INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Como podem a aprendizagem automática e a música ajudar a detetar e prevenir episódios de ansiedade

João Rodrigues Carvalho

Mestrado em Informática e Gestão,

Orientador(a):
Doutora Catarina Ferreira da Silva, Professora Associada com Agregação,
Iscte – Instituto Universitário de Lisboa

Novembro, 2022

TECNOLOGIAS
E ARQUITETURA

Departamento de Ciências e Tecnologias da Informação

Como podem a aprendizagem automática e a música ajudar a detetar e prevenir episódios de ansiedade

João Rodrigues Carvalho

Mestrado em Informática e Gestão,

Orientador(a):
Doutora Catarina Ferreira da Silva, Professora Associada com Agregação,
Iscte – Instituto Universitário de Lisboa

Novembro, 2022

Direitos de cópia ou Copyright
©Copyright: João Rodrigues Carvalho

O Iscte - Instituto Universitário de Lisboa tem o direito, perpétuo e sem limites geográficos, de arquivar e publicitar este trabalho através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, de o divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

À minha orientadora, Professora Catarina Ferreira da Silva, agradeço toda a sua orientação ao longo do trabalho desenvolvido, ajuda na resolução dos problemas encontrados e acompanhamento de todo o processo.

À empresa, genesis.studio, em especial ao Sr. Luís que se revelou uma grande ajuda nas decisões tomadas e na perceção do caminho certo a seguir, dando a entender quais os problemas que se depararam e suas possíveis soluções.

À família, e em especial aos meus pais que sempre me ajudaram e apoiaram em todas as minhas decisões e me forneceram tudo o que era possível de modo a desenvolver este projeto, e ao meu irmão que sempre esteve ao meu lado em todas as alturas da minha vida.

Aos colegas, em especial ao Miguel, que sempre se revelou uma ajuda em todo o meu percurso académico e sempre esteve ao meu lado em muitas decisões e trocas de ideias.

Aos amigos, em especial ao Luís e ao António por me terem feito sempre companhia e estarem sempre presentes, principalmente quando mais necessitava, apoiando-me e dando força nos momentos mais complicados, quer na vida pessoal como académica.

Agradeço também à Raquel por toda a ajuda no desenvolvimento deste trabalho, por toda a atenção prestada, pela companhia e motivação dada e principalmente pela sua alegria, tornando todo o processo mais fácil.

A todos os que enumerei o meu sincero “Obrigado”.

Resumo

O avanço e a descoberta de novos métodos de aprendizagem automática e a evolução na realização de algoritmos fazem com que as áreas de investigação e de previsão sejam cada vez mais vastas e, com o aumento da capacidade de armazenamento de dados, é possível com maior eficácia prever cenários da vida real. Nos últimos anos tem-se também verificado um aumento do número de episódios de ansiedade na sociedade, assim como o número de pessoas diagnosticadas com este problema.

Assim, este trabalho tem como objetivo utilizar os métodos e algoritmos de aprendizagem automática, com base nos dados fornecidos pelos utilizadores, de forma a desenvolver um modelo de previsão de ansiedade, descobrindo padrões da presença de episódios de ansiedade, utilizando como ferramenta auxiliar a música. Será feita uma análise da importância da música na vida das pessoas e perceber qual o impacto desta em situações de ansiedade. Em resultado do desenvolvimento deste projeto, foi desenvolvido um modelo cujo objetivo é prever se uma certa pessoa tem ansiedade, ou não, consoante as suas respostas, que foram obtidas por meio de um questionário.

Palavras-Chave: música, ansiedade, aprendizagem automática.

Abstract

The advancement and discovery of new machine learning methods and the evolution in the realisation of algorithms mean that the areas of research and prediction are increasingly vast and, with the increase in data storage capacity, it is possible to predict real-life scenarios more effectively. In recent years there has also been an increase in the number of anxiety episodes in society, as well as the number of people diagnosed with this problem.

Thus, this work aims to use machine learning methods and algorithms, based on data provided by users, to develop an anxiety prediction model, discovering patterns of the presence of anxiety episodes, using music as an auxiliary tool. An analysis will be made of the importance of music in people's lives and to understand its impact in anxiety situations. As a result of the development of this project, a model was developed whose objective is to predict whether a certain person has anxiety or not, depending on their answers, which were obtained through a questionnaire.

Keywords: music, anxiety, machine learning.

Índice

Agradecimentos ..... **i**

Resumo ..... **iii**

Abstract ..... **v**

Índice ..... **vii**

Índice de Tabelas ..... **ix**

Índice de Figuras ..... **xi**

Lista de Abreviaturas e Siglas ..... **xiii**

Capítulo 1 – Introdução **1**

1.1. Enquadramento do tema 1
1.2. Motivação e relevância do tema 2
1.3. Conceitos Fundamentais 3
1.3.1. Aprendizagem Automática 3
1.3.2. Ansiedade 4
1.4. Questões e objetivos de investigação 5
1.5. Hipóteses de investigação 5
1.6. Abordagem metodológica 6
1.7. Estrutura e organização da dissertação 10

Capítulo 2 – Revisão da literatura **13**

2.1. Importância da Música no Tratamento da Ansiedade 13
2.1.1. Efeitos Sonoros nas Emoções 13
2.1.2. Efeitos da Música nas Emoções 14
2.1.3. Tratamento de Ansiedade com Música 15
2.2. Utilização da Aprendizagem Automática no Tratamento e Descoberta da Ansiedade. 17
2.2.1. Ferramentas Auxiliares na Descoberta de Ansiedade 17
2.2.2. Importância da Aprendizagem Automática no Tratamento da Ansiedade 20
2.2.3. Tratamento de diferentes tipo de ansiedade com Aprendizagem Automática 21

Capítulo 3 - Metodologia **25**

3.1. Objetivos de investigação 25
3.2. Desenho de investigação 25

3.3. Recolha de dados 25
3.4. Compreensão dos dados e análise demográfica 26
3.5. Preparação dos dados 35
3.6. Análise exploratória dos dados 43
3.7. Modelação dos dados 46
3.8. Avaliação do modelo 49
Capítulo 4 - Resultados e Discussões 51
4.1. Resultados da análise exploratória dos dados 51
4.2. Resultados da modelação dos dados 52
4.3. Avaliação do modelo 54
4.4. Teste das hipóteses de estudo 56
Capítulo 5 - Conclusões e recomendações 59
5.1. Principais conclusões 59
5.2. Contribuições a nível empresarial e académico 60
5.3. Limitações do trabalho realizado 61
5.4. Propostas de investigação futura 61
Referências bibliográficas 63
A. Apêndices 70
Apêndice A 70
Apêndice B 76
Apêndice C 82

Índice de Tabelas

Tabela 3.1: Estatística das perguntas de problemas de ansiedade generalizada 31
Tabela 3.2: Lista de transformações em “Que plataforma usa para ouvir música?” 37
Tabela 3.3: Lista de transformações em “Que plataforma usa para ouvir música?” 38
Tabela 3.4: Lista de transformações em “Caso sim, qual?” 40
Tabela 3.5: Lista de transformações em “Que tipo de música ouve?” 41
Tabela 4.1: Avaliação dos resultados da variável alvo 53

Índice de Figuras

Figura 1.1: Cross-Industry Standard Process for Data Mining (Moro, 2015) 6
Figura 1.2: Níveis Hierárquicos da CRISP-DM (Pete et al., 2000) 7
Figura 1.3: Ciclo de Vida da CRISP-DM (Pete et al., 2000) 8
Figura 2.1: Russell's Circumplex Model of Affect Internet (Russel, 1980) 13
Figura 3.1: Metodologia CRISP-DM aplicada 25
Figura 3.2: Resultados da pergunta "Qual o seu género?" 27
Figura 3.3: Resultados da pergunta "Qual é a sua atividade atual?" 28
Figura 3.4: Resultados da pergunta "Qual é o seu nível de escolaridade?" 28
Figura 3.5: Resultados da pergunta "Qual é a sua idade?" 29
Figura 3.6: Resultados da pergunta "Qual é a sua área de residência?" 29
Figura 3.7: Resultados das perguntas de GAD no questionário 30
Figura 3.8: Resultados da pergunta "Quantas horas ouve música por semana?" 32
Figura 3.9: Resultados da pergunta "Quanto se sente ansioso, ouve música?" 34
Figura 3.10: Resultados da pergunta "Quanto se sente ansioso…" 34
Figura 3.11: Resultados da pergunta "Sente que ouvir música diminui o seu estado de ansiedade?" 35
Figura 3.12: Workflow utilizado na modelação e avaliação dos modelos, fluxo com leitura da esquerda para a direita. 48
Figura 4.1: Matriz de confusão do algoritmo Naive Bayes 55
Figura A.1: Exemplo de uma Decision Tree (Bi et al., 2019) 70
Figura A.2: Exemplo de uma Artificial Neural Networks (Bi et al., 2019) 71
Figura A.3: Exemplo de um Support Vector Machines (Bi et al., 2019) 71
Figura A.4: Exemplo de uma Logistic Regression (Bi et al., 2019) 72
Figura A.5: Exemplo de um Naive Bayes Algorithm (Bi et al., 2019) 72
Figura A.6: Exemplo de um kNN Algorithm (Bi et al., 2019) 73
Figura A.7: Exemplo de uma Random Forest (Bi et al., 2019) 73
Figura A.8: Exemplo de um AdaBoost (Bi et al., 2019) 74
Figura A.9: Exemplo de um SGD (Bi et al., 2019) 74
Figura A.10: Exemplo de um Gradient Boosting (Bi et al., 2019) 75
Figura A.11: Introdução do questionário. 76
Figura A.12: Perguntas GAD. 76

Figura A.13: Segunda pergunta do questionário. 77
Figura A.14: Terceira pergunta do questionário. 77
Figura A.15: Quarta pergunta do questionário. 77
Figura A.16: Quinta pergunta do questionário. 77
Figura A.17: Sexta pergunta do questionário. 78
Figura A.18: Sétima pergunta do questionário. 78
Figura A.19: Oitava pergunta do questionário. 78
Figura A.20: Nona pergunta do questionário. 79
Figura A.21: Décima pergunta do questionário. 79
Figura A.22: Décima primeira pergunta do questionário. 79
Figura A.23: Décima segunda pergunta do questionário. 79
Figura A.24: Décima terceira pergunta do questionário. 80
Figura A.25: Décima quarta pergunta do questionário. 80
Figura A.26: Décima quinta pergunta do questionário. 80
Figura A.27: Décima sexta pergunta do questionário. 80
Figura A.28: Décima sétima pergunta do questionário. 81
Figura A.29: Imagem utilizada na partilha do questionário. 82

Lista de Abreviaturas e Siglas

ML	Machine Learning
IA	Inteligência Artificial
SVM	Support Vector Machines
kNN	k-Nearest Neighbours
ANN	Artificial Neural Networks
AdaBoost	Adaptive Boosting
CRISP-DM	CRoss-Industry Standard Process for Data Mining
DSRM	Design Science Research Methodogy
ECG	Eletrocardiograma
EEG	Eletroencefalograma
GAD	Generalized Anxiety Disorder
CA	Classification Accuracy
VP	Verdadeiros Positivos
VN	Verdadeiros Negativos
FP	Falsos Positivos
FN	Falsos Negativos

Capítulo 1 – Introdução

1.1. Enquadramento do tema

A ansiedade é um problema muito comum na sociedade, sendo esta caracterizada por fortes sentimentos de medo e receio (Pintelas et al., 2018). É considerada como uma resposta cognitiva, comportamental e biológica que o corpo nos fornece para a pessoa se preparar e ficar alerta de forma a conseguir lidar com situações de stress e conflitos. Uma pessoa que receba um excesso destes sinais corporais como forma de aviso aparentemente desnecessário, pode ser considerada como ansiosa ou com problemas de distúrbios de ansiedade (George et al., 2021).

Com a recente evolução das técnicas e mecanismos de Aprendizagem Automática, ou "Machine Learning" (ML), viu-se uma grande oportunidade de desenvolver várias ferramentas que consigam ajudar os médicos a prever e ajudar na resolução de problemas de saúde mental tais como a ansiedade (Pintelas et al., 2018). As tecnologias digitais vieram melhorar e até resolver grandes problemas e obstáculos no que toca ao diagnóstico e tratamento da ansiedade, fornecendo mais e melhores serviços médicos. Nestes serviços estão incluídos aparelhos médicos, como sensores, e aplicações digitais móveis, que serviram para monitorizar, prevenir e ajudar a descobrir sintomas relacionados com a ansiedade (Thieme et al., 2020).

Para além da deteção de episódios de ansiedade em pacientes já diagnosticados com ansiedade, a ML pode ter uma abordagem mais extensiva, de forma a conseguir diagnosticar problemas de ansiedade em indivíduos que nunca relataram tais problemas, assim como utilizar ferramentas, como por exemplo a música, de forma a aconselhar e/ou prevenir novos episódios de ansiedade. (Rahman et al., 2021).

A música tem um papel muito importante no dia a dia das pessoas. Esta tem o poder de afetar, positivamente ou negativamente, o estado emocional das pessoas. Existem estudos que evidenciam que os estudantes conseguem concentrar-se mais enquanto ouvem música, ou que corredores conseguem correr durante mais tempo enquanto ouvem música (Chiu and Ko, 2017).

Com a utilização da música como ferramenta de auxílio no combate às doenças de saúde mental, é possível que as pessoas diagnosticadas com ansiedade possam saber quando poderão vir a ter um episódio de ansiedade e saber que músicas deve ouvir de forma a conseguir evitar ter esse episódio de ansiedade ou, pelo menos, conseguir diminuir os sintomas e consequências do mesmo (Quasim et al., 2021).

Este trabalho foi realizado em colaboração com a empresa genesis.studio que pretende analisar dados de indivíduos que apresentam sintomas de episódios de ansiedade e de que forma estes se relacionam com a frequência e tipo de música que ouvem.

1.2. Motivação e relevância do tema

A ansiedade e os distúrbios mentais na sua generalidade têm sido temas cada vez mais tidos em consideração, principalmente com a vinda do vírus COVID-19. Com o aparecimento da pandemia causada pela COVID-19, muitas pessoas ficaram fechadas em casa, isoladas e com medo do que a pandemia irá causar ao mundo. Para além dos graves problemas de saúde física que as pessoas possam ter se forem contagiadas com este vírus, estas podem também sofrer de graves problemas de saúde mental tais como depressão relacionada com as consequências das medidas impostas devido ao COVID-19 e ansiedade do que pode vir a acontecer. Viu-se então uma necessidade cada vez maior de tratar e resolver problemas derivados do campo emocional e mental. (N. Al-Wesabi et al., 2022).

A utilização de dados de qualidade relativamente aos episódios de ansiedade são um ponto chave para a elaboração de boas análises e conclusões relativamente ao diagnóstico de problemas mentais (Hirsch et al., 2017). A utilização de ferramentas que possam capturar dados biométricos ou respostas dos pacientes de forma mais conceptual, por base em perguntas, mas com uma maior frequência, irá aumentar o número de dados, garantindo uma das principais condições da utilização das técnicas de ML: a utilização de um grande número de dados na amostra. (Thieme et al., 2020).

Com o aumento da utilização de técnicas de ML, muitos cientistas procuram inovar o avanço no campo da inteligência emocional. A música é uma das áreas de investigação que suscita a curiosidade dos cientistas. Estes conseguem perceber o impacto da música no estado emocional nas pessoas. Assim como a música pode criar sentimentos de suspense, como nos filmes de terror, esta também pode criar sentimentos de alívio e relaxamento. (Quasim et al., 2021).

O combinar da preocupação com a saúde mental, com o avanço da tecnologia, principalmente na parte da IA, leva a um ramo de investigação bastante importante, o estudo da utilização de ML no tratamento e descoberta de problemas de saúde mental (Chiu and Ko, 2017).

1.3. Conceitos Fundamentais

1.3.1. Aprendizagem Automática

A aprendizagem automática, ou aprendizagem da máquina, é um ramo da IA que procura aprender e imitar comportamentos humanos com base nos dados adquiridos no passado (Zhang, 2020). Esta utiliza algoritmos computacionais criados pelo homem que permite emular a inteligência humana, tendo em conta o seu meio envolvente. (El Naqa and Murphy, 2015). A utilização destes algoritmos faz com que o desempenho das funcionalidades da máquina seja melhor consoante os resultados e dados passados, e que esta aprenda com a experiência (Yao and Liu, 2013).

As técnicas de ML podem ser ramificadas em conjuntos diferentes de acordo com o tipo de dados e estudos a realizar. Esta pode ser então definida por "supervisionada" e "não supervisionada" (Zhao and Liu, 2007).

As técnicas de ML supervisionadas constroem algoritmos capazes de produzir informação padronizada com base em dados já categorizados e identificados de forma a prever resultados futuros de novas instâncias com o mesmo padrão (Singh et al., 2016). O seu objetivo é aprender um modelo preditivo que estuda todas as características de todos os dados de treino para comparar com todas as características de dados futuros e podê-los mapear consoante os seus valores e características (Molnar, 2019).

Esta classificação é feita em duas fases distintas. A primeira é a aplicação dos algoritmos aos dados de treino de forma a este perceber como os dados já categorizados estão agrupados e definidos. A segunda é a utilização do algoritmo já treinado com os dados de treino em novos dados teste, também categorizados, com o objetivo de perceber o desempenho do algoritmo treinado e a sua eficácia na prevenção de novos dados (Singh et al., 2016).

A técnica de ML não supervisionada diferencia-se da supervisionada na medida em que esta não utiliza dados de treino por forma a treinar o algoritmo, não descobrindo uma relação das características de dados anteriores com os novos dados (Zanero and Serazzi, 2008). Em vez disso, o algoritmo irá então descobrir-se a si mesmo primeiro de forma a perceber os dados fornecidos e agrupá-los por semelhança de características. Desta forma, para que o algoritmo consiga padronizar os dados em categorias, é preciso um grande conjunto de dados, normalmente maior que com a técnica supervisionada, de modo a conseguir ter uma eficácia no desempenho elevada (Molnar, 2019).

A aprendizagem automática recorre à utilização de vários algoritmos e modelos, gerados pelo ser humano, de forma a facilitar a vida das pessoas aquando da tomada de decisões. Segundo Bi et al. (2019), existem vários algoritmos comummente utilizados nas práticas de ML, tais como o "Decision Tree", "Artificial Neural Networks" (ANN), "Support Vector Machines" (SVM), "Logistic Regression", "Naive Bayes", "K-Nearest Neighbors" (kNN) e "Random Forest". Mais informações sobre estes algoritmos encontram-se no apêndice A.

1.3.2. Ansiedade

A ansiedade é um tipo de doença mental, sendo considerada a mais comum registada na sociedade (Pintelas et al., 2018). Esta pode ser definida como uma resposta do corpo humano de forma a alertar e reagir a uma ameaça ou perigo. Apesar de ser um estímulo cerebral natural, o seu excesso pode ser maligno ao sistema nervoso, havendo um excesso de ansiedade face a uma situação de perigo mínima, podendo criar sintomas físicos desnecessários e crenças não reais (Gask et al., 2018). Se os sintomas se prolongarem no tempo e começarem a ser mais intensos, os episódios de ansiedade vão interferir na realização de tarefas da vida quotidiana e, consequentemente, originar problemas de ansiedade (Shri, 2010).

Os sintomas de ansiedade são despertados por um estímulo produzido pelo cérebro que liberta hormonas como adrenalina e cortisol no sistema nervoso central humano, que nos prepara para enfrentar ou fugir de uma dada situação de perigo (McHugh et al., 2004). Estas hormonas comportam-se como reatores, ou sensibilizadores, do organismo de forma a alertar o corpo a reagir. Os sintomas relacionados com a ansiedade estão, normalmente, relacionados com duas componentes. Uma componente física, onde é produzido reações no corpo e sensações de mau estar fisiológico tais com a falta de ar, aumento do batimento cardíaco, aumento da pressão arterial ou náuseas, e uma componente emocional, onde são sentidas emoções negativas afetantes do bem-estar mental, com por exemplo o medo, insegurança sensação de morte iminente, entre outros (Ihmig et al., 2020).

Ao contrário da depressão, a ansiedade é caracterizada por uma preocupação excessiva face a uma situação futura. Esta é prolongada ao longo do tempo e sem razão aparente. A depressão é também caracterizada por uma preocupação excessiva, mas face a uma situação passada (Richter et al. 2020). A depressão não é apenas uma tristeza profunda face a uma situação marcante no passado. Tal como a ansiedade, a depressão tem efeitos físicos no cérebro, não sendo produzida adrenalina suficiente capaz de fazer a pessoa reagir face a determinadas situações (Kingsbury et al., 1987).

Pode então considerar-se que uma pessoa ansiosa vive no futuro, tendo esta receio de situações e eventos futuros, e uma pessoa depressiva vive no passado, presa a situações e eventos passados, em que cada uma tem sintomas físicos e reações químicas cerebrais anormais (Kalin, 2020).

Os problemas de ansiedade têm origens diferentes consoante o tipo de situação enfrentada. Os sintomas e reações gerados pelo organismo serão diferentes consoante as pessoas e consoante o tipo de receio ou medo. Pode-se então dividir a ansiedade em seis tipos de classificação: Problemas de Ansiedade Geral ou "Generalized Anxiety Disorder" (GAD), Problemas de Ansiedade Social, Agorafobia, Ataques de Pânico, Perturbações Obsessivo-Compulsivas e Stress Pós-Traumático (Shikha et al., 2021 and Sharma and Verbeke, 2021).

1.4. Questões e objetivos de investigação

O principal objetivo de investigação do tema em questão abordado neste estudo é a perceção da importância da utilização da música e ferramentas e sistemas de ML de forma a ajudar a detetar distúrbios de ansiedade.

Considerou-se então relevante definir os seguintes objetivos específicos:

- Definir estratégias de levantamento de dados.
- Perceber a importância da música, e o seu impacto nas emoções, quando relacionada com sintomas de ansiedade.
- Desenvolver um sistema que tenta analisar a importância da música na diminuição de episódios de ansiedade.
- Aplicar algoritmos capazes de prever ou detetar sintomas de ansiedade.

Com estes objetivos, pretende-se conseguir responder à seguinte questão de investigação:

- Como utilizar ML sobre dados de ansiedade e música para ajudar a detetar e evitar episódios de ansiedade?

1.5. Hipóteses de investigação

Dos objetivos específicos decorrem as seguintes hipóteses de investigação:

- Hipótese 1: A música tem um impacto positivo na redução da ansiedade.
- Hipótese 2: A aplicação de algoritmos de ML revela-se importantes na identificação de episódios da ansiedade.

- Hipótese 3: A caracterização sociodemográfica auxilia na identificação de padrões de ansiedade.
- Hipótese 4: É possível descobrir padrões de ansiedade com base no estilo de música que se ouve em conjunto com os restantes dados.

1.6. Abordagem metodológica

Neste estudo combinou-se duas metodologias de forma a atingir os objetivos propostos, i.e. a "Design Science Research Methodology" (DSRM), e a Cross-Industry Standard Process for Data Mining (CRISP-DM), podendo ser ambas visualizadas na Figura 1.1.

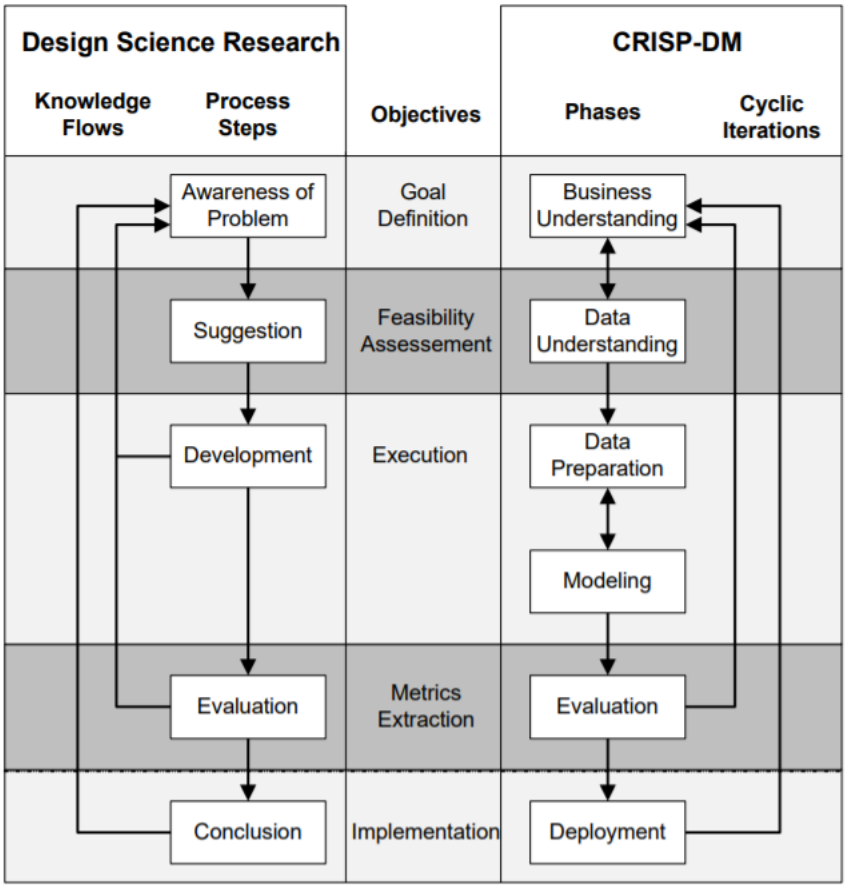


Figura 1.1: Cross-Industry Standard Process for Data Mining (Moro, 2015)

Esta metodologia foi escolhida por produzir rigor científico na construção de conhecimento, utilizando métodos de ML e abordagens rigorosas para produzir um artefacto baseado em modelação de dados. A abordagem metodológica CRISP-DM foca-se em desenhar soluções que possam vir a resolver problemas registados na área científica em questão, através de um artefacto inovador, com base num modelo.

Apesar da escolha da metodologia ser a CRISP-DM, podem verificar-se semelhanças entre esta metodologia e a DSRM (Moro, 2015). A DSRM foca-se em desenhar soluções que possam vir a resolver problemas registados na área científica em questão, através de um artefacto inovador, enquanto a CRISP-DM se foca no desenvolvimento de soluções propostas no processo de desenho do artefacto, tendo por base métodos de ML.

A CRISP-DM é uma metodologia com seis fases, desenhada de forma a representar o ciclo de vida de um projeto de Ciência de Dados ou pode ser utilizada na aplicação de algoritmos de ML. A metodologia CRISP-DM pode ser descrita em quatro níveis hierárquicos, como pode ser visualizado na Figura 1.2.

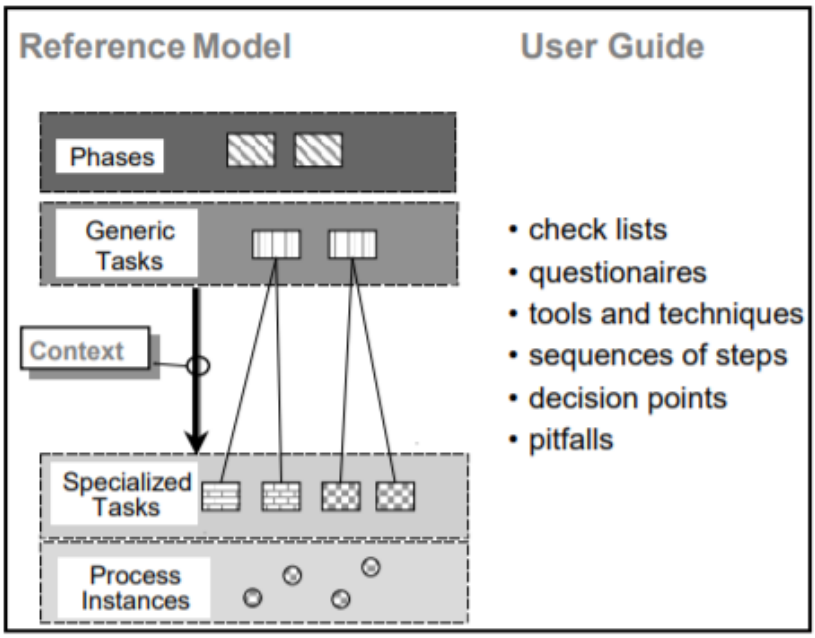


Figura 1.2: Níveis Hierárquicos da CRISP-DM (Pete et al., 2000)

Segundo Pete et al. (2000), o primeiro nível hierárquico representa as seis fases representadas pela metodologia, sendo estas a "Perceção do Negócio", "Perceção dos Dados", "Preparação dos Dados", "Modelação", "Avaliação" e "Implantação", representado na Figura 1.3. O segundo nível corresponde às tarefas genéricas representativas de todas as situações possíveis no desenvolvimento do modelo. As tarefas especializadas constituem o terceiro nível hierárquico e representam os cenários específicos mencionados pelo nível dois. O nível quatro da hierarquia da CRISP-DM é o nível mais específico e representa todas as ações e desenvolvimento dos processos das tarefas especificas, ou seja, representam os dados e os acontecimentos do nível acima (Pete et al. 2000).



Figura 1.3: Ciclo de Vida da CRISP-DM (Pete et al., 2000)

A primeira fase do ciclo de vida do modelo é a "Perceção do Negócio". Esta é uma fase fulcral na metodologia, uma vez que é necessária uma boa compreensão do negócio, assim como dos objetivos e requisitos do projeto de forma a cumprir com todos os requisitos e necessidades. Serão feitas as designações de cada fase assim como todos os objetivos e requisitos do modelo e os possíveis problemas que podem aparecer no decorrer de cada tarefa (Schäfer et al., 2018). Segundo Pete et al. (2000), as tarefas genéricas representativas "Perceção do Negócio" são as seguintes:

1. Determinação dos objetivos de negócio: Perceção das necessidades do projeto e os seus critérios de sucesso.
2. Avaliação da situação: Determinação dos requisitos, riscos e recursos disponíveis do projeto.
3. Determinação dos objetivos de data mining: Perceção dos objetivos do modelo, assim como os seus critérios de sucesso.
4. Estruturação do plano do projeto: Definição de cada fase do projeto.

A segunda fase é a "Perceção dos Dados", onde é feita recolha de dados e o seu estudo, com o objetivo de perceber os seus valores e reter informação relativa ao seu conteúdo, estudando todas as variáveis e possíveis. A representação gráfica e o estudo das correlações das variáveis são importantes na determinação dos critérios de qualidade de dados, de forma a perceber quais as variáveis importantes e de que maneira podem ajudar na construção do modelo (Schäfer et al., 2018). Segundo Pete et al. (2000), as tarefas genéricas representativas da "Perceção dos Dados" são as seguintes:

1. Recolha dos dados: Recolha dos dados essenciais à construção do modelo.
2. Descrição dos dados: Identificação do tipo de dados utilizados.
3. Exploração dos dados: Análise dos dados e das suas correlações e valores.
4. Verificação dos critérios de qualidade dos dados: Identificação da qualidade dos dados recolhidos.

A terceira fase corresponde à "Preparação dos Dados". Após a recolha dos dados iniciais e do seu estudo de valores e correlações, é feito o processamento dos dados, tais como a remoção de valores ou variáveis desnecessárias ao desenvolvimento do modelo, de forma a preencher os critérios de qualidade definidos na fase anterior (Schäfer et al., 2018). Segundo Pete et al. (2000), as tarefas genéricas representativas da "Preparação dos Dados" são as seguintes:

1. Seleção dos dados: Determinação dos dados relevantes para o modelo.
2. Limpeza de dados: Eliminação dos dados não necessários à realização do projeto.
3. Construção dos dados: Criação de novos dados consoante as necessidades encontradas na análise feita.
4. Integração dos dados: Junção dos dados existentes, criados e alterados.
5. Formatação dos dados: Ajustar o tipo de dados às características das variáveis.

A quarta fase do modelo é a "Modelação", onde é feita a escolha dos algoritmos a serem utilizados e construção de modelos de ML, utilizando os dados preparados na fase anterior, assim como o registo dos seus resultados (Schäfer et al., 2018). Segundo Pete et al. (2000), as tarefas genéricas representativas da "Preparação dos Dados" são as seguintes:

1. Seleção das técnicas de modelação: Determinação dos algoritmos a utilizar.
2. Criação de dados teste: Criação de dados de teste de forma a treinar o algoritmo.
3. Construção do modelo: Definição dos algoritmos a utilizar e construção do modelo com base nos dados recebidos e treinados.

4. Avaliação do modelo: Interpretação dos resultados do modelo com base nos objetivos definidos.

A quinta fase do ciclo de vida é representada pela "Avaliação", onde é feita a avaliação dos resultados obtidos na fase de modelação, tendo em conta todos os objetivos do negócio e a necessidade do projeto. É também feita uma análise de possíveis melhoramentos do modelo e se este corresponde aos critérios de qualidade definidos (Schäfer et al., 2018). Segundo Pete et al. (2000), as tarefas genéricas representativas da "Avaliação" são as seguintes:

1. Avaliação dos resultados: Perceção da correspondência entre os resultados obtidos e os objetivos definidos.
2. Revisão do processo: Revisão do trabalho desenvolvido e verificação das tarefas realizadas.
3. Determinação dos Próximos passos: Determinação do processo de implantação.

A sexta fase, e última da metodologia CRISP-DM é designada como "Implantação". (Schäfer et al., 2018). Segundo Pete et al. (2000), as tarefas genéricas representativas da "Implantação" são as seguintes:

1. Planeamento da implantação: Criação do plano de implantação do modelo.
2. Planeamento da monitorização e manutenção: Criação do plano de monitorização e manutenção para a fase seguinte da implantação do modelo.
3. Produção do relatório final: Criação do relatório final do projeto, onde é feito um resumo do projeto e dos resultados obtidos.
4. Revisão do projeto: Verificação das tarefas e dos objetivos cumpridos no projeto e diagnóstico do seu desenvolvimento.

1.7. Estrutura e organização da dissertação

O presente estudo está organizado em cinco capítulos, correspondentes a cada fase de desenvolvimento do projeto, desde uma fase mais introdutória, onde se apresenta a introdução, revisão de literatura e metodologia, até à parte de desenvolvimentos e análise e discussão de resultados obtidos, acabando na conclusão.

O primeiro capítulo, introdução, introduz o tema em estudo, fazendo uma contextualização da área de investigação, indicando porque este é relevante e quais as suas motivações. É feita uma abordagem aos conceitos teóricos importantes da ansiedade e ML, de forma a fazer um enquadramento teórico das matérias utilizadas no desenvolvimento do processo. Por fim são definidos quais os objetivos, questões da investigação presente e hipóteses de investigação, bem como a identificação da metodologia utilizada no desenvolvimento do estudo.

O segundo capítulo, revisão da literatura, reflete o enquadramento teórico, onde é definido o estado da arte. É feito o estudo sobre as matérias envolventes de forma a fazer um enquadramento teórico das necessidades de aprendizagem na realização do modelo pretendido. Neste contexto, é importante perceber a importância da música no tratamento da ansiedade, percebendo o efeito dos sons e da música nas emoções humanas, assim como as formas de tratamento da ansiedade existentes relacionadas com a música. A utilização de ML no tratamento e descoberta da ansiedade é também uma matéria interessante no estudo do tema, percebendo quais são as ferramentas auxiliares existentes no tratamento da ansiedade, assim como a importância destas ferramentas e quais é que podem ajudar nos diferentes tipos de ansiedade.

O terceiro capítulo, metodologia, descreve os procedimentos de realização de todos os processos de recolha, tratamento e modelação dos dados, assim como todos os passos definidos de forma a atingir os objetivos do estudo, e a explicação e desenvolvimento da metodologia utilizada. Serão explicadas todas as fases da CRISP-DM assim como as suas tarefas e ações associadas em cada fase do desenvolvimento do modelo. Todas estas fases e etapas, que serão seguidas no desenvolvimento do modelo criado, correspondem ao desenho de investigação escolhido neste trabalho. Também serão desenvolvidas as hipóteses de investigação consoante os objetivos definidos.

O quarto capítulo, análise e discussão de resultados, demonstra todos os procedimentos práticos utilizados no processo de desenvolvimento do modelo, explicando todos os passos e ações feitas de forma a chegar aos resultados pretendidos, bem como a sua análise. Todos estes passos terão em conta a metodologia adotada e quais os procedimentos necessários de forma a chegar ao desenvolvimento do modelo. Será também feita uma avaliação e testes das hipóteses do presente estudo, de forma a confirmar, ou não, o resultado de cada hipótese.

O quinto capítulo, conclusões e recomendações, representa a parte final da dissertação, onde será feita uma reflexão do trabalho desenvolvido, apresentando as principais conclusões do estudo, assim como as implicações do mesmo a nível académico e empresarial. Serão também apresentadas as limitações encontradas no decorrer do processo e quais as possíveis melhorias e novos estudos relacionados com o tema em investigações futuras.

Capítulo 2 – Revisão da literatura

2.1. Importância da Música no Tratamento da Ansiedade

2.1.1. Efeitos Sonoros nas Emoções

A audição é um sentido muito importante na vida humana, sendo esta uma necessidade essencial à existência do Ser Humano. Como todos os outros sentidos, a audição, mais especificamente o sistema auditivo, está ligado ao sistema nervoso. As emoções são uma característica fisiológica humana que reflete o estado de espírito humano, tais como sensações de prazer ou tristeza (Nardelli et al., 2015). Alguns sons podem desencadear estados emocionais a partir do Sistema Nervoso Autónomo. Este sistema é responsável pela normal função de processos autónomos do corpo humano tais como a frequência respiratória ou cardíaca (Angelis et al., 2004).

James Russel criou um modelo bidimensional, chamado "Russell's Circumplex Model", de forma a classificar e quantificar as emoções humanas. Este modelo utiliza uma combinação linear entre as dimensões de valência, sendo esta caracterizada pela quantidade de prazer que um certo som produz, e da excitação, caracterizada pelo nível de excitação do som, como pode ser verificado na Figura 2.1 (Nardelli et al., 2015).

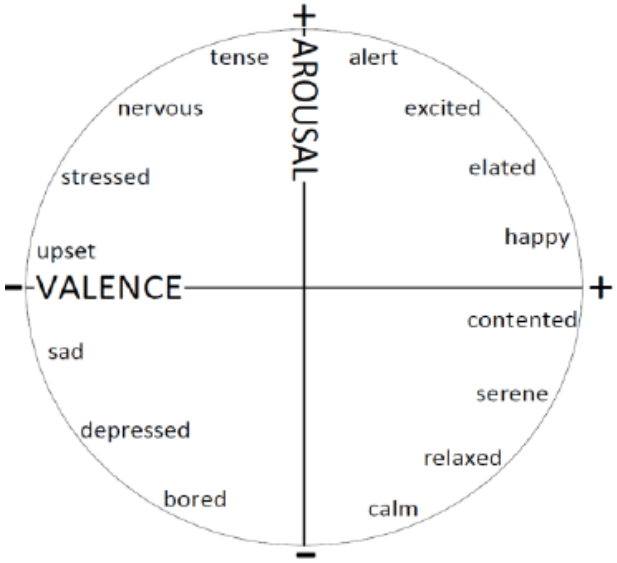


Figura 2.1: Russell's Circumplex Model of Affect Internet (Russel, 1980) [1]

[1] Imagem retirada do site: https://www.researchgate.net/figure/Russells-Circumplex-Model-of-Affect_fig2_258273035

A partir de certos sons, podem-se gerar certas reações emocionais. Este efeito pode ser visualizado no dia-a-dia das pessoas. Existem sons que fazem as pessoas estarem sossegadas e felizes, sendo estes considerados agradáveis, provocando um estado emocional positivo, outros sons têm reações negativas ao estado emocional humano (McGeoch, 2020). Estudos científicos, como o eletroencefalograma (EEG) ou eletrocardiograma (ECG), provam que a atividade cerebral é alterada consoante a o tipo de sons envolventes, podendo-se observar uma variação da atividade cerebral ou ritmo cardíaco na presença de diferentes sons. (Zoefel and Heil, 2013).

2.1.2. Efeitos da Música nas Emoções

A música tem um papel fundamental no dia-a-dia das pessoas, provocando alterações no seu estado emocional, como por exemplo no trabalho, em viagens de carro, no ginásio ou até mesmo para adormecer. A música pode ser utilizada para provocar efeitos positivos ou negativos às pessoas, consoante a sua personalidade e as suas experiências emocionais (Taruffi et al., 2017).

O stress provoca o aumento dos níveis de cortisol, enquanto a presença da música, normalmente, provoca a diminuição dos seus níveis. Segundo Yehuda (2011), diferentes tipos de música provam diferentes reações emocionais, como por exemplo, músicas clássicas estão associadas a menores níveis de stress, originando uma descida dos níveis de cortisol. Por outro lado, a música eletrónica está associada a músicas mais agressivas ao sistema nervoso, aumentando a produção de hormonas de stress como o cortisol.

A utilização da música na atividade regular da prática de exercício físico tem sido estudada nos últimos anos e, segundo Hallet and Lamont (2015), a utilização de música na prática desportiva pode ter um efeito positivo no sistema hormonal humano, podendo ser possível verificar uma melhoria de resultados na prática desportiva quando se ouve música, comparativamente aos resultados do exercício quando não se ouve música.

No trabalho também pode ser visível a importância da música de forma a melhorar os níveis de concentração das tarefas. Músicas com letra associada não são aconselhadas na prática das tarefas que necessitem concentração, fazendo perder a concentração na realização das mesmas. Já as músicas sem letra, ou seja, apenas com sons de instrumentos musicais, são mais aconselhadas como ferramenta de concentração na realização das tarefas (Shih et al., 2009).

Para além dos efeitos positivos que a música pode provocar nas emoções humanas, podem também ser criados efeitos negativos. Muitas músicas têm o poder de lembrar certas situações ou acontecimentos passados. Quando uma música é ouvida no acontecimento de situações negativas, esta pode ser relacionada com esse acontecimento, fazendo com que, quando esta música for ouvida novamente, desenvolva emoções negativas (Thompson and Biddle, 2013). As músicas que provocam sentimentos negativos podem, por vezes, criar um contexto paradoxal, ou seja, podem provocar emoções positivas, mesmo sendo estas caracterizadas como músicas tristes (Garrido and Schubert, 2011). Apesar de a reação natural humana ser a fuga das emoções negativas, algumas pessoas conseguem encontrar prazer nas músicas consideradas como tristes, tendo em contas os seus traços de personalidade e associações com a letra da música, criando sensações de relação e familiarização com o sentido transmitido (Vuoskoski et al., 2012).

2.1.3. Tratamento de Ansiedade com Música

O tratamento da ansiedade é, normalmente, baseado na mudança da vida quotidiana das pessoas que experienciam episódios de ansiedade, a recorrência a um psicólogo ou até mesmo de um psiquiatra (Englander, 2016). É importante denotar a diferença entre a função de um psicólogo comparativamente à de um psiquiatra no que toca ao desempenhar de funções no tratamento de doenças mentais. Esta diferença na perspetiva de visão da ciência e formas de tratamento cria muitas dificuldades de comunicação entre psicólogos e psiquiatras (Kingsbury et al., 1987).

Pode-se considerar então que os psicólogos utilizam ferramentas mais invasivas, entrando no espaço pessoal da pessoa, de modo a perceber as suas rotinas e o seu comportamento face a várias situações do dia-a-dia, criando uma ligação pessoal com o paciente de modo a perceber quais os seus gostos e ocupações de forma a conseguir aconselhar atividades e mudanças na rotina diária da pessoa. Os psiquiatras são médicos capazes de fazer uma análise à saúde mental de forma mais científica, percebendo os sintomas e podendo prescrever medicação de forma a aliviar ou tratar qualquer reação do corpo face a cenários de stress. A utilização de medicamentos como calmantes é bastante conhecida na prescrição feita por psiquiatras (Kanas and Manzey, 2004).

Uma vez que a música pode ter efeitos positivos no sistema hormonal humano, é possível utilizar este fenómeno no tratamento de problemas mentais como a depressão ou a ansiedade (Brugués, 2011). A utilização de música como forma de alívio ou redução de sintomas de ansiedade é uma prática comum, utilizada de forma autónoma e, por vezes, inconsciente. A música pode ajudar na redução do stress com a libertação de endorfina, hormona responsável pelas sensações de bem-estar e prazer (Yehuda, 2011).

Para além da utilização da música como prática individual no tratamento da ansiedade, existem técnicas de contexto clínico para o seu controlo e tratamento como a musicoterapia. Esta técnica é utilizada em todas as idades e com efeitos e objetivos diferentes consoante a necessidade do paciente. O musicoterapeuta irá comunicar verbalmente com o paciente de forma a perceber os seus gostos musicais, conhecer a pessoa e construir uma relação de confiança pessoal com o objetivo de perceber os seus problemas e origem da ansiedade (Zhou et al., 2009). Os pacientes irão então ouvir música que faça a sua mente relaxar, consoante os gostos e estilos de música que ouve. Estes são também incentivados a praticar exercícios simples relacionados com a música como cantar, dançar ou tocar instrumentos (Junior, 2018).

A meditação é uma prática bastante conhecida pelo poder de controlo do corpo e da mente, muitas vezes recomendada por especialistas médicos no combate à ansiedade. Para além do seu recurso na tentativa de redução de sentimentos de ansiedade, esta é muitas vezes utilizada por pessoas não diagnosticadas com ansiedade, sendo então esta prática de uso comum geral (Delmonte, 1985). A atividade cerebral, na prática da meditação, diminui consideravelmente, causada pela concentração na atividade física e tentativa de relaxamento dos músculos e abstração de pensamentos. É feita a concentração na respiração e foco no momento presente. A música é um fator bastante importante na meditação, utilizando músicas sem letra e sons referentes à natureza e instrumentos harmoniosos de forma a atingir um estado de concentração e relaxamento praticamente total (Junior, 2018).

2.2. Utilização da Aprendizagem Automática no Tratamento e Descoberta da Ansiedade.

2.2.1. Ferramentas Auxiliares na Descoberta de Ansiedade

Existem várias ferramentas alternativas capazes de identificar problemas de ansiedade e fazer o diagnóstico dos seus sintomas. A psicologia é uma das ferramentas mais conhecidas no tratamento da ansiedade, recorrendo a psicólogos capazes de perceber e tratar pacientes com vários mecanismos de suporte à ultrapassagem de qualquer um dos tipos de ansiedade presentes (Kingsbury et al., 1987). A técnica mais comum utilizada de forma a obter dados sobre os pacientes diagnosticados com ansiedade, ou pessoas que estão a ser alvo de estudo em relação a esta matéria, requer horas de perguntas e questionários e estruturados de forma clínica durante vários dias ou semanas. É um processo demorado e complexo (McGinnis et al., 2018).

Para além da interação humana, com os psicólogos, existem também vários aparelhos médicos (sensores) preparados para receber e registar todos os dados necessários ao estudo da relação do comportamento fisiológico humano com a ansiedade. Estes aparelhos podem ser considerados de utilização pessoal no dia-a-dia como aplicações de telemóvel de registo de dados, ou então aparelhos médicos específicos mais complexos, onde o paciente comum não entende o resultado do diagnóstico realizado (Thieme et al., 2020).

No intuito de detetar os sintomas relatados pelos doentes de ansiedade, os sensores irão medir e registar o funcionamento do corpo do paciente de forma a conseguir explicar cientificamente os sintomas ocorridos, a sua origem e quais as suas consequências na vida presente e futura do paciente (Manocha et al., 2021).

Um dos sintomas mais recorrentes na ansiedade é o sentimento enganador de falta de ar. Este fenómeno acontece devido ao facto de o batimento cardíaco e a pressão arterial se alterarem, obrigando o sistema respiratório a adaptar-se às novas condições impostas pelo coração. O Eletrocardiograma pode então ser útil na descoberta de problemas de ansiedade, percebendo a atividade do coração registado em certas ocasiões e atividades (George et al., 2021).

A utilização de parâmetros físicos como a variação do batimento cardíaco e da tensão arterial provindos dos resultados deste exame em conjunto com o registo da atividade que o individuo está a realizar pode detetar indícios de ansiedade. A classificação destes é bastante importante na medida que fornece resultados suficientes e de qualidade ao médico que irá avaliar a condição psicológica do paciente (Srividya et al., 2018). No entanto, obter uma classificação suficientemente fiável pode ser difícil na medida que existe um leque de características humanas em termos de saúde física que pode comprometer os valores, ou seja, uma pessoa com tendências para ter o nível de ritmo cardíaco bastante variado, apresentando episódios de taquicardia, e não estando diagnosticado com ansiedade, pode ter os mesmos resultados no exame ECG comparativamente a uma pessoa que tenha ansiedade e não tenha taquicardia (Jambukia et al., 2015).

Para além da atividade cardiorrespiratória e dos seus sintomas associados, a atividade cerebral também pode ser alvo de investigação no que toca à deteção de problemas de ansiedade. A aplicação de técnicas de ML no tratamento da saúde mental, utilizando dados provenientes de EEG pode ajudar no diagnostico de doenças mentais e correlacionar a ansiedade com outros comportamentos físicos (Gore and Rathi, 2019). Este exame mede o estado mental e cognitivo humano, bem como a sua atividade cerebral. A análise dos resultados dos sinais registados no EEG verificam a diferença da atividade cerebral e qual a parte do cérebro é alterada quando o examinando está numa fase ansiosa, comparativamente ao seu estado numa fase de relaxamento (Shikha et al., 2021). Segundo Wang et al. (2014), a atividade elétrica da parte frontal do cérebro fornece informações sobre emoções positivas e negativas. Apesar do EEG poder registar problemas de ansiedade, este exame tem problemas na captura e análise de dados, uma vez que é necessário o recurso de vários exames EEG para fazer uma média e poder retirar dados conclusivos, e também apenas uma pequena quantidade destes dados é que podem ser utilizados na modelação dos modelos. Estas barreiras criam graves problemas na construção de modelos de ML.

Outra ferramenta utilizada, mas não tão comum, é a foto pletismografias. Esta é uma técnica ótica não-invasiva que regista a iluminação na superfície da pele, utilizando raios de luz de díodo e mede a quantidade de luz refletida. Como a iluminação está dependente da quantidade de sangue (tensão arterial) e o número de batimentos cardíacos, estes irão afetar os resultados do valor de iluminação da pele quando os valores se alterarem (Udovičić, 2017). Estas características estão associadas à atividade do sistema nervoso central e ao estado psicológico da pessoa. Percebendo quando estes sintomas ocorrem e em que ocasiões, causadores da variação dos valores normais de pressão arterial e variação cardíaca, e percebendo a forma de alteração dos registos feitos pela foto pletismografia, pode-se entender quando ocorrem episódios de ansiedade. Este exame pode obter resultados fiáveis e fáceis de obter, mas precisa de dados de pessoas diagnosticadas com um nível de ansiedade elevado de forma a se poderem correlacionar estas duas matérias (Perpetuini et al., 2021).

O recurso de aplicações móveis tem sido também uma ferramenta de auxílio na deteção da ansiedade. Esta funciona de forma a receber dados das pessoas de forma menos invasiva em comparação com sensores de utilização médica, o que ajuda na recolha de dados e aceitação das pessoas, por ser bastante mais fácil (Thieme et al., 2020). A utilização de aplicações direcionadas à saúde mental é uma técnica cada vez mais posta em prática, utilizando aparelhos de monitorização de dados em tempo real, como relógios inteligentes que medem a variação da pulsação ou o rácio de batimentos cardíacos. Estes aparelhos conseguem, de forma autónoma, registar os dados provenientes de condições físicas que podem estar relacionadas com a ansiedade e, com base em modelos de ML e IA, podem indicar e comunicar episódios de ansiedade (Miranda et al., 2014). Outras aplicações utilizam parâmetros que não estão relacionados com sintomas físicos humanos, como por exemplo, a localização do paciente por via GPS do telemóvel. Este tipo de aplicações correlaciona os lugares que os pacientes frequentaram, com o hiato de valores do nível de ansiedade entre cada localização, podendo afirmar que uma certa localização provoca episódios de ansiedade ou variação nas suas emoções (Saeb et al., 2017).

É também possível, através da fala, detetar problemas de ansiedade. Tal como é possível perceber sinais e informações importantes relativamente ao estado mental de uma pessoa através de expressões físicas, é também possível, através da fala, perceber algumas emoções transmitidas. A comunicação verbal faz parte de um processo de reconhecimento do estado mental, através da expressão da fala, intensidade e ritmo (Laukka et al., 2008). Através de dados obtidos relativos à comunicação feita por uma pessoa diagnosticada com ansiedade, é possível comparar a intensidade do som, a vibração das palavras e o seu ritmo de forma a identificar a presença de ansiedade (McGinnis et al., 2019). A comunicação e a perceção das palavras utilizadas em certos discursos também podem indicar sinais de ansiedade. A utilização de certas palavras em determinadas frases e contextos pode ser reveladora da presença de episódios de ansiedade. A descoberta deste tipo de correlações tem por base métodos de Mineração de texto, por forma a descobrir padrões de utilização de certas palavras em determinados contextos de frases. Utilizando métodos de ML combinados com Mineração de Texto podem então descobrir-se padrões de ansiedade (Rheault, 2016).

2.2.2. Importância da Aprendizagem Automática no Tratamento da Ansiedade

A utilização de ferramentas médicas, como sensores, no processo de descoberta de sintomas de ansiedade é feita de forma a fornecer dados clínicos aos médicos no seu rastreamento e controlo. As técnicas de ML revelam-se ferramentas importantes no suporte à tomada de decisão no contexto de avaliação clínica com base na análise de dados fornecidos por estes equipamentos (Pintelas et al., 2018).

A aplicação destas técnicas de aprendizagem em conjunto com a saúde mental, pode ser a ponte de ligação entre os psiquiatras e os pacientes que nunca relataram os seus episódios de ansiedade ou depressão (Gore and Rathi., 2019). Normalmente, os dados relativos aos sintomas de ansiedade são fornecidos pelo paciente, ao médico, tendo por base a sua perceção do que sentiu, mas existem muitas formas de quantificar o nível de stress e ansiedade com base em dados recebidos por sensores. Estes sensores podem medir o batimento cardíaco e a sua variabilidade, ou a atividade do cérebro, utilizando EEG (Shikha et al., 2021).

A utilização de processos de automação na capacidade de tomada de decisão tem sido alvo de investigação nos últimos anos, sendo esta ainda pouca explorada e utilizada em contextos reais. Ao contrário de algoritmos de ML de Caixa-Preta (“Black-Box”), os modelos de ML utilizados pela IA que justificam o resultado da sua decisão podem aumentar a confiança dos médicos na qualidade de resposta dos modelos (Tutul et al., 2021).

O avanço da tecnologia e o aumento da procura de dados relativos a sintomas de ansiedade ou depressão, permitem a construção de modelos de ML de forma a desenvolver aplicações de telemóvel que, com a ajuda de sensores moveis, fazem a recolha de dados de forma a monitorizar, avaliar e prever problemas de saúde mental futuros (Thieme et al., 2020).

Com a utilização de dados suficientes e de qualidade, será então possível prever e ajudar um grupo de indivíduos a prevenir episódios de agravamento do seu estado mental, direcionando o melhor caminho de forma a evitar que estes episódios ocorram de novo, sugerindo qual o trabalho futuro que têm de fazer de forma a melhor a sua condição mental (Gore and Rathi., 2019).

2.2.3. Aplicação da Aprendizagem Automática a diferentes tipos de ansiedade

A utilização de técnicas de ML no estudo de problemas de ansiedade pode ajudar na descoberta destes problemas assim como a diferença entre cada patologia, tendo por base diferentes sintomas e valores diferentes de características fisiológicas consoante cada tipo de ansiedade (Sharma and Verbeke, 2021).

Os Problemas de Ansiedade Geral são causados pelo excesso de preocupação em tarefas normais do dia-a-dia, sem aparente razão ou sinais de perigo. As preocupações são normalmente irrealistas e desnecessárias ao bom funcionamento do corpo humano. A vida normal, tal como trabalhar, estudar, conduzir ou cozinhar, pode ser posta em causa devido ao facto de haver uma preocupação constante com os acontecimentos exteriores de forma corrente (Wittchen and Hoyer, 2001). Segundo um estudo feito por Hilbert (2017), a utilização do algoritmo SVM demonstra-se eficaz na deteção deste tipo de ansiedade, na utilização de parâmetros de entrada como questionários clínicos, níveis de cortisol, matéria cinzenta e matéria branca. De forma a detetar que um individuo está na presença deste tipo de ansiedade, o autor reporta a recolha dos dados, e posteriormente tabelados, de pessoas diagnosticadas com problemas de ansiedade geral, outras com depressão, pessoas com ambas as patologias assim como sem patologia nenhuma diagnosticada.

Os Problemas de Ansiedade Social, também conhecido por fobia social, são os problemas de ansiedade mais comuns de todos os tipos de ansiedade. Este é causado pela preocupação de se sentir observado e julgado pelas pessoas em redor face a situações de exposição social, como apresentações de trabalhos escolares (Gask et al., 2018). Este problema é normalmente experienciado por adolescentes, tendo medo da opinião social envolvente (Stein and Stein, 2008). Segundo o estudo feito por Demetriou et al. (2020), a utilização de algoritmos baseados no modelo Random Forest consegue demonstrar a diferença de resultados entre pessoas diagnosticadas com problemas de ansiedade social, autismo e psicose, e identificar a sua patologia com base nesses resultados. Os pacientes demonstram diferentes níveis de atenção, processo de informação, cognição social e desempenho visuomotor.

A Agorafobia é designada como o medo de situações difíceis de escapar ou que a pessoa não tenha ajuda imediata por perto. É também conhecida pelo medo de sair de casa, algo que acontece em estados mais avançados deste tipo de ansiedade, onde uma pessoa não sai de casa de forma a evitar situações de pânico ou desconforto social. Agorafobia deriva da palavra grega "agora" que que significa espaço público. Esta surge quando uma pessoa experienciou algum ataque de ansiedade ou pânico num certo local e, a partir desse momento, evita frequentar ou permanecer nesse mesmo local com medo de que algo de semelhante volte a acontecer (Kwon et al., 1990). Num estudo feito por Hahn et al. (2015), a utilização do algoritmo SVM demonstra-se eficaz na deteção de problemas de agorafobia. Este estudo utilizou, como valores de entrada, o nível de oxigénio no sangue no decorrer de diferentes tarefas que provocassem medo ou receio ao paciente, alterando a atividade cerebral em diferentes partes do cérebro. Nenhuma parte individual do cérebro foi alterada de forma a encontrar um padrão na previsão de sinais de agorafobia, mas o conjunto de certas partes do cérebro, resultando num grupo regional cerebral, foram identificadas como boas previsoras deste tipo de ansiedade.

Os Ataques de Pânico são tipos de ansiedade mais intensos, cujos sintomas são mais agressivos e não tão recorrentes como nos outros tipos de ansiedade, tais como: o corpo a tremer, dor no peito, desespero e taquicardia acentuada (Shri, 2010). Os ataques de pânico têm origem no culminar de sensações de medo que originam uma combinação extrema de sintomas físicos e psicológicos. São normalmente associados ao ponto extremo da ansiedade ou sentimentos de ansiedade (Sharma and Verbeke, 2021).

A utilização de Regressão Logística, em comparação com outros algoritmos como o ANN, SVM, Random Forest e Gradient Boosting, demonstrou o melhor desempenho na previsão deste tipo de ansiedade, utilizando a variação do rácio do batimento cardíaco como variável fundamental na distinção dos ataques de Pânico com os outros tipos de ansiedade. Apesar de se poder observar a variação dos resultados gerados da variação do rácio do batimento cardíaco, esta sozinha não é suficiente para a distinção de ataques de pânico com outros tipos de ansiedade, sendo necessário recorrer à utilização de diagnósticos clínicos personalizados e à intervenção de campos da psiquiatria (Na et al., 2021).

As Perturbações Obsessivo-Compulsivas são tipos de ansiedade que, normalmente, não geram sintomas físicos aparentes como os outros tipos de ansiedade (Shikha et al., 2021). É importante salientar o significado e origem das palavras, realçando a utilização de dois acontecimentos, a obsessão e a compulsão, de forma a demonstrar a ocorrência das patologias verificadas (Abriel, 2010). As obsessões originam-se com base em pensamentos estranhos e sem sentido de forma repetida, criando um conflito psicológico, tais como o medo de perder o controlo do corpo e da mente, visualização de imagens perturbadoras e medo de causar danos físicos próprios. As compulsões têm origem em ações e comportamentos repetidos, desnecessários à vida normal humana, tais como a verificação repetida se algo está arrumado ou se o fogão está apagado (Leckman et al., 1997). Askland (2015) desenvolveu um estudo baseado em metodologias de ML com o objetivo de identificar problemas mentais relacionados com obsessões obsessivo-compulsivas e fazer os primeiros possíveis diagnósticos. A utilização do algoritmo Random Forest mostrou o mais revelador no que toca à deteção de primeiros sinais deste tipo de perturbações. O estudo revelou um problema inicial grave na teoria da modelação. Os dados utilizados como valores de treino do modelo foram tomados por métodos de testagem baseado em hipóteses, não sendo estes dados de treino fiáveis.

O Stresse Pós-Traumático é a ansiedade associada ao medo e stress causado por uma ou várias situações traumáticas, tais como situações de morte iminente, medo e terror extremo ou de lesões graves, episódios que podem demorar anos a ultrapassar (Pintelas et al., 2018). As pessoas que padecem de stresse Pós-Traumático, têm um medo extremo quando vivenciam situações idênticas às que originaram um elevado nível de stress e medo (Beck et al., 2009). No estudo realizado por Wshah et al. (2019), a utilização de técnicas de ML mostrou-se importante na previsão de Stress Pós-Traumático. Este estudo teve como dados de entrada, sintomas relatados pelas pessoas que sofreram um trauma por via de uma aplicação de telemóvel.

Capítulo 3 - Metodologia

3.1. Objetivos de investigação

A presente investigação tem como objetivo criar um sistema de identificação da presença de episódios de ansiedade e perceber se a música poderá ter a influência na redução destes. Este sistema tem como objetivo tentar prever se uma pessoa tem, ou não, episódios de ansiedade, assim como compreender qual o impacto da música na redução da ansiedade. Para tal, foi desenvolvido um questionário com o objetivo de adquirir dados que permitam alimentar modelos gerados por métodos de aprendizagem automática.

Desta forma, com o desenvolvimento deste sistema pretende-se melhorar a vida das pessoas que potencialmente tenham episódios de ansiedade. Este trabalho é realizado no contexto de um protocolo de colaboração com a empresa genesis.studio e pretende-se que os resultados venham a ser utilizados por profissionais na área da saúde mental, potencialmente no quadro de um sistema de recomendações de músicas para minorar sintomas de ansiedade.

3.2. Desenho de investigação

Na Figura 3.1 está representada a metodologia CRISP-DM aplicada à presente investigação, onde as fases e tarefas descritas na figura têm por base a metodologia CRISP-DM genérica apresentada na secção 1.6.

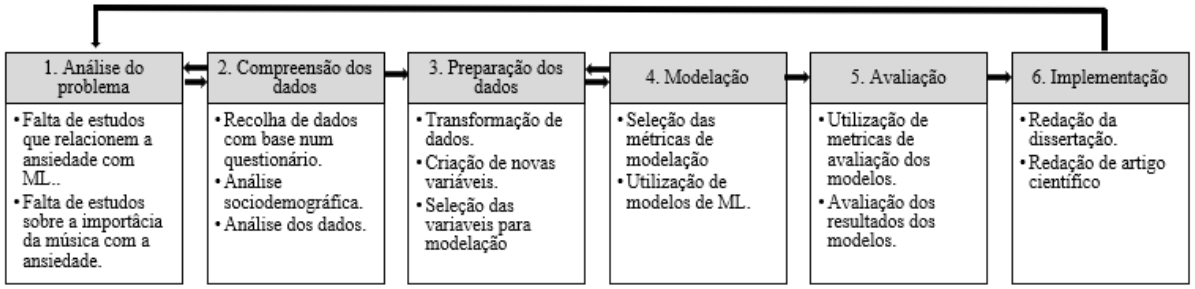


Figura 3.1: Metodologia CRISP-DM aplicada

3.3. Recolha de dados

De forma a conseguir obter os dados necessários para a realização do estudo da importância da música na ansiedade e perceber qual o impacto de um sistema de recomendações de música no seu tratamento, foi desenvolvido um questionário que contém perguntas relativas a problemas de ansiedade generalizada, perguntas sobre música e qual o seu impacto na vida das pessoas, assim como perguntas relativas ao impacto da música em possíveis situações de ansiedade.

A participação no questionário é anónima, de forma a garantir que as respostas sejam imparciais e que seja preservada a confidencialidade dos participantes.

Quase todas perguntas do questionário foram desenvolvidas de forma a ser obrigatória a sua resposta, não sendo possível a sua entrega na falta de alguma resposta obrigatória.

O questionário tem quatro secções principais para a aquisição de informação. A primeira secção questiona sobre Generalized Anxiety Disorder (GAD). Esta abordagem sobre as perguntas GAD inspira-se num estudo já realizado por Rahman et al. (2021), tendo este sido adaptado para a presente tese. Esse questionário tem como objetivo determinar níveis de ansiedade com base em perguntas de GAD em conjunto com perguntas relacionadas com jogos online, tendo sido este adaptada para a música. Este utiliza as mesmas sete perguntas GAD, assim como a sua utilização na definição dos níveis de ansiedade por base na soma e divisão em quatro grupos. A segunda secção questiona sobre a importância da música na vida dos respondentes. A terceira foca-se na importância da música em momentos de ansiedade. Por fim, a quarta secção pretende caracterizar os inquiridos em termos sociodemográficos. O questionário utilizado encontra-se no apêndice B. Este questionário foi disponibilizado online através do Google Forms[2] entre 19 de agosto e 6 de setembro de 2022, fazendo-se chegar às pessoas através da partilha do mesmo pelas redes sociais Facebook, Instagram e WhatsApp por via do link mencionado em nota de rodapé e pela Figura A.21 que pode ser verificado no Apêndice C.

3.4. Compreensão dos dados e análise demográfica

Após a recolha de dados, provenientes do questionário, foi feita uma análise inicial dos mesmos, de forma a conseguir perceber qual o tipo de respostas que foram obtidas e a população respetiva. Esta compreensão dos dados existentes numa fase inicial serviu para perceber a qualidade dos dados e as características da população em estudo, assim como quais os ajustes necessários à base de dados e as possíveis alterações e transformações iniciais de forma a melhor a qualidade dos dados tendo em conta o objetivo do estudo. A ferramenta utilizada nesta análise e nos passos futuros foi python, utilizando a plataforma Orange[3] de forma a simplificar todo o processo com widgets.

[2] Link: https://forms.gle/HJYgVHhhgYnBwpxbA
[3] Link: https://orangedatamining.com/download/#windows

De forma a facilitar a compreensão dos dados existentes provenientes das respostas, foram utilizados gráficos de barras e gráficos circulares de modo a facilitar a leitura generalizada da população e as suas características.

Numa fase inicial, foram recolhidas 500 respostas de pessoas diferentes e únicas de forma aleatória. Nesta fase do estudo todos os dados foram considerados, visto que as perguntas têm um carácter de resposta obrigatório, e as que não têm, não são necessárias respostas consoante as características envolventes.

A última fase do questionário representa perguntas sociodemográficas, para perceber e compreender melhor a população em estudo, sabendo as suas características de caracter mais pessoal, sem nunca comprometer a privacidade e anonimidade no inquirido. Estas perguntas servem também para conseguir compreender e descobrir padrões nos inquiridos com base nas suas respostas às perguntas de ansiedade e música. Foram então utilizadas perguntas para saber o género, idade, atividade atual, nível de escolaridade e área de residência.

Dos 500 inquiridos, pode afirmar-se que a maioria é do sexo feminino, representando 70% da população. Apesar desta pergunta ter uma opção de resposta aberta, não houve respostas diferentes de "Masculino" e "Feminino", como se pode observar na Figura 3.2. Foi também utilizado os modos de proteção de privacidade, dando opção ao inquirido de preferir não responder a esta pergunta.

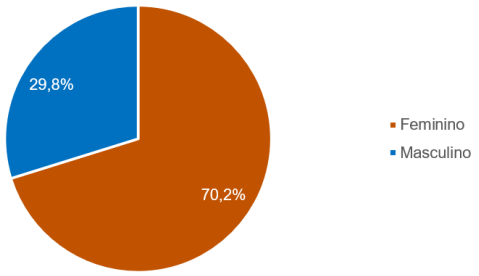


Figura 3.2: Resultados da pergunta "Qual o seu género?"

É possível verificar na Figura 3.3 que a grande maioria da população está no mundo do trabalho, sendo "trabalhador" ou "trabalhador-estudante", correspondendo a 82.8% dos inquiridos. Apenas 11.2% dos participantes se encontram apenas a estudar.

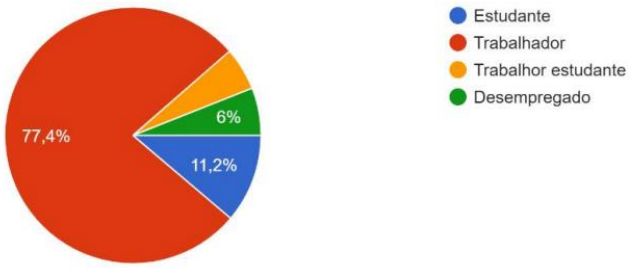


Figura 3.3: Resultados da pergunta "Qual é a sua atividade atual?"

Em relação ao nível de escolaridade apresentado pelos inquiridos, pode verificar-se uma predominância equitativa entre o "Ensino Secundário" e a "Licenciatura", representando, respetivamente, 44% e 35.2% da população. É de notar que apenas 10% da população tem o grau de mestre e que, existe 7.6% da população com o ensino básico, como se pode verificar na Figura 3.4.

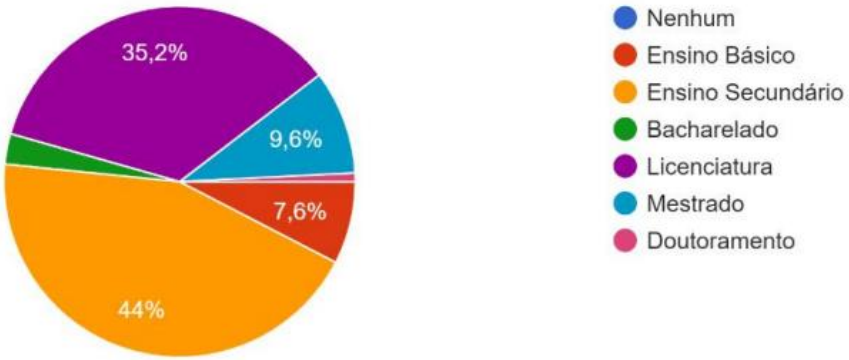


Figura 3.4: Resultados da pergunta "Qual é o seu nível de escolaridade?"

A idade dos inquiridos foi relevante no estudo de forma a perceber qual o significado das restantes perguntas sociodemográficas, como por exemplo o nível de escolaridade e atividade pessoal, assim como para perceber qual o estilo de música que ouvem e a plataforma utilizada para ouvir música. A idade dos inquiridos também foi relevante para perceber a presença de sintomas de ansiedade em cada faixa etária.

Pode verificar-se que neste estudo as idades variam entre os 14 e os 88 anos, sendo a média de idades de 39 anos. Dos 500 inquiridos, 100 tinham idade até aos 23 anos, correspondentes a 20% da população, 122 com idades compreendidas entre os 24 e os 39 anos, correspondente a 24.4% da população, 216 entre os 40 e os 54 anos, que corresponde a 43.2% da população e as restantes 62 com mais de 54 anos, que representam 12.4% da população. Como se pode verificar na Figura 3.5, houve uma maior concentração de dados nas idades compreendidas entre os 40 e os 54 anos.

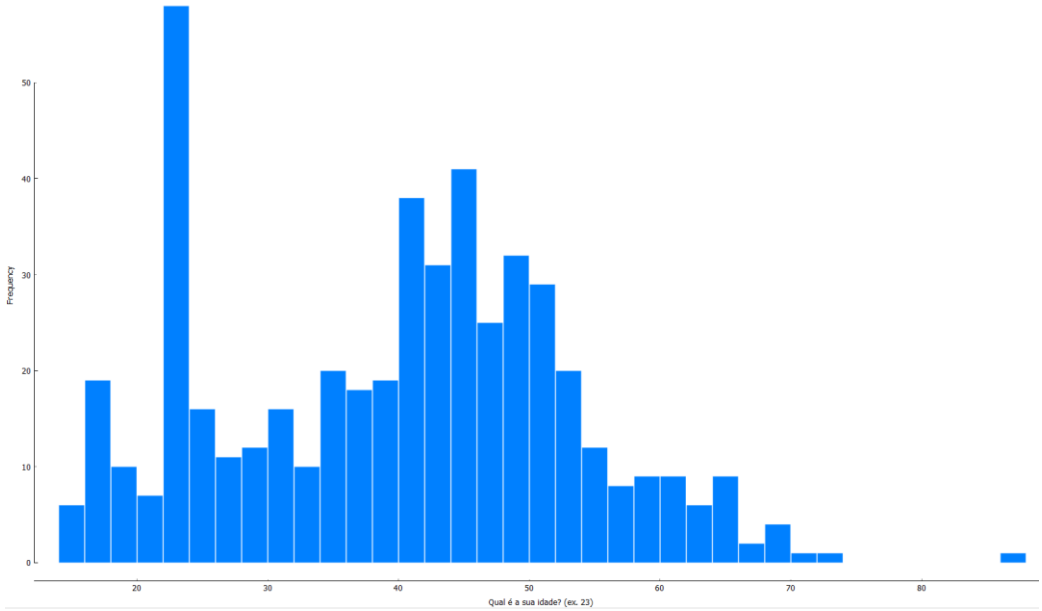


Figura 3.5: Resultados da pergunta "Qual é a a sua idade?"

Os inquiridos foram também questionados sobre a sua área de residência, onde os resultados desta pergunta não se diferenciam muito. Pode-se verificar que 57.4% dos inquiridos reside no meio urbano, correspondendo à maioria, apesar de haver também bastantes inquiridos, 42.6% a residir no meio rural, como é representado na Figura 3.6.

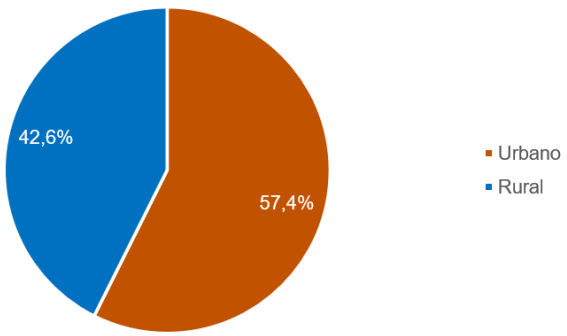


Figura 3.6: Resultados da pergunta "Qual é a sua área de residência?"

Na Figura 3.7 pode-se observar um conjunto de 7 gráficos de barras, correspondentes às 7 perguntas relacionadas com problemas de ansiedade generalizada. Cada gráfico contém 4 colunas correspondentes a cada resposta possível, como se pode observar na legenda da figura.

Pode-se afirmar que a resposta predominante em todas as perguntas é "alguns dias", correspondente à barra vermelha, tendo esta uma média 235 participantes. A resposta menos escolhida pelos inquiridos foi a "Quase todos os dias", correspondente à barra verde, tendo esta uma média de respostas de 48 participantes. Por norma, houve sempre mais participantes a escolher a opção "Mais de metade dos dias" em comparação com a resposta "Nunca", apesar de que, na quinta pergunta, estes valores se tenham alterado. Nesta quinta pergunta "Nas últimas duas semanas, com que frequência se sentiu afetado pelo seguinte problema: Estar tão inquieto que é difícil ficar parado", os resultados mostraram-se diferentes em comparação com as restantes perguntas. Cento e setenta e sete participantes responderam "Nunca" a esta pergunta, mais de o dobro em comparação com a média desta resposta nas outras perguntas, que é noventa. Esta diferença de respostas fez diminuir, por consequência o número de respostas das restantes opções, mas esta diminuição foi equilibrada, ou seja, diminui, em média, o mesmo número de respostas em cada opção. Pode-se verificar também que a resposta "Mais de metade dos dias" foi a segunda opção mais escolhida pelos inquiridos, com a exceção da segunda e quinta pergunta, onde a segunda opção mais escolhida foi "Nunca".

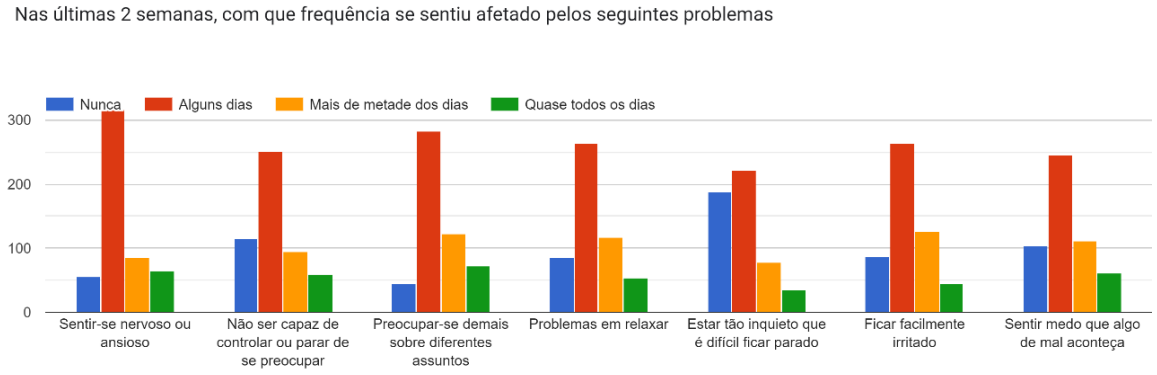


Figura 3.7: Resultados das perguntas de GAD no questionário

Na Tabela 1 pode-se observar, em forma numérica, todos os valores concretos relativos ao número de respostas às sete perguntas de problemas de ansiedade generalizada, assim como a média relativa a cada opção escolhida pelos inquiridos.

Tabela 3.1: Estatística das perguntas de problemas de ansiedade generalizada

	"Nunca"	"Alguns dias"	"Mais de metade dos dias"	"Quase todos os dias"
GAD 1	54	301	83	62
GAD 2	110	224	91	55
GAD 3	42	272	115	71
GAD 4	85	252	113	50
GAD 5	181	213	73	33
GAD 6	81	256	120	43
GAD 7	101	234	106	59
MÉDIA	92	253	99	52

Pode-se verificar que 49.6% dos inquiridos deste estudo considera que tem ansiedade, 25% não tem a certeza que tem e 25.4% considera que não tem ansiedade. Tais valores não são representantes da pergunta "Já foi diagnosticado profissionalmente com ansiedade?", uma vez que apenas um quarto dos inquiridos admite terem sido diagnosticados.

Na segunda parte do questionário, os inquiridos foram questionados com perguntas sobre música, sendo uma delas "Que plataforma usa para ouvir música?". Esta pergunta, não sendo obrigatória, houve 21 inquiridos a não responder. Pode-se verificar a existência de 2 plataformas muito utilizadas pelos inquiridos, tendo uma diferença bastante significativa em comparação com as outras plataformas escolhidas, sendo estas o "Youtube", existindo 348 respondentes a utilizar esta plataforma para ouvir música, o que corresponde a 72.7% da população, e o "Spotify" com 238, correspondendo a 49.7%. Outras plataformas foram também escolhidas pelos inquiridos, como o "SoundCloud", com 5 respostas (1%), "Apple Music", com 15 respostas (3.1%) e Deezer com 4 respostas (0.8%).

Como se pode verificar, estas plataformas tem um peso muito menor na escolha dos inquiridos, abrangendo apenas, em conjunto, 5% da população. Outras plataformas também foram escolhidas, na opção "Outro", sendo esta uma opção de resposta aberta, havendo 57 respostas. Estas opções extra escolhidas pelos ouvintes serão analisadas, transformadas e agrupadas na secção seguinte "Preparação de dados".

Os inquiridos foram também questionados sobre "Quantas horas ouve música por semana?", havendo a possibilidade de o inquirido responder "Não oiço música". Neste caso, houve 14 participantes a escolher esta opção, correspondendo a 2.8% da população. A distribuição do número de respostas entre perguntas não variou muito, não havendo uma opção com demasiado distanciamento entre as outras. A resposta mais escolhida pelos inquiridos foi "1-3" horas por semana, correspondendo a 28.6% dos inquiridos, seguindo-se a opção "3-5" horas por semana, correspondendo a 17%. Estas duas opções representam perto de metade da população. Cerca de um quarto da população do estudo ouve "5-10" horas de música por semana, fazendo com que as restantes perguntas (metade das opções), representem um quarto da população, como se pode verificar na Figura 3.8.

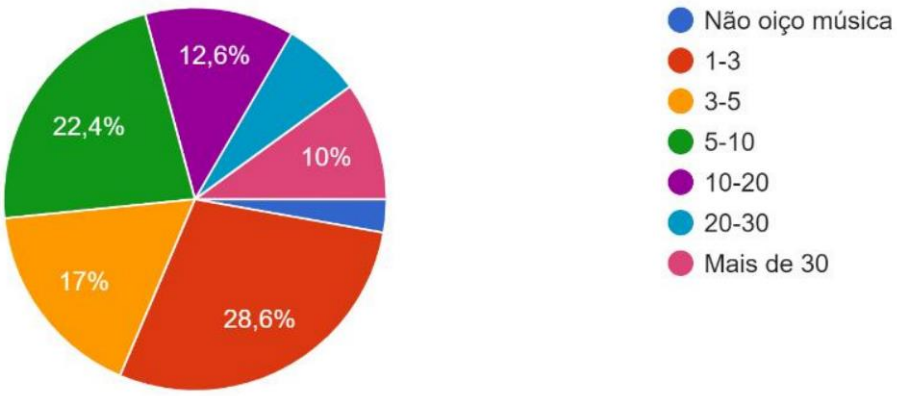


Figura 3.8: Resultados da pergunta "Quantas horas ouve música por semana?"

À pergunta "Ouve música enquanto faz outra atividade", pode-se verificar que a maioria da população respondeu que "Sim", correspondendo a 94.4% dos inquiridos.

Em sequência da pergunta anterior, o inquirido foi questionado sobre qual a atividade que praticava enquanto ouvia música (no caso de ouvir música enquanto faz outra atividade). Sendo esta uma pergunta de escolha múltipla e não obrigatória, houve muitos inquiridos a responder a mais do que uma atividade, e alguns a não escolher atividade nenhum, neste caso, 21 inquiridos.

Pode-se verificar que grande parte da população (79.3%) ouve música enquanto anda de carro, enquanto faz tarefas domésticas (65.1%) e enquanto trabalha. Verifica-se que cerca de 40% dos inquiridos ouve música enquanto faz desporto. Outras duas atividades foram também consideradas relevantes como "Jogar videojogos" e "Estudar", apesar de não ter tanto significado no espaço de amostra. Os inquiridos também tinham a opção de responder a esta pergunta de forma aberta, escrevendo a atividade que pratica enquanto ouve música. Estas opções extra (17 opções) escolhidas pelos ouvintes serão analisadas, transformadas e agrupadas na secção seguinte "Preparação de dados".

Os inquiridos foram também questionados sobre que tipo de música ouvem. Sendo esta uma pergunta de escolha múltipla e não obrigatória, houve muitos inquiridos a responder a mais do que um tipo de música, e alguns a não escolher nenhum tipo, neste caso, 25 inquiridos. Foram considerados alguns tipos de música mais relevantes para a população alvo, sendo estas "Pop"; "Rock", "Techno", "Eletronic", "Jazz", "Classic", "Hip pop", "Funk" e "Reggaeton", sendo os que obtiveram mais respostas o "Pop", com 80% da população a ouvir este tipo de música, o "Rock", com 49.7% e o "Hip pop", com 34.3%. Todos os outros estilos de música considerados relevantes obtiveram também resultados consideráveis, mas não tão elevados, variando estes entre 13.1% e 29.5%. O estilo de música menos ouvido das opções predefinidas foi o "Techno" apenas com 13.1% da população a ouvir. Os inquiridos também tinham a opção de responder a esta pergunta de forma aberta, escrevendo o tipo de música que ouve. Estas opções extras (79) escolhidas pelos ouvintes serão analisadas, transformadas e agrupadas na secção seguinte "Preparação de dados".

Na terceira fase do questionário, foram desenvolvidas perguntas relativas à importância da música na ansiedade, assim como perguntas que reúnem informações comuns às duas áreas de investigação. Foi então desenvolvida também uma pergunta relativa a qual o estilo de música que ouve quando o inquirido se sente ansioso (no caso de ouvir), onde esta mantém os mesmos padrões e estilo de pergunta que a anterior, apesar de haver menos inquiridos a responder, cerca de 58%, tendo em conta que o inquirido só deveria responder no caso de ouvir música quando sente ansiedade. Pode-se verificar que nas opções de resposta aberta, houve mais respostas relacionadas com as opções de música de relaxamento, tais como "Sons da Natureza", músicas relaxantes no modo geral ou músicas mais calmas.

Foi também considerado relevante perguntar aos inquiridos se "Quando se sente ansioso, ouve música?", onde um terço da população respondeu que "Sim", como se pode verificar na Figura 3.9. Esta informação é relevante para o estudo de forma a fazer a distinção entre os inquiridos que ouvem música e não têm ansiedade, as que não ouvem música quando têm ansiedade e as que utilizam a música por forma a tentar aliviar os sentimentos de ansiedade.

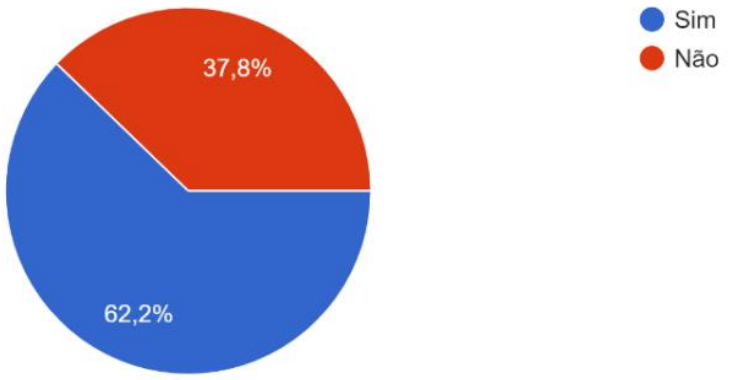


Figura 3.9: Resultados da pergunta "Quanto se sente ansioso, ouve música?"

Para além de perceber a importância da música nos inquiridos, quando estes se sentem em momentos de ansiedade, notou-se a necessidade de perceber qual a importância e o nível de impacto que a música tem, tentando quantificar o tempo que ouve músicas nestes momentos, comparativamente aquando se sente bem. Foi então utilizada a pergunta "Quando se sente ansioso…", onde 87% dos inquiridos ouve a mesma ou mais quantidade de música, sendo que 32.8% ouvem mais música, como se pode verificar na Figura 3.10.

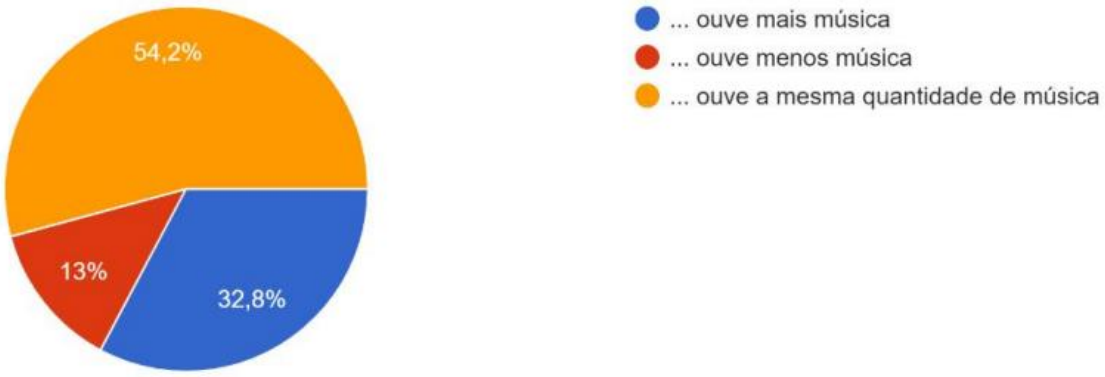


Figura 3.10: Resultados da pergunta "Quanto se sente ansioso…"

Por fim, foi necessário perceber qual foi o resultado final da utilização da música na diminuição dos sentimentos de ansiedade, percebendo se o inquirido sente que a música tem um impacto positivo na atenuação deste problema e se é uma boa técnica para a resolução do seu problema ou pelo menos, para a sua melhoria. Foi então feita a pergunta "Sente que ouvir música diminui o seu estado de ansiedade?", onde 78.2% da população respondeu que "Sim", como se pode verificar na Figura 3.11.

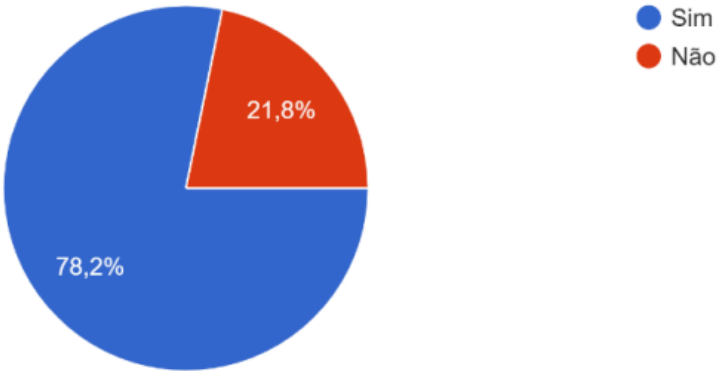


Figura 3.11: Resultados da pergunta "Sente que ouvir música diminui o seu estado de ansiedade?"

3.5. Preparação dos dados

Após a recolha dos dados e a sua análise, foi feita uma transformação aos que se consideraram irrelevantes ao estudo, assim como a sua eliminação. O objetivo desta preparação é eliminar dados irrelevantes que não trazem informação útil ao caso, bem como valores que podem distorcer os resultados, ou transformá-los, de forma a adaptá-los de forma adequada ao estudo.

As perguntas da primeira fase do questionário, ou seja, as relacionadas com os problemas de ansiedade generalizada, de forma a melhor analisar e contabilizar os níveis de ansiedade de cada individuo, o tipo de dados foi alterado de categórico para número, de forma a facilitar a soma dos valores das 7 perguntas. Esta transformação tem como objetivo a criação de uma nova variável numérica "GAD Total", que representa a soma dos valores das sete perguntas relativas a problemas de ansiedade geral.

Foram, portanto, transformados os valores “Nunca” para o valor “0”, “Alguns dias” para “1”, Mais de metade dos dias” para “2” e “Quase todos os dias” para “3”. Será também criada uma variável “Nível de ansiedade”, que corresponde ao agrupamento dos valores da variável “GAD Total” em 4 grupos. O primeiro “Mínima ansiedade”, correspondendo aos valores entre 0 e 4; o segundo “Ligeira ansiedade, correspondendo aos valores entre 5 e 9. O terceiro “Média ansiedade”, correspondendo aos valores entre 10 e 14. E o Quarto grupo “Severa ansiedade”, correspondendo aos valores entre 15 e 21. Desta variável nova criada, viu-se a necessidade de criar uma variável “Tem ansiedade”, definindo esta se o inquirido tem ou não ansiedade, tendo os valores possíveis “Sim”, correspondentes aos valores “Mínima ansiedade” e “Ligeira Ansiedade”, e “Não”, correspondentes aos valores “Média ansiedade” e “Severa ansiedade”. Esta variável, apesar de ter o mesmo significado que a outra, não se notou redundante, uma vez que “Tem ansiedade” será a variável alvo final, mas também foram realizados testes utilizando a variável “Níveis de ansiedade”, como será possível verificar nas secções 4.5 e 4.6.

Não se viu a necessidade de eliminar nenhuns dados uma vez que a resposta a estas perguntas é independente das respostas a outras perguntas, ou seja, o seu resultado não se verificará incoerente com o decorrer do questionário, sendo todos os conjuntos de resposta possíveis.

Já na segunda fase de perguntas foram feitas alterações, quer a nível de transformações quer a nível de eliminações de dados. Encontrando padrões, foram feitos também agrupamentos de dados nas perguntas de resposta aberta tentando criar grupos, de forma a reduzir as possibilidades de diferenciação excessiva e desnecessária dos dados.

Neste caso foram feitos agrupamentos e eliminações na pergunta “Que plataforma usa para ouvir música”. Para além das respostas predefinidas no questionário, “Youtube”, “Spotify”, “SoundCloud”, “Apple Music” e “Deezer”, foi criada uma opção de resposta aberta, o que resultou em 66 novas respostas diferentes.

Com base na análise feita, verificou-se que apenas um inquirido respondeu “Amazon Music” e dois responderam “Tidal”, o que não se demonstrou um número suficientemente grande de forma a justificar a criação de um novo grupo, o que pode comprometer a qualidade dos resultados após a aplicação do algoritmo. Foram também eliminadas as opções “Deezer” e “SoundCloud”, por representarem apenas 2% da população e, com a escassez de dados, não ser possível encontrar um padrão. Verificou-se também que houve 64 inquiridos a responder na resposta aberta “Radio”, ou plataformas relacionadas com “Radio”, pelo que se considerou importante guardar este padrão e criar este novo grupo. Foram também encontradas duas plataformas que não foram identificadas como tal, tais como “Q” e “Amerigo”, o que resultou na sua eliminação. Por último, constatou-se também a presença de respostas relacionadas com hardwares utilizados para a reprodução da música tais como “MP3”, “CD’s” ou “Aparelhagem”, o que resultou também na criação de um novo grupo chamado “Private”. Na tabela 2 pode-se verificar as opções existentes e as novas opções criadas, apresentando-se entre parêntesis o número de recorrências de cada opção e, se for o caso, a informar se foi eliminada.

Tabela 3.2: Lista de transformações em “Que plataforma usa para ouvir música?”

EXISTENTES	NOVAS
Youtube (352)	Amazon (1) - Eliminado
Spotify (242)	Tidal (2) - Eliminado
SoundCloud (5) - Eliminado	Radio (64)
Apple Music (18)	Private (13)
Deezer (4) - Eliminado	

Todo o agrupamento pode ser visualizado na Tabela 3.2, assim como a explicação de alguns dados transformados. Alguns dos dados desta tabela encontram-se repetidos, uma vez que há respostas que se encontram em ambos os novos grupos criados.

Tabela 3.3: Lista de transformações em "Que plataforma usa para ouvir música?"

Nova plataforma	Respostas originais
Radio	"Rádio", "Radio", "Rádio", "Radio", "rádio", "Rádio Online", "Radio comercial", "Young Radio", "auto rádio", "rádio no carro", "Canais música", "Radio, cds e vinil", "Carro", "Rádios Online", "Rádio mp3", "CDs, rádio", "Progamas de rádio (Radio France; antena 2 e outras)", "Radio online", "Radio, aparelhagem", "Rádio e TV", "Rádio no carro", "Rádio (o tradicional...)", "RADIO"
Private	"Música guardada na hard drive", "Play List própria", "CD", "Radio, cds e vinil", "Carro", "Mp3", "Rádio mp3", "CDs, rádio", "Radio, aparelhagem",
Dados eliminados	"Q", "Amerigo", "Tide", "Tidal", "Amazon"

Relativamente à pergunta "Quantas horas ouve música por semana?", notou-se a necessidade de comparar as respostas desta pergunta com as restantes perguntas relativas à ansiedade e à música, uma vez que a escolha de cada resposta pode criar dificuldades na aprendizagem do modelo, para além da ambiguidade em termos teóricos. Esta pergunta foi, portanto, comparada com as seguintes perguntas: "Que plataforma usa para ouvir música?"; "Ouve música enquanto faz outra atividade?"; "Caso sim, qual?"; "Que tipo de música ouve?"; "Quando se sente ansioso, ouve música?"; "Caso sim, que tipo de musica?"; "Quando se sente ansioso…"; "Sente que a musica diminui o seu estado de ansiedade?". Todas as respostas que se apresentavam ambíguas estavam relacionadas com o facto de o inquirido ter respondido "Não oiço música", mas responder às outras perguntas, presumindo que ouve música, o que se viu apropriado alterar os dados nas respostas "Não oiço música" para "1-3" horas. Desta forma mantém se a coerência dos dados, não perturbando o desempenho do algoritmo e a veracidade das respostas.

Foi também necessário proceder à alteração de dados relativos à pergunta "Ouve música enquanto faz outra atividade?", uma vez que foram encontradas ambiguidades relativas à pergunta "Caso sim, qual?". Foi também encontrado que há inquiridos a responder "Não" à primeira pergunta, mas depois dizem qual a atividade que praticam enquanto ouvem música, o que levou à necessidade de transformação dos dados de "Não" para "Sim".

Em relação à pergunta "Quando se sente ansioso, ouve música?", foram descobertas incoerências em relação à pergunta "Caso sim, que tipo de música?", sendo que o inquirido respondeu que ouve música quando se sente ansioso, mas não disse qual o tipo de música que ouve. Nos casos em que o inquirido respondeu que "Não" à primeira pergunta, mas disse qual o tipo de música, transformou-se os dados de "Não" para "Sim". Foram também alterados os dados em que os inquiridos responderam que ouviam música quando se sentiam ansiosos, mas que não responderam qual o tipo de música, pelo que se alterou a resposta de "Sim" para "Não".

Foram também feitos agrupamentos em relação às perguntas de escolha aberta relativas às atividades praticadas quando o inquirido ouve música e em relação ao tipo de música que ouvem.

Sendo assim, em resposta à pergunta "Caso sim, qual?", não foram considerados nenhuns dados relativos à resposta aberta, uma vez que alguns se apresentavam abusivos, não sendo estes considerados, ou muito generalistas, não se enquadrando em nenhum grupo de atividades, sendo, portanto, eliminados. Foram também inicialmente criados três novos grupos, para além dos já existentes por definição, mas não se demonstraram em número relevante para o estudo, o que poderá comprometer a aprendizagem da ferramenta, não sendo estes considerados. Os grupos criados foram "Ler", "Tomar banho" e "Relaxar", apresentando, respetivamente, dois, quatro e quatro resultados. Apenas as respostas "Nos transportes" e "Caminhadas" foram consideradas, uma vez que estas se podem encontrar nas opções predefinidas "Andar de carro" e "Fazer desporto", respetivamente. Todos os resultados obtidos na resposta de opção aberta podem ser verificados na Tabela 4.

Tabela 3.4: Lista de transformações em "Caso sim, qual?"

Transformação feita	**Razão da transformação**	**Respostas originais**
Eliminar	Atividades abusivas tendo em conta o objetivo deste estudo, não apresentando relevância	"Rolling a joint", "Fumar charros", "Sexo"
Eliminar	Atividades muito generalistas	"Múltiplas atividades", "Hobbies",
Eliminar	Número insuficiente para criar grupo "Ler"	"Ler", "Ler", "Ler, Simplesmente relaxar"
Eliminar	Número insuficiente para criar grupo "Relaxar"	"Na cama", "Pintar, desenhar", "Antes de dormir", "Meditação"
Eliminar	Número insuficiente para criar grupo "Tomar banho"	"banho", "Banho/vestir-me, "Tomar banho", "Tomar duche"
Transformar	Inserir em "Fazer desporto"	"Caminhadas"
Transformar	Inserir em "Andar de carro"	"Nos transportes"

Em relação às transformações que foram feitas aos tipos de música que os inquiridos ouvem, foram criados diversos grupos em que, cada grupo, corresponde a um novo estilo de música. Nenhum dos novos estilos obteve grande representação da população tendo, no máximo, oito inquiridos a ouvi-los. Apesar de na pergunta anterior os novos grupos terem sido eliminados por fraca representação da população, estes novos estilos não serão eliminados, uma vez que a representação de cada estilo é única e necessária para a execução do algoritmo. Foram, portanto, criados os estilos: "Indie"; "Metal"; "Portuguesa"; "Fado"; "MPB"; "Drill"; "R&B"; "Lo-fi"; "Fi-bum"; "World music"; "Mantras"; "Folk"; "Afro"; "Latino"; "Hard Rock"; "Grunge"; "Relaxar", como se pode verificar na Tabela 5.

Esta tabela tem uma lista de notas[4] que contém informações relevantes para perceber a razão da transformação.

Tabela 3.5: Lista de transformações em "Que tipo de música ouve?"

Original	Transformação	Original	Transformação
A que estiver a passar na radio	Nota 1	Louvores e reflexões	Nota 6
Alternativa	Alternativa	Mantras	Mantras
Bandas sonoras de filmes	Nota 1	Mantras	Mantras
De tudo um pouco	Nota 1	Meditação	Relaxar (Nota 3)
Reagge	Juntar a "Reaggaeton"	Metal	Metal
Fado	Fado	Metal	Metal
Instrumental de diversos géneros	Nota 1	O que passa na radio	Nota 1
Músicas mais calmas	Relaxar (Nota 3)	Música Portuguesa	Portuguesa (Nota 2)
O que estiver a dar na radio	Nota 1	Pimba	Portuguesa (Nota 2)
O que estiver a dar	Nota 1	Comercial	Nota 6
O que estiver a passar na radio	Nota 1	Kizomba	Kizomba
Pimba	Portuguesa (Nota 2)	Etnica	Nota 6
Bossa Nova	Brasileira (Nota 5)	Folk	Folk
Indie Pop	Indie (Nota 4)	Mantras	Mantras
Heavy Metal	Metal (Nota 4)	Alternativa	Alternativa
Portuguesa	Portuguesa (Nota 2)	City Pop	Pop (Nota 4)
Portuguesa de preferência	Portuguesa (Nota 2)	Fado	Portuguesa (Nota 2)
Qualquer uma, mas se for dos anos 80 melhor	Nota 1	Indie	Indie
Relaxamento	Relaxar (Nota 3)	Indie	Indie
Música Portuguesa	Portuguesa (Nota 2)	Kizomba	Kizomba
Radio	Nota 1	R&B	R&B

[4] Notas:
Nota 1: Este tipo de músicas são muitos gerais, tendo sido eliminadas
Nota 2: Todas as respostas relativas a músicas portuguesas como por exemplo pimba e opções que continham "Portuguesa", foram consideradas do mesmo grupo de música portuguesa.
Nota 3: Opções relativas a natureza e que contenham a palavra "Relaxar", foram todas agrupadas.
Nota 4: Estes estilos foram alterados para o estilo mais próximo que seja predefinido, como por exemplo "Indie Pop" foi agrupado em "Indie".
Nota 5: Sertanejo e MPB (música popular brasileira) e bossa nova foram agrupados no grupo "MPB".
Nota 6: Não são considerados estilos de música, sendo estes eliminados

Sou adepta de radio por isso, ouço um pouco de tudo	Nota 1	Afro	Afro
Tudo	Nota 1	Sertanejo	Brasileira (Nota 5)
80s	Nota 1	Relaxante	Relaxar (Nota 3)
Bossa nova	Brasileira (Nota 5)	Latino	Brasileira (Nota 5)
R&B	R&B	Bossa Nova	Brasileira (Nota 5)
Kpop	Pop (Nota 4)	Bossa Nova/musica brasileira	Brasileira (Nota 5)
Um pouco de tudo	Nota 1	Metal	Metal
Música portuguesa	Nota 2	Indie	Indie
Sertanejo	Brasileira (Nota 5)	Portuguesa	Portuguesa (Nota 2)
Drill	Drill	Hard Rock	Rock (Nota 4)
Fado	Portuguesa (Nota 2)	MPB	Brasileira (Nota 5)
Heavy Metal	Metal (Nota 4)	Grunge	Nota 1
R&B	Brazileira (Nota 5)	Indie Rock	Indie (Nota 4)
Kizomba	Kizomba	Indie	Indie
Lo-fi	Lo-fi	MPB	Brasileira (Nota 5)
Calma e relaxante	Relaxar (Nota 3)	Pimba	Portuguesa (Nota 2)
Sons da natureza	Relaxar (Nota 3)	Bossa Nova	Brasileira (Nota 5)
Instrumental	Nota 1	Chill out	Relaxar (Nota 3)
Fit-bum	Nota 1	Sons da natureza	Relaxar (Nota 3)
World music	World music	Metal	Metal
Programas de autores e temáticos	Nota 6	Todas as que passam na radio	Nota 1
Metal (Heavy, trash, gothic)	Metal (Nota 4)	Todas	Nota 1
Kizomba	Kizomba	Lo-fi-, Techno	Lo-fi, techno

As mesmas transformações e opções de agrupamento foram tomadas para a pergunta “Caso sim, que tipo de música”. Foram também criados novos grupos representantes de novos estilos de música, mantendo o mesmo processo de transformação.

Após todas estas transformações e agrupamentos relativos às perguntas da música, foram criadas novas variáveis (binárias) de forma a dizer se o inquirido ouve cada estilo de música. Esta técnica é denominada por "one-hot encode" em ML, sendo esta muito utilizada em variáveis categóricas que não têm uma relação ordinal, sendo a sua representação por números inteiros insuficiente. Neste caso, viu-se a necessidade de criar vinte novas variáveis, dez para os dez estilos de música relativos à música que o inquirido ouve na generalidade, e outras dez relativas aos mesmos dez estilos de música que este ouve em situações de ansiedade. Todas estas variáveis têm os valores possíveis "Sim" e "Não".

3.6. Análise exploratória dos dados

Após as transformações todas feitas aos dados, foi feita uma análise exploratória, onde serão estudadas as correlações entre perguntas e descoberta de alguns padrões. Este passo revela-se importante de forma a perceber quais as características e relações importantes entre variáveis para aplicar na fase de modelação.

Notou-se relevante explorar a correlação entre a variável nova criada "Nível de ansiedade" com as perguntas relativas a ansiedade, de modo a perceber, consoante o seu valor, se o inquirido acha que tem ansiedade, ou se já foi diagnosticado profissionalmente com ansiedade. Em primeira análise, foi verificada a correlação entre os níveis de ansiedade com a variável "Considera que tem ansiedade?". Verificou-se que, dos 61 inquiridos com um severo nível de ansiedade, 54 destes consideram-se ansiosos, mas apenas metade foi diagnosticada profissionalmente como tal. Mais de metade dos inquiridos que não têm a certeza se têm ou não ansiedade, apresentam níveis de ansiedade mínimos e ligeiros, apesar de um quarto apresentar sintomas médios presentes.

Dos inquiridos que consideram que não têm ansiedade, dois responderam que já foram diagnosticados profissionalmente como tendo ansiedade, o que pode, por exemplo, corresponder a já terem sido diagnosticados no passado, mas já não se consideram ansiosos. O mesmo acontece com os inquiridos que responderam que não tinham a certeza se tinham ou não ansiedade, cinco deles já foram diagnosticados profissionalmente.

De todos os inquiridos, os homens são os que apresentaram menos resultados positivos à perceção de presença de ansiedade severa comparativamente com as mulheres, apresentando metade dos resultados (em termos percentuais, ou seja, tendo em conta que a amostra tem mais mulheres que homens). Nota-se, por consequência, que os homens apresentam o dobro das respostas de ansiedade mínima face às mulheres, mantendo os outros valores de níveis de ansiedade similares. É interessante também perceber que as mulheres admitem mais sentir ansiedade, em comparação com os homens.

Os inquiridos mais velhos do estudo, com idades acima dos 55 anos, tendem em apresentar menos sintomas de ansiedade. Por outro lado, os jovens com idades compreendidas entre os 15 e os 20 anos, em conjunto com os adultos com idades compreendidas entre os 35 e os 40 anos, apresentam resultados mais indicativos da presença de ansiedade, em comparação com as outras faixas etárias.

Os níveis de ansiedade não apresentam nenhum padrão em relação à atividade atual do inquirido, nível de escolaridade ou área de residência, mantendo-se equitativamente distribuídos.

Existe uma minoria de inquiridos que sente que ouvir música não diminui o seu estado de ansiedade, mas quando se sentem ansiosos ouvem música. Deste mesmo grupo, existem respondentes que ouvem mais música quando se sentem ansiosos em comparação ao seu estado normal, podendo a quantidade de música ser um possível fator de impacto nos sentimentos de ansiedade.

Foi feita uma análise relativa aos estilos de música que os respondentes ouvem na sua generalidade, quando se sentem ansiosos e, no caso de não ouvirem o mesmo estilo de música quando sentem ansiedade, que estilos ouvem. Por falta de dados de caracterização de música, apenas foram considerados os estilos predefinidos nas opções do questionário, sendo estes “Rock”, “Hip Pop”, “Techno”, “Electronic”, “Jazz”, “Classic”, “Funk”, “Reggaeton” e “Pop”. A adicionar a estes estilos de música, foi também considerado o estilo “Relaxar”, de forma a perceber qual o tipo de música que ouve quando relaxa.

Foi descoberto que, dos inquiridos que ouvem música para relaxar e ouvem “Rock”, “Pop”, “Jazz” e “Classic”, cerca de 60% continuam a ouvir estes estilos quando se sentem ansiosos, tendo em especial atenção o “Classic” que 80% dos respondentes que ouvem este estilo, também o ouvem nos momentos de ansiedade. Pode-se também verificar que, deste 20% restantes, mais de metade prefere ouvir o estilo “Relaxar”. É de notar que para os estilos “Techno”, “Electronic”, “Hip Pop”, “Funk” e “Reggaeton”, as percentagens dos que ouvem música quando relaxam e ouvem o mesmo estilo de música em momentos de ansiedade ronda apenas os 40% e os 50%. Dos inquiridos que ouvem os estilos “Techno”, “Electronic, “Hip Pop” e “Rock” e, quando se sentem ansiosos, não ouvem os respetivos estilos, mas ouvem em momentos normais, metade deles demonstra preferir ouvir os estilos “Classic”, “Relaxar” e “Jazz”, sendo estas músicas mais calmas. No “Techno” e “Electronic” também se verificou que um terço desses inquiridos prefere ouvir “Pop”, sendo este valor inferior comparativamente ao “Hip Pop”, onde metade prefere ouvir também “Pop”. Já no “Hip Pop” demonstra-se também que um terço dos respondentes prefere ouvir “Reaggaeton” em vez de “Hip Pop”. O estilo “Jazz” é o que demonstra estilos de músicas diferentes em comparação com os restantes, conseguindo visualizar três estilos predominantes: “Rock”, “Pop” e “Relaxar/Classic”, ou seja, dos inquiridos que se sentem ansiosos e ouvem “Jazz” no geral, mas depois não ouvem quando se sentem ansiosos, os inquiridos tendem em ouvir os três estilos mencionados anteriormente. Dos respondentes que ouvem “Funk” mais de metade prefere ouvir os estilos “Pop” ou “Reggaeton” em momentos de ansiedade e o “Reggaeton” apresenta respostas parecidas ao “Hip Pop”. O estilo que se viu mais dificuldades em perceber quais os estilos tendenciais de troca em momentos de ansiedade foi o “Pop”, uma vez que todos os estilos foram escolhidos de forma parecida pelos inquiridos, apesar de se denotar que houve uma maior consistência no estilo “Classic”, que apresentou resultados na ordem dos 30%.

A pergunta “Sente que a música diminui o seu estado de ansiedade?” foi também comparada entre os estilos de música diferentes, verificando a percentagem de inquiridos que acha que a música diminui o seu estado de ansiedade e ouve música quando se sente ansioso, com cada estilo de música. Não foi demonstrada grande diferença entre uns e outros. Apesar desta parecença, reparou-se que o estilo “Classic” foi o estilo de música que apresentou menos resultados, ou seja, 90% dos inquiridos que ouvem música quando se sente ansiosos e ouvem música clássica, mencionaram que sentem que ouvir música diminui o seu estado de ansiedade, em comparação com os outros estilos, que representam percentagens na ordem dos 93 a 95 pontos percentuais.

É também importante notar que o estilo “Classic” é o que apresenta maior percentagem de inquiridos que ouve este estilo em qualquer estado emocional, quer estejam ansiosos ou não. Pode-se também notar que o “Funk” é o estilo de música que apresenta resultados mais positivos quando se compara o número de inquiridos que ouvem música quando se sentem ansiosos e que sentem que a música diminui o seu estado de ansiedade.

3.7. Modelação dos dados

Após a transformação dos dados e feita uma análise exploratória de forma a entender quais os possíveis padrões e variáveis dependentes, foram utilizados diversos modelos de ML com o objetivo de prever se o inquirido tem ou não ansiedade, utilizando algumas variáveis provenientes do questionário. Foi então criada a variável alvo “Tem ansiedade”, com os valores possíveis “Sim” e “Não”, tendo em conta que a variável alvo “Tem ansiedade” resulta de um subagrupamento dos valores da variável “Níveis de ansiedade” onde a resposta “Sim” é o agrupamentos dos níveis “Severa ansiedade” e “Média ansiedade”, e a resposta “Não” é o agrupamento dos níveis “Ligeira ansiedade” e “Mínima ansiedade”.

Os algoritmos utilizados são todos supervisionados, podendo estes serem divididos em três grupos distintos. O primeiro, referente a técnicas de regressão linear tais como o SVM, Logistic Regression, Naive Bayes e SGD. O segundo grupo consiste nos modelos de regressão não lineares, sendo estes o kNN, ANN, e Decision Tree os escolhidos a utilizar neste estudo. Por último, o terceiro grupo é formado por modelos que representam conjuntos de outros modelos num só, tais como o AdaBoost, Random Forest e Gradient Boosting.

De forma a operacionalizar estes modelos, foi utilizada a técnica de retenção que consiste em dividir aleatoriamente os dados do estudo em dois grupos distintos, um para treino e outro para teste. Foram então utilizados dois terços dos dados para treino, correspondentes a 330 dos 500 dados totais, e um terço dos dados para teste, correspondendo a 170 dos 500 dados totais.

Foi também aplicada a técnica de validação cruzada estratificada para cada algoritmo, onde é feita a divisão dos dados em 10 grupos. Cada modelo será testado utilizando exemplos de cada grupo individualmente, não utilizando exemplos de grupos diferentes ao mesmo tempo, sendo este processo repetido para cada grupo.

O valor da variável alvo de previsão deste estudo resulta da combinação de outras variáveis, nomeadamente das que são relativas ao conjunto das primeiras sete perguntas do questionário "Nas últimas 2 semanas, com que frequência se sentiu afetado pelos seguintes problemas?" e a variável "GAD Total". Uma vez que "GAD Total" resulta da soma dos valores de sete perguntas, a utilização desta variável não é exequível com a utilização do conjunto das primeiras sete perguntas do questionário, uma vez que estas são dependentes entre si. Tendo em conta que a variável alvo "Tem ansiedade" resulta de um subagrupamento dos valores da variável "Níveis de ansiedade", esta não pode ser utilizada. O mesmo acontece com a variável "GAD Total", uma vez que "Tem ansiedade" tem por base os valores de "GAD Total", sendo estas dependentes uma da outra, não podendo "GAD Total" também ser utilizada.

Recorreu-se à ferramenta Orange de forma a realizar toda a modelação feita, definição da variável alvo e apresentação de resultados de cada modelo, assim como a divisão da amostra nos grupos de teste e aprendizagem e escolha de variáveis para os modelos. Foi escolhida esta ferramenta de modo a facilitar todo o processo de manipulação de dados, assim como a sua análise e utilização de modelos, uma vez que o Orange é uma ferramenta "low-code"/"no-code", que utiliza pyhton, agilizando todo o processo e fluxo de trabalho. Esta pode ser obtida de forma gratuita no site da anaconda[5]

Esta ferramenta utiliza widgets de forma a definir todo o processo recorrendo a um "workflow". O widget "File" representa a escolha do ficheiro que contém os dados, neste caso em formato ".xlsx". O widget "Select Columns" foi utilizado de forma a escolher quais as variáveis a serem consideradas para os modelos e a definição da variável alvo, neste caso, "Tem ansiedade". O widget "Data Sampler" foi o responsável pela divisão da amostra nos grupos de teste e aprendizagem, passando para a modelação os dados de aprendizagem e para a previsão os dados de teste. Este widget também irá fornecer os dados de aprendizagem ao(s) modelo(s) escolhidos, após a avaliação dos modelos.

[5] https://orangedatamining.com/blog/2020/2020-02-24-anaconda-navigator/

De modo a utilizar os modelos e avaliá-los, foi utilizado o widget “Test and Score”, que irá receber os dados de aprendizagem e todos os modelos que foram alvo de estudo. Após a avaliação de cada modelo, tendo em conta as métricas definidas na secção 3.8, foi utilizada uma matriz de confusão (confusion matrix) para verificar todos os resultados obtidos por cada modelo. Recorreu-se a esta matriz de forma a visualizar todos os casos previstos por cada modelo, tendo em conta os valores verdadeiros e previstos. Após a escolha dos modelos a serem utilizados, foi então utilizado o widget “Predictions” de modo a testar os seus resultados. Este widget irá então receber os dados teste e os modelos escolhidos e avaliar qual a percentagem de previsão dos modelos, utilizando também as métricas definidas. Os resultados também são possíveis de serem observados numa matriz de confusão. Todo este workflow pode ser observado na Figura 3.12.

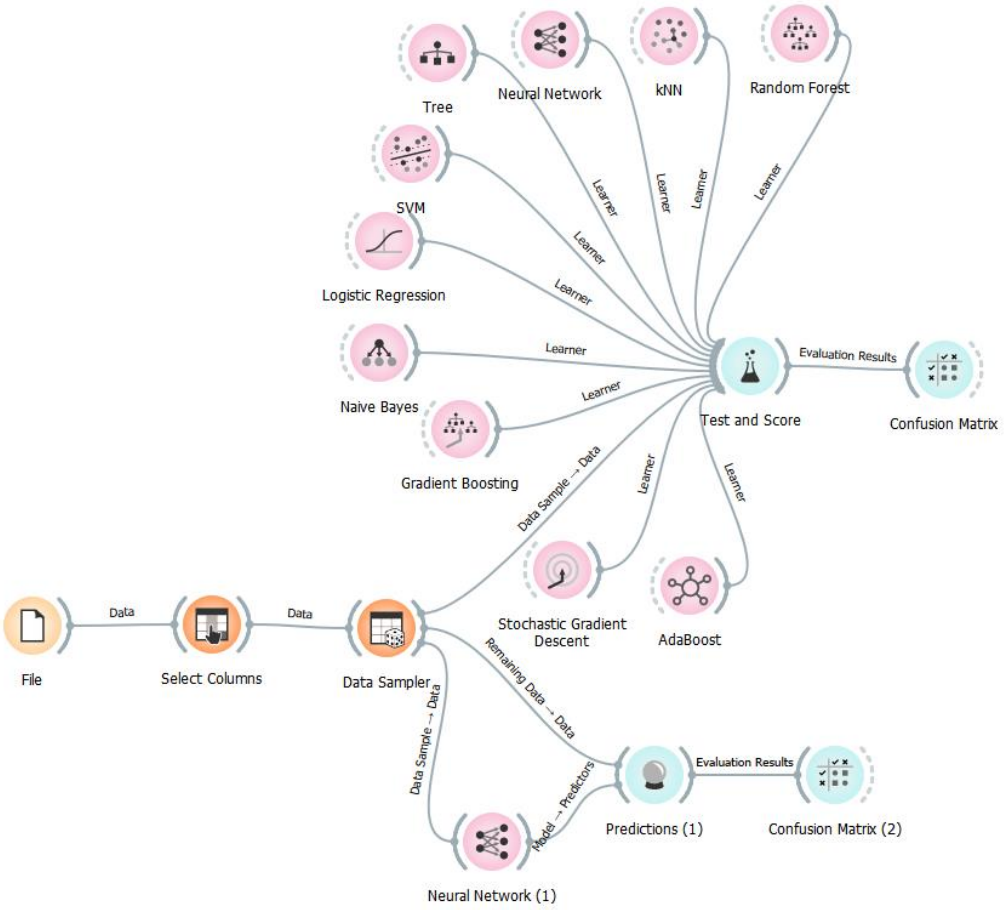


Figura 3.12: Workflow utilizado na modelação e avaliação dos modelos, fluxo com leitura da esquerda para a direita.

Foi feita também uma experiência parecida com a anterior, mas agora com uma variável alvo diferente, sendo esta "Níveis de ansiedade" em vez de "Tem ansiedade". Todo o processo manteve-se o mesmo à exceção da utilização da variável "Tem ansiedade", que se notou diretamente dependente da nova variável alvo, sendo esta excluída do estudo. Apesar de ser interessante o estudo dos níveis de ansiedade de cada individuo, os resultados obtidos relevam a pequena dimensão dos dados recolhidos, não permitindo retirar conclusões relevantes em termos de modelação. Esta segunda experiência categorizou então os 500 inquiridos em quatro níveis de ansiedade "Mínima ansiedade", "Ligeira ansiedade", "Média ansiedade" e "Severa ansiedade".

3.8. Avaliação do modelo

De forma a avaliar os modelos utilizados no estudo, foram aplicadas quatro métricas de avaliação sendo estas a "Precision", "F1", "Recall" e "Classification accuracy" (CA). Estas métricas são o resultado do modelo, que têm origem na classificação dos dados utilizados para treino, que utilizam quatro valores diferentes representados por:

- Verdadeiros positivos (VP) – valores positivos à questão que o modelo previu corretamente.
- Verdadeiros negativos (VN) – valores negativos à questão que o modelo previu corretamente.
- Falsos positivos (FP) - valores positivos à questão que o modelo não previu corretamente.
- Falsos negativos (FN) - valores negativos à questão que o modelo não previu corretamente.

Estes valores serão então utilizados pelas seguintes métricas:

- Precision – Representa o rácio entre os VP e todos os positivos, ou seja, quando o modelo prevê um certo valor, por exemplo "Sim", quantas vezes este prevê "Sim".
- CA – Representa o rácio entre os valores verdadeiros e todos os valores possíveis, ou seja, quantas vezes o modelo acerta na resposta, qualquer que seja a resposta, por exemplo, quantas vezes o modelo acerta a dizer "Sim" e "Não".
- Recall – Representa o rácio entre os VP e a soma entre os VP e os FN, ou seja, quando o valor verdadeiro é "Sim", quantas vezes o modelo vai prever "Sim"
- F1 – Representa a média harmónica entre a Precision e o Recall.

De forma a avaliar os resultados do modelo e perceber quais os algoritmos que melhor se adequam ao estudo e à previsão da variável alvo, será verificado o valor resultante F1, uma vez que este representa a média harmónica entre os dois valores considerados da precisão e do Recall. Este valor, assim como o CA, Recall e Precision, varia entre 0 e 1, representando o rácio de acerto consoante as métricas acima explicadas. Quanto mais o valor se aproxima de 1, maior será a percentagem de acerto da variável alvo feita pelo modelo. Desta maneira, os algoritmos que obtiverem maiores valores de F1 serão considerados como mais eficazes.

Capítulo 4 - Resultados e Discussões

4.1. Resultados da análise exploratória dos dados

Notou-se que, dos inquiridos que consideram que têm ansiedade, 49% não foi diagnosticado profissionalmente como ansiosos, o que poderá ser um sinal de possível necessidade de contacto com um médico profissional.

Os inquiridos que ouvem mais música demonstram menores níveis de ansiedade, sendo 80% dos inquiridos a terem resultados de presença de níveis de ansiedade ligeiros e mínimos, comparativamente aos que têm resultados de presença de ansiedade superiores (20%).

Independentemente do nível de ansiedade dos inquiridos, quando se sentem ansiosos, 54% sente a necessidade de ouvir a mesma quantidade de música e 33% sente a necessidade ouvir mais música, mostrando que, na generalidade, os inquiridos gostam de ouvir música quanto se sentem ansiosos.

Cinquenta e seis porcento dos inquiridos admite ouvir música quando se sente ansioso e sente que esta ação diminui o seu estado de ansiedade, correspondendo a 280 inquiridos. Destes, apenas 13 admitem ouvir menos música nestes momentos, apresentando também sintomas de ansiedade mais agravados. Pode-se verificar também que, dos 280 inquiridos, uma grande parte admite a ouvir mais música quando se sentem ansiosos, correspondendo a 42%, em comparação com os que ouvem a mesma quantidade de música, correspondendo a 54%.

Dos inquiridos que ouvem música quando se sentem ansiosos e sentem que ouvir música diminui o seu estado de ansiedade, independentemente dos seus níveis de ansiedade diagnosticados, estes vão ouvir a mesma, ou mais, quantidade de música.

Existe também um universo de inquiridos que não pertencem aos padrões resultantes da análise feita, podendo-se denotar de outliers, mas também com algum interesse de investigação. Por exemplo, os respondentes admitem poder ter ansiedade, mas ouvirem “menos música”, “a mesma quantidade de música” ou “mais música”. O efeito da música pode ser positivo no seu estado ansioso, independentemente da quantidade de música que ouvem. Existem também inquiridos que dizem não terem ansiedade, na generalidade, mas quando se sentem ansiosos ouvem mais música, mostrando a importância da música nestas situações.

Foram apresentadas situações em que os inquiridos dizem ouvir música quando se sentem ansiosos, mas que ouvem menos música, podendo este caso representar inquiridos que se sentem melhor a ouvir música, mas não tanta como antes, ou inquiridos que quando se sentem ansiosos, esquecem-se de ouvir música devido ao seu estado de ansiedade, no entanto quando ouvem música sentem-se melhores. O mesmo acontece no grupo de inquiridos que não ouvem música quando se sentem ansiosos, mas sentem que ouvir música diminui o seu estado de ansiedade.

Todos os estilos de música foram também comparados com os níveis de ansiedade presentes em cada inquirido, para verificar a existência de possível relação entre os estilos de música ouvidos pelos respondentes e o seu nível de ansiedade. Em quase todos os estilos não se verificaram padrões de alteração do número de inquiridos mais, ou menos, ansiosos, mantendo-se os quatro grupos de níveis de ansiedade entre todos os estilos. Apenas os estilos "Hip Pop", "Funk" e "Electronic" demonstraram ter mais respondentes com níveis de ansiedade inferiores comparativamente com os outros estilos musicais, havendo cerca de 20% a mais de respondentes com menores níveis de ansiedade.

4.2. Resultados da modelação dos dados

Na secção 3.7. foi explicada a modelação dos dados, quais os algoritmos a utilizar, quais as técnicas de modelação utilizadas e todo o workflow da modelação. Na presente secção são apresentados os resultados da modelação tais como o desempenho dos modelos e quais as variáveis a ser utilizadas.

Foi verificado que as variáveis "Considera que tem ansiedade?" e "Já foi diagnosticado profissionalmente como ansioso?" são as que mais influenciam o modelo, verificando-se uma diminuição dos valores de F1. Retirando a primeira variável agora mencionada e mantendo as restantes, os modelos baixaram em média cinco pontos percentuais. Retirando as duas variáveis, os modelos mostraram baixar em média dez pontos percentuais. Notam-se estas variáveis como as mais importantes para o modelo, sendo, portanto, estas consideradas para o modelo. Todas as outras demonstram também relevância, apesar de a diferença da sua utilização não se mostrar tão significante como estas duas variáveis agora analisadas.

As variáveis relativas às atividades que os inquiridos praticam enquanto ouvem música "Ouve música enquanto faz outra atividade?" e "Caso sim, qual?" não demonstraram um impacto positivo no modelo. Como estas não foram também consideradas muito relevantes no estudo em resultado da análise exploratória, não serão então considerados para a aprendizagem dos modelos.

Apesar de a diferença entre os modelos não ser considerável, pode-se verificar no teste realizado que os modelos de regressão lineares, à exceção do Naive Bayes, apresentaram resultados inferiores em comparação com modelos de regressão não linear, com exceção do modelo kNN que apresentou resultados inferiores de entre o grupo. Os modelos representativos do terceiro grupo apresentam também resultados inferiores aos modelos não lineares, tendo estes resultados parecidos aos lineares. À exceção dos modelos kNN e AdaBoost que obtiveram os resultados de F1 inferiores, respetivamente, a diferença entre os resultados dos modelos que obtiveram valores de acerto de previsão superiores e inferiores não chega a oito pontos percentuais, como se pode verificar na Tabela 4.1.

Tabela 4.1: Avaliação dos resultados da variável alvo

Algoritmo	**CA**	**Precision**	**Recall**	**F1**
Naive Bayes	0.673	0.671	0.673	0.672
SVM	0.642	0.645	0.642	0.643
Gradient Boosting	0.630	0.620	0.630	0.623
SGD	0.624	0.615	0.624	0.618
Random Forest	0.624	0.610	0.624	0.613
ANN	0.615	0.608	0.615	0.611
Logistic Regression	0.612	0.600	0.612	0.60.
AdaBoost	0.576	0.577	0.576	0.576
Decision Tree	0.576	0.567	0.576	0.570
kNN	0.558	0.533	0.558	0.540

Os resultados dos modelos utilizando como variável alvo "Níveis de ansiedade" demonstram-se muito inferiores em comparação com a utilização da variável alvo definida "Tem ansiedade", reduzindo as percentagens de acerto das previsões para metade, tendo em conta os valores de F1. Analisando a matriz de confusão para cada algoritmo, verificou-se que, no estudo feito, as percentagens de acerto do modelo nos níveis de ansiedade superiores, média e severa ansiedade, rondam os 25% e os 30%, à exceção do Naive Bayes (que acertou em 51.2% dos casos de severa ansiedade) e do Decision Tree, kNN e SVM (que apenas acertaram em média 7% dos casos de severa ansiedade). Tendo em contas estas resultados, a utilização da variável "Níveis de ansiedade" como variável alvo foi excluída deste projeto sendo, apesar de tudo, uma variável interessante a utilizar em estudos futuros.

4.3. Avaliação do modelo

Na avaliação dos resultados dos algoritmos pode-se verificar que os valores de F1 variam entre os 0.54 e os 0.672 correspondendo ao kNN e Naive Bayes. Este último, em conjunto com o SVM, são os que mostram serem mais eficazes na previsão dos resultados de uma forma geral. Todos os outros algoritmos apresentam valores de F1 entre os 0.603 e os 0.623, à exceção dos algoritmos AdaBoost, Decision Tree e kNN que apresentaram os valores de F1 de 0.576, 0.570 e 0.540, respetivamente. Estes resultados podem ser verificados na Tabela 4.1.

Foi também verificado o valor de F1 para os resultados de avaliação do caso "Sim" da variável alvo, onde os resultados, como era expectável, baixaram, tendo em conta o tamanho da amostra e a percentagem de inquiridos que foram considerados que têm ansiedade ser mais baixa em comparação com os que não têm ansiedade.

Em média, o valor de F1 diminui em 20% na totalidade para todos os algoritmos, estando então estes compreendidos agora entre os valores 0.32 e 0.56, sendo que os algoritmos que os representam se mantêm iguais. Apenas os dois primeiros algoritmos apresentam agora resultados acima dos 50%, o que não indica resultados positivos aos algoritmos para prever inquiridos que têm ansiedade.

Através das matrizes de confusão, pode-se verificar todos os valores VP, VN, FP e FN para cada algoritmo, tendo sido considerados os quatros que apresentaram resultados superiores de F1 sendo estes, por ordem decrescente, o Naive Bayes, SVM, Gradient Boosting e SGD. Através da matriz de confusão do Naive Bayes, pode-se verificar que a percentagem de VN é bastante superior à percentagem de VP sendo estes 74.6% e 55.2%, respetivamente. Já os FN apresentam também uma maior percentagem em comparação com os FP sendo estes 44.8% e 25.4%, respetivamente. Na Figura 4.1 pode observar-se, como exemplo, a matriz de confusão do Naive Bayes.

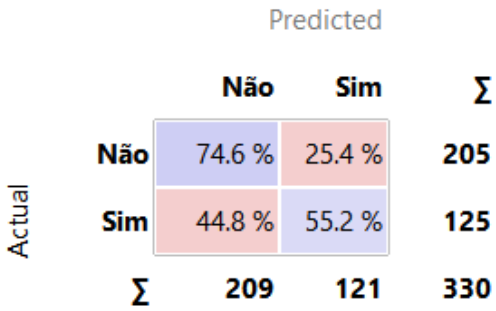


Figura 4.1: Matriz de confusão do algoritmo Naive Bayes

Apesar de o SVM apresentar resultados de F1 inferiores ao Naive Bayes, a percentagem de VP apenas difere por menos de um ponto percentual, sendo este também considerado como interessante na medida de descoberta de ansiedade. Já nos VN apresenta uma diminuição de perto de 5% em relação ao Naive Bayes. Esta diferença fez-se notar no resultado de FP, aumentando a mesma percentagem.

O Gradient Boosting, assim como o SGD, apesar de apresentarem resultados parecidos com o Naive Bayes comparando o F1, este não se notou tão importante utilizar em comparação com o SVM. Apesar deste obter resultados superiores de TN, a percentagem de TP é bastante inferior, na ordem dos 12%. Agravando a esta condicionante, reparou-se que no estudo em questão, a percentagem de FP aumentou também na ordem dos 12%, errando de uma forma mais frequente quando diz que o inquirido não tem ansiedade, quando este tem, algo que se considera grave consoante os objetivos. Desta forma, estes dois algoritmos não serão considerados no presente estudo.

Por último, foi verificado os resultados dos algoritmos SVM e Naive Bayes, agora utilizando os dados de teste. Os valores de F1, em ambos os casos, mantiveram-se parecidos com os valores de F1 dos dados de aprendizagem, aumentando 0.012 e 0.005, respetivamente. Foi também utilizada uma matriz de confusão de forma a verificar os resultados dos dados de teste e pode-se verificar que, neste estudo, as percentagens de FN e VP mantiveram-se iguais em ambos os algoritmos, apresentando apenas diferenças nos VN e FP, sendo que o Naive Bayes, como seria expectável devido aos resultados de F1, obteve maiores percentagens de VN em comparação com o SVM.

4.4. Teste das hipóteses de estudo

Tendo em conta as perspetivas iniciais do trabalho, foram desenvolvidas hipóteses de investigação, tendo em conta os objetivos iniciais. Após o trabalho realizado, verificou-se que algumas destas hipóteses foram confirmadas e outras não.

Hipótese 1: A música tem um impacto positivo na redução da ansiedade.

Ao longo desta dissertação pode-se verificar a importância da música na vida das pessoas e como esta tem um impacto positivo no seu estado emocional. Com base na revisão de literatura feita nesta dissertação pode-se verificar que os efeitos sonoros e da música nas emoções são práticas utilizadas no tratamento da ansiedade e redução dos seus sintomas.

Tendo em conta o questionário desenvolvido e a análise dos seus resultados pode-se verificar que 95% dos inquiridos ouvem música enquanto fazem qualquer outra atividade, notando-se a importância da música na vida quotidiana nas pessoas. Cerca de 80% dos inquiridos sente que ouvir música diminui o seu estado de ansiedade, onde dois terços utiliza a música como tentativa de diminuição do seu estado de ansiedade. Pode-se então confirmar que a música tem um impacto positivo na redução dos sintomas de ansiedade.

Hipótese 2: A aplicação de algoritmos e ML revelam-se importantes na identificação de episódios da ansiedade.

A utilização de algoritmos e técnicas de ML tem-se revelado uma prática cada vez mais recorrente na ajuda à dentificação de diversos problemas a nível da saúde, de modo a ajudar no processo de tomada de decisão dos médicos e especialistas, apresentando resultados consoante o objetivo dos modelos. Tendo em conta os resultados obtidos na revisão de literatura, pode-se verificar que existem diversas formas de aplicar o ML na deteção de episódios de ansiedade, utilizando diversos modelos médicos tais como o ECG, EEG, foto pletismografia, entre outros, sendo o ML aplicado a dados recolhidos de sintomas fisiológicos e não emocionais.

Por outro lado, com a aplicação de algoritmos e ML na deteção de episódios de ansiedade com base no questionário desenvolvido, é possível também verificar a importância da utilização de algoritmos e ML na deteção de episódios de ansiedade, aplicando-se este a dados recolhidos de sintomas emocionais, apesar de o modelo desenvolvido apresentar fracos resultados (na ordem dos 70%) devido à falta de dados.

Hipótese 3: A caracterização sociodemográfica auxilia na identificação de padrões de ansiedade.

A caracterização sociodemográfica da população revelou diferenças no que toca à presença de episódios de ansiedade nos inquiridos, consoante ao seu género e idade. Apesar de ser possível verificar grupos diferentes consoante estas características sociodemográficas, não foi possível (possivelmente devido à falta de dados) identificar padrões consoante a atividade atual do inquirido, assim como o seu nível de escolaridade e área de residência. Apesar de não ter sido confirmado neste estudo, é de notar que estas três características poderão vir a ser reveladas importantes na identificação de padrões de ansiedade em estudos futuros, onde seja possível obter mais dados.

Hipótese 4: É possível descobrir padrões de ansiedade com base no estilo de música que se ouve em conjunto com os restantes dados do questionário.

Para além do estudo da importância da música na vida dos inquiridos, foi também verificado quais os tipos de música que as pessoas ouviam no seu estado emocional normal e quando se sentiam ansiosas. Revelou-se notório o número de inquiridos que ouve música quando se sentem ansiosos. Verificou-se, por exemplo, que existem estilos de música que agrupam inquiridos que apresentam menores níveis de ansiedade, tais como “Hip Pop”, “Funk” e “Electronic”. Foram também encontrados estilos de música que os inquiridos recorrem mais quando se sentem ansiosos e, principalmente quais as diferenças entre os estilos que ouvem no estado não ansioso em comparação com quando se sentem ansiosos. É, portanto, de notar que, devido ao tamanho da população de dados, todos estes grupos contêm pequenos aglomerados de dados, não tendo um grande grau de certeza da forma de cada grupo.

Capítulo 5 - Conclusões e recomendações

5.1. Principais conclusões

A ansiedade é um problema que tem cada vez mais sido alvo de estudo e preocupação, sendo esta ainda muito desprezada da atenção das pessoas. A utilização de métodos de ML tem-se revelado uma ferramenta muito utilizada na descoberta e previsão de cenários da vida real, podendo esta ajudar a prevenir muitas situações. A música tem uma grande influência na vida das pessoas, podendo esta transmitir emoções mais positivas ou negativas, dependente do estado emocional atual da pessoa, dos seus gostos musicais e da reação de cada pessoa a cada música. A combinação destas três áreas de estudo demonstrou ter um grande potencial no que toca à descoberta de sintomas e padrões de ansiedade, assim como a descoberta desta em inquiridos não diagnosticados como ansiosos, utilizando a música como ferramenta auxiliar no seu tratamento e descoberta de padrões.

Foi então desenvolvido um modelo de previsão de ansiedade nas pessoas, com base num questionário. Este continha perguntas relativas à ansiedade generalizada, em relação à música e ao impacto desta na vida dos inquiridos. O objetivo deste modelo é prever se o inquirido tem ou não ansiedade tendo em conta as respostas de cada individuo ao questionário. Foi também feito um estudo sobre a importância da música na vida das pessoas e uma análise de possíveis padrões, apesar desta ter sido comprometido pela falta de dados.

Apesar dos níveis de certeza dos resultados do modelo utilizado serem reduzidos devido à escassez de dados, foi possível prever, com um grau de certeza de cerca de 67%, a presença de ansiedade nos inquiridos, respondendo “Sim” ou “Não” à variável alvo “Tem ansiedade”. Verificou-se os níveis de eficácia de cada modelo e constatou-se que estes, possivelmente devido à fraca qualidade dos dados e dimensão da amostra, demonstraram resultados semelhantes. O Naive Bayes foi considerado, de entre todos, o algoritmo a utilizar no modelo, uma vez que este apresentou os resultados mais positivos na previsão da presença de ansiedade, consoante os dados presentes do estudo e toda a sua escolha e transformação.

Contudo, todo o processo de preparação dos dados dos questionários, de execução dos algoritmos e de avaliação dos resultados mostra que é possível utilizar ML sobre dados de ansiedade e música para ajudar a detetar e evitar episódios de ansiedade, o que responde à questão de investigação colocada na seção 1.4. Estes resultados preliminares poderão ser melhorados futuramente com a aplicação do questionário a mais respondentes para obtenção de mais respostas.

A análise feita e tentativa de descoberta de padrões também foi realizada, mas com um nível de certeza bastante reduzido, uma vez que a quantidade de dados não permite dividir a população de dados em grandes grupos com padrões semelhantes. Foi verificada a presença de ansiedade nos inquiridos e qual a importância da música na tentativa de diminuição dos sintomas dos episódios de ansiedade. Conclui-se que dois terços dos inquiridos deste estudo sentem que a música tem um efeito positivo na diminuição dos sentimentos de ansiedade. Nota-se também que os inquiridos por vezes ouvem outro estilo de música quando se sentem ansiosos, em comparação com o seu estado não ansioso.

5.2. Contribuições a nível empresarial e académico

Pretende-se que esta análise venha a ser usada pela empresa genesis.studio, adaptando futuramente o modelo de previsão a um sistema de recomendações e fazer novas investigações sobre o estudo. Os dados poderão ser utilizados quer para a descoberta de sintomas de ansiedade, quer para a utilização dos dados adquiridos para o impacto da música nas emoções das pessoas. O modelo poderá também ser utilizado como ferramenta auxiliar aos médicos e psicólogos na descoberta da presença de sintomas de ansiedade. Este estudo poderá também ser utilizado a nível empresarial, nomeadamente no que toca à descoberta de episódios de ansiedade na vida profissional de cada colaborador de uma certa empresa, de modo a ajudar os indivíduos cujo modelo previu existência de ansiedade, melhorando o seu estado de espírito e o seu rendimento no trabalho.

A aplicação do questionário do presente estudo à população, fazendo perguntas relacionadas com música, ansiedade e a combinação destas duas, demonstrou-se relevante no sentido de perceber a importância da ansiedade no mundo atual e como o ML pode impactar de forma positiva a vida das pessoas. Neste estudo demonstra-se a relevância da utilização de práticas de ML no decorrer normal da vida das pessoas e como este pode ter um impacto positivo na saúde mental destas, notando-se a necessidade de desenvolvimento de mais estudos destas práticas. É também importante notar que a música pode ser um motivo de estudo na descoberta de episódios de ansiedade. Este estudo pode então contribuir para a partilha de informação e conhecimento para a comunidade científica sobre a importância da deteção e prevenção de episódios de ansiedade com base em aplicações de ML, tendo como ferramenta auxiliar a música.

5.3. Limitações do trabalho realizado

A grande limitação encontrada neste estudo foi a falta de dados, limitando o tamanho da amostra e a eficácia dos algoritmos condicionando a qualidade do modelo. Não foi possível descobrir muitos padrões nos dados, dificultando a descoberta e definição de grupos. Mesmo tendo sido feita alguma análise com base nos padrões descobertos, esta metodologia falha devido a esta escassez de dados, tendo sido criados grupos com apenas algumas dezenas de inquiridos (o que, considerando o tamanho da amostra, já equivale a uma boa percentagem dos dados).

Estando o presente estudo limitado a uma análise de resultados de aplicação de algoritmos de ML e a análise exploratória dos dados, a análise de resultados das respostas ao questionário poderá ser alvo de uma análise estatística, nomeadamente utilizando uma ferramenta como o SPSS (Statistical Package for the Social Science).

5.4. Propostas de investigação futura

Para futuros trabalhos apresentam-se as seguintes propostas, que poderiam ser realizadas a curto e médio prazo, e a longo prazo.

A curto e médio prazos, a proposta seria por exemplo a execução da análise estatística acima mencionada e relançar o questionário para ter mais resultados. Uma vez que as limitações encontradas no trabalho desenvolvido foram a falta de uma análise estatística dos dados e a escassez destes, propõe-se o relançamento do mesmo questionário, desta vez tentando recolher uma maior quantidade de respostas e uma maior diversificação de inquiridos, principalmente quanto à idade e género, não havendo uma grande diferenciação entre estes grupos. Pretende-se, com a utilização do mesmo estudo com uma maior quantidade de dados, que o modelo melhore os seus resultados na tentativa de descoberta da presença de ansiedade, assim como uma possível divisão em diferentes níveis da variável alvo e descoberta de novos padrões.

A longo prazo a proposta passa pelo desenvolvimento de um sistema de recomendação de músicas em função da aprendizagem dos sintomas de ansiedade. Este sistema necessita da colaboração dos utilizadores por via de uma aplicação móvel onde estes iriam, de forma recorrente, transmitir o seu estado de espírito. Consoante o resultado do modelo face às respostas do utilizador e, tendo em conta os seus estilos de música preferidos, a aplicação recomendaria uma música para este ouvir, com o objetivo de tentar melhorar o seu estado emocional atual. Desta maneira, o modelo necessitaria do feedback dos utilizadores face a cada música e perceber o impacto de cada música recomendada. A recolha de dados sobre a localização e atividade atual do utilizador seria também interessante neste sistema, uma vez que os locais e atividades influenciam o estado emocional do ser humano.

Referências bibliográficas

Abriel, S. M. B. A. S. (2010). *Perturbação Obsessivo-Compulsiva: Conceitos, Limites E Controvérsias*.

Angelis, K. De, Do, M., Brasileiro, S., Irigoyen, M. C., Humano, M., São, U., Tadeu, J., Paulo, S., Hipertensão, U. De, Coração, I., Clínicas, H., Medicina, F. De, De, U., Nefrologia, D. De, & São, F. De. (2004). Sistema nervoso autônomo e doença cardiovascular. *Revista Da Sociedade de Cardiologia Do Rio Grande Do Sul*, *13*(03), 1–7.

Askland, K., Garnaat, S., Sibrava, N., Boisseau, C., Strong, D., Mancebo, M., Greenberg, B., Rasmussen, S., & Eisen, J. (2015). Prediction of remission in obsessive compulsive disorder using a novel machine learning strategy. Int. J. Methods in Psychiatr Res. 24(2): 156-169. doi:10.1002/mrp.1463

Beck, J. G., Coffey, S. F., Foy, D. W., Keane, T. M., & Blanchard, E. B. (2009). Group Cognitive Behavior Therapy for Chronic Posttraumatic Stress Disorder: An Initial Randomized Pilot Study. *Behavior Therapy*, *40*(1), 82–92. https://doi.org/10.1016/j.beth.2008.01.003

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239. https://doi.org/10.1093/aje/kwz189

Brugués, Ariadna O (2011). Medical Problems of Performing Artists, Volume 26, Number 3, 1 September 2011, pp. 164-171(8). https://doi.org/10.21091/mppa.2011.3026

Chiu, M. C., & Ko, L. W. (2017). Develop a personalized intelligent music selection system based on heart rate variability and machine learning. *Multimedia Tools and Applications*, *76*(14), 15607–15639. https://doi.org/10.1007/s11042-016-3860-x

Delmonte, M.M. (1985). Meditation and anxiety reduction: A literature review, Clinical Psychology Review, 91-102. https://doi.org/10.1016/0272-7358(85)90016-9.

Demetriou, E. A., Park, S. H., Ho, N., Pepper, K. L., Song, Y. J. C., Naismith, S. L., Thomas, E. E., Hickie, I. B., & Guastella, A. J. (2020). Machine Learning for Differential Diagnosis Between Clinical Conditions With Social Difficulty: Autism Spectrum Disorder, Early Psychosis, and Social Anxiety Disorder. *Frontiers in Psychiatry*, *11*(June), 1–16. https://doi.org/10.3389/fpsyt.2020.00545

Englander, M. (2016). The phenomenological method in qualitative psychology and psychiatry. *International Journal of Qualitative Studies on Health and Well-Being*, *11*. https://doi.org/10.3402/qhw.v11.30682

Garrido, S., & Schubert, E. (2011). Negative Emotion in Music: What is the Attraction? A Qualitative Study. *Empirical Musicology Review*, *6*(4), 214–230. https://doi.org/10.18061/1811/52950

Gask, L., Kendrick, T., Peveler, R., & Chew-Graham, C. (Eds.). (2018). Primary Care Mental Health (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/9781911623038

George, A. S., Kurup, A. V., Balachandran, P., Nair, M., Gopinath, S., Kumar, A., & Parasuram, H. (2021). Predicting Autonomic Dysfunction in Anxiety Disorder from ECG and Respiratory Signals Using Machine Learning Models. *International Journal of Online and Biomedical Engineering*, *17*(7), 143–155. https://doi.org/10.3991/ijoe.v17i07.22581

Gore, E., & Rathi, S. (2019). Surveying Machine Learning Algorithms on Eeg Signals Data for Mental Health Assessment. *2019 IEEE Pune Section International Conference, PuneCon 2019*, 1–6. https://doi.org/10.1109/PuneCon46936.2019.9105749

Hahn, T., Kircher, T., Straube, B., Wittchen, H. U., Konrad, C., Ströhle, A., Wittmann, A., Pfleiderer, B., Reif, A., Arolt, V., & Lueken, U. (2015). Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry*, *72*(1), 68–74. https://doi.org/10.1001/jamapsychiatry.2014.1741

Hilbert, K., Lueken, U., Muehlhan, M., & Beesdo-Baum, K. (2017). Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. Brain and behavior, 7(3), e00633. https://doi.org/10.1002/brb3.633

Hirsch, T., Merced, K., Narayanan, S., Imel, Z. E., & Atkins, D. C. (2017). Designing contestability: Interaction design, machine learning, and mental health. *DIS 2017 - Proceedings of the 2017 ACM Conference on Designing Interactive Systems*, 95–99. https://doi.org/10.1145/3064663.3064703

Ihmig, F. R., Antonio Gogeascoechea, H., Neurohr-Parakenings, F., Schäfer, S. K., Lass-Hennemann, J., & Michael, T. (2020). On-line anxiety level detection from biosignals: Machine learning based on a randomized controlled trial with spider-fearful individuals. *PLoS ONE*, *15*(6), 1–20. https://doi.org/10.1371/journal.pone.0231517

Jambukia, S. H., Dabhi, V. K., & Prajapati, H. B. (2015). Classification of ECG signals using machine learning techniques: A survey. *Conference Proceeding - 2015 International Conference on Advances in Computer Engineering and Applications, ICACEA 2015*, 714–721. https://doi.org/10.1109/ICACEA.2015.7164783

Junior, D. B. S. (2018). Estudo Comparativo Entre Dois Modelos De Improvisação Rítmica Musical Em Musicoterapia E Ansiedade. http://repositorio.bc.ufg.br/tede/handle/tede/9685

Kalin, N. H. (2020). The critical relationship between anxiety and depression. *American Journal of Psychiatry*, *177*(5), 365–367. https://doi.org/10.1176/appi.ajp.2020.20030305

Kanas, N., & Manzey, D. (2004). Chapter 3. Human Performance. In *Space psychology and psychiatry* (Vol. 53).

Kingsbury, S. J. (1987). Cognitive Differences Between Clinical Psychologists and Psychiatrists. *American Psychologist*, *42*(2), 152–156. https://doi.org/10.1037/0003-066X.42.2.152

Kwon, S. M., Evans, L., & Oei, T. P. S. (1990). Factor structure of the mobility inventory for agoraphobia: A validational study with australian samples of agoraphobic patients. *Journal of Psychopathology and Behavioral Assessment*, *12*(4), 365–374. https://doi.org/10.1007/BF00965990

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., Michelgård, Å., Appel, L., Fredrikson, M., & Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior*, *32*(4), 195–214. https://doi.org/10.1007/s10919-008-0055-9

Leckman, J. F., Grice, D. E., Boardman, J., Zhang, H., Vitale, A., Bondi, C., Alsobrook, J., Peterson, B. S., Cohen, D. J., Rasmussen, S. A., Goodman, W. K., McDougle, C. J., & Pauls, D. L. (1997). Symptoms of obsessive-compulsive disorder. *American Journal of Psychiatry*, *154*(7), 911–917. https://doi.org/10.1176/ajp.154.7.911

Manocha, A., Kumar, G., Bhatia, M., & Sharma, A. (2021). IoT-inspired machine learning-assisted sedentary behavior analysis in smart healthcare industry. *Journal of Ambient Intelligence and Humanized Computing*, *0123456789*. https://doi.org/10.1007/s12652-021-03371-x

McGeoch, P. D., & Rouw, R. (2020). How everyday sounds can trigger strong emotions: ASMR, misophonia and the feeling of wellbeing. *BioEssays*, *42*(12), 1–10. https://doi.org/10.1002/bies.202000099

McGinnis, E. W., Anderau, S. P., Hruschak, J., Gurchiek, R. D., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., Muzik, M., & McGinnis, R. S. (2019). Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood. *IEEE Journal of Biomedical and Health Informatics*, *23*(6), 2294–2301. https://doi.org/10.1109/JBHI.2019.2913590

McGinnis, R. S., McGinnis, E. W., Hruschak, J., Lopez-Duran, N. L., Fitzgerald, K., Rosenblum, K. L., & Muzik, M. (2018). Wearable sensors and machine learning diagnose anxiety and depression in young children. *2018 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2018*, *2018-Janua*(March), 410–413. https://doi.org/10.1109/BHI.2018.8333455

McHugh, S. B., Deacon, R. M. J., Rawlins, J. N. P., & Bannerman, D. M. (2004). Amygdala and Ventral Hippocampus Contribute Differentially to Mechanisms of Fear and Anxiety. *Behavioral Neuroscience*, *118*(1), 63–78. https://doi.org/10.1037/0735-7044.118.1.63

Miranda, D., Calderón, M., & Favela, J. (2014). Anxiety detection using wearable monitoring. *ACM International Conference Proceeding Series*, *03-05-Nove*, 34–41. https://doi.org/10.1145/2676690.2676694

Moro, S. M. C. (2015). Feature Selection Strategies for Improving Data-Driven Decision Support in Bank Telemarketing. *PQDT - Global*, 114.

https://login.pallas2.tcl.sc.edu/login?url=https://search.proquest.com/docview/2033382463?accountid=13965%0Ahttp://resolver.ebscohost.com/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rfr_id=info:sid/ProQuest+Dissertations+%26+Theses+Global&rft_v

Molnar, C. (2020). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable.

N. Al-Wesabi, F., Alsolai, H., Mustafa Hilal, A., Ahmed Hamza, M., Al Duhayyim, M., & Negm, N. (2022). Machine Learning Based Depression, Anxiety, and Stress Predictive Model During COVID-19 Crisis. *Computers, Materials & Continua*, *70*(3), 5803–5820. https://doi.org/10.32604/cmc.2022.021195

Na, K. S., Cho, S. E., & Cho, S. J. (2021). Machine learning-based discrimination of panic disorder from other anxiety disorders. *Journal of Affective Disorders*, *278*(February 2020), 1–4. https://doi.org/10.1016/j.jad.2020.09.027

El Naqa, I., Murphy, M.J. (2015). What Is Machine Learning?. In: El Naqa, I., Li, R., Murphy, M. (eds) Machine Learning in Radiation Oncology. Springer, Cham. https://doi.org/10.1007/978-3-319-18305-3_1

Nardelli, M., Valenza, G., Greco, A., Lanata, A., & Scilingo, E. P. (2015). Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, *6*(4), 385–394. https://doi.org/10.1109/TAFFC.2015.2432810

Perpetuini, D., Chiarelli, A. M., Cardone, D., Filippini, C., Rinella, S., Massimino, S., Bianco, F., Bucciarelli, V., Vinciguerra, V., Fallica, P., Perciavalle, V., Gallina, S., Conoci, S., & Merla, A. (2021). Prediction of state anxiety by machine learning applied to photoplethysmography data. *PeerJ*, *9*, 1–20. https://doi.org/10.7717/peerj.10448

Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S., & Wirth, R. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, 76.

Pintelas, E. G., Kotsilieris, T., Livieris, I. E., & Pintelas, P. (2018). A review of machine learning prediction methods for anxiety disorders. *ACM International Conference Proceeding Series*, 8–15. https://doi.org/10.1145/3218585.3218587

Quasim, M. T., Alkhammash, E. H., Khan, M. A., & Hadjouni, M. (2021). Emotion-based music recommendation and classification using machine learning with IoT Framework. *Soft Computing*, *25*(18), 12249–12260. https://doi.org/10.1007/s00500-021-05898-9

Rahman, J. S., Gedeon, T., Caldwell, S., Jones, R., & Jin, Z. (2021). Towards Effective Music Therapy for Mental Health Care Using Machine Learning Tools: Human Affective Reasoning and Music Genres. *Journal of Artificial Intelligence and Soft Computing Research*, *11*(1), 5–20. https://doi.org/10.2478/jaiscr-2021-0001

Rheault, L. (2016). *Expressions of Anxiety in Political Texts*. 92–101. https://doi.org/10.18653/v1/w16-5612

Richter, T., Fishbain, B., Markus, A., Richter-Levin, G., & Okon-Singer, H. (2020). Using machine learning-based analysis for behavioral differentiation between anxiety and depression. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-72289-9

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178. https://doi.org/10.1037/h0077714

Saeb, S., Lattie, E. G., Kording, K. P., & Mohr, D. C. (2017). Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR MHealth and UHealth*, *5*(8). https://doi.org/10.2196/mhealth.7297

Schäfer, F., Zeiselmair, C., Becker, J., & Otten, H. (2018). Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes. *2018 IEEE International Conference on Technology Management, Operations and Decisions, ICTMOD 2018*, 190–195. https://doi.org/10.1109/ITMC.2018.8691266

Sharma, A., & Verbeke, W. J. M. I. (2021). Understanding importance of clinical biomarkers for diagnosis of anxiety disorders using machine learning models. *PLoS ONE*, *16*(5 May 2021), 1–16. https://doi.org/10.1371/journal.pone.0251365

Shih, Yi-Nuo, Huang, Rong-Hwa, and Chiang, Hsin-Yu. 'Background Music: Effects on Attention Performance'. 1 Jan. 2012 : 573 – 578. https://doi.org/10.3233/WOR-2012-1410

Shikha, S., Agrawal, M., Anwar, M. A., & Sethia, D. (2021). Stacked Sparse Autoencoder and Machine Learning Based Anxiety Classification Using EEG Signals. *The First International Conference on AI-ML-Systems*, 1–7. https://doi.org/10.1145/3486001.3486227

Shri, R. (2010). Anxiety : causes and management symptoms of anxiety disorders. *The Journal of Behavioral Science*, *5*(1), 100–118.

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, 1310–1315.

Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral Modeling for Mental Health using Machine Learning Algorithms. *Journal of Medical Systems*, *42*(5). https://doi.org/10.1007/s10916-018-0934-5

Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. *The Lancet*, *371*(9618), 1115–1125. https://doi.org/10.1016/S0140-6736(08)60488-2

Taruffi, L., Pehrs, C., Skouras, S., & Koelsch, S. (2017). Effects of Sad and Happy Music on Mind-Wandering and the Default Mode Network. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-14849-0

Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine Learning in Mental Health: A systematic review of the HCI literature to support the development of effective and implementable ML Systems. In *ACM Transactions on Computer-Human Interaction*

(Vol. 27, Issue 5). Association for Computing Machinery. https://doi.org/10.1145/3398069

Thompson, M, & Biddle, M. (2013). Sound, Music, Affect: Theorizing Sonic Experience. (pp. 18-21). Bloomsbury Publishing USA.

Tutul, A. A., Nirjhar, E. H., & Chaspari, T. (2021). Investigating Trust in Human-Machine Learning Collaboration: A Pilot Study on Estimating Public Anxiety from Speech. *Proceedings of the 2021 International Conference on Multimodal Interaction*, 288–296. https://doi.org/10.1145/3462244.3479926

Udovičić, G., Derek, J., Russo, M., & Sikora, M. (2017). Wearable Emotion Recognition system based on GSR and PPG signals. *MMHealth 2017 - Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care, Co-Located with MM 2017*, 53–59. https://doi.org/10.1145/3132635.3132641

Vuoskoski, J., Thompson W., McIlwain D., Eerola T. (2012); Who Enjoys Listening to Sad Music and Why?. 29 (3): 311–317. doi: https://doi.org/10.1525/mp.2012.29.3.311

Wang, X. W., Nie, D., & Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. In *Neurocomputing* (Vol. 129, pp. 94–106). https://doi.org/10.1016/j.neucom.2013.06.046

Wittchen, H. (2001). *Wittchen.Hoyer.2001.Prev.Epi.Risk*. *62*(suppl 11), 15–19.

Wshah, S., Skalka, C., & Price, M. (2019). Predicting posttraumatic stress disorder risk: A machine learning approach. *JMIR Mental Health*, *6*(7). https://doi.org/10.2196/13946

Yao, X., Liu, Y. (2014). Machine Learning. In: Burke, E., Kendall, G. (eds) Search Methodologies. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-6940-7_17

Yehuda, N. (2011). Music and Stress. *Journal of Adult Development*, *18*(2), 85–94. https://doi.org/10.1007/s10804-010-9117-4

Zanero, S., & Serazzi, G. (2008). Unsupervised learning algorithms for intrusion detection. *NOMS 2008 - IEEE/IFIP Network Operations and Management Symposium: Pervasive Management for Ubiquitous Networks and Services*, 1043–1048. https://doi.org/10.1109/NOMS.2008.4575276

Zhang, X. Da. (2020). A matrix algebra approach to artificial intelligence. In *A Matrix Algebra Approach to Artificial Intelligence*. https://doi.org/10.1007/978-981-15-2770-8

Zhao, Z., & Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. *ACM International Conference Proceeding Series*, *227*, 1151–1157. https://doi.org/10.1145/1273496.1273641

Zhou, P., Lin, D., He, W., Li, G., & Shang, K. (2009). Influence of musicotherapy on mental status and cognitional function of patient with depression disease. *Proceedings of the 2009 2nd International Conference on Biomedical Engineering and Informatics, BMEI 2009*, 17–20. https://doi.org/10.1109/BMEI.2009.5303000

Zoefel, B., & Heil, P. (2013). Detection of near-threshold sounds is independent of eeg phase in common frequency bands. *Frontiers in Psychology*, *4*(MAY), 1–17. https://doi.org/10.3389/fpsyg.2013.00262

A. Apêndices

Apêndice A

O ML recorre à utilização de vários algoritmos e modelos, gerados pelo ser humano, de forma a facilitar a vida das pessoas aquando da tomada de decisões. Segundo Bi et al. (2019), existem vários algoritmos comummente utilizados nas práticas de ML:

1) Decision Tree – Divide a população em dois ou mais grupos homogéneos, onde é usado um modelo preditivo de forma que se consiga, a partir de cada observação sobre dada característica (ramo), chegar a uma conclusão relativamente à variável alvo (folha), como pode ser observado na Figura A.1.

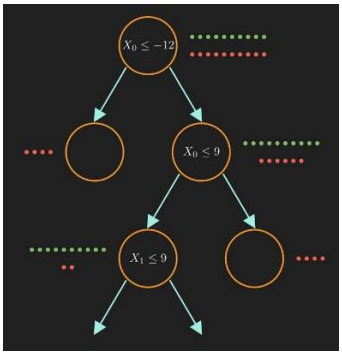


Figura A.1: Exemplo de uma Decision Tree (Bi et al., 2019) [6]

2) Artificial Neural Networks – Utiliza uma população de neurónios ligados entre si através de vias de sinalização. Cada neurónio artificial recebe um sinal, processa a informação e transmite aos neurónios que estão ligados a si, através de um canal de entrada e saída, sendo estes independentes e dependentes, respetivamente, como pode ser observado na Figura A.2.

[6] Imagem retirada do site: https://pythonkai.org/2021/12/20/machine-learning-for-beginners-project-4-decision-tree-classifier/

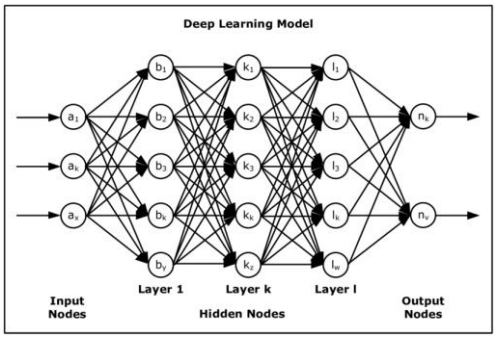


Figura A.2: Exemplo de uma Artificial Neural Networks (Bi et al., 2019)[7]

3) Support Vector Machines (SVM) – Método de classificação supervisionado em que cada dado é identificado como um ponto no espaço com n dimensões, sendo "n" o número de características que cada dado tem. Este algoritmo constrói um modelo que atribui novos dados a uma característica com base em duas possíveis classes de entrada, como pode ser observado na Figura A.3.

