



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

***Machine Learning* como ferramenta de apoio à prevenção: identificação de níveis de ansiedade**

Carolina Micaela Domingos Marques

Mestrado em Métodos Analíticos para Gestão

Orientadora:

Doutora Catarina Ferreira da Silva, Professora Associada com Agregação
ISCTE - Instituto Universitário de Lisboa

Co-Orientadora:

Doutora Carla Moleiro, Professora Associada
ISCTE - Instituto Universitário de Lisboa

Novembro, 2022



BUSINESS
SCHOOL

Departamento de Métodos Quantitativos para Gestão e Economia

***Machine Learning* como ferramenta de apoio à prevenção: identificação de níveis de ansiedade**

Carolina Micaela Domingos Marques

Mestrado em Métodos Analíticos para Gestão

Orientadora:

Doutora Catarina Ferreira da Silva, Professora Associada com Agregação
ISCTE - Instituto Universitário de Lisboa

Co-Orientadora:

Doutora Carla Moleiro, Professora Associada
ISCTE - Instituto Universitário de Lisboa

Novembro, 2022

Agradecimentos

Durante a elaboração da presente dissertação foram diversas as pessoas que estiveram, direta ou indiretamente, presentes e às quais quero expressar o meu profundo agradecimento:

Primeiramente, às minhas orientadoras, Professora Catarina Ferreira da Silva e Professora Carla Moleiro, por toda a sua orientação, paciência, disponibilidade e voto de confiança. Sem o vosso apoio não teria conseguido alcançar esta meta.

À *Genesis Studio*, em particular, ao Luís e à Isabel, por todo o incentivo, disponibilidade e abertura. Por me ensinarem a ver o “Mundo dos dados” nas várias perspetivas e realidades. Por todo o tempo facultado para a concretização desta dissertação.

Ao Professor Raul Laureano, por ter sido um ótimo diretor de curso. Por toda a disponibilidade, acompanhamento e conselhos. Por ter sido um elemento fulcral em todo o percurso académico.

Aos meus pais, por nunca me deixarem desistir dos meus sonhos. Tudo o que tenho e sou hoje, devo-vos a vocês. Obrigada por me proporcionarem a oportunidade de poder estudar no curso e cidade que sempre desejei, sei que foi necessário um esforço imensurável de ambos. Agradeço-vos por me ensinarem que com esforço e dedicação é possível alcançar qualquer objetivo e por serem um exemplo a seguir. Um obrigada nunca vai chegar para demonstrar a gratidão que tenho por tudo aquilo que fizeram e fazem por mim.

À minha irmã, por me apoiar e ouvir em todos os momentos, tanto na realização desta dissertação como na vida. Por seres o meu lado mais descontraído e leve. Por carregares as minhas vitórias como tuas.

Ao Bernardo, por ser o melhor companheiro. Por todo o apoio incondicional que deste na realização deste projeto. Por seres o meu equilíbrio e me mostrares o lado bom da vida.

À minha família e amigos, por acreditarem sempre em mim.

Ao ISCTE-IUL, por ter sido a minha casa durante estes dois anos. Por me proporcionar boas amizades e momentos. À equipa de basquetebol desta Instituição por me ter acolhido e pelo companheirismo.

A todos os que mencionei o meu sincero “Obrigada”.

Resumo

A pandemia COVID-19, que surgiu em Portugal no ano 2020, teve um impacto a nível mundial quer em termos de saúde pública, por se tratar de uma doença infecciosa de trato respiratório, como em termos de saúde mental. Posto isto, devido à crescente evolução da utilização de *Machine Learning* em problemas da vida real, é questionada a eficácia da utilização desta ferramenta na aplicação ao estudo da saúde mental, no caso concreto da previsão de níveis de ansiedade num determinado período, a pandemia.

Neste sentido, esta investigação avalia a previsão, através de *Machine Learning*, de níveis de ansiedade de estudantes e trabalhadores/as portugueses/as e o impacto que a pandemia teve na saúde mental dos mesmos.

O presente estudo foi suportado por uma Revisão Sistemática da Literatura e por técnicas de análise de dados recorrendo a algoritmos de *Machine Learning*, nomeadamente árvores de decisão. Para tal, a partir da recolha de dados através de um questionário *online*, que corresponde a 597 pessoas inquiridas elegíveis para integrar a análise, foram criados dois modelos distintos, assim como foram determinadas diversas variáveis independentes necessárias para os modelos preditores dos níveis de ansiedade.

Assim, os resultados obtidos permitiram identificar os perfis de estudantes e trabalhadores/as com maior suscetibilidade a agravar o nível de ansiedade associado, como também as variáveis com maior importância no modelo. Deste modo, o principal contributo desta investigação é apoiar a tomada de decisão de profissionais de saúde, de forma que estes tenham uma maior capacidade de realizar diagnósticos de perturbações de ansiedade.

Palavras-Chave: *Machine Learning*, Árvores de Decisão, Saúde Mental, Perturbação de Ansiedade, Portugal, COVID-19

JEL Classification System: I10, C83

Abstract

The COVID-19 pandemic, which emerged in Portugal in 2020, had a worldwide impact both in terms of public health, as it is an infectious disease of the respiratory tract, and regarding mental health. Thus, due to the growing evolution in the use of Machine Learning in real life problems, the effectiveness of this tool in the study of mental health is questioned, specifically on predicting anxiety levels particular period - the pandemic.

In this regard, this research evaluates the prediction, through Machine Learning, of anxiety levels of portuguese students and workers and the impact that the pandemic had on their mental health.

This study was supported by a Systematic Literature Review and by data analysis techniques using Machine Learning algorithms, in particular decision trees. For this purpose, based on data collection through an online survey, corresponding to 597 surveyed people eligible to be included in the analysis, two distinct models were created as well as several independent variables necessary for the predictive models of anxiety levels.

Therefore, the results obtained allowed us to identify the profiles of students and workers who are more likely to aggravate the associated anxiety level, as well as the variables with greater importance in the model. Thus, the main contribution of this research is to support decision-making by health professionals, so that they have a greater ability to diagnose anxiety disorders.

Keywords: Machine Learning, Decision Trees, Mental Health, Anxiety Disorder, Portugal, COVID-19

Classification JEL: I10, C83

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 E RELEVÂNCIA DO TEMA	1
1.2. PROBLEMA E QUESTÃO DE INVESTIGAÇÃO	2
1.3. OBJETIVOS E CONTRIBUTOS DA INVESTIGAÇÃO	4
1.4. ABORDAGEM METODOLÓGICA	5
1.5. ESTRUTURA E ORGANIZAÇÃO DA DISSERTAÇÃO	6
CAPÍTULO 2 – REVISÃO DA LITERATURA.....	7
2.1. PROTOCOLO PARA A REVISÃO SISTEMÁTICA	7
2.2. MODELOS DE <i>MACHINE LEARNING</i>	13
2.2.1. CONTEXTO DE APLICAÇÃO	13
2.2.2. TIPOLOGIA DOS MODELOS.....	17
2.2.3. AVALIAÇÃO DOS RESULTADOS	19
CAPÍTULO 3 – METODOLOGIA.....	21
3.1. POSICIONAMENTO DA INVESTIGAÇÃO	21
3.2. AMOSTRA E DADOS	21
3.2.1. PROCESSO DE SELEÇÃO DA AMOSTRA	21
3.3. CONSTRUÇÃO DO QUESTIONÁRIO	23
3.4. METODOLOGIA CRISP-DM	24
3.4.1. COMPREENSÃO DOS DADOS	25
3.4.2. PREPARAÇÃO DOS DADOS	29
3.4.3. MODELAÇÃO E AVALIAÇÃO	32
CAPÍTULO 4 – RESULTADOS E DISCUSSÃO	37
4.1. MODELO PREDITIVO PARA PREVISÃO DE NÍVEIS DE ANSIEDADE.....	37
4.2. MODELO PREDITIVO DE AVALIAÇÃO DO IMPACTO DA PANDEMIA COVID-19 NO NÍVEL DE ANSIEDADE .	41
4.3. SUSCETIBILIDADE AO AGRAVAMENTO DO NÍVEL DE ANSIEDADE E VARIÁVