do coração é bloqueado por um coágulo, acidentes vasculares cerebrais (AVC) que ocorrem quando um vaso sanguíneo que alimenta o cérebro fica bloqueado, também por um coágulo, e Arritmias caraterizadas por um ritmo cardíaco anormal [1]. A razão mais comum para o acontecimento destas doenças é a acumulação de depósitos de gordura nas paredes internas dos vasos sanguíneos que abastecem estes órgãos. Os principais fatores de risco estão relacionados com a prática de dietas pouco saudáveis, inatividade física, pobreza, fatores hereditários, consumo de tabaco e uso nocivo de álcool. Como forma de reduzir a incidência destas doenças a nível mundial, a prevenção tem sido o principal foco.

Existem três tipos de mecanismos de prevenção que podem reduzir os impactos da doença. A prevenção primária refere-se às medidas tomadas por um indivíduo para prevenir o aparecimento da doença. Estas medidas passam por manter um estilo de vida saudável, equilibrado por exercício físico e uma dieta saudável. A prevenção secundária foca-se na redução do impacto da doença através de um diagnóstico precoce antes do aparecimento de qualquer dano crítico e permanente, evitando situações de risco de vida e de incapacidade devido a uma doença cardíaca. A prevenção terciária é aplicada depois da manifestação da doença e dos efeitos a longo prazo que esta implica, ajudando os pacientes a gerir a sua condição, aumentando a esperança e a qualidade de vida [2].

O passo mais crítico da prevenção secundária é o diagnóstico precoce, que permite aos profissionais de saúde prestar os cuidados necessários aos pacientes e melhorar a sua qualidade de vida. Isto requer a identificação dos fatores de riscos e a forma como a variação desses fatores está relacionada com o aparecimento da doença. Em caso de um diagnóstico precoce, os pacientes poderão ser encaminhados para tratamentos que garantam uma qualidade de vida mais elevada. A principal razão para apostar mais na prevenção secundária provém de dois fatores. O primeiro está relacionado com o custo, uma vez que o custo da prevenção secundária é muito menor em relação à terciária. E em segundo lugar, tem efeitos sobre a qualidade de vida dos pacientes. A prevenção terciária envolve procedimentos mais complexos que podem causar desconforto ao paciente, bem como perturbar as suas atividades diárias, enquanto a prevenção secundária se concentra em tratamentos menos intensos que podem incluir alguns medicamentos e mudança de estilo de vida [3].

Por outro lado, a constante aplicação da tecnologia à área da medicina tem permitido aos profissionais de saúde tratar os seus pacientes de forma mais eficaz. Com o decorrer do tempo, muitos dados de investigação e registos de doentes provenientes dos hospitais estão agora disponíveis. Os constantes avanços na inteligência artificial e na extração de conhecimento a partir de dados, têm permitido aos profissionais de saúde o alargamento das suas capacidades desde o diagnóstico até à deteção da própria doença. Uma análise completa dos dados de um paciente pode ser facilmente realizada através de modelos de *Machine learning* (ML). Desta forma, muitos estudos têm sido realizados e vários modelos de ML são utilizados para fazer a classificação e a previsão para o diagnóstico de doenças cardíacas [4].

Mas, em domínios mais críticos como é o da saúde, é necessário que os sistemas ML sejam capazes de responder a outros critérios como a segurança, não discriminação e a explicação do modelo para uma tomada de decisão. Em relação à segurança, para se compreender melhor, dá-se um exemplo conhecido da *IBM Watson for Oncology.* Esta empresa utiliza algoritmos de AI (Inteligência Artificial) para avaliar as informações dos registos médicos dos pacientes e auxiliar médicos a explorar opções de tratamento de cancro para novos pacientes. No entanto, foi bastante criticada por dar recomendações inseguras para tratamentos. O problema ocorreu devido ao facto de o algoritmo usar apenas dados gerados de outros pacientes. É compreensível que a obtenção de dados de pacientes reais nem sempre é fácil, mas é imprescindível para transmitir a segurança necessária aos utilizadores dos modelos. Desta forma, é fundamental ser transparente no desenvolvimento dos algoritmos, quanto ao tipo de dados usados e em relação a quaisquer falhas do *software*, porque só assim se consegue obter a confiança dos clínicos e dos pacientes.

Além da segurança, surge também um novo paradigma nestes modelos que reforça a importância do conceito de interpretabilidade. Este conceito pretende-se com o facto de que as abordagens ML têm de ser simples de interpretar e ao mesmo manter um bom nível de desempenho que permita aos utilizadores entender e ao mesmo tempo confiar nos modelos apresentados. Este trabalho explora alguns dos modelos atualmente disponíveis, tentando alcançar uma boa relação entre a performance e a interpretabilidade, através da extração de regras explicáveis.

1.1 Contexto do problema

A avaliação da condição de um paciente pode ser uma tarefa desafiadora na medida em que existem algumas situações complexas em que a decisão do profissional de saúde não é simples. A quantidade de informação é muita e o tempo para obter o diagnóstico é relativamente limitado. Isto faz com que os profissionais de saúde precisem de ajuda para chegar a um diagnóstico e posteriormente identificar o melhor tratamento a aplicar a cada paciente.

Com recurso à tecnologia, existem muitas oportunidades para desenvolver modelos que possam ajudar os médicos no diagnóstico, tratamento ou mesmo na prevenção

de doenças, com o objetivo de reduzir a sobrecarga verificada nos serviços de saúde. Neste sentido, nos últimos anos tem vindo a surgir o impulso de desenvolver soluções de saúde com recurso a modelos de *Machine learning, o*s quais permitem a representação de conhecimento com base em conjuntos de dados previamente tratados.

As ferramentas que existem atualmente para ajudar a prever o risco são importantes para a avaliação do risco Cardiovascular e disponibilizam aos clínicos uma informação adicional para otimização dos tratamentos a aplicar aos seus pacientes. Um exemplo da aplicação da inteligência artificial à área médica é um modelo desenvolvido pela Google que é capaz de detetar cancro da mama em mamografias, com uma melhor precisão e menos falsos positivos do que os humanos. Alguns investigadores avaliaram o modelo com um conjunto de mamografias de mais de 3000 mulheres dos Estados Unidos da América (EUA) e 25000 do Reino Unido (UK) e verificaram uma redução de 5.7% na deteção de falsos positivos nos EUA e de 1.2% no Reino Unido [5].

No entanto, tem vindo a verificar-se que estas ferramentas ainda apresentam algumas limitações. Em alguns casos, estes modelos fornecem apenas resultados, sem justificar devidamente as conclusões. Por isso, um modelo ML deverá ser interpretável, para que se consigam extrair explicações confiáveis acerca dos resultados e ganhar a confiança dos profissionais de saúde. A interpretabilidade é definida como a capacidade de explicar ou apresentar de forma compreensível os resultados obtidos pelos modelos. Por essa razão, quanto maior a interpretabilidade de um modelo, mais fácil será para alguém compreender o porquê de certas decisões ou previsões serem feitas e, no caso médico, mais facilmente se consegue captar a confiança dos clínicos.

Com o presente trabalho, o que se pretende é extrair interpretabilidade através da análise da influência de dados de pacientes nos modelos ML, de modo a melhorar a avaliação do risco cardiovascular e aumentar a sua interpretabilidade com o objetivo de potenciar a utilização destas ferramentas na área clínica.

1.2 Objetivos

Como foi mencionado anteriormente, um dos objetivos deste trabalho é explorar alguns dos modelos *Machine learning* atualmente disponíveis, tendo em consideração a respetiva interpretabilidade. Numa fase inicial o objetivo é analisar o conjunto de dados utilizado para o trabalho e aplicar-lhe o tratamento necessário.

Posteriormente, pretende efetuar-se uma análise comparativa da interpretabilidade de um conjunto de modelos de ML, de acordo com três perspetivas diferentes, baseadas em:

- Regras
- Probabilidades
- Distâncias

Para explorar a interpretabilidade através de regras, recorre-se a Árvores de decisão, por fornecerem uma estrutura em árvore da qual são facilmente extraídas regras. Por sua vez, para o estudo com base em probabilidades, utiliza-se o *Naïve Bayes*, uma vez que analisa as probabilidades de ocorrência de doença na presença de sintomas, e para a análise com base em distâncias é aplicado o *Clustering,* porque agrupa pacientes de acordo com as suas semelhanças. Para além dos modelos referidos anteriormente, explora-se também o GRACE, um modelo desenvolvido em específico para a análise do risco cardiovascular. Os resultados obtidos por cada um dos modelos são também analisados com base em algumas métricas de performance, que vão ser exploradas mais à frente no documento.

Desta forma, como objetivo final do trabalho pretende-se obter um conjunto de condições, probabilísticas ou regras, para cada modelo analisado, que devem corresponder o mais possível à realidade, tentando garantir um compromisso entre a interpretabilidade dos resultados e ao mesmo tempo um bom desempenho do modelo.

1.3 Desafios

Com a aplicação de algoritmos *Machine learning,* tem tentado chegar-se a um *trade-off* entre a interpretabilidade e performance dos modelos preditivos. Isto é, modelos mais interpretáveis como modelos de regressão e árvores de decisão oferecem, regra geral, piores performances quando comparados com outros modelos menos interpretáveis como as redes neuronais, por exemplo. Na área médica, a falta de interpretabilidade pode representar consequências muito adversas nas decisões tomadas para o tratamento de um doente. Segundo um estudo realizado por Caruana [6], o qual tenta identificar doentes com pneumonia como sendo de alto ou baixo risco hospitalar, uma rede neuronal não consegue assegurar interpretabilidade, mas revelou ser o melhor classificador em termos de performance para este problema. O modelo previu que, um doente com asma, corre um risco mais baixo de morte hospitalar quando admitido por pneumonia. Mas na realidade, é o contrário que se verifica, ou seja, os doentes com asma correm um maior risco de sofrer complicações graves e sequelas devido a uma doença pulmonar infeciosa como a pneumonia. Este exemplo revela que, apesar de um modelo ter bons resultados, isso nem sempre significa que dá as previsões mais corretas, porque o modelo pode também errar nas previsões que faz [7]. Consegue perceber-se assim porque é que dar uma explicação de um modelo a quem o vai utilizar é tão importante. Quando o modelo é explicado, os clínicos podem questionar e perceber o que está por detrás de uma previsão. Caso contrário, torna-se muito difícil confiar no modelo.

Como tal, o desafio deste trabalho é alcançar o compromisso entre os modelos explorados e a interpretabilidade que cada um deles pode fornecer e, como tal, ser aplicada no âmbito do risco cardiovascular.

1.4 Tarefas realizadas

O projeto foi desenvolvido ao longo de aproximadamente 12 meses, podendo ser identificadas as tarefas principais:

- Tarefa 1 – Levantamento do estado da arte e compreensão das temáticas em estudo.
- Tarefa 2 – Análise dos dados utilizados no estudo e aplicação do tratamento necessário aos mesmos.
- Tarefa 3 – Seleção de classificadores *Machine learning* a utilizar no projeto e início da exploração dos mesmos.
- Tarefa 4 – Realização de testes com cada um dos modelos, com o objetivo de perceber em que condições cada um deles obtém melhores valores de performance.
- Tarefa 5 – Extração de intepretabilidade de cada um dos algoritmos.
- Tarefa 6 – Comparação de resultados entre algoritmos, como objetivo de perceber qual consegue alcançar a melhor relação entre a interpretabilidade e a performance.
- Tarefa 7 – Escrita da dissertação.

A Figura 1, apresenta um cronograma com a organização das tarefas anteriores ao longo do período de desenvolvimento do projeto.

	2020		2021											
	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
T1														
T2														
T3														
T4														
T5														
T6														
T7														

Figura 1 - Cronograma com as tarefas realizadas.

1.5 Estrutura do documento

O documento é constituído por 6 capítulos, sendo os mesmos descritos da seguinte forma:

O Capítulo 2 explora o conceito de interpretabilidade, numa visão mais aprofundada. É clarificada a definição de interpretabilidade e o compromisso que deve existir entre esta e o desempenho do modelo. No final deste Capítulo, são dados alguns exemplos de trabalhos desenvolvidos na área em estudo neste trabalho.

O Capítulo 3 aborda os modelos de classificação explorados no âmbito do trabalho, nomeadamente GRACE, Árvore de decisão, *Naïve Bayes* e *Clustering*. Neste capítulo são abordadas também as técnicas de validação dos modelos, mais concretamente as métricas.

O Capítulo 4 apresenta o *dataset* utilizado para o estudo e todas as técnicas de pré-processamento que lhe foram aplicadas, desde as estratégias de balanceamento, à seleção e descrição das *features* a utilizar pelos modelos e a eliminação dos valores em falta.

No Capítulo 5 é dada uma definição inicial da metodologia utilizada para o treino dos algoritmos. Posteriormente, são apresentados os resultados obtidos para os modelos: GRACE, Árvores de decisão, *Naïve Bayes* e *Clustering*. No final deste capítulo, é realizada uma comparação entre todos os modelos e a indicação de qual seria o ideal a aplicar na previsão do risco cardiovascular, tendo em consideração as respetivas performance e interpretabilidade.

No Capítulo 6 são apresentadas algumas conclusões relativas ao trabalho e uma retrospetiva geral do mesmo.

CAPÍTULO 2

INTERPRETABILIDADE

Os sistemas desenvolvidos em *Machine learning* têm vindo a ganhar cada vez mais relevo nas mais diversas áreas da sociedade. Nos cuidados de saúde, os desafios são únicos porque um modelo tem de conseguir aliar interpretabilidade, confiança e desempenho, uma vez que qualquer decisão baseada nestes modelos pode ter uma implicação imediata no bem-estar ou mesmo na vida de uma pessoa. Por tudo isso, um modelo tem de ser interpretável para permitir aos utilizadores entender os resultados das previsões. Como tal, a interpretabilidade é o foco principal deste capítulo, no qual serão apresentadas diferentes interpretações para este conceito e é também apresentado algum do trabalho já desenvolvido nesta área.

2.1 Definição de Interpretabilidade

Apesar do reconhecimento que já é dado ao ML nos cuidados de saúde, ainda existem impedimentos para uma maior adoção em ambientes reais. Neste tipo de ambientes, que incluem a tomada de decisão clínica, existe bastante hesitação na utilização destes modelos, uma vez que o custo de uma classificação errada é demasiado elevado. Por essa razão, para que um modelo possa ser aceite e utilizado por um profissional de saúde, este deve conseguir perceber quais as razões que levam a uma previsão e se os resultados do modelo são realmente fiáveis. Aqui entra-se na questão da interpretabilidade de um modelo. Mas, se por um lado existe um consenso geral de que a interpretabilidade é necessária nos modelos *Machine learning*, por outro lado, existe muito desacordo em relação ao que constitui a interpretabilidade. Segundo Lipton [3], os investigadores têm tentado encontrar inúmeras noções e definições para o conceito. Já foi definida como transparência, confiança e compreensão do modelo. Mas, uma objeção comum a este tipo de definições é que não é colocada a enfâse necessária no utilizador final destes sistemas de ML, ou seja, os modelos e explicações produzidos pelos modelos não facilitam as necessidades dos utilizadores finais. Existe uma grande dificuldade por parte destes, neste caso os clínicos, em perceber as explicações dadas pelo modelo. Para colmatar esta dificuldade dos utilizadores, o sentimento primário de interpretabilidade deve ser a explicação do modelo, isto é, o modelo deve fornecer uma justificação do porquê de uma previsão ou de uma recomendação. Isto é frequentemente chamado de *user trust*, ou confiança do utilizador. Um modelo interpretável deve permitir ao utilizador final interrogar, perceber e até mesmo melhorar o sistema antes de tomar qualquer tipo de ação [3]. No fundo, um modelo é mais interpretável do que outro quando as suas decisões são mais fáceis de compreender por um ser humano. Verifica-se também que algumas destas abordagens de ML, como as árvores de decisão e os modelos de regressão, a explicação por si só já faz parte do modelo.

Também algumas das definições de interpretabilidade referem a transparência dos componentes e dos algoritmos e o uso de campos compreensíveis na construção dos modelos. No que diz respeito às *features* do *dataset*, a sua semântica deve também ser clara e fácil de entender. É preciso considerar que a interpretabilidade dos modelos de ML no domínio da saúde é sempre dependente do contexto, mesmo ao nível da função do utilizador. O mesmo modelo pode exigir gerar diferentes explicações para diferentes utilizadores finais.

É importante compreender também que ter um modelo interpretável, não é sinónimo de ter um modelo correto e confiável. As regras fornecidas pelo modelo podem ser erradas e este pode errar nas previsões que dá. Assim, o modelo deve ser devidamente validado pelos profissionais de saúde. Neste trabalho, esta validação por parte de clínicos não é realizada. As regras obtidas para cada modelos apenas foram comparadas com alguns dados reais para tentar entender cada uma delas e perceber se poderiam fazer sentido no contexto do risco cardiovascular, mas não foram objetivo de validação clínica.

2.2 Compromisso entre interpretabilidade e desempenho do modelo

Tal como foi referido anteriormente, uma das questões a ter em conta quando se desenvolve um modelo de *Machine learning*, é tentar alcançar ao mesmo tempo a interpretabilidade e o bom desempenho do modelo. No entanto, nem todos os modelos conseguem alcançar esse compromisso. Como se pode observar pela Figura 2 [8], as melhores técnicas em termos de *accuracy*, são tendencialmente piores no que toca à interpretabilidade. Como se verifica, os modelos mais interpretáveis, mas que tendencionalmente apresentam pior performance são a Árvore de decisão e a Regressão linear. Por outro lado, as redes neuronais bastante conhecidas pela sua boa performance, são regra geral muito pouco interpretáveis.

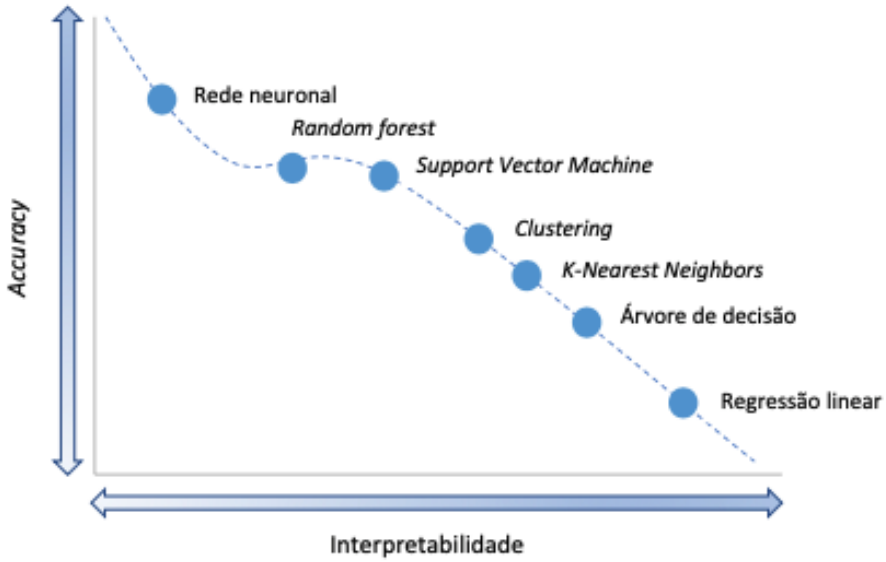

Figura 2 - Relação entre interpretabilidade e accuracy. Adaptado de [8].

Esta informação também é importante para tomar uma decisão em relação aos modelos a explorar consoante o caso de uso que se está a tentar analisar. No contexto do trabalho, esta relação entre o modelo, o desempenho e interpretabilidade é também tida em consideração na seleção dos modelos.

2.3 Modelos interpretáveis

Atualmente, um dos maiores problemas dos modelos baseados em inteligência artificial prende-se com o facto de alguns destes não serem interpretáveis. Estes modelos, *black-box,* apesar dos bons resultados, têm o grande desafio dar sentido às soluções que apresentam. Devido a esta falta de transparência nas previsões, tem-se tentado que os modelos sejam interpretáveis, e desta forma se tornem cada vez mais em modelos *white-box*, cujos principais objetivos são permitir aos utilizadores entender o comportamento dos modelos e predições, utilizar modelos complexos e conseguir interpretabilidade, aumentando a confiança dos utilizadores nas soluções propostas pelos modelos [9]. A transparência de um modelo *white-box* significa que a lógica e o comportamento necessários para chegar a um resultado final são facilmente determinados e compreensíveis. Além disso, é muito importante também que as *features* usadas sejam interpretáveis e existe uma grande necessidade em explicar as variáveis de um modelo, para que dessa forma consigam ganhar mais confiança perante os utilizadores finais. Classificadores como Árvores de decisão são *white-box,* ao contrário de Redes neuronais que são *black-box.*

Numa área tão importante como a saúde, é fundamental que um modelo seja transparente nas suas decisões e por isso, neste trabalho, foram selecionados apenas modelos *white-box*.

2.4 Trabalho desenvolvido na área

Tal como foi referido anteriormente, este trabalho tem como finalidade avaliar a interpretabilidade de um conjunto de modelos que possam auxiliar o diagnóstico clínico e o prognóstico no contexto das doenças cardiovasculares. Como tal, é necessária a extração de informação relevante diretamente de conjuntos de dados reais de pacientes, de modo a encontrar regras e padrões.

Para se compreender melhor o conceito do trabalho, é importante entender o que tem sido desenvolvido através de técnicas de ML como auxílio para a área médica, em particular na área das doenças cardiovasculares. Apesar de muito já ter sido feito, nem todos os trabalhos desenvolvidos até ao momento são interpretáveis, sendo que muitas vezes estes apenas fornecem a predição do modelo sem explicações do mesmo. São apresentados de seguida alguns exemplos de trabalhos realizados ao longo dos últimos anos.

Em 2012, Roohallah Alizadehsani e Mohammad Javad Hosseini, juntamente com outros investigadores, aplicaram algoritmos de *data mining* ao *dataset* Z-Alizadeh

Sani, constituído por um total de 303 pessoas que visitaram o Hospital *Shaheed Rajaei*, no Irão. Este *dataset* continha informações demográficas, histórico e exames das pessoas, perfazendo um total de 54 *features*. A primeira fase do trabalho passou por encontrar o melhor conjunto de *features* através de técnicas de seleção, que revelaram quais as mais importantes para o modelo. Depois desta análise, selecionaram-se os métodos de classificação a explorar: Árvores de decisão, modelo *Naïve Bayes*, entre outros. O trabalho aborda também como é realizada a medição das performances dos algoritmos, nomeadamente, através da matriz de confusão e das métricas que podem ser extraídas a partir desta. Quanto aos resultados, a árvore de decisão registou uma *accuracy* de 83.82%, sensibilidade de 93.98%, e especificidade de 85.06%, enquanto o Naive Bayes teve *accuracy* de 69.91%, sensibilidade de 63.89% e especificidade de 85.06%. Ou seja, obtiveram-se melhores resultados para a Árvore de decisão do que para o *Naïve Bayes* [10].

Em 2018, Uma N. Dulhare realizou um estudo no qual aplicou o *Naïve Bayes* a várias *features* de um *dataset* clínico, com o objetivo de prever o risco cardíaco. Fez uma seleção de *features* de entre todas e aplicou-lhe o modelo Bayesiano. Mas, para perceber a influência da dimensionalidade do *dataset* na performance do modelo, aplicou também o modelo ao conjunto de dados original, com todas as *features*. Concluiu que a *accuracy* do modelo para todas as variáveis foi de 79.12% e que, ao utilizar apenas as selecionadas, obteve um valor de 87.91% para a mesma métrica [11].

Em 2014, Pawel Marciniak, juntamente com outros investigadores, apresentou um estudo referente à implementação de métodos de *Machine learning* como apoio à previsão de doenças cardiovasculares. Aplicaram, entre outros, os modelos Bayesianos e as árvores de decisão a um *dataset* proveniente da *UCI Machine Learning Repository*. Trata-se de um conjunto de dados público, que contém informação relativa a 303 pacientes, distribuída por 13 atributos. Utilizaram apenas 297 instâncias do *dataset*, após a realização de uma limpeza aos dados em falta. Pelos resultados obtidos pode concluir-se que, em relação à árvore de decisão, esta obteve melhores resultados quanto maior é a profundidade da árvore, tendo sido, neste caso, os melhores resultados de *accuracy* de 75.26%. Para o *Naïve Bayes*, os melhores resultados para esta métrica foram de 64.95% [12].

Em relação aos primeiros dois exemplos dados, estes foram relevantes para perceber a importância que a redução do número de *features* e a remoção dos valores em falta tem na performance do modelo. Permitiram também compreender as métricas utilizadas para o cálculo do desempenho do modelo, através da matriz de confusão e da sua análise detalhada. Por sua vez, o trabalho de Pawel Marciniak, foi essencial para perceber como é que a profundidade de uma Árvore de decisão pode influenciar os resultados e isso vai ser aplicado também ao trabalho desenvolvido neste projeto.

CAPÍTULO 3

ALGORITMOS DE *MACHINE LEARNING*

Os algoritmos de *Machine learning* estão organizados por taxonomia, baseada no *output* que devolvem. Destacam-se três tipos de aprendizagem em modelos de ML: supervisionada, não supervisionada e semi-supervisionada.

Na aprendizagem supervisionada, o algoritmo gera uma função que consegue mapear os dados de entrada com os dados de saída desejados. Constrói um modelo a partir de conjuntos de dados com o objetivo de fazer previsões em novos dados, nunca antes vistos pelo modelo. Existem dois tipos principais de algoritmos supervisionados, a classificação e a regressão. A classificação tem como objetivo classificar um resultado de entre um conjunto de categorias como {doença, não doença} ou {vermelho, verde}. A regressão também prevê um resultado, mas este é um valor real e não uma categoria.

Por outro lado, na aprendizagem não supervisionada, não existem resultados pré-definidos para o modelo utilizar como referência para aprender. Ou seja, os modelos pertencentes a esta categoria analisam um conjunto de dados e, de acordo com algum mecanismo/padrão subjacente e criam grupos de instâncias. Os problemas de aprendizagem não supervisionada são agrupados em *Clustering* e associação. Em relação aos primeiros, estes são utilizados quando se pretendem encontrar agrupamentos inerentes aos dados. Nos algoritmos de associação, pretende-se descobrir regras que descrevam grandes quantidades de dados, como por exemplo, pacientes com o problema X tendem a sofrer Y [13]. De entre todos os algoritmos que seguem esta aprendizagem, alguns são o *K-Means* e o *Clustering* hierárquico.

Por sua vez, a aprendizagem semi-supervisionada é uma combinação de aprendizagem supervisionada e não supervisionada, na qual alguns dados estão identificados e outros não. Tipicamente, esta combinação tem uma quantidade muito pequena de dados identificados e uma quantidade muito grande de dados não identificados. Isto é útil porque o processo de identificar quantidades enormes de dados para aprendizagem supervisionada é muitas vezes dispendioso, mas também porque ter uma quantidade muito grande de classes identificadas pode trazer *bias* para o modelo [14].

Com base no tipo de dados disponíveis e no tipo de investigação a realizar, podem escolher-se os modelos de aprendizagem que mais se identifiquem com cada situação. Neste trabalho, são selecionados alguns modelos de entre os que foram mencionados anteriormente para aplicar à avaliação do risco cardiovascular.

3.1 Perspetiva global

Neste trabalho, os algoritmos explorados no âmbito da Interpretabilidade são a Árvores de decisão, o modelo Bayesiano e o *Clustering*.

As Árvores de decisão são exploradas porque, como resultam num conjunto de regras, são perfeitamente interpretáveis. Para além disso, fornecem como resultado uma classe categórica, que é o que se pretende neste caso de estudo, isto é, são particularmente bem adaptadas a problemas de classificação.

Por outro lado, também os modelos Bayesianos são baseados na evidência e por isso bastante alinhados com o que se pratica na área clínica. Têm como premissa a independência entre as variáveis do problema e realizam uma classificação probabilística de observações.

Por sua vez, os modelos de *Clustering* permitem criar grupos de instâncias com caraterísticas semelhantes entre si e, a partir destes, estabelecer um conjunto de regras com base em medidas de semelhança. Isso também é o que se verifica na parte clínica, ou seja, os novos casos também são comparados com anteriores para realizar um diagnóstico.

3.2 Abordagens interpretáveis

A forma mais simples de alcançar a interpretabilidade é através da utilização de um subconjunto de algoritmos capazes de criar modelos interpretáveis.

Desta forma, para que sejam aceites e úteis para os clínicos, as técnicas de ML usadas no apoio à deteção de doenças cardiovasculares devem demonstrar um bom desempenho e ser interpretáveis. Estas abordagens são importantes para um clínico, uma vez que adicionam um grau de confiança nas decisões tomadas. Mas o que é concretamente a confiança num modelo? Traduz-se apenas na confiança de que o modelo terá bom desempenho? Se fosse assim tão linear, o facto de o modelo ser suficientemente preciso podia torná-lo confiável e não seria necessário definir a interpretabilidade do modelo.

Como tal, este projeto tenta apresentar modelos onde seja possível entender o porquê de uma determinada conclusão. Estas conclusões podem ser baseadas em regras que podem vir a ser usadas por especialistas em aplicações práticas, uma vez que fornecem um modelo mais próximo da linguagem humana.

Tal como já foi mencionado anteriormente, neste projeto são utilizados os modelos Bayesianos, as árvores de decisão e os modelos baseados em técnicas de *Clustering*. Os algoritmos de árvores de decisão, por permitirem extrair interações entre variáveis através de regras (*if/then)* tornam-se relativamente fáceis na extração de conclusões. Também os modelos Bayesianos, que aplicam o Teorema de *Bayes*, calculam a probabilidade condicional para as variáveis e tomam as suas decisões com base nessas probabilidades condicionais extraídas diretamente dos dados. Os algoritmos

de *Clustering* foram escolhidos dado que são escaláveis para grandes conjuntos de dados e adaptam-se facilmente a novos exemplos conseguindo generalizar para clusters de diferentes formas e tamanhos.

De seguida, cada um dos algoritmos referidos anteriormente é explorado com mais profundidade, explicando-se como se conseguirá alcançar interpretabilidade através da sua aplicação ao problema.

3.2.1 Árvores de decisão

As árvores de decisão são modelos preditivos, usados principalmente em problemas de classificação e de regressão, que utilizam estratégias para aprender padrões de um conjunto de dados. A sua principal vantagem é a grande legibilidade que apresentam, ao ser representadas em forma de árvore. Esta é desenhada de cima para baixo, com a sua raiz no topo. Na estrutura da árvore, cada condição é representada por um nó interno ao qual é atribuída uma classe, com base no qual a árvore se divide em diferentes nós, ou *edges*, que são possíveis decisões sobre o atributo. Estes nós folha classificam as instâncias que os atingem de acordo com a classe associada a eles. Portanto, o conhecimento de uma árvore de decisão é representado por cada nó, que ao ser testado procura um nó filho até chegar ao nó folha. Um nó de teste calcula um resultado com base nos valores do atributo de uma instância, em que cada resultado possível é associado a uma das subárvores. O último nó de uma árvore, ou nó terminal, que já não se divide em nada, é a decisão [15].

A Figura 3 é um exemplo de uma árvore de decisão simples, implementada no âmbito da área da saúde [16].

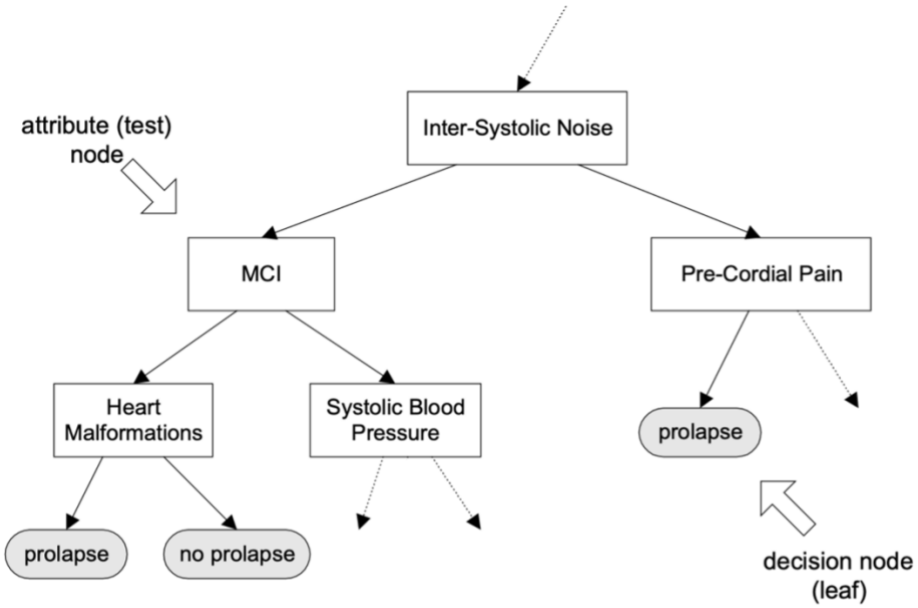

Figura 3 - Exemplo de Árvore de decisão [16].

Uma árvore de decisão pode ser de fácil interpretação, como se pode observa pela Figura 3. Neste caso é possível extrair, por exemplo, as duas seguintes regras:

1. *IF Inter-Systolic Noise AND Milld Cognitive Impairment (MCI) AND Heart Malformations THEN Prolapse.*
2. *IF Inter-Systolic Noise AND Milld Cognitive Impairment (MCI) AND No Heart Malformations THEN No Prolapse.*

Uma decisão pode ser construída a partir de um conjunto de dados de treino através do princípio "dividir e conquistar". Quando os objetos têm a mesma classe de decisão, isto é, o valor do output é o mesmo, a árvore consiste apenas num único nó que contém a decisão apropriada. Caso contrário, é selecionado um atributo cujo valor é de pelo menos duas classes diferentes e um conjunto de objetos é dividido de acordo com a categoria do atributo selecionado. Os atributos selecionados constroem um nó atributo (*test*) numa árvore de decisão em crescimento, para cada ramo a partir desse nó e o processo de indução é repetido até que uma folha (decisão) seja encontrada.

A árvore de decisão é uma das ferramentas mais amplamente usada para a tomada de uma decisão. Dependendo da situação e do resultado desejado existem vários tipos de árvores de decisão que se podem utilizar, sendo os principais [17]:

- *Árvore de classificação***:** Permite obter o resultado mais previsível de diferentes informações, através da sua aplicação em probabilidade e estatística. Esta árvore é construída por meio de um processo iterativo que divide os dados em partições e, em seguida, os vai dividindo ainda mais em cada uma das ramificações.
- *Árvore de regressão*: Estas são árvores de decisão nas quais a variável de destino pode assumir valores contínuos (normalmente números reais). Aplicado para determinar um único resultado predeterminado de diferentes partes da informação.

Ao implementar a árvore de decisão, o principal desafio surge ao selecionar a relevância e importância de cada classe. À medida que descemos na árvore, o nível de impureza diminui, levando assim a uma melhor classificação ou melhor divisão de cada nó. A identificação da divisão mais adequada, a forma como se processa o crescimento da árvore, tem necessariamente de ser baseado num critério, tipicamente o *information gain* ou o índice de Gini.

O *information gain* é utilizado para identificar a *feature* que fornece mais informações sobre uma classe, baseando-se na noção de entropia. A entropia indica o grau de incerteza e impureza de uma classe, e é representada pela equação (1). Se as observações de subconjuntos de dados de um *dataset* forem homogéneas, não há impureza no conjunto de dados, ou seja, quando todas as observações pertencem à mesma classe a entropia é 0 [18].

$$Entropy(L) = -\sum_{i=1}^{j} p_i log_2(p_i) \qquad (1)$$

Em que *L* é uma amostra de aprendizagem $L = \{(x_i, c_1), (x_2, c_2), \ldots, (x_j, c_j)\}$, onde $x_1, x_2, \ldots, x_i$, são vetores de medida, $c_1, c_2, \ldots, x_j$ são classes e p_i é a frequência relativa da classe i.

Assim, *information gain* corresponde à diferença entre a entropia de uma classe e entropia condicional. Mede a utilidade de uma *feature* f na classificação, ou seja, qual foi a diferença no valor de entropia de antes e depois da divisão do conjunto de dados. Se o valor aumentar, isto significa que f é mais útil para a classificação. Assumindo que há V diferentes valores para a *feature* f, $|L^V|$ representa um subconjunto de L com $f = v$, a *information gain* depois de dividir L na *feature* f é medida pela equação (2).

$$IG(L, f) = Entropy(L) - \sum_{v=1}^{v} \frac{|L^V|}{|L|} \left(Entropy(L^V)\right) \qquad (2)$$

Por sua vez, o índice de gini, ou impureza de gini, calcula a probabilidade de uma amostra ser classificada incorretamente quando selecionada aleatoriamente. Os valores deste índice variam entre 0 e 1 e uma classe é considerada pura quando o seu índice é 0. Nestes casos, todos os elementos estão classificados corretamente como pertencendo à classe a que foram atribuídos. Se L for um *dataset com j* classes diferentes*,* GINI é definido por (3) [18]*:*

$$GINI(L) = 1 - \sum^{j} {p_i}^2 \qquad (3)$$

Onde p_i é a frequência relativa se a classe i pertence a L.

Mas, apesar de todos os pontos positivos, estes algoritmos apresentam também alguns pontos fracos. Uma das desvantagens da árvore de decisão é que pode facilmente causar situações de *overfitting*. Se a árvore crescer sem um limite pode dividir-se tantas vezes quantas conseguir, e pode aprender com os dados e generalizar de forma errada na presença de novas informações. Para evitar situações destas, existem algumas técnicas que ser aplicadas, sendo algumas delas [19]:

Pruning: Esta é uma abordagem que corrige a árvore depois da sua adaptação ao conjunto de dados de treino. O algoritmo começa pelos nós folha e remove esses ramos que, depois da remoção não afetam a performance geral da árvore no conjunto de dados de teste. Este método preserva um alto desempenho ao mesmo tempo que reduz a complexidade do modelo.

Métodos *Ensemble*: Estes métodos combinam várias árvores num só algoritmo que utiliza o 'voto' em árvore como mecanismo para determinar a verdadeira resposta.

Desta forma, quando se treina um conjunto de dados através de uma árvore de decisão as questões referidas anteriormente têm de ser tidas em conta. Neste trabalho, alguns destas técnicas são aplicadas, como vai ser possível perceber posteriormente no Capítulo 5.

3.2.2 Modelos Bayesianos

Alguns dos algoritmos de ML são capazes de produzir estimativas de probabilidade, ou seja, para cada classe estimam a probabilidade de uma dada amostra pertencer a essa classe. O modelo Bayesiano é uma dessas abordagens de classificação. Tem como principais vantagens o facto de apresentar métodos bem fundamentados para representar distribuições de classes de probabilidade de forma compreensível.

A estimação Bayesiana, não tem como intuito encontrar um conjunto de parâmetros, θ, que expliquem os dados, mas sim fornecer uma possível distribuição sobre o valor dos parâmetros que quantifiquem a incerteza dos valores [20]. Muitas escolhas de parâmetros são possíveis, mas algumas têm maior probabilidade anterior (*priori*) de ocorrer do que ocorrer [21]. Desta forma, a abordagem Bayesiana combina o conhecimento anterior sobre os parâmetros com a distribuição anterior de uma nova variável D, dado θ.

A regra de Bayes é representada por (4):

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{4}$$

Onde $P(D|\theta)$ é a função *likelihood,* $P(\theta)$ é a distribuição anterior de θ que representa o conhecimento anterior sobre os vários valores que os parâmetros podem tomar. $P(D)$ representa a probabilidade anterior da observação dos dados obtidos. Pode ver-se a representação de P(D) na equação (5).

$$P(D) = \int_{\theta} P(D|\theta)P(\theta)d\theta \tag{5}$$

Os classificadores Bayesianos são classificadores probabilísticos que implementam estruturas particulares de redes Bayesianas, uma vez que o seu objetivo é classificar as instâncias descritas por um conjunto de dados. A classificação baseia-se na regra de Bayes (4) para prever a classe de D que tem a maior probabilidade, dado um conjunto de dados $\{\theta_1, \ldots \theta_p\}$.

O Naïve Bayes (NB) implementa uma estrutura de rede Bayesiana, em que cada nó corresponde uma variável que toma um valor dentro de um intervalo e cada arco representa uma relação probabilística entre as variáveis. Para cada atributo do gráfico são definidas as probabilidades condicionais. Como se percebe pela estrutura da rede Bayesiana da Figura 4 [22], todos os atributos são assumidos como condicionalmente independentes. A independência entre as classes ignora algum tipo de correlação que possa existir entre elas.

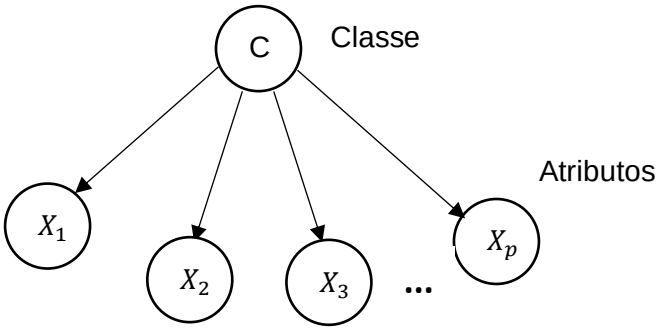

Figura 4 - Representação da rede Naïve Bayes.

O classificador é constituído por apenas um pai (C), e vários filhos ($\mathrm{X_1}, \ldots \mathrm{X_p}$}. Apesar da sua configuração simples, o *Naïve Bayes* é computacionalmente eficiente, consegue lidar com informação incompleta e tem uma performance competitiva quando comprado com outros classificadores [23].

Considerando uma estrutura de gráfico *G* que contenha a variável que seja o nó raiz e um conjunto de variáveis aleatórias (atributos) $\{\mathrm{X_1}, \ldots \mathrm{X_p}\}$, onde cada atributo tem a variável do nó raiz como único pai, chamada $Pa_i = C$ para todos os $1 \leq i \leq p$.

De acordo com a equação (6), a probabilidade conjunta das variáveis que pertencem a *G* é dada por:

$$P(X_I, \ldots, X_p, C) = P(C) \prod_{i=1}^{p} P(X_i|C) \tag{6}$$

As probabilidades condicionais são calculadas por (7):

$$\mathrm{P(X|Y)} = \frac{\mathrm{P(X,Y)}}{\mathrm{P(Y)}} \tag{7}$$

Na qual,

$P(X,Y) = P \cap Y$

Tem por base a definição de probabilidade condicional, dada pela equação (8), pode calcular-se a probabilidade de C, sabendo que se verifica $\mathrm{X_1, X_2, \ldots, X_n}$, ou seja, $\mathrm{P(C|X)}$.

$$P(C|X) = P(C|X_1, X_2, \ldots, X_n) = \alpha P(C) \prod_{i=1}^{n} P(X_i|C) \tag{8}$$

Em que,

$P(C|X)$ corresponde à probabilidade de acontecer C, sabendo que acorrem os eventos X, nomeadamente, $\mathrm{X_1, X_2, \ldots, X_n}$.

$P(X_i|C)$ representa a probabilidade condicional de cada um dos atributos X_i se verificar, sabendo que acontece C.

$P(C)$ é a probabilidade do evento C acontecer.

α é a constante de normalização. Assim, a classificação final é dada pela equação (9):

$$c = argmax_{cj}(\alpha P(c_j) \prod_{i=1}^{p} P(x_i \mid c_j)) \quad (9)$$

Onde c_j é uma classe de C, x_i é o valor do atributo $\mathrm{X_i}$ que pertence instância $\mathrm{X_p} = \mathrm{X_1}, \ldots \mathrm{X_p}$ e $\alpha = 1/P(X = x_p)$.

Desta forma, sabendo o valor de probabilidade condicional para cada variável e a probabilidade de um evento C acontecer, torna-se simples prever a probabilidade o evento C, sabendo que $X_1, X_2, \ldots, X_n$ fatores se verificam.

No caso concreto de estudo deste projeto, o objetivo é calcular as probabilidades de Morte e de Sobrevivência, sabendo que se verificam um conjunto de caraterísticas no paciente. Sabendo que o resultado é representado por M, e M = {M, $\overline{\mathrm{M}}$}, o objetivo é calcular a probabilidade *posterior* de $P(M|X)$ e de $P(\overline{\mathrm{M}}|X)$, sendo X um conjunto de sintomas.

3.2.3 *Clustering*

Tal como referido anteriormente, no contexto de *Machine learning* existem métodos supervisionados e não-supervisionados. O *Clustering* enquadra-se nos não supervisionados, uma vez que neste método não existe uma saída desejada para o treino do modelo, sendo o objetivo a interpretação dos dados (entradas) e a capacidade de os explicar [24]. Esta capacidade de encontrar similaridade entre objetos de dados através da estrutura do *dataset,* torna-se muito útil no contexto da previsão do risco cardiovascular. Ao aplicar este modelo, a população é dividida em grupos ou *clusters* de modo que, os pontos de dados sejam mais semelhantes a outros pontos de dados no mesmo grupo e diferentes dos pontos de outros grupos.

Dentro dos métodos de *Clustering*, existem algumas classificações possíveis, entre os quais a Hierárquica e a Particionada. Em métodos baseados em hierarquia, os novos *clusters* são criados com base nos anteriores e em métodos particionados, os dados são divididos em k clusters, com base na similaridade entre eles [25]. Para aplicar o *Clustering* à identificação do risco cardiovascular, utilizaram-se métodos de particionamento, uma vez que o objetivo seria criar dois grupos entre os dados. Um para os utentes que sofreram a morte e outro para os sobreviventes.

De entre os métodos baseados em partição de dados, foi selecionado para aplicar ao projeto o K-Means. A ideia principal deste algoritmo é que os dados sejam organizados em k agrupamentos (*clusters*), em que k é um valor definido à priori. Cada um destes agrupamentos tem um centro, ou centróide. Desta forma, cada cluster tem sempre pelo menos um elemento, sendo que cada um deles pertence a apenas um grupo,

sendo este aquele cujo centro lhe ficar mais próximo, portanto, que for mais similar. Os conceitos de similaridade são definidos com base em diferentes métricas de distância de acordo com o tipo de dados. As métricas de distância são organizadas pelo tipo de atributo, como se pode observar na Tabela 1.

Tabela 1 – Distâncias entre instâncias de dados [26].

Atributos escalonados por intervalos	Euclidiana	$d(u,v) = \sqrt{\sum_{i=1}^{p}(u_i - v_i)^2}$
	Minkowsky	$d(u,v) = \sqrt[r]{\sum_{i=1}^{p}\lvert u_i - v_i\rvert^r}$
	Manhattan	$d(u,v) = \sum_{i=1}^{p}\lvert u_i - v_i\rvert$
	Chebychev	$d(u,v) = max_{i=1}^{n}\lvert u_i - v_i\rvert$
	Camberra	$d(u,v) = \sum_{i=1}^{p}\frac{\lvert u_i - v_i\rvert}{\lvert u_i + v_i\rvert}$
Atributos binários {0,1}	Coeficiente de correspondência simples	$d(u,v) = \frac{n_1 + n_2}{p}$
	Coeficiente de semelhança *Jaccard*	$d(u,v) = \frac{n_1}{n_1 + n_3 + n_4}$
	Distância *Hamming*	$d(u,v) = n_3 + n_4$
Atributos nominais em escala	Coeficiente de dissimilaridade	$d(u,v) = \frac{p - o}{p}$
Atributos ordinais à escala	Semelhança de atributos baseados em intervalos	$z_i = \frac{u_i - 1}{U_k - 1}$

No contexto do trabalho, recorreu-se à distância euclidiana para calcular a proximidade entre os pontos, uma vez que esta permite o cálculo de distâncias entre pontos num espaço euclidiano n-dimensional. A distância entre os pontos $P = (p_1, p_2, \dots, p_n)$ e $Q = (q_1, q_2, \dots, q_n)$ é representada pela equação (10) [27]:

$$d(P,Q) = \sqrt{(P_1 - Q_1)^2 + (P_2 - Q_2)^2 + \dots + (P_n - Q_n)^2} = \sqrt{\sum_{i=1}^{n} (P_i - Q_i)^2} \qquad (10)$$

Desta forma, o que se pretende com a aplicação do *Clustering* é tentar agrupar os pacientes para os quais se verificou morte e sobrevivência em *clusters* e calcular os seus centróides através do K-Means. Posteriormente, é calculada a distância euclidiana de cada paciente a cada um dos centróides (Morte e Sobrevivência), para que se tente perceber em qual deles é que se verifica uma menor distância. Desta forma, consegue perceber-se de que diagnóstico o paciente está mais próximo, ou seja, quando menor a distância euclidiana do paciente ao centróide, mais próximo este se encontra do respetivo diagnóstico. Esta representação pode ser visualizada na Figura 5.

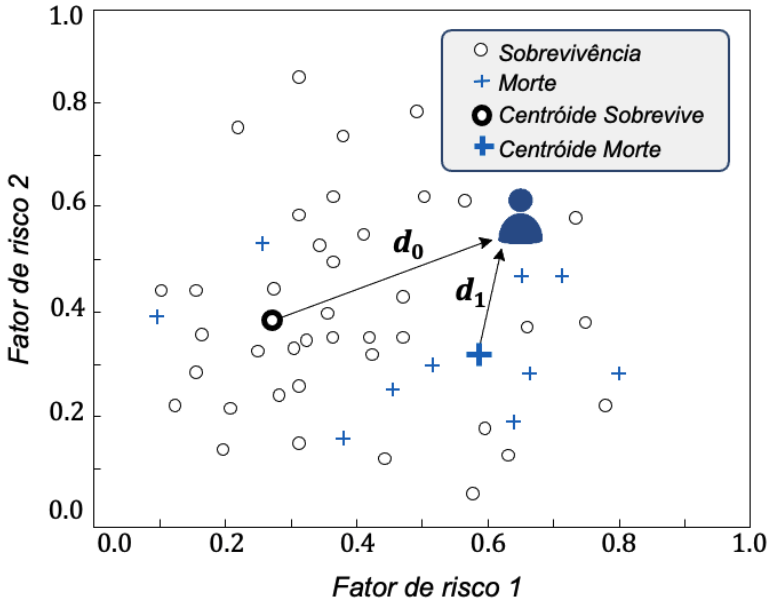

Figura 5 - Representação do Clustering.

3.3 Validação

Avaliar um classificador é uma parte muito importante no processo de classificação. São as metodologias de validação escolhidas que permitem avaliar o desempenho do modelo, uma vez que desta forma ele pode ser constantemente melhorado para conseguir alcançar a melhor performance possível. Existem várias formas de avaliar

um algoritmo e as métricas de desempenho são uma delas. Estas são medidas numéricas que quantificam a performance de um classificador.

A matriz de confusão possibilita a derivação de um conjunto de métricas que permitem aferir a performance de um modelo. Ela apresenta os resultados sob a forma de uma tabela com duas entradas, uma com as classes desejadas e outro com as classes previstas pelo modelo. O número de instâncias que correspondem ao cruzamento das entradas é o que preenche as células da tabela. Uma representação da matriz de confusão é dada pela Figura 6.

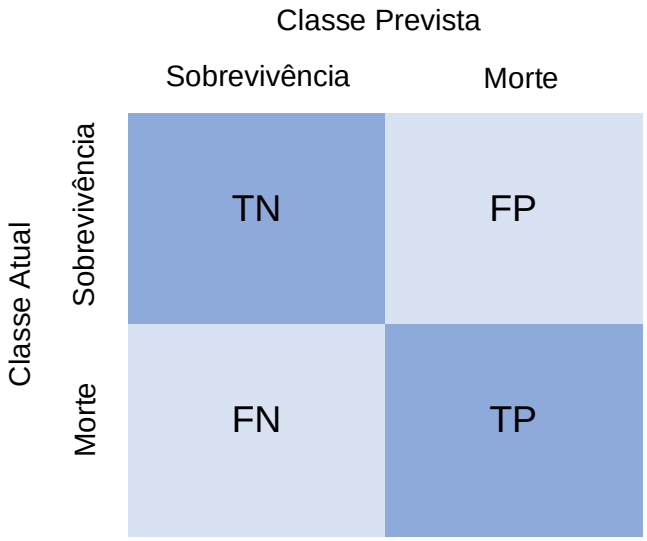

Figura 6 - Representação da matriz de confusão.

Onde,

TP (*True positives*) são os verdadeiros positivos e correspondem às instâncias positivas classificadas de forma correta. A classe que resulta da classificação do modelo coincide com o valor real.

FP (*False positives*) são os falsos negativos e correspondem às instâncias negativas classificadas de forma incorreta

TN (*True negatives*) são os verdadeiros negativos e corresponde às instâncias negativas classificadas de forma correta. O valor previsto corresponde com o atual.

FN (*False negatives*) são os falsos positivos e correspondem às instâncias positivas classificadas de forma incorreta.

No contexto do trabalho, os TP são as instâncias classificadas como morte e em que realmente isso aconteceu. Os FP são os casos em que as classes são previstas como morte, mas na realidade o que se verificou foi sobrevivência. Os TN correspondem aos casos em que as instâncias foram classificadas como sobrevivência e isso realmente aconteceu e os FN são as situações em que as classes são identificadas como sobrevivência, mas na realidade ocorre a morte.

A partir da matriz de confusão é possível retirar informação sobre cada uma das métricas, sendo elas:

Accuracy representa a razão entre o número de previsões positivas e negativas e o número total de amostras. Dá a percentagem de previsões corretas e é representada pela equação (11):

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

Precisão que fornece o número de resultados positivos corretos dividido pelo número de resultados positivos previstos pelo classificador, ou seja, quantifica o número de previsões da classe positiva que realmente pertencem a essa classe. É dada pela equação (12):

$$Precisão = \frac{TP}{TP + FP} \tag{12}$$

Sensibilidade representa a medida da proporção de casos positivos reais que foram previstos corretamente como positivos (ou verdadeiros positivos). Esta medida é também conhecida como Taxa de verdadeiros positivos, sendo representada pela equação (13):

$$Sensibilidade = \frac{TP}{TP + FN} \tag{13}$$

Especificidade é definida como a proporção de negativos reais, que foram previstos de forma correta como os negativos (ou verdadeiros negativos). É também conhecida como Taxa de verdadeiros negativos e dada pela equação (14):

$$Especificidade = \frac{TN}{TN + FP} \tag{14}$$

G-mean (Métrica geométrica) é uma métrica que mede o equilíbrio entre as performances da classificação. Um baixo valor de G-mean é uma indicação de um fraco desempenho na classificação dos casos positivos, mesmo que os casos negativos sejam classificados corretamente como tal. É representada pela equação (15):

$$G - mean = \sqrt{sensibilidade \times especificidade} \tag{15}$$

De entre todas as métricas referidas anteriormente, apenas algumas delas são utilizadas para avaliar a performance dos modelos no trabalho. A *accuracy* é usada uma vez que permite identificar o número de instâncias positivas classificadas corretamente. No entanto, não deve ser tida em conta em *datasets* não balanceados, porque pode aprender com os dados em maioria e inflacionar os resultados, como se vai perceber mais à frente no documento. A sensibilidade é importante, e usada neste

trabalho, porque indica qual a proporção com que o modelo deteta corretamente uma instância como positiva, sendo que a especificidade representa o mesmo, mas para casos negativos. A métrica geométrica, por representar o equilíbrio do modelo é também fundamental para dar uma perceção geral sobre a performance do mesmo. Como tal, estas serão as quatro métricas utilizadas na análise dos modelos.

CAPÍTULO 4

CASO DE ESTUDO: AVALIAÇÃO DE RISCO CARDIOVASCULAR

Foram propostos anteriormente diferentes algoritmos *white-box* que podem ser utilizados para auxiliar na previsão do risco cardíaco. Para o estudo de interpretabilidade realizado no trabalho, foram usados dados reais de pacientes para aplicar os modelos escolhidos.

4.1 Descrição do *dataset*

O *dataset* utilizado resulta de uma compilação de dois conjuntos de dados provenientes de dois hospitais portugueses. Um deles contém informação sobre pacientes do Hospital de Santa Cruz, em Lisboa, e o outro é proveniente do Hospital de Leiria. No total, o *dataset* contém informação sobre 1185 pacientes.

Os dados incluem características demográficas, fatores de risco prévios, antecedentes médicos, antecedentes (problemas cardíacos que se tenham verificado no passado), caraterização de episódios agudos que possam ter ocorrido, resultados de análises laboratoriais, tratamento hospital e até possível medicação prévia à qual tenha sido necessário recorrer. Para cada paciente é dada também informação relativa ao respetivo acompanhamento clínico (*follow-up*) que tiveram. Ou seja, é fornecido um registo que indica se ocorreu o evento (morte), durante um determinado período de tempo em que o doente foi seguido.

Na Tabela 2 encontram-se representados todos os atributos que constituem no *dataset*. Para cada um deles é apresentada a sua frequência relativa, nos casos em que foi registada a sobrevivência e nos casos em que ocorreu morte. Na tabela, para variáveis numéricas, como Hemoglobina, Pressão sistólica ou plaquetas, é dada a média de valores que se verificaram nessa variável, seguindo-se os valores mais baixo e mais elevado que foram registados. No caso das variáveis binárias e categóricas é dado o número de registos verificados para a situação de ocorrência (“Sim”) e de não ocorrência (“Não”), seguido da percentagem que esse valor representa no total dos dados.

Tabela 2 – Descrição de features do dataset em estudo.

		Total	Sobrevivência	Morte
Número de pacientes		1185 (100%)	1070 (90.3%)	115 (9.7%)
Diagnóstico	NSTEMI STEMI UA	523 (42.1%) 452 (36.1%) 274 (21.8%)	435 (40.7%) 373 (34.9%) 262 (24.4%)	52 (45.2%) 49 (42.9%) 14 (11.9%)
Caraterísticas demográficas				
Sexo	Mulher Homem	960 (75.9%) 304 (24.1%)	831 (77.7%) 239 (22.3%)	74 (54.4%) 41 (45.6%)
Idade		64 (23-96)	63 (23-94)	71 (33-96)
Peso		75.3 (40-155)	75.6 (43-135)	73.8 (42-122)
IMC		27.1 (17.6-53)	27 (17.6-49.8)	27.1 (18.9-50.7)
Fatores de risco prévios				
Diabetes	Sim Não	321 (25.6%) 935 (74.4%)	257 (24.1%) 808 (75.9%)	43 (37.7%) 71 (62.3%)
Fumador	Sim Não	325 (25.8%) 933 (74.2%)	292 (27.3%) 776 (72.7%)	14 (87.5%) 98 (12.5%)
Ex-fumador	Sim Não	285 (22.9%) 960 (77.1%)	260 (24.6%) 797 (75.4%)	15 (13.4%) 97 (86.7%)
Hipertensão	Sim Não	793 (63.1%) 464 (36.9%)	410 (38.5%) 656 (61.5%)	85 (75.2%) 28 (24.8%)
Colesterol	Sim Não	396 (57.7% 290 (42.3%)	369 (59.4%) 252 (40.6%)	27 (41.5%) 38 (58.6%)
Antecedentes				
AVC/AIT	Sim Não	93 (7.4%) 1166 (92.6%)	66 (6.2%) 1001 (93.8%)	23 (20.4%) 90 (79.6%)
AI	Sim Não	182 (14.47%) 1076 (85.53%)	157 (14.71%) 910 (85.29%)	14 (12.39%) 99 (87.61%)
PTCA	Sim Não	261 (20.7%) 1001 (79.3%)	228 (21.4%) 841 (78.6%)	94 (83.19%) 19 (18.81%)
CABG	Sim Não	132 (10.5%) 1129 (89.53%)	120 (11.2%) 949 (89.8%)	12 (10.6%) 101 (89.38%)
Caraterização de episódio agudo				
Pressão sistólica		140.9 (50-240)	141.9 (50-231)	133.1 (70-240)
Pressão diastólica		81.7 (30-160)	82.4 (30-160)	76.6 (40-140)
Frequência cardíaca		71.6 (20-160)	69.3 (20-153)	88.4 (25-160)
Classe Killip	Classe I Classe II Classe III Classe IV	1047 (84.6%) 109 (8.8%) 70 (5.7%) 12 (1.0%)	914 (87.0%) 87 (8.3%) 44 (4.2%) 6 (0.5%)	63 (56.8%) 24 (21.6%) 19 (17.1%) 5 (4.5%)
Angina na admissão	Sim Não	859 (70.7%) 355 (29.2%)	737 (71.3%) 297 (28.72%)	66 (58.93%) 46 (41.07%)
Desvios ST	Sim Não	822 (66.4%) 417 (33.6%)	86 (76.8%) 26 (23.2%)	86 (76.8%) 26 (23.2%)
Depressão ST	Sim Não	515 (50.7%) 500 (49.3%)	46 (55.4%) 37 (44.6%)	46 (55.4%) 37 (44.6%)
Análises laboratoriais				
Creatinina		1.4 (0.4–14)	1.3 (0.44-11.5)	1.8 (0.45-14.0)
Biomarcadores	Sim Não	989 (78.24%) 275 (21.76%)	817 (76.36%) 253 (23.64%)	102 (88.70%) 13 (11.30%)
Hemoglobina		13.8 (1.4-30.2)	13.9 (1.4-30.2)	12.9 (6.5-17.6)
HB Glicada		6.0 (3.7-13.1)	6.0 (3.8-13.1)	6.4 (3.7-11.1)
Plaquetas		207.2 (0.2-793)	207 (0.22-793)	215.8 (15-472)
Triglicerídeos		154.1 (27-1204)	154.5 (27-1204)	150.8 (35-777)
Medicação Prévia				
Aspirina	Sim Não	618 (50.37%) 610 (40.63%)	527 (49.52%) 518 (50.38%)	57 (53.77%) 49 (46.23%)
Betabloqueadores	Sim Não	326 (27.65%) 853 (72.35%)	275 (27.45%) 727 (72.55%)	31 (30.69%) 70 (69.31%)
Anti-cálcio	Sim Não	254 (21.6%) 924 (78.4%)	229 (22.9%) 772 (77.1%)	21 (20.8%) 80 (79.2%)
Estatina	Sim Não	289 (24.5%) 890 (75.5%)	246 (24.6%) 755 (75.4%)	27 (26.7%) 74 (73.3%)
Anti-hipertensivos	Sim Não	366 (31.1%) 811 (68.9%)	305 (30.5%) 695 (69.5%)	42 (41.6%) 59 (58.4%)
Nitratos	Sim Não	358 (62.59%) 599 (37.41%)	300 (37.45%) 501 (62.55%)	51 (64.56%) 28 (35.44%)
Reperfusão realizada	Sim Não	660 (52.22%) 604 (47.78%)	590 (55.14%) 480 (44.86%)	50 (43.48%) 65 (56.52%)

Como é possível verificar pela observação da Tabela 2, o *dataset* contém dados sobre pacientes que tiveram entrada hospitalar e foram diagnosticados com enfarte do miocárdio com elevação de ST (STEMI), enfarte do miocárdio sem elevação de ST (NSTEMI) e angina instável (UA). O segmento ST é um elemento muito importante na avaliação das síndromes coronárias agudas, uma vez que a oclusão total ou parcial do vaso coronário se traduz na redução do fluxo sanguíneo para determinada região do coração, alterando a atividade elétrica dessa região gerando alterações na eletrocardiograma. O STEMI e o NSTEMI ocorrem principalmente devido à aterosclerose que corresponde à acumulação de placas de colesterol nas paredes das artérias, sendo que no STEMI se verifica uma maior oclusão que provoca o desvio do segmento ST, ao contrário do que acontece no NSTEMI. A angina instável (UA) provoca sintomas sugestivos de isquemia cardíaca sem biomarcadores.

Tal como também é possível verificar, de um total de 1185 de pacientes, 1070 desses tiveram como resultado final a sobrevivência, sendo que em apenas 115 se registou a morte. Como tal, o *dataset* é pouco balanceado, uma vez que regista uma menor incidência de uma classe (classe minoritária / morte) em relação a outra (classe maioritária / sobrevivência). Esta situação faz com que um algoritmo aprenda que uma determinada classe é mais comum do que outra, acabando por ter uma maior tendência para classificar futuros registos como pertencentes à classe maioritária. Assim, ao prever sempre a classe maioritária, os modelos podem registar valores de *accuracy* elevados que não correspondem à verdade [28]. Para contornar estas situações há algumas técnicas que podem ser aplicadas.

4.2 Pré-processamento dos dados

Balanceamento do *dataset*

O problema das classes balanceadas verifica-se quando a distribuição das classes não é uniforme. A existência de dados não balanceados pode dificultar a robustez da decisão a ser tomada. Como exemplo prático, num conjunto de dados, em que 95% da população esteja classificada como X e apenas 5% como Y, caso o objetivo seja obter um modelo de previsão para classificar a população em X ou Y, esta tarefa pode ser complexa uma vez que os algoritmos podem ser induzidos a classificar tudo como X devido à superioridade de casos X. Neste caso, se o modelo classificar tudo como X apresentará 95% de previsões corretas, apesar de ser um modelo com elevada taxa de previsões corretas, não tem valor preditivo pois nunca classifica casos Y. Para evitar situações como a descrita é necessário reduzir estas diferenças entre as classes e para isso já algumas técnicas têm sido estudadas. Existem várias abordagens que podem ser aplicadas nestes casos, e algumas passam pela duplicação de dados de treino da classe menos representada ou pela remoção de dados de treino classe mais representada. Isto é, como o problema se encontra na

diferença do número de exemplos entre as classes procede-se a um balanceamento aleatório, que pode ser realizado através de várias técnicas, como as seguintes [29]:

- *Undersampling* consiste em reduzir a classe com maior número de amostras até esta ter o mesmo número de amostras da classe em minoria. A principal desvantagem deste método é que pode descartar uma grande quantidade de informação relevante ao eliminar os casos da classe maioritária.
- *Oversampling* é o contrário da técnica anterior, ou seja, tenta-se aumentar o número de amostras da classe em minoria, através da sua replicação. Alguns dos algoritmos mais utilizados e que seguem esta metodologia são o *Variational Autoencoders* (VAE) ou o *Synthetic Minority Over-sampling Technique* (SMOTE). Este último, que é aplicado no trabalho, tem como objetivo criar instâncias sintéticas de dados a partir dos existentes.
- Com Métodos avançados de *sampling* consegue obter-se uma combinação das duas técnicas anteriores. A abordagem *Boosting*, por exemplo, é iterativa e em cada classificação os resultados são analisados para de seguida ser aumentado o peso da classe que apresentar menos resultados corretos e diminuído o peso da classe com mais resultados corretos. Desta forma, na iteração seguinte o algoritmo apresenta melhores resultados na previsão da classe com pior classificação.

Tal como foi referido anteriormente, estas técnicas serão aplicadas para balancear os dados e desta forma, conseguir obter resultados que não tendam para a *feature* maioritária e sejam imparciais na classificação.

Seleção de *features*

Como é percetível pela Tabela 2, o *dataset* possui 34 *features*. Esta dimensionalidade dos dados traduz-se no aumento da complexidade do modelo e faz com que seja mais difícil alcançar uma predição quando a quantidade de dados para analisar é tão grande. Desta forma, é necessário reduzir a dimensionalidade do *dataset* para reduzir consequentemente o custo da aprendizagem e permitir resolver problemas complexos com modelos simples [30].

Para fazer a seleção das *features* a aplicar aos algoritmos *Machine learning*, recorreu-se a alguns modelos atualmente existentes.

A maioria dos métodos de previsão de risco utiliza a opinião de especialistas para identificar um número de variáveis clínicas com prognóstico significativo. Os modelos *Global Registry of Acute Coronary Events* (GRACE), *The thrombolysis In Myocardial Infarction* (TIMI) e *Platelet glycoprotein IIb/IIIa in unstable agina: Receptor Suppression Using Integrilin* (PURSUIT) são *risk scores* usados para o diagnóstico a curto prazo da Síndrome Coronária Aguda (NSTE-ACS). Os pacientes com esta síndrome representam uma população heterogénea com diferentes riscos de morte e eventos cardíacos recorrentes. Por isso, nestes pacientes a estratificação precoce do

risco é fundamental, uma vez que o benefício de estratégias de tratamento mais recentes parece ser proporcional ao risco de eventos clínicos adversos. Foi realizado um estudo clínico [31] com o objetivo de comparar o desempenho dos modelos TIMI, PURSUIT e GRACE e extrair um conjunto de regras capazes de avaliar o estado de um paciente. Estes modelos foram criados através da informação de pacientes, como o histórico clínico, eletrocardiograma (ECG) e valores laboratoriais recolhidos no ato da admissão hospitalar. Na Tabela 3 é possível ver qual a informação relativa aos pacientes utilizada neste estudo para cada um dos modelos.

Tabela 3 - Features usadas pelos modelos PURSUIT, TIMI e GRACE.

Modelo de risco	Informação de pacientes
PURSUIT	Idade Sexo Classificação Angina (Sem Angina ou CCS I/II ou CCS III/IV) Sinais de falha cardíaca Depressão ST ECG
TIMI	Idade Fatores de risco de CADs Uso de ASA (Aspirina) nos últimos 7 dias CADs conhecidas Episódio de repouso angina em menos de 24h Desvio do segmento ST Marcadores cardíacos elevados
GRACE	Idade Ritmo cardíaco (bpm) Pressão arterial sistólica (mmHg) Creatinina (mg/dL) Classificação Killip Paragem cardíaca na admissão Marcadores cardíacos elevados Desvio do segmento ST

Este trabalho, depois de avaliar a performance de cada um dos modelos, revelou que o modelo GRACE é aquele que apresenta a melhor performance ao prever o risco de morte ou MI (*Myocardial Infarction*), o enfarte de miocárdio. Desta forma, a seleção de *features* para este projeto é baseada no modelo GRACE.

De entre todas as *features* do *dataset* inicial, como forma de encontrar um paralelismo com as que são utilizadas pelo GRACE, alguns ajustes foram feitos. Ao comparar a Tabela 3 com a Tabela 2, que se encontra no início deste capítulo, pode verificar-se que praticamente todas as variáveis do modelo estão presentes no *dataset*, à exceção da Paragem cardíaca na admissão. Esta variável foi derivada a partir da classe Killip,

uma vez que quando esta se verifica num paciente como classe IV, pode caraterizar-se por episódios de paragem cardíaca.

Valores em falta

Para além da dimensionalidade já referida, um dos desafios de utilizar um *dataset*, principalmente quando os dados são reais, é a ausência de valores. Estes valores em falta podem impactar na performance do modelo ao criar *bias* no *dataset.* Existem várias alternativas para contornar estas situações, mas inicialmente é necessário perceber que dados estão efetivamente ausentes.

A Tabela 4*Tabela 4*, revela a percentagem de valores em falta existente em cada uma das *features* do *dataset* inicial.

Tabela 4 - Percentagem de dados em falta em variável.

Feature	**% de dados em falta**
Idade	0%
Ritmo cardíaco	0%
Pressão arterial sistólica	0.08%
Creatinina	10.6%
Classificação Killip	2.06%
Marcadores cardíacos	0%
Paragem cardíaca na admissão	0%
Desvio do segmento ST	2.07%

Como se pode observar pela Tabela 4, em algumas colunas não se verificavam valores em falta, mas noutras essa situação ocorre. A *feature* com maior percentagem de dados em falta é a creatinina, seguida pelo desvio do segmento ST, classe Killip e pressão sistólica. Para fazer face a esta situação, foi realizada uma limpeza dessas mesmas *features* que apresentavam valores em falta.

Descrição das *features*

A Tabela 5 contém uma representação de todas variáveis utilizadas. No que diz respeito às variáveis numéricas, como idade ou ritmo cardíaco, é apresentado o valor médio seguido do desvio padrão. Quanto às variáveis categóricas, como classe Killip ou paragem cardíaca na admissão, é disponibilizado o valor total de instâncias em que se ocorre cada uma das categorias, seguido da percentagem correspondente.

Tabela 5 - Descrição de features selecionadas para treinar os modelos.

Feature		**Evento**
Idade *(em anos)*		63.8±12.5
Ritmo/Frequência Cardíaca *(bpm)*		71.8±20.6
Pressão Sistólica *(mmHg)*		141.7±27.9
Creatinina *(mg/dL)*		1.4±2.2
Classe Killip	Classe 1 Classe 2 Classe 3 Classe 4	874 (83.7%) 97 (9.3%) 64 (6.1%) 9 (0.9%)
Paragem cardíaca na admissão	Sim Não	1035 (90.7%) 9 (9.3%)
Marcadores cardíacos	Sim Não	862 (77.7%) 247 (22.3%)
Desvios ST	Sim Não	729 (65.3%) 380 (34.3%)
Morte	Sim Não	947 (90.2%) 97 (9.2%)

Para que melhor se compreendam os valores do *dataset* e os que, posteriormente, serão fornecidos pelos modelos, seguem-se alguns dados sobre os valores limite de algumas das variáveis que correspondem a valores clínicos.

Em relação à frequência cardíaca, esta indica o número de vezes que o coração bate por minuto, podendo sofrer variações consoante a idade ou o esforço físico da pessoa num determinado momento. Os valores de frequência cardíaca mediante a idade da pessoa podem ser observados na Tabela 6.

Tabela 6 - Valores para frequência cardíaca por idade [32].

Tabela 7 - Valores para frequência cardíaca por idade [32].

Idade	Frequência máxima	50-70% do Máximo
20	200 bpm	100-140 bpm
30	190 bpm	95-133 bpm
40	180 bpm	90-126 bpm
50	170 bpm	85-119 bpm
60	160 bpm	80-112 bpm
70	150 bpm	75-98 bpm
80	140 bpm	70-98 bpm
90	130 bpm	65-91 bpm

Quanto à pressão sistólica, esta ocorre quando há contração dos músculos cardíacos de forma a esvaziar e impulsionar o sangue até às artérias. É representada por mm/Hg, o milímetro de mercúrio, que é uma unidade pressão. Existem também para esta medida um limite de valores dentro dos quais esta pressão se deve enquadrar. Como se consegue ver pela Tabela 7, são alguns os estados que se podem verificar para a pressão sistólica. O estado Normal é o desejável e aquele a que geralmente está associada uma pessoa saudável. O estado Elevado revela algum alerta e deve ter siso em atenção pelo profissional de saúde. Nos estados 1 e 2 de hipertensão é necessário ter bastante cuidado e pode ser recomendado iniciar medicação para a baixa pressão sanguínea. A crise hipertensa é o cenário mais grave, em que pode ocorrer dor no peito, fraqueza ou até mesmo problemas de visão.

Tabela 8 - Limite de valores para pressão sistólica [33].

Estado	Valores (em mm/Hg)
Normal	<120
Elevado	120-129
Estado 1 (hipertensão)	130-139
Estado 2	>140
Crise hipertensa	>180

A creatinina, por sua vez, é uma substância produzida pelos músculos e eliminada pelos rins. Quando em excesso ou em falta no organismo pode ser um sinal de alerta para problemas renais que estão bastantes ligados a problemas cardíacas. As

doenças relacionadas com o coração são mais comuns em pessoas com doenças renais. A creatinina apresenta também alguns valores limite, representados na Tabela 8.

Tabela 9 - Limite de valores para creatinina [34].

Indivíduo	Valores de Creatinina
Homem adulto	0.6 – 1.2 mg/dL
Mulher adulta	0.5 – 1.1 mg/dL
Homem idoso	0.6 – 1.7 mg/dL
Mulher idosa	0.6 – 1.6 mg/dL

Em relação à classificação Killip, esta foi proposta por Thomas Killip III and John T. Kimball em 1967. Foi baseada num exame físico de pacientes com um possível enfarte agudo do miocárdio (AMI) e usada para identificar os de maior risco de morte e com potenciais benefícios dos cuidados especializados em unidades de cuidados continuados. A partir deste estudo [35], obtiveram quatro tipos de classificação Killip, sendo eles:

- Killip 1: Sem evidências de insuficiência cardíaca.
- Killip 2: Descobertas de insuficiência cardíaca ligeira e moderada (pressão venosa jugular elevada) e som do terceiro coração (S3).
- Killip 3: APE (Edema pulmonar agudo).
- Killip 4: CHF (*Congestive heart failure*) com choque cardíaco definido por pressão arterial sistólica < 90 mmHg e vasoconstrição periférica.

Quanto aos marcadores cardíacos, esta variável representa a presença ou ausência de diversos biomarcadores de lesão cardíaca, como troponina e outros. Uma vez que a insuficiência circulatória provoca alterações celulares, algumas substâncias intracelulares aumentam o seu consumo, como é o caso da mioglobina, creatina quinase, troponinas, entre outras [36]. Desta forma, a presença destes biomarcadores pode revelar a existência de problemas cardíacos.

No que diz respeito aos desvios ST, esta é uma variável que indica a ausência ou presença dos mesmos no paciente. Tal como foi referido anteriormente, o desvio segmento ST ocorre em situações de síndrome coronária aguda, uma vez que a oclusão do vaso sanguíneo pode fazer diminuir o fluxo sanguíneo para o coração.

CAPÍTULO 5

RESULTADOS E DISCUSSÃO

Após a seleção das *features,* remoção dos valores em falta e de selecionados os algoritmos *Machine learning* a explorar, foram então aplicados os modelos com o objeto de perceber qual deles consegue assegurar melhor interpretabilidade, mas ao mesmo tempo uma boa performance.

5.1 Definição inicial

Antes de iniciar a aplicação dos algoritmos foi definida uma metodologia geral a seguir para cada um deles. Para a execução do modelo GRACE foram utilizados todos os dados do *dataset*, ou seja, a informação de todos os pacientes, uma vez que se trata de um modelo de pontuação e os resultados não são influenciados pelo balanceamento ou não do conjunto de dados utilizado. O mesmo se verificou com o *Clustering,* no qual não se verificou a necessidade de balanceamento.

Para os modelos Árvore de decisão e *Naive Bayes* os dados foram divididos em treino e teste, em proporções de 70% e 30%, respetivamente. Os dados de treino foram utilizados para treinar o modelo e os de teste usados para simular as previsões. É realizada esta divisão entre treino e teste porque se forem usados os dados na totalidade o algoritmo pode aprender a partir dos dados. Além disso, aos modelos de Árvore e Bayesiano, foram treinados sem balanceamento e com balanceamento. Este último conseguiu-se ao aplicar a técnica *Synthetic Minority Oversampling Technique* (SMOTE), com o objetivo de demostrar a influência do não balanceamento na performance dos modelos. O SMOTE é uma técnica de *oversampling*, no qual algumas amostras sintéticas são geradas para a classe minoritária, tendo com referência os dados existentes. Neste procedimento, pode escolher-se o nível de balanceamento pretendido sendo possível configurar o modelo para gerar apenas uma determinada proporção de instâncias [37].

5.2 Modelo GRACE

O GRACE segue um sistema de pontuação, baseado num registo internacional de uma população com Síndrome coronária aguda [31]. Foi selecionado para exploração no presente trabalho, uma vez que é interpretável e aplicado na prática clínica. Devido à atribuição de pontos que faz para cada *feature*, permite avaliar a gravidade de cada uma delas e do seu todo, através da soma dos pontos individuais de cada uma das variáveis.

Para a classificação, utiliza as variáveis Idade, ritmo cardíaco, pressão sistólica, creatinina, classe Killip, paragem cardíaca na admissão, marcadores elevados e desvio do segmento ST. No cálculo da pontuação de cada umas das classes referidas

anteriormente, é atribuída uma pontuação consoante os respetivos valores. Esta pontuação é atribuída tendo em consideração os scores da Tabela 9.

Tabela 10 - Variáveis de especificação de pontuação de risco do GRACE [31].

Variável	**Valores**	**Pontuação / *score***
Idade (anos)	<40 40 – 49 50 – 59 60 – 69 70 – 79 >= 80	0 18 36 55 73 91
Ritmo cardíaco (bpm)	<70 70 – 89 90 – 109 110 – 149 150 – 199 >200	0 7 13 23 36 46
Pressão Sistólica (mmHg)	<80 80 – 99 100 – 119 120 – 139 140 – 159 160 – 199 >200	63 58 47 37 26 11 0
Creatinina (mg/dL)	0 – 0.39 0.4 – 0.79 0.8 – 1.19 1.2 – 1.59 1.6 – 1.99 2 – 3.99 >4	2 5 8 11 14 23 31
Classe Killip	I II III IV	0 21 43 64
Paragem cardíaca na admissão	Não Sim	0 43
Marcadores cardíacos elevados	Não Sim	0 15
Desvio do segmento ST	Não Sim	0 30

Para calcular a pontuação de risco de cada paciente, é necessário verificar qual a pontuação correspondente a cada uma das variáveis que apresenta. Posteriormente, calcula-se a soma de todas as pontuações e obtém-se a pontuação final.

No final, depois de obter a pontuação total da instância valida-se esse valor de acordo com:

- Se pontuação < 141, o paciente corre um baixo risco cardíaco (Morte = 0).
- Se pontuação ≥ 141, o paciente corre um alto risco cardíaco (Morte = 1).

Desta forma, obtém-se uma previsão de risco cardíaco para cada um dos pacientes do *dataset*.

5.2.1 Resultados do método GRACE

Uma vez calculada a previsão de risco para cada paciente, fez-se uma comparação destes valores com os valores reais e desta forma obteve-se a matriz de confusão da Figura 7.

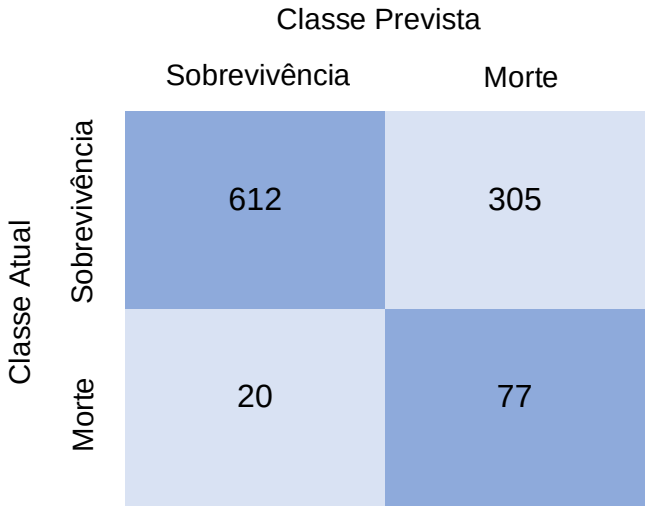

Figura 7 - Matriz de confusão obtida para o modelo GRACE.

Com base na matriz de confusão obtida, foram obtidos os resultados de sensibilidade, especificidade, *G-mean* e *accuracy* para o modelo, representados na Tabela 10.

Tabela 11 - Resultados obtidos usando o modelo GRACE.

Métrica	Valor
Sensibilidade	79.4%
Especificidade	64.6%
G-mean	71.6%
Accuracy	65.9%

Em relação à Sensibilidade verificou-se um valor de 79.4%, que indica uma boa proporção de casos de morte que efetivamente foram calculados como morte e a

Especificidade foi de 64.6%, o que indica também uma percentagem de casos de sobrevivência classificados como tal. Por sua vez, para a métrica *G-mean* obteve-se um valor de 71.6%.

Para o modelo GRACE os valores de *G-mean* variam entre 60% e 70% e, como o valor obtido no trabalho não se distancia muito dos valores normais, pode considerar-se que o modelo está a prever corretamente. Verifica-se também, pela *accuracy*, que o modelo prevê corretamente 65.9% dos casos.

5.3 Resultados do modelo Árvore de decisão

Como foi referido do Capítulo 3, existem alguns parâmetros que podem ser alterados no modelo de árvore de decisão. Um deles é a profundidade da árvore, o *max_depth*. O objetivo foi perceber de que forma varia a performance com ou aumento ou diminuição da profundidade da árvore. Por outro lado, também é importante perceber como é que estes fatores podem influenciar a interpretabilidade da árvore. Como tal, fez-se variar o tamanho da árvore (*max-depth*) entre 4 e 10. Para dados não balanceados, e para o conjunto de dados de teste, foram obtidos os resultados das Tabelas 11 (profundidade 4) e 12 (profundidade 10).

*Tabela 12 - Resultados obtidos com profundidade de **4**, SEM balanceamento.*

Sensibilidade	**Especificidade**	***G-mean***	***Accuracy***
30.3%	98.9%	54.8%	91.7%

*Tabela 13 - Resultados obtidos com profundidade de **10**, SEM balanceamento.*

Sensibilidade	**Especificidade**	***G-mean***	***Accuracy***
54.2%	99.7%	73.5%	96.2%

Como é percetível pelos resultados anteriores, quanto maior é a profundidade da árvore, melhores são os resultados. No entanto, apesar dos elevados valores de *accuracy,* que podem ser justificados pelo não balanceamento dos dados, e dos tambémelevados valores de especificidade, a sensibilidade obteve resultados inaceitáveis, que se refletem nos valores de G-mean.

Como foi referido anteriormente, a utilização de um *dataset* não balanceado faz com que o modelo tenda a classificar as instâncias como pertencentes à classe maioritária. Após a aplicação da técnica SMOTE para balanceamento dos dados, foram obtidos os resultados das Tabelas 13 e 14.

*Tabela 14 - Resultados obtidos para profundidade de **4**, COM balanceamento.*

Sensibilidade	Especificidade	*G-mean*	*Accuracy*
55.6%	92.3%	40.5%	46.8%

*Tabela 15 - Resultados obtidos para profundidade de **10**, COM balanceamento.*

Sensibilidade	Especificidade	*G-mean*	*Accuracy*
77.8%	94.4%	71.2%	89.2%

Ao aplicar o balanceamento aos dados obtiveram-se melhores valores de sensibilidade e valores mais realísticos também para a *accuracy*. A especificidade manteve-se alta em ambos os casos, mas a alta sensibilidade verificada nos dados balanceados fez com que a G-mean nestes casos seja também melhor e bastante satisfatória (71.2%).

Na sequência da aplicação da árvore de decisão, obteve-se ainda a árvore em si. Para uma profundidade 4 a árvore correspondente é a da Figura 8.

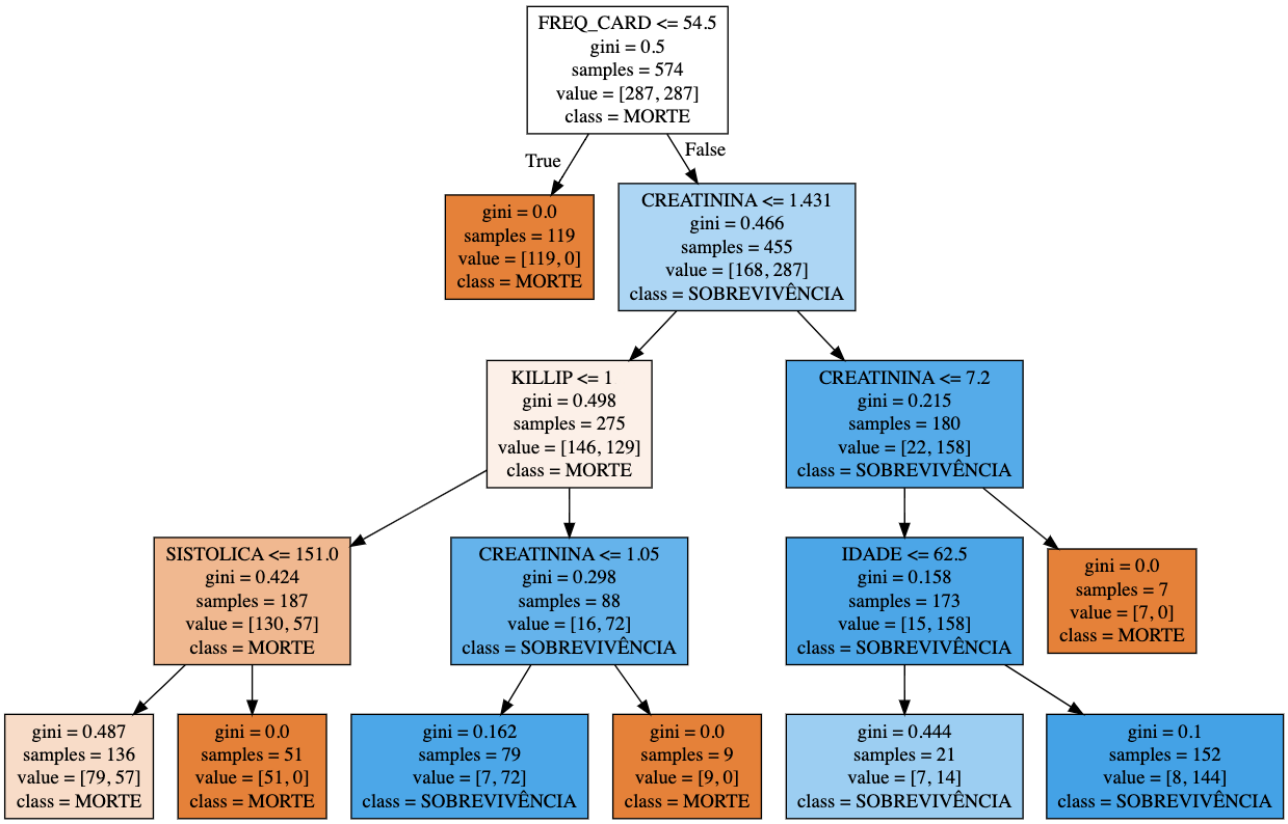

Figura 8 - Árvore de decisão com profundidade de 4.

Pela observação da árvore anterior, é possível retirar 8 regras:

1. **SE** Frequência cardíaca < 54.5 **ENTÃO** Morte.

2. **SE** Frequência cardíaca ≥ 54.5 **E** Creatinina ≤ 1.431 **E** Killip ≤ 1 **E** Pressão Sistólica ≤ 151 **ENTÃO** Morte.

3. **SE** Frequência cardíaca ≥ 54.5 **E** Creatinina ≤ 1.431 **E** Killip ≤ 1 **E** Pressão Sistólica ≥ 151 **ENTÃO** Morte.

4. **SE** Frequência cardíaca ≥ 54.5 **E** 1.05 < Creatinina ≤ 1.43 **E** Killip ≥ 1 **ENTÃO** Sobrevivência.

5. **SE** Frequência cardíaca ≥ 54.5 **E** 1.05 < Creatinina ≤ 1.43 **E** Killip ≥ 1 **ENTÃO** Morte.

6. **SE** Frequência cardíaca ≥ 54.5 **E** 1.43 < Creatinina ≤ 7.2 **E** Idade ≤ 62.5 **ENTÃO** Sobrevivência.

7. **SE** Frequência cardíaca ≥ 54.5 **E** 1.43 < Creatinina ≤ 7.2 **E** Idade ≥ 62.5 **ENTÃO** Sobrevivência.

8. **SE** Frequência cardíaca ≥ 54.5 **E** Creatinina ≥ 7.2 **ENTÃO** Morte.

As regras fornecidas pela árvore de decisão permitem, perante um novo conjunto de dados de um paciente, retirar algumas conclusões sobre o seu risco cardíaco.

Neste caso, por exemplo, a primeira revela que um individuo que apresente uma frequência cardíaca menor do que 54.5 bpm pode sofrer Morte. Pela Tabela 5, correspondente aos valores limite de frequência cardíaca consoante a idade, percebe-se que uma frequência cardíaca mais baixa pode ocorrer em pessoas com idade mais avançada, mas valores menores do que 54.5 bpm estão fora dos limites e representam perigo.

Por outro lado, pela segunda regra, se a frequência cardíaca for maior ou igual a 54.5 bpm, se a classe Killip for 1 (sem evidências de insuficiência cardíaca) e ao mesmo tempo a pressão sistólica for menor ou igual 151.0 mm/Hg, também existe uma grande probabilidade de Morte. Isto pode dever-se aos valores de pressão sistólica poderem ser altos, uma vez que devem estar abaixo de 120 mm/Hg. Também o valor de índice gini no nó final que classifica este resultado é de 0.487. Tal como foi mencionado anteriormente no capítulo 4, este valor traduz a existência de uma distribuição equilibrada das classes Morte e Sobrevivência neste nó e, desta forma, a classificação acaba por não ser 100% confiável. Nestas condições, o indivíduo poderá morrer ou sobreviver.

Segundo a terceira regra, pode acontecer que se um indivíduo apresentar condições idênticas às anteriores, mas uma pressão sistólica maior do que 151 mm/Hg, tem grande probabilidade de Morte. A pressão sistólica, quando superior a 140, começa a ser preocupante e por essa razão, os valores indicados por esta regra são preocupantes e podem estar associados ao perigo de vida.

Observando a quarta regra, percebe-se que se os valores de frequência cardíaca forem maiores do que 54.5 bpm, os valores de creatinina forem menores do que 1.431 mg/dL ou 1.05 mg/dL, mas as funções Killip foram as 2, 3 ou 4 (situações mais graves como edema pulmonar ou falha cardíaca), pode verificar-se alguma probabilidade de sobrevivência. Neste caso, o *gini index* foi de 0.162, o que significa que em algumas situações pode ocorrer Morte. Isto poderá acontecer provavelmente devido às funções Killip presentes serem as mais graves neste caso.

Quanto à regra 5, esta indica que um indivíduo com estas caraterísticas pode ser classificado como Morte, uma vez que as condições são muito semelhantes às anteriores, mas os valores de creatinina podem ser ainda mais elevados e dessa forma aumentar o risco.

Por outro lado, a sexta regra indica-nos que se uma pessoa apresentar frequência cardíaca maior do que 54.5 bpm, valores de Creatinina entre 1.43 mg/dL e 7.2 mg/dL e se a idade for menor do que 62.5, então pode verificar-se a sobrevivência. No entanto, esta classe tem *gini index* de 0.44. Este valor intermédio, tal como foi referido anteriormente, pode indicar uma incerteza na classificação da classe e por isso, tanto pode identificar este caso como Morte ou Sobrevivência.

De acordo com a sétima regra, se as condições reunidas forem semelhantes às anteriores, mas se a idade for superior a 63 anos, é registada também a sobrevivência do indivíduo. Isto pode dever-se ao facto de os idosos apresentarem naturalmente valores mais altos de Creatinina e esses valores não estarem a ter grande influência nas amostras classificadas como sobrevivência.

Por último, a oitava regra pode classificar como Morte todos os indivíduos com valores de creatinina superiores a 7.2 mg/dL. Tal como se pode validar pela Tabela 7, indicativa dos valores limite para esta variável, valores muito altos de creatinina são bastante fora dos parâmetros e por isso representam um perigo de vida para o paciente.

Após a validação anterior, seria também interessante recorrer à validação através do *input* por parte de uma equipa clínica perita no domínio do problema, com o objetivo de avaliarem as regras extraídas da árvore.

Também para o modelo executado com o *dataset* balanceado através do SMOTE e com uma profundidade de 10, foi gerado uma árvore, representada na Figura 9.

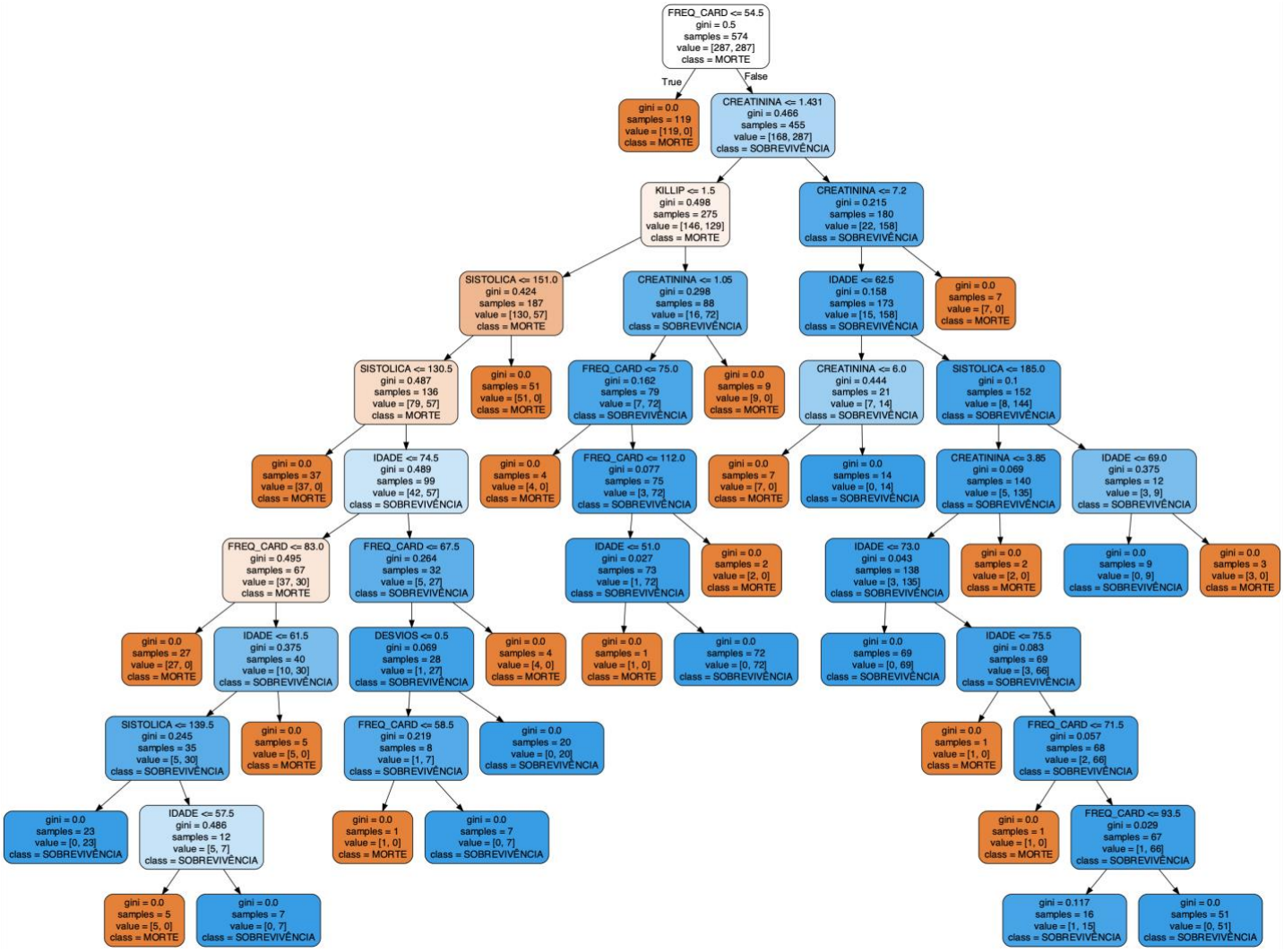

Figura 9 - Árvore de decisão com profundidade de 10.

Tal como é percetível pela Figura 9, uma árvore de profundidade maior, apesar de ter resultados melhores em termos de performance, tem muito pouca interpretabilidade. A árvore torna-se muito mais complexa e como tal difícil de analisar.

5.4 Resultados *Naïve Bayes*

Neste subcapítulo avalia-se o desempenho do classificador Bayesiano na previsão do risco cardiovascular. A Figura 10 representa o mecanismo de decisão Bayeasiano, no contexto do tema em estudo [38].

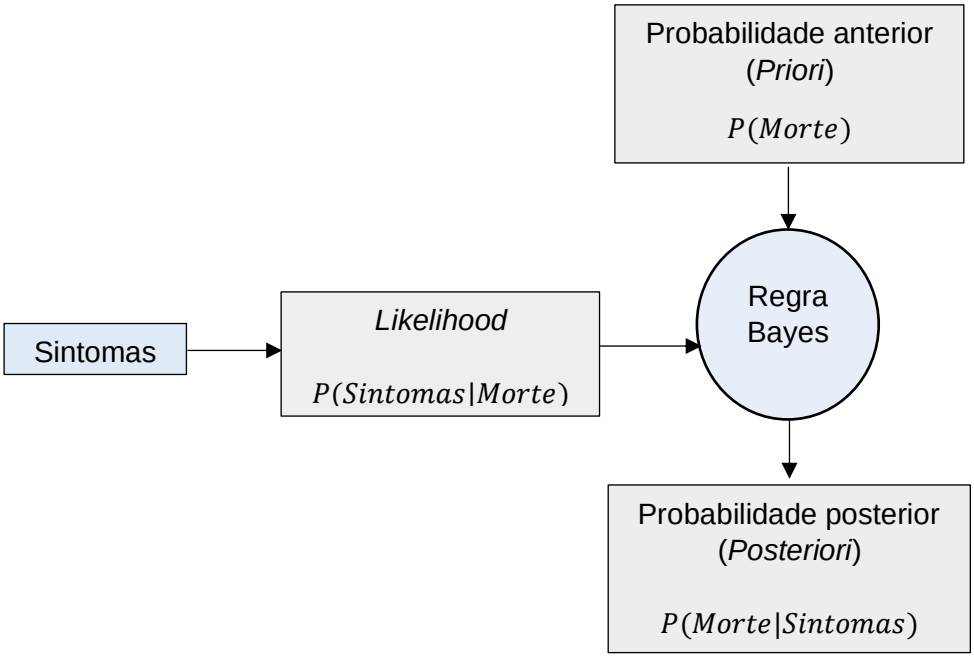

Figura 10 - Representação esquemática da rede Bayesiana. Adaptado de [38].

Pela Figura 10, os dados, que são os sintomas dos pacientes, são usados para encontrar uma probabilidade (*Likelihood*), que é a probabilidade de os sintomas ocorrerem, dado que ocorre a Morte ou a Sobrevivência do paciente. É também fundamental calcular as probabilidades anteriores, de Morte e Sobrevivência e combinar tudo, de forma a obter a probabilidade Posteriori. Portanto, ao usar o modelo Bayesiano pretende-se combinar a probabilidade anterior com o conhecimento já adquirido e obter a probabilidade (*Posteriori*) de o doente sofrer ou não a morte, consoante os sintomas observados.

Para simplificar, o resultado morte vai ser representada por M, sendo M = {M, $\overline{\mathrm{M}}$}, em que M corresponde à morte e $\overline{\mathrm{M}}$ à sobrevivência do paciente. Também X são as variáveis correspondentes aos sintomas dos pacientes.

Tal como referido, foi realizada a separação dos dados em treino e teste, com uma proporção de 70-30%. Foram usados os dados de treino e, uma vez que estes não se encontravam balanceados, aplicou-se o SMOTE para tentar obter melhores resultados. O SMOTE pretende criar mais instâncias de teste como base as que já existem. Este algoritmo foi aplicado com uma proporção de 60% e no final obteve-se a seguinte quantidade de dados, num total de 464 amostras:

Morte = 174

Sobrevivência = 290

Assim, as probabilidades de Morte e Sobrevivência são, respetivamente:

$P(\mathrm{M}) = 0.375$

$P(\overline{\mathrm{M}}) = 0.625$

As probabilidades anteriores são denominadas como *Priori*, uma vez que representam o que é conhecido sobre o resultado antes da observação. E, depois de calculada a *Priori*, foi necessário calcular a probabilidade *Likelihood*, que representa a probabilidade condicional de cada variável acontecer sendo que um evento se verifica, neste caso Morte ou Sobrevivência).

A probabilidade condicional de cada variável pode ser deduzida através da observação ou pelo cálculo distribuição normal que apresentam. Quando uma probabilidade é calculada através da observação, o que acontece é que, por exemplo para o cálculo de uma probabilidade $P(X_i = y|\mathrm{M})$, são verificados todos os registos onde ocorre Morte e, dentro desses, aqueles em que X_i tem o valor y. Realiza-se a divisão entre o total de registos em que X_i=y pelo total de casos de morte e obtém-se a probabilidade. Quanto à distribuição normal ou gaussiana, esta é usada para dar uma equação (16) de probabilidade para cada variável a que se aplica. Através da definição da função densidade de probabilidade para cada *feature*, com base na média e desvio padrão, é possível calcular a probabilidade de qualquer sintoma ocorrer, sabendo que ocorreu morte ou sobrevivência.

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \qquad (16)$$

Onde σ corresponde ao desvio padrão e μ é a média.

Para as *features* numéricas Idade, Frequência cardíaca, Pressão sistólica e Creatinina, as suas probabilidades foram calculadas através de uma função de distribuição normal.

Idade

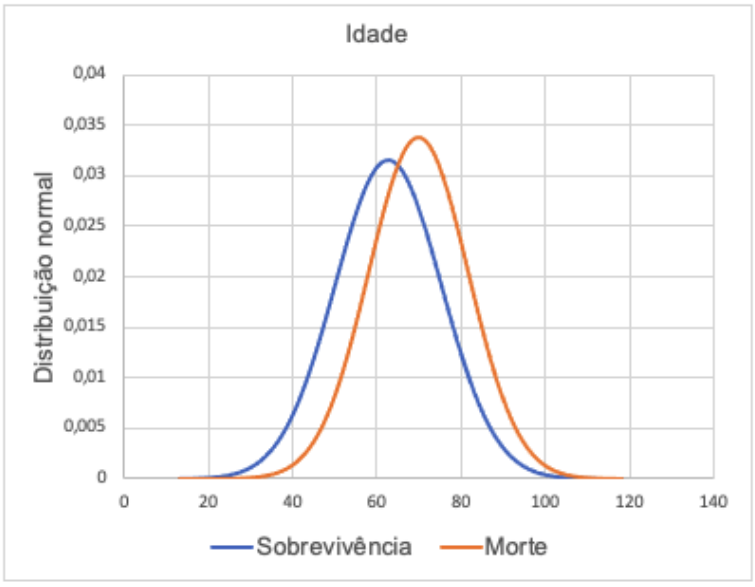

Figura 11 - Distribuição normal de Idade para Morte e Sobrevivência.

Em relação à idade, o gráfico que relaciona a idade dos pacientes com a distribuição normal, em caso de morte ou de sobrevivência é representado na Figura 11.

Segundo o gráfico anterior, é percetível que nos casos em que se registou a morte, a idade é também mais avançada do que nos casos de sobrevivência. Da mesma forma, também a idade média é mais elevada nos casos de morte. Através do gráfico, foi possível calcular os valores de Idade média e desvio padrão para os pacientes que morreram e que sobreviveram (Tabela 15).

Tabela 16 - Valores de média e desvio padrão para Idade.

Morte		**Sobrevivência**	
μ	σ	μ	σ
69.9	11.8	62.8	12.6

Tendo já os valores de idade média e desvio padrão, foi possível definir uma função gaussiana de idade para as situações em que ocorre a morte (17).

$$f(X) = \frac{1}{11.8\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\mathrm{X}-69.9}{11.8}\right)^2} \quad (17)$$

E também uma função para as situações em que não ocorre morte (18):

$$\mathrm{f(X)} = \frac{1}{12.6\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{\mathrm{X}-62.8}{12.6}\right)^2} \quad (18)$$

X corresponde a um valor de idade.

Para uma idade, por exemplo, de 65 anos, podem assim aplicar-se as equações (17) e (18) e calcular as probabilidades condicionais desta variável, da seguinte forma:

$$P(Idade = 65|\mathrm{M}) = 0.035$$

$$P(Idade = 65|\overline{\mathrm{M}}) = 0.021$$

Pressão sistólica

Como aconteceu com a Idade, também para a Pressão sistólica foi possível obter o gráfico da que relaciona a distribuição normal com a pressão, em casos de morte e de sobrevivência.

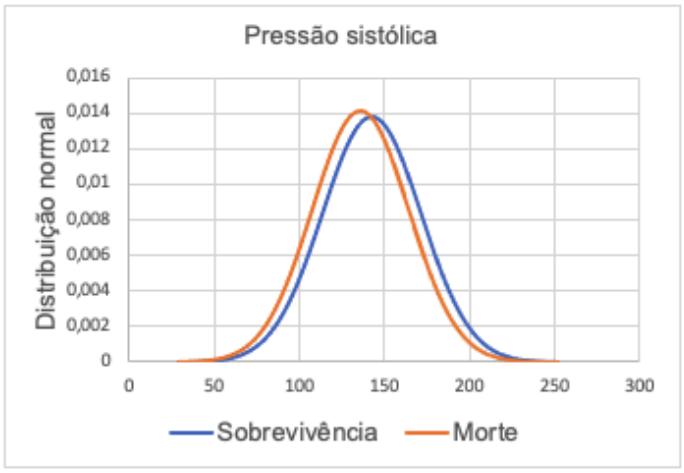

Figura 12 - Distribuição normal de Pressão sistólica para Morte e Sobrevivência.

De acordo com o gráfico da Figura 12, os valores de pressão sistólica podem ser menores para pacientes em que aconteceu morte do que para aqueles que sobreviveram. Este é um caso assumido em que uma suposição obtida através de um modelo interpretável pode não ser a mais correta. Neste caso, o modelo previu que a morte está associada a valores de pressão sistólica mais baixos, enquanto a sobrevivência se relaciona com valores superiores, mas o cenário correto seria o contrário. Por norma, quanto maior a pressão sistólica maior é o risco que lhe está associado.

Através do gráfico foi também possível extrair a média e o desvio padrão para esta variável, presentes na Tabela 16.

Tabela 17 - Valores de média e desvio padrão para Pressão sistólica.

Morte		**Sobrevivência**	
μ	σ	μ	σ
135.7	28.3	142.3	28.9

Com base nos valores de pressão média e desvio padrão da Tabela 16, foi definida uma função gaussiana para as situações em que ocorre a morte (19) e sobrevivência (20).

$$f(X) = \frac{1}{28.31\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-135.69}{28.31}\right)^2} \quad (19)$$

$$f(X) = \frac{1}{28.85\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-142.29}{28.85}\right)^2} \quad (20)$$

Onde X corresponde a um valor de pressão sistólica.

Para uma pressão sistólica de 130 mm/Hg, como forma de calcular $P(X = 130|M)$ basta utilizar a equação (19) e para o cálculo de $P(X = 130|\bar{M})$ aplicar (20). Desta forma, obtém-se:

$$P(X = 130|\mathrm{M}) = 0.014$$

$$P(X = 130|\bar{\mathrm{M}}) = 0.013$$

Frequência cardíaca

Em relação à Frequência cardíaca, também foi possível relacionar a distribuição normal desta variável com os casos de morte e de sobrevivência, como se pode ver pela Figura 13.

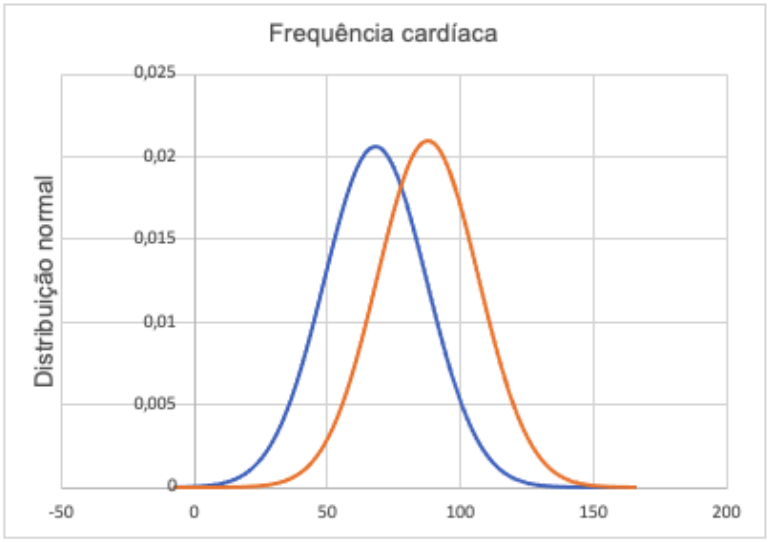

Figura 13 - Distribuição normal de Frequência cardíaca para Morte e Sobrevivência.

Pelo gráfico anterior é percetível que os valores de frequência cardíaca são maiores nos casos de morte do que nos casos de sobrevivência.

Tabela 18 - Valores de média e desvio padrão para Frequência cardíaca.

Morte		**Sobrevivência**	
μ	σ	μ	σ
87.9	19.0	68.2	19.3

Depois de ter os valores de média e de desvio padrão, presentes na Tabela 17, é possível definir as funções normais para as situações em que ocorre morte (21) e sobrevivência (22).

$$f(X) = \frac{1}{19.0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-87.9}{19.0}\right)^2} \qquad (21)$$

$$f(X) = \frac{1}{19.3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-68.2}{19.3}\right)^2} \quad (22)$$

Onde X corresponde a um valor de frequência cardíaca.

Assim, para um valor de frequência cardíaca de 60 bpm, para calcular $P(X = 60|M)$ basta utilizar a equação (21) e para calcular $P(X = 60|\bar{M})$ aplicar a equação (22). Obtém-se assim:

$$P(X = 60\ |\mathrm{M}) = 0.075$$

$$P(X = 60\ |\bar{\mathrm{M}}) = 0.019$$

Creatinina

Para a Creatinina também foi possível gerar um gráfico que relaciona a distribuição normal desta variável com as situações de Morte e de Sobrevivência, como se vê pela Figura 14.

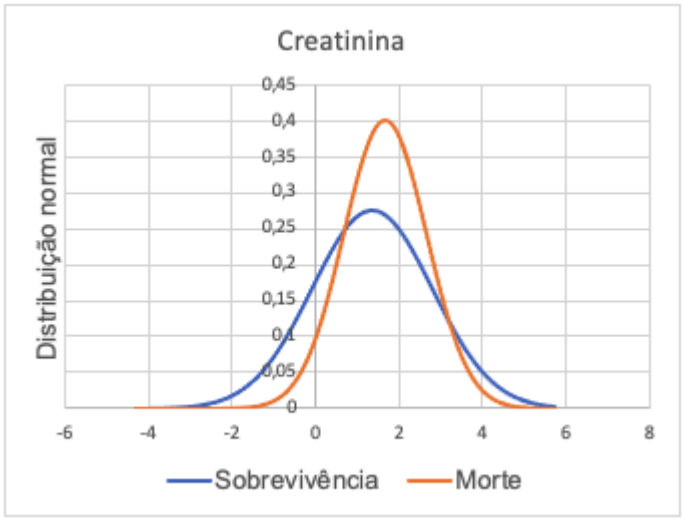

Figura 14 - Distribuição normal de Creatinina para Morte e Sobrevivência.

Pelo gráfico da Figura 14, pode perceber-se que os pacientes com valores de creatinina mais elevados têm uma maior de morte, ao contrário daqueles que têm valores de creatinina mais baixos. Tiraram-se através do gráfico os valores de média e desvio padrão para a Creatinina, representados na Tabela 18.

Tabela 19 - Valores de média e desvio padrão para Creatinina.

Morte		**Sobrevivência**	
μ	σ	μ	σ
0.9	1.7	1.3	1.4

Tendo já os valores de idade média e desvio padrão, foi possível definir uma função gaussiana para as situações em que ocorre a morte (23).

$$f(X) = \frac{1}{0.9\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-1.7}{0.9}\right)^2} \quad (23)$$

E também uma para as situações de sobrevivência (24):

$$f(X) = \frac{1}{1.4\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-1.3}{1.4}\right)^2} \quad (24)$$

Para um valor de Creatinina de 1.5 mg/dL, quando se pretende calcular $P(X = 1.5|M)$ basta utilizar a equação (23) e para calcular $P(X = 1.5|\bar{M})$ aplicar a equação (24). Obtém-se assim:

$$P(X = 1.5|\mathrm{M}) = 0.39$$

$$P(X = 1.5|\bar{\mathrm{M}}) = 0.28$$

Em relação às variáveis categóricas como classe Killip, biomarcadores, desvios ST e paragem, foi deduzida a probabilidade condicional tendo em conta a observação e obtiveram-se os valores seguintes:

Classe Killip

$$P(\mathrm{Killip} = 1|\mathrm{M}) = 0.59$$

$$P\left(\mathrm{Killip} = 1\middle|\overline{M}\right) = 0.85$$

$$P(\mathrm{Killip} = 2|\mathrm{M}) = 0.22$$

$$P\left(\mathrm{Killip} = 2\middle|\overline{\mathrm{M}}\right) = 0.09$$

$$P(\mathrm{Killip} = 3|\mathrm{M}) = 0.17$$

$$P\left(\mathrm{Killip} = 3\middle|\overline{M}\right) = 0.05$$

$$P(\mathrm{Killip} = 4|\mathrm{M}) = 0.02$$

$$P\left(\mathrm{Killip} = 4\middle|\overline{\mathrm{M}}\right) = 0.01$$

Biomarcadores

$$P(Biomarcadores = 1|\mathrm{M}) = 0.87$$

$$P\left(Biomarcadores = 1\middle|\overline{M}\right) = 0.78$$

$$P(Biomarcadores = 0|\mathrm{M}) = 0.13$$

$$P\left(Biomarcadores = 0\middle|\overline{M}\right) = 0.22$$

Desvios ST

$$P(Desvios\ ST = 1|\mathrm{M}) = 0.75$$

$$P\left(Desvios\ ST = 1\middle|\overline{M}\right) = 0.65$$

$$P(Desvios\ ST = 0|\mathrm{M}) = 0.25$$

$$P\left(Desvios\ ST = 0\middle|\overline{M}\right) = 0.35$$

Paragem

$$P(Paragem = 1|\mathrm{M}) = 0.02$$

$$P\left(Paragem = 1\middle|\overline{M}\right) = 0.01$$

$$P(Paragem = 0|\mathrm{M}) = 0.98$$

$$P\left(Paragem = 0\middle|\overline{M}\right) = 0.99$$

Depois de calcular a probabilidade condicional de cada uma das variáveis, é necessário multiplicar todos os valores, de forma a obter a probabilidade de todos os sintomas, sabendo que ocorre morte e sobrevivência. Desta forma, $P(X|M)$ e $P(X|\overline{M})$:

$$P(X|M) = 0.035 \times 0.014 \times 0.075 \times 0.49 \times 0.59 \times 0.87 \times 0.75 \times 0.02 = 0.00000014$$

$$P(X|\overline{M}) = 0.021 \times 0.013 \times 0.019 \times 0.28 \times 0.85 \times 0.78 \times 0.65 \times 0.01 \ = \ 0.0000000063$$

No final, é necessário calcular as probabilidades *Posteriori*, P(M|X) e P(M|$\overline{\mathrm{X}}$), que vão dar informação sobre a probabilidade do paciente com as caraterísticas conhecidas X sofrer a morte ou sobreviver.

Esta probabilidade é dada por:

$$P(M|X) \ = \ P(X|M)\ P(M)$$

Substituindo os valores já obtidos anteriormente, verifica-se que:

$$P(M|X) = 0.00000014 \times 0.375 = \ 0.000000052$$

$$P(\overline{M}|X) = 0.0000000063 \times 0.625 = \ 0.0000000039$$

Como o objetivo é chegar a uma decisão que classifique um doente como Morte ou Sobrevivência, é necessário comparar a probabilidade de morte com a de sobrevivência da seguinte forma:

- Se $P(M|X) > P(\bar{M}|X)$, então Morte (M)
- Se $P(\bar{M}|X) > P(M|X)$, então Sobrevivência ($\bar{M}$)

Neste caso do exemplo dado, como $P(\bar{M}|X) > P(M|X)$ pode concluir-se que, com os sintomas considerados, é mais provável que não ocorra a morte.

No contexto do trabalho, calculou-se esta previsão para cada um dos pacientes do conjunto de dados de teste obteve-se a matriz de confusão da Figura 15 e os resultados da Tabela 19.

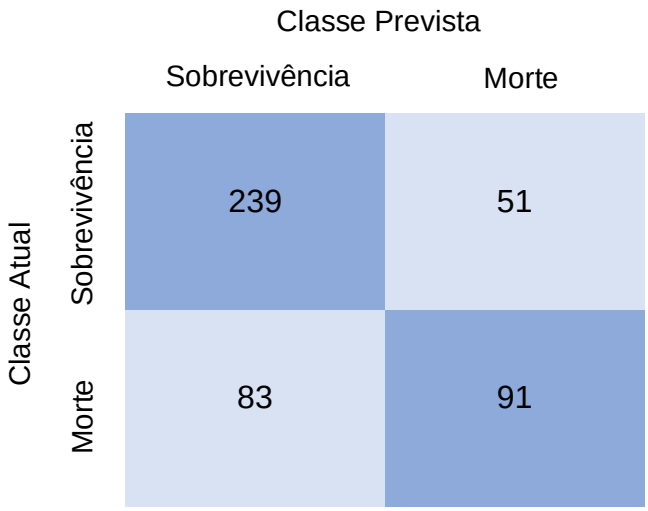

Figura 15 - Matriz de confusão do modelo Bayesiano, COM dados balanceados.

Tabela 20 - Resultados do modelo Bayesiano, COM dados balanceados.

Sensibilidade	**Especificidade**	***G-mean***	***Accuracy***
59.3%	82.4%	65.7%	71.1%

No que diz respeito aos resultados do modelo, presentes na Tabela 19, a Sensibilidade foi de 59.3%, o que revela que mais de metade dos dados de teste foram classificados corretamente como Morte. A Especificidade foi bastante elevada, o que revela uma maior capacidade deste modelo em avaliar casos reais de sobrevivência como pertencentes a essa mesma classe. A média geométrica é de 65.7%, que revela um bom equilíbrio entre as classificações de ambas as classes. A *accuracy* regista também valores bastante aceitáveis que indicam uma boa percentagem de previsões corretas atribuídas por este modelo.

Também foram realizados testes experimentais, como aconteceu na árvore de decisão, para o *dataset* original, sem qualquer balanceamento aplicado aos dados.

Os valores obtidos na matriz de confusão são os da Figura 16 e os resultados são os da Tabela 20.

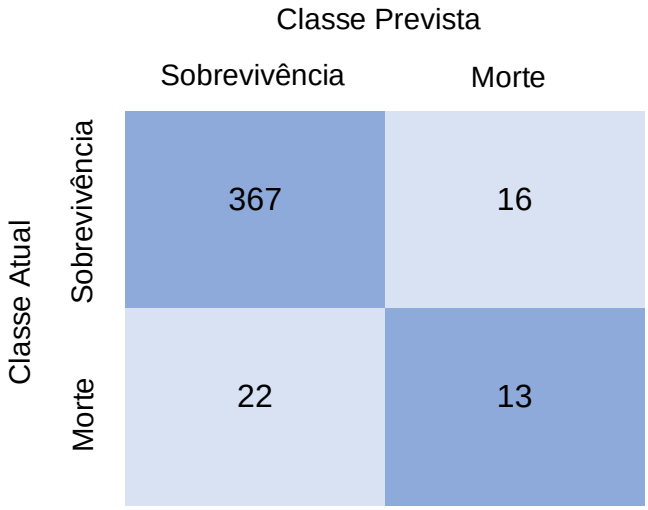

Figura 16 - Matriz de confusão do modelo Bayesiano, SEM dados balanceados.

Tabela 21 - Resultados obtidos para modelo Bayesiano, SEM dados balanceados.

Sensibilidade	**Especificidade**	***G-mean***	***Accuracy***
37.1%	95.8%	59.7%	90.9%

Os resultados apresentados na Tabela 20 comprovam que, mais uma vez, o facto de os resultados serem pouco balanceados, afeta a performance do *dataset*. No entanto, no classificador *Naive Bayes*, essa diferença não foi muito notada, uma vez que este apresentou resultados satisfatórios.

A Sensibilidade em classificar resultados de Morte como sendo efetivamente pertencentes a essa classe é relativamente baixa, ao contrário do que acontece com a Especificidade. O modelo acabou por classificar corretamente as instâncias como pertencentes à classe maioritária, Sobrevivência. Quanto à média geométrica, esta não se revelou muito baixa, revelando algum equilíbrio na classificação por parte do modelo. A *accuracy* é muito elevada, uma vez que houve uma elevada percentagem de instâncias classificadas como Sobrevivência, mas por outro lado, a Precisão é mediana uma vez que as classes que foram classificadas corretamente como Morte foram relativamente poucas.

5.5 Resultados *Clustering*

Neste subcapítulo, tenta perceber-se qual o desempenho do *Clustering*, quando aplicado no contexto da avaliação do risco cardiovascular. Tal como referido anteriormente, o grande objetivo do *Clustering* é que os dados sejam agrupados entre si com base na sua semelhança. Desta forma, neste estudo começaram por se dividir

os dados em dois grandes grupos: os pacientes que sofreram morte e aqueles que sobreviveram. Para cada um destes grupos, foi gerado um centro/*cluster* através do método K-means. Cada cluster é representado por 8 valores, sendo cada um deles correspondente a uma *feature* (idade, frequência cardíaca, pressão sistólica, creatinina, classe killip, biomarcadores, desvios ST, paragem na admissão. São eles:

$$C1 = [70.5, 89.2, 136.5, 1.7, 1.8, 0.8, 0.78, 0.03]$$

$$C0 = [63.2, 70.9, 142.3, 1.3, 1.2, 0.7, 0.67, 0.01]$$

Sendo que $C1$ corresponde ao cluster dos pacientes que morreram (1) e $C0$ o cluster dos pacientes que sobreviveram (0). Os *clusters* criados permitem determinar, para cada paciente, qual a classificação associada às suas caraterísticas. Através do cálculo da distância euclidiana entre o *cluster* e os dados do paciente, é possível determinar se este poderá ser classificado como morte ou como sobrevivência.

Relembrando a equação da distância euclidiana (25):

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)} \quad (25)$$

A título de exemplo, considerando um paciente com 68 de anos de idade, frequência cardíaca de 52 bpm, pressão sistólica de 100 mmHg, creatinina de 1.3 mg/dL, classe Killip 1, presença de biomarcadores, registo de desvios do segmento ST e sem registo de paragem cardíaca na admissão, representado por:

$$P(Paciente) = [68, 52, 100, 1.3, 1, 1, 1, 0]$$

Como forma de classificar o paciente anterior é realizado o cálculo da distância euclidiana entre cada cluster, $C1$ e $C0$, e os dados do paciente $P(Paciente)$:

$$d1(P, C1) = \sqrt{(70.5 - 68)^2 + (89.2 - 52)^2 + (136.5 - 100)^2 + (1.7 - 1.3)^2 + (1.8 - 1)^2 + (0.8 - 1)^2 + (0.78 - 1)^2 + (0.03 - 0)^2} = 52.18$$

$$d0(P, C0) = \sqrt{(63.2 - 68)^2 + (70.9 - 52)^2 + (142.3 - 100)^2 + (1.3 - 1.3)^2 + (1.2 - 1)^2 + (0.7 - 1)^2 + (0.67 - 1)^2 + (0.01 - 0)^2} = 46.58$$

Para eliminar a importância do valor absoluto das *features* e dessa forma evitar o enviesamento no cálculo das distâncias, optou-se pela normalização das *features*. Assim, de forma a determinar uma distância entre 0 e 1, que tornasse mais fácil a análise dos resultados, optou-se pela normalização definida em (26). Desta forma, garante-se que o 0 é o mais próximo dos pacientes que sobreviveram e o 1 é o mais perto daqueles que morreram. Existem vários métodos de normalização, sendo que neste trabalho se optou por esta estratégia, uma vez que tira partido diretamente da distância dos centróides (morte / sobrevivência).

$$d = 1 - \frac{d1}{d1 + d0}, \text{ sendo } d = [0,1]. \quad (26)$$

Neste contexto, se a distância for $d \geq 0.5$, o paciente é classificado como Morte (1) e se $d < 0.5$ o paciente é classificado como Sobrevivência (0).

Ao aplicar a normalização ao exemplo anterior:

$$d = 1 - \frac{52.18}{52.18 + 46.58} = 0.53$$

Nesta situação, como $d \geq 0.5$, então o paciente é classificado como Morte (1).

O método descrito anteriormente foi aplicado para todos os pacientes, tendo-se obtido os resultados da Figura 17.

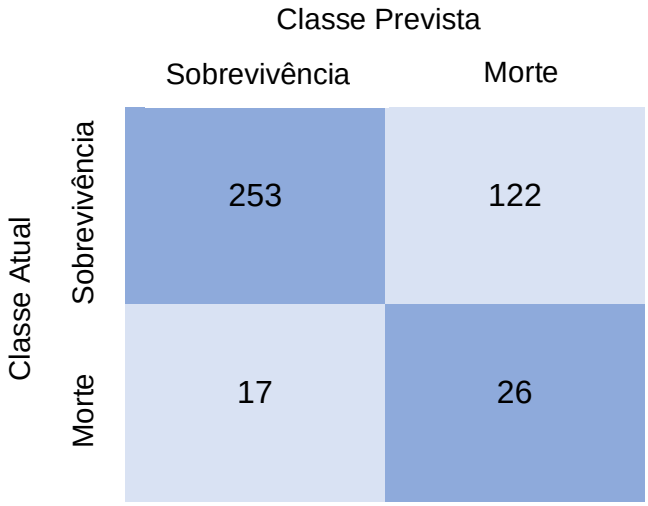

Figura 17 - Matriz de confusão obtida para Clustering.

Tabela 22 - Resultados do Clustering.

Sensibilidade	**Especificidade**	***G-mean***	***Accuracy***
60.5%	67.4%	63.9%	66.7%

Ao analisar os resultados da Tabela 21, conclui-se que o *Clustering* consegue obter bons resultados. A média geométrica, de 63.9%, revela que existe um equilíbrio satisfatório na classificação por parte do modelo. De entre todos os valores obtidos, a especificidade foi aquela que apresentou os melhores resultados, o que revela uma boa capacidade deste modelo em avaliar casos de sobrevivência como pertencentes realmente a essa mesma classe.

5.6 Considerações adicionais

Depois de testados todos os algoritmos, é necessário perceber qual deles consegue o melhor compromisso entre a performance e a interpretabilidade. A Tabela 22 resulta da compilação de todos os resultados obtidos. Não se consideraram para esta análise dos resultados, aqueles que foram obtidos ao realizar testes com *dataset* não balanceado, uma vez que estes se revelarem piores em relação aos obtidos para o *dataset* balanceado.

Tabela 23 - Compilação de resultados de todos os modelos.

		Sensibilidade	**Especificidade**	***G-mean***	***Accuracy***
GRACE		79.4%	64.6%	71.6%	65.6%
Árvore de decisão	Profundidade 4	55.6%	92.3%	40.5%	46.8%
	Profundidade 10	77.8%	94.4%	71.2%	89.2%
Modelo Bayesiano		59.3%	82.4%	65.7%	71.1%
Clustering		60.5%	67.4%	63.9%	66.7%

A Árvore de decisão é, de entre os três classificadores em estudo, aquele que parece ter tido os piores resultados, quando a sua profundidade foi restringida a 4. Apesar da grande interpretabilidade que demostra, pela extração intuitiva de regras, para conseguir obter um melhor desempenho tem de comprometer a interpretabilidade ao aumentar a profundidade. Os resultados para uma árvore mais extensa foram bastante melhores, mas essa árvore não consegue ser tão interpretável para o utilizador. O aumento da performance é conseguido à custa de uma degradação muito acentuada da interpretabilidade.

Em relação ao modelo Bayesiano, este obteve valores bastantes bons de performance. Com uma média geométrica de 65.7%, que combina a sensibilidade e especificidade, é percetível que este modelo tem uma boa capacidade em identificar corretamente casos positivos, de morte, e também casos negativos, de sobrevivência. O *Naive Bayes*, para além dos bons resultados que apresenta, consegue ser bastante interpretável uma vez que, através do cálculo de probabilidades, ajuda a classificar os pacientes.

Por fim, ao analisar a Tabela 22, é percetível que o modelo que revelou ser o melhor ao nível da performance foi o GRACE. Esta evidência pode ser justificada pelo facto de este ser um modelo que foi desenvolvido propositadamente para a identificação do risco cardiovascular. O GRACE é resultado de vários anos de pesquisa e de trabalho dos clínicos. Os parâmetros usados por este modelo têm vindo a ser otimizados,

através do conhecimento clínico e, por essa razão, é espectável que dê bons resultados. Para além disso, é um método desenvolvido especificamente para a previsão do risco cardiovascular, enquanto os outros são modelos genéricos que foram aplicados neste problema, mas que podem ser usados em outros também. No entanto, como usa o sistema de pontos para a classificação de pacientes, torna-se menos interpretável, apesar dos bons resultados que apresenta.

É também importante destacar que, com os resultados e regras obtidas, se comprovou o que foi referido inicialmente. Nem sempre os modelos, apesar do bom desempenho, fornecem as regras mais corretas e interpretáveis. Devem sempre ser postas em questão essas mesmas regras, porque o modelo pode errar. Ou seja, não é por um modelo ser interpretável, que é igualmente correto, podendo este apresentar falhas. Depois da extração de interpretabilidade de um modelo *Machine learning*, é sempre importante fazer a respetiva validação.

CAPÍTULO 6

CONCLUSÕES

O principal objetivo do trabalho era explorar a interpretabilidade de modelos *Machine learning,* e tentar encontrar um equilíbrio entre esta e o desempenho de cada modelo, no âmbito do risco cardiovascular.

Começou por ser realizada a análise de um *dataset* com dados reais de pacientes clínicos provenientes dos hospitais de Santa Cruz e de Leiria. Foram-lhes aplicadas algumas técnicas de pré-processamento, entre as quais a limpeza de valores em falta e a seleção de *features*. Posteriormente, foram selecionados os modelos *Machine learning* sobre os quais incidiu o estudo e realizada uma investigação sobre cada um deles. De seguida, após a aplicação dos dados aos classificadores, foram realizados vários testes que possibilitaram a descoberta de regras e probabilidades condicionais para um dos modelos, com o objetivo de explorar a interpretabilidade dada pelos mesmos. Na fase final do trabalho, todos os modelos foram comparados em relação aos resultados de performance que obtiveram, tendo-se concluído que o modelo GRACE foi aquele que conseguiu a melhor relação entre a interpretabilidade e o desempenho.

O presente trabalho permitiu adquirir uma série de conhecimentos na área do *Machine learning* que anteriormente eram desconhecidos. Para além disso foi possível compreender também como podem ser criados sistemas mais interpretáveis, na área do risco cardiovascular, mas não só. A temática da inteligência artificial é bastante atual e é fundamental que todos os sistemas inteligentes utilizados no dia-a-dia sejam transparentes nas previsões que fornecem aos utilizadores. E isso é muito deve ser mais importante em áreas como a da saúde.

No decorrer do trabalho surgiram alguns desafios, muitas vezes relacionados com os algoritmos. Em certos momentos existiu alguma dificuldade em perceber como é que estes poderiam ser aplicados para no final conseguirem ter como resultado a interpretabilidade, mas tudo foi superado com sucesso.

Por todas estas razões, o trabalho desenvolvido foi acima de tudo desafiador, uma vez que permitiu a exploração de áreas totalmente desconhecidas, mas que são bastantes atuais.

No futuro, seria importante recorrer à validação das condições obtidas através dos modelos por parte de profissionais de saúde especializados na área, de forma a confirmar a sua veracidade.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] American Heart Association. What is Cardiovascular Disease? https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease. (2017, acedido em Novembro 2021).

[2] World Health Organization. Cardiovascular diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). (2021, acedido em Novembro 2021).

[3] Zachary Lipton. The Mythos of Model of Interpretability. https://dl.acm.org/doi/pdf/10.1145/3236386.3241340. (2018, acedido em Setembro de 2021).

[4] R. Bharti et Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Hindawi,* 2021.

[5] Jessica Kent. Health IT Analytics. https://healthitanalytics.com/news/google-uses-artificial-intelligence-to-detect-breast-cancer. (2020, acedido em Setembro de 2021)

[6] Rich Caruana, et. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. Microsoft, 2015.

[7] Muhammad A. Ahmad. Interpretable Machine Learning in Healthcare. https://www.researchgate.net/publication/328416903_Interpretable_Machine_Learning_in_Healthcare (2018).

[8] Sharayu Rane. The balance: Accuracy vs. Intepretability. https://towardsdatascience.com/the-balance-accuracy-vs-interpretability-1b3861408062 (2018, acedido em Setembro de 2021).

[9] Milton G. Neto. Explainable AI: extraindo explicações e aumentando a confianã dos modelos de ML. https://medium.com/data-hackers/explainable-ai-extraindo-explica%C3%A7%C3%B5es-e-aumentando-a-confian%C3%A7a-dos-modelos-de-ml-3a89b7b5a584 (2021, acedido em Outubro 2021).

[10] Roohallah Alizadehsani et al. A data mining approach for diagnosis of coronary artery disease. Computer Programs and methods in biomedicine, 2013, 111; 52-61.

[11] U. N. Dulhare, “Prediction system for heart disease using Naive Bayes and particle swarm optimization,” Muffakham Jah College of Engineering and Technology, India.

[12] Pawel Marciniak. Implementation of artificial intelligence methods on example of cardiovascular diseases risk stratification. https://ieeexplore.ieee.org/document/6872251. (2015, acedido em Novembro de 2021).

[13] Taiwo Ayodele, et Types of Machine Learning Algorithms. New Advances in Machine Learning, 2010.

[14] Y. Hajaj. Baeldung, Semi-supervised, Unsupervised and Reinforcement Learning. https://www.baeldung.com/cs/machine-learning-intro (2014, acedido em Octubro 2021).

[15] P. Gupta. Decision trees in Machine Learning. https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052. (2017, acedido em Novembro de 2021).

[16] V. Podgorelec, et al. Decision Trees: An Overview and Their Use in Medicine, Journal of Medical Systems, 2002.

[17] A. Chakure. Decision Tree Classification. https://medium.com/swlh/decision-tree-classification-de64fc4d5aac. (2018, acedido em Novembro 2021).

[18] T. Suryakanthi, et al. Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm. International Journal of Advanced Computer Science and Applications, 2020.

[19] Keboola. The Ultimate Guide to Decision Trees for Machine Learning. https://www.keboola.com/blog/decision-trees-machine-learning. (2020, acedido em Novembro de 2021).

[20] S. Visweswaran, et al. Learning patient-specific models from clinical data. University of Pittsburgh, 1987.

[21] R. S. Niculescu, et al. Exploiting Parameter Domain Knowledge for Learning in Bayesian Networks. Pittsburgh, 2005.

[22] W. Ali, et al. Intelligent Naïve Bayes-based approaches for Web proxy caching. Knowledge-Based Systems, 2012.

[23] N. Friedman, et al. Bayesian Network Classifiers. Kluwer Academic Publishers, 1997.

[24] Onesmus Mbaabu. Clustering in Unsupervised Machine Learning. https://www.section.io/engineering-education/clustering-in-unsupervised-ml/ (2020, acedido em Janeiro de 2021).

[25] S. Priy. Clustering in Machine Learning. https://www.geeksforgeeks.org/clustering-in-machine-learning/ (2021, acedido em Janeiro de 2021).

[26] P. Andritsos, et al. Data Clustering Techniques. University of Toronto, Department of Computer Science, Toronto, 2002.

[27] F. Bonesi. Euclidian distance: distance definition. http://mathematics-in-europe.eu/?p=835 (2017, acedido em Dezembro 2021).

[28] G. Lahera. Unbalanced Datasets & What To Do. https://medium.com/strands-tech-corner/unbalanced-datasets-what-to-do-144e0552d9cd (2018, accessed in November 2021).

[29] M. Altini. Dealing with imbalanced data: undersampling, oversampling and proper cross validarion. https://www.marcoaltini.com/blog/dealing-with-imbalanced-data-undersampling-oversampling-and-proper-cross-validation (2015, acedido em Novembro 2021).

[30] Softtek. Redução da dimensionalidade na Machine Learning. https://softtek.eu/pt-pt/tech-magazine-pt/artificial-intelligence-pt/reducao-da-dimensionalidade-na-machine-learning/ (2021, acedido em Outubro de 2021).

[31] P. A. Gonçalves, et al. TIMI, PURSUIT, and GRACE risk scores: sustained prognostic value and interaction with revascularization in NSTE-ACS. European Heart Journal, 2005; 26, 865–872.

[32] H. Foundation, How to check your pulse (heart rate). https://www.heartfoundation.org.nz/wellbeing/managing-risk/how-to-check-your-pulse-heart-rate. (2007, acedido em Novembro de 2021).

[33] M. Nayana Ambardekar. Diastole vs. Systole: Know Your Blood Pressure Numbers. https://www.webmd.com/hypertension-high-blood-pressure/guide/diastolic-and-systolic-blood-pressure-know-your-numbers. (2021, acedido em Novembro de 2021).

[34] C. P. Davis. Creatinine Blood Test (Normal, Low, High Levels). https://www.medicinenet.com/creatinine_blood_test/article.html (2021, acedido em Outubro de 2021).

[35] H. G. Mello. Validation of the Killip–Kimball Classification and Late Mortality after Acute Myocardial Infarction. Instituto Dante Pazzanese de Cardiologia, São Paulo, 2014.

[36] Fleury. Marcadores bioquímicos de lesão cardíaca. https://www.fleury.com.br/medico/artigos-cientificos/marcadores-bioquimicos-de-lesao-cardiaca. (2006, acedido em Novembro de 2021).

[37] R. B. Lusa. SMOTE for high-dimensional class-imbalanced data. https://bmcbioinformatics.biomedcentral.com/track/pdf/10.1186/1471-2105-14-106.pdf. (2013, acedido em Novembro de 2021).

[38] M. F. Alaa Alsayad, et al. “Diagnosis of Cardiovascular Diseases with Bayesian Classifiers, Journal of Computer Science, 2015.