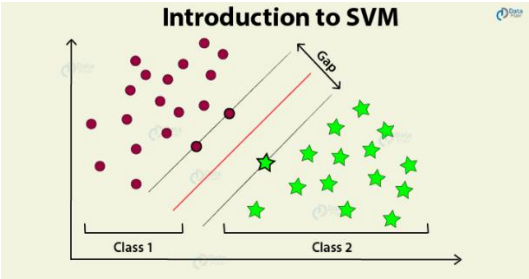


Figura A.3: Exemplo de um Support Vector Machines (Bi et al., 2019)[8]

4) Logistic Regression – Modelo de classificação utilizado para estimar valores discretos binários. Será medida a relação entre variáveis categóricas dependentes com uma ou mais variáveis independentes, como pode ser observado na Figura A.4.

[7] Imagem retirada do site: https://www.researchgate.net/figure/Artificial-Neural-Network-Deep-Learning-model_fig8_313408173
[8] Imagem retirada do site: https://ichi.pro/pt/introducao-ao-support-vector-machine-svm-225431442261105

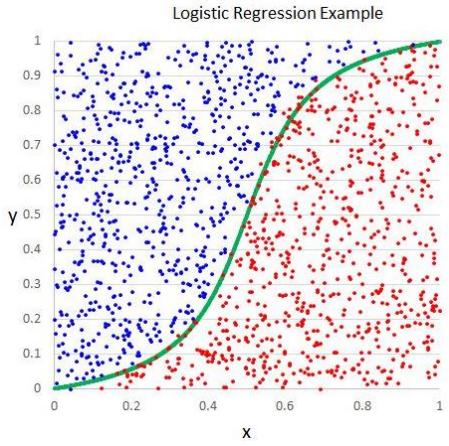


Figura A.4: Exemplo de uma Logistic Regression (Bi et al., 2019) [9]

5) Naive Bayes Algorithms – Utiliza uma técnica de classificação onde assume a independência entre as características dos dados. Esta técnica é bastante utilizada em modelos onde existe independência entre as características de cada dado, não havendo perda de informação entre a relação dessas mesmas características. Pode ser utilizada a fórmula matemática como forma de exemplo da sua independência: P(A|B) = [P(B|A) * P(A) / P(B)], como pode ser observado na Figura A.5.

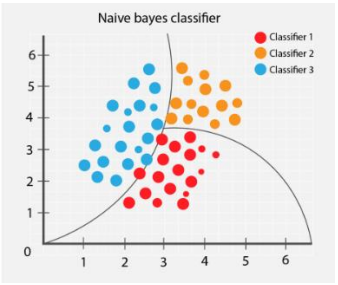


Figura A.5: Exemplo de um Naive Bayes Algorithm (Bi et al., 2019) [10]

6) K-nearest neighbors (kNN) – Utiliza um método de classificação não paramétrico, ou seja, não está direcionado a utilizar o valor da característica do dado em questão, mas sim em avaliar e fazer a média da característica dos dados vizinhos ao dado em questão. As funções que medem a distância entre os dados são normalmente Euclidianas, de Manhattan ou Minkowski para variáveis numéricas, ou então Hamming para variáveis categóricas, como pode ser observado na Figura A.6.

[9] Imagem retirada do site: https://helloacm.com/a-short-introduction-logistic-regression-algorithm/
[10] Imagem retirada do site: https://ichi.pro/pt/algoritmo-naive-bayes-83062940267289

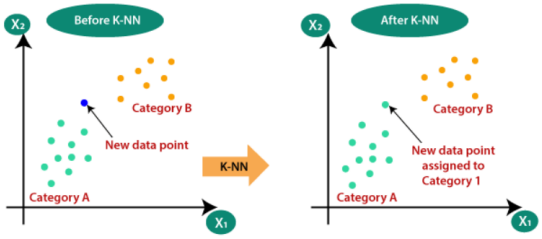


Figura A.6: Exemplo de um kNN Algorithm (Bi et al., 2019) [11]

7) Random Forest – Tem por base a utilização de múltiplas Decison Tree's no modelo de treino. O valor previsto numa determinada característica de um determinado dado do conjunto de dados de testes será o valor mais comum previsto pelas Decision Tree's, ou seja, o valor modal, como pode ser observado na Figura A.7.

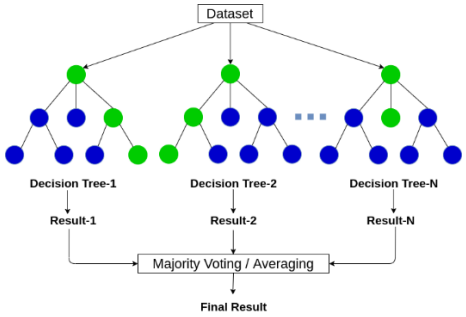


Figura A.7: Exemplo de uma Random Forest (Bi et al., 2019) [12]

8) Adaptive Boosting (AdaBoost) – É um algoritmo que resulta na utilização de vários algoritmos. A técnica mais conhecida do AdaBoost é a utilização de Decision Trees de apenas um nível, criando vários modelos e distribuindo todos os dados de forma equitativa entre cada um. Após esta distribuição, irá ser atribuído um nível de importância a cada modelo. Quanto mais este apresentar erros, maior vai será o seu nível de importância, de forma que o próximo modelo tenha em consideração esta secção, como pode ser observado na Figura A.8.

[11] Imagem retirada do site: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning
[12] Imagem retirada do site: https://ai-pool.com/a/s/random-forests-understanding

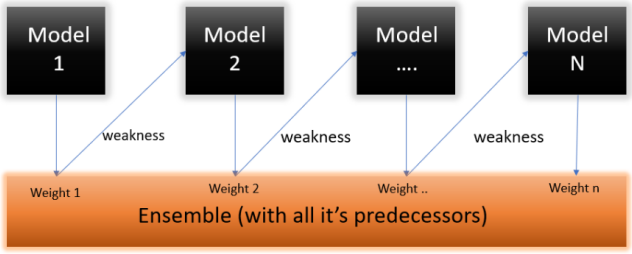


Figura A.8: Exemplo de um AdaBoost (Bi et al., 2019) [13]

9) Stochastic Gradient Descent (SGD) – É um método que liga modelos de classificação lineares tais como o SVM e Logistic Regression. Este fará com que o sistema fique ligado a uma probabilidade aleatória, ou seja, em vez da utilização de todos os dados em cada iteração, serão apenas utilizados alguns dados como exemplo. Este serve para encontrar um mínimo local da função, normalmente repetindo todo o processo em direção oposta ao declive, como pode ser observado na Figura A.9.

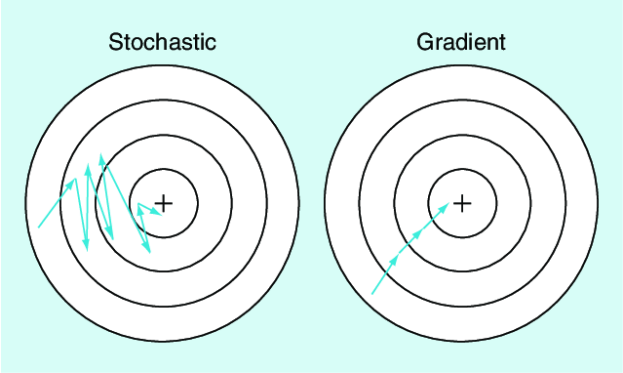


Figura A.9: Exemplo de um SGD (Bi et al., 2019) [14]

10) Gradient Boosting – É uma técnica de ML que combina outros algoritmos mais fracos (normalmente árvores de decisão) num forte por etapas, combinando-os de forma iterativa. Este tenta reduzir o erro médio quadrático, adicionando um novo estimador em cada iteração, com o objetivo de corrigir erros do algoritmo anterior, como pode ser observado na Figura A.10.

[13] Imagem retirada do site: https://www.analyticsvidhya.com/blog/2021/09/adaboost-algorithm-a-complete-guide-for-beginners/

[14] Imagem retirada do site: https://www.researchgate.net/figure/Stochastic-gradient-descent-compared-with-gradient-descent_fig3_328106221

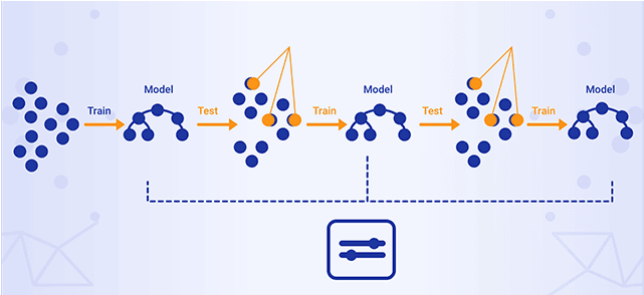
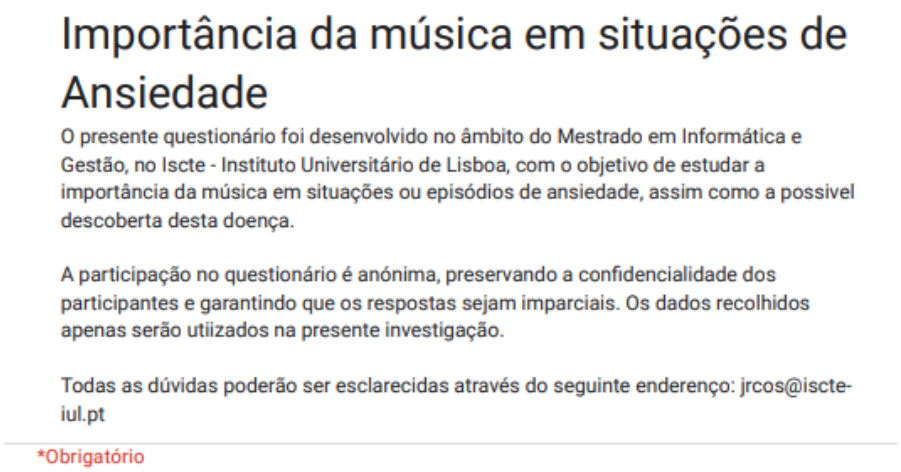


Figura A.10: Exemplo de um Gradient Boosting (Bi et al., 2019) [15]

[15] Imagem retirada do site: https://www.akira.ai/glossary/gradient-boosting

Apêndice B

O questionário desenvolvido no âmbito desta dissertação pode ser visualizado nas figuras seguintes:



Importância da música em situações de Ansiedade

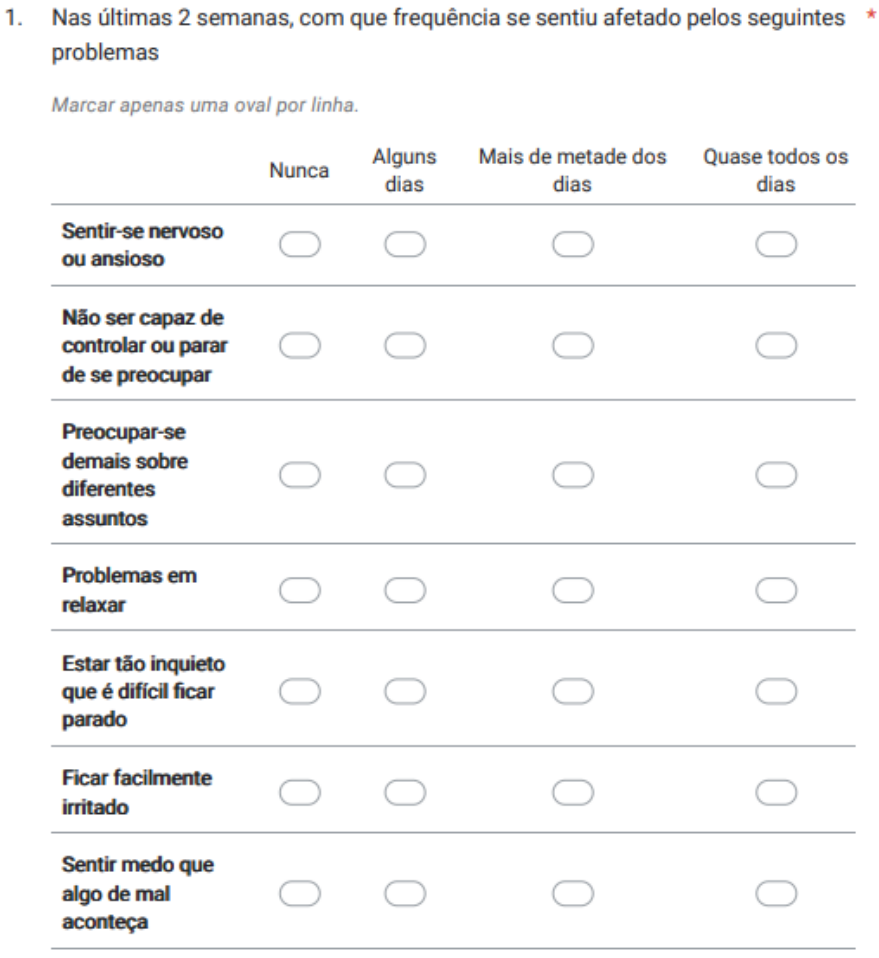
O presente questionário foi desenvolvido no âmbito do Mestrado em Informática e Gestão, no Iscte - Instituto Universitário de Lisboa, com o objetivo de estudar a importância da música em situações ou episódios de ansiedade, assim como a possivel descoberta desta doença.

A participação no questionário é anónima, preservando a confidencialidade dos participantes e garantindo que os respostas sejam imparciais. Os dados recolhidos apenas serão utiizados na presente investigação.

Todas as dúvidas poderão ser esclarecidas através do seguinte enderenço: jrcos@iscte-iul.pt

*Obrigatório

Figura A.11: Introdução do questionário.



1. Nas últimas 2 semanas, com que frequência se sentiu afetado pelos seguintes problemas *

Marcar apenas uma oval por linha.

	Nunca	Alguns dias	Mais de metade dos dias	Quase todos os dias
Sentir-se nervoso ou ansioso				
Não ser capaz de controlar ou parar de se preocupar				
Preocupar-se demais sobre diferentes assuntos				
Problemas em relaxar				
Estar tão inquieto que é difícil ficar parado				
Ficar facilmente irritado				
Sentir medo que algo de mal aconteça				

Figura A.12: Perguntas GAD.

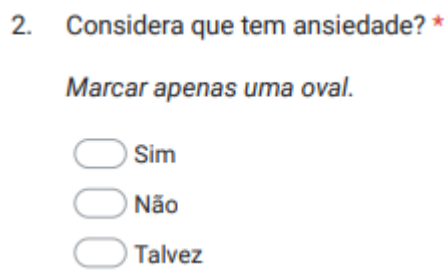


Figura A.13: Segunda pergunta do questionário.

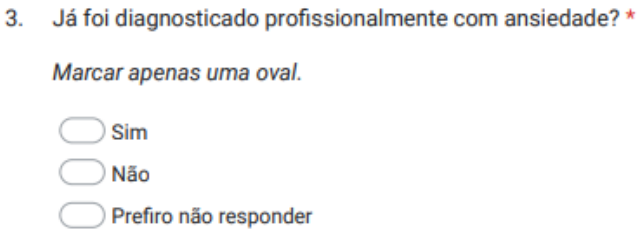


Figura A.14: Terceira pergunta do questionário.

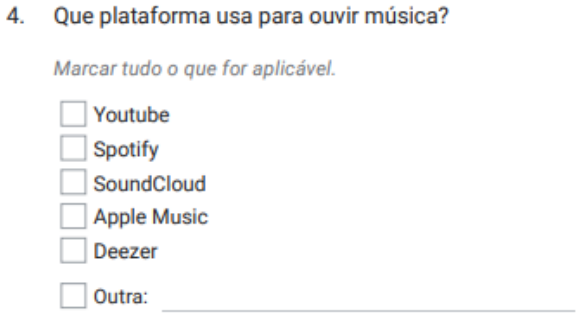


Figura A.15: Quarta pergunta do questionário.

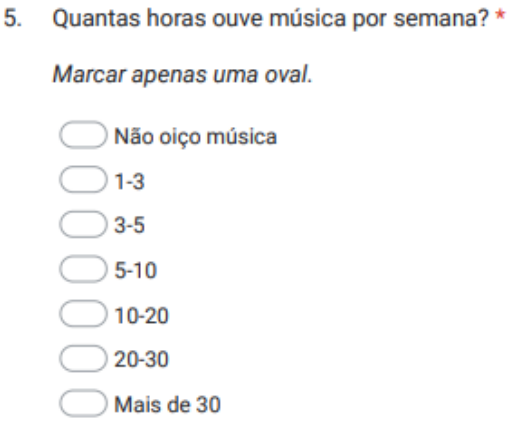


Figura A.16: Quinta pergunta do questionário.

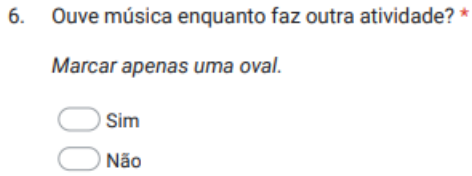


Figura A.17: Sexta pergunta do questionário.

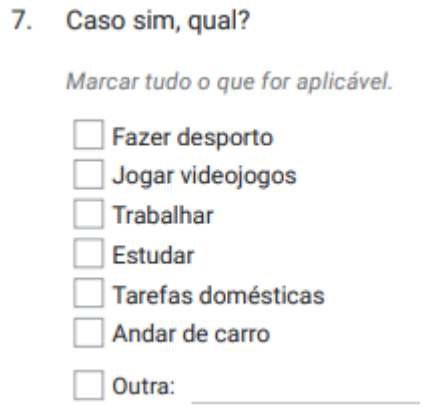


Figura A.18: Sétima pergunta do questionário.

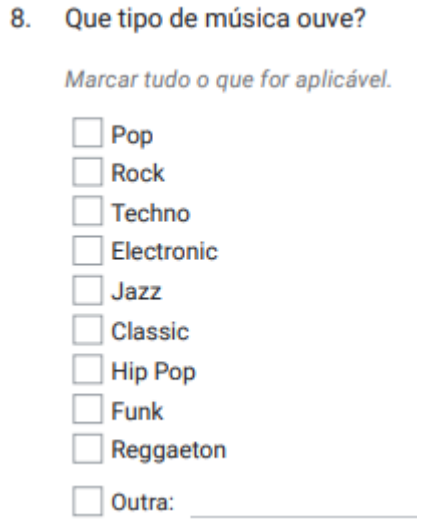


Figura A.19: Oitava pergunta do questionário.

9. Quando se sente ansioso, ouve música? *

Marcar apenas uma oval.

- Sim
- Não

Figura A.20: Nona pergunta do questionário.

10. Caso sim, que tipo de música?

Marcar tudo o que for aplicável.

- [ ] Pop
- [ ] Rock
- [ ] Techno
- [ ] Electronic
- [ ] Jazz
- [ ] Classic
- [ ] Hip Pop
- [ ] Funk
- [ ] Reggaeton
- [ ] Outra: _____

Figura A.21: Décima pergunta do questionário.

11. Quando se sente ansioso... *

Marcar apenas uma oval.

- ... ouve mais música
- ... ouve menos música
- ... ouve a mesma quantidade de música

Figura A.22: Décima primeira pergunta do questionário.

12. Sente que ouvir música diminui o seu estado de ansiedade? *

Marcar apenas uma oval.

- Sim
- Não

Figura A.23: Décima segunda pergunta do questionário.

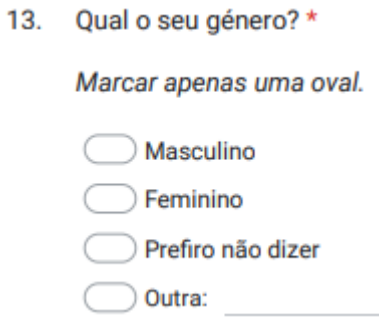


Figura A.24: Décima terceira pergunta do questionário.

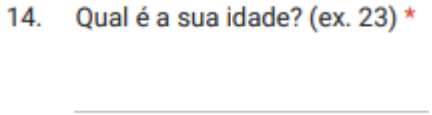


Figura A.25: Décima quarta pergunta do questionário.

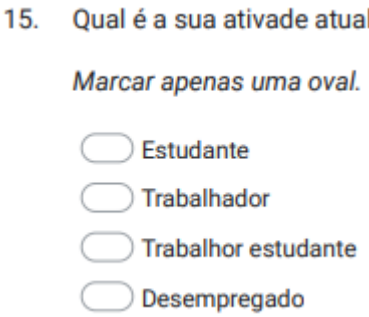


Figura A.26: Décima quinta pergunta do questionário.

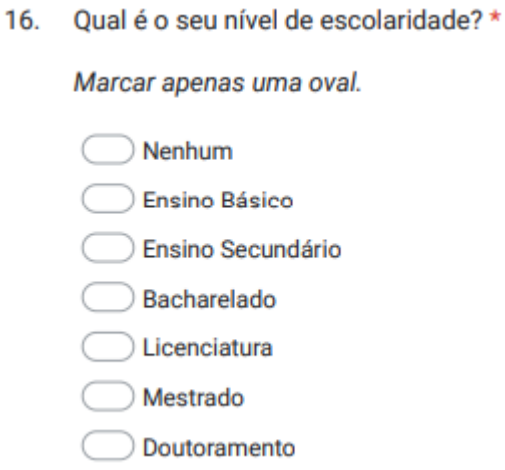


Figura A.27: Décima sexta pergunta do questionário.

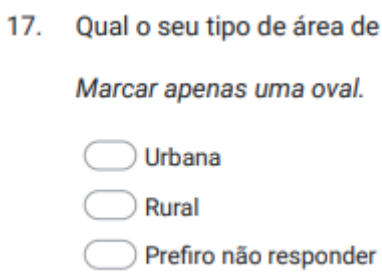


Figura A.28: Décima sétima pergunta do questionário.

Apêndice C

Na Figura A.29 pode-se verificar a imagem utilizada de modo a partilhar o questionário pelas redes sociais Instagram, Facebook e WhatsApp.

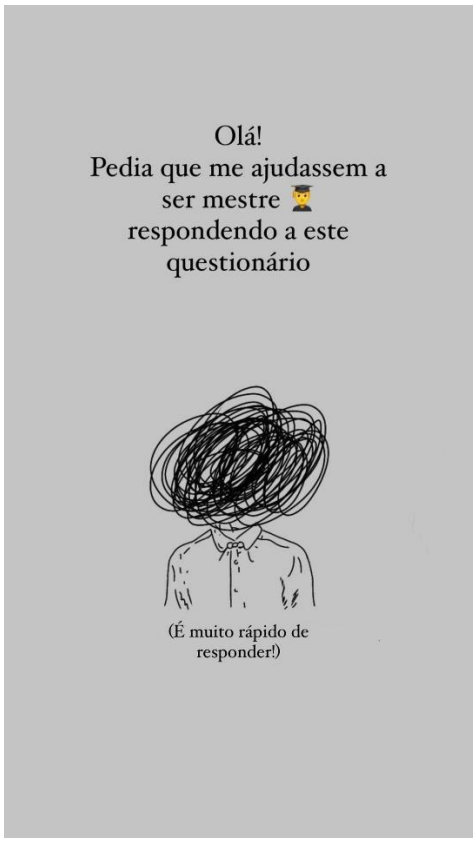


Figura A.29: Imagem utilizada na partilha do questionário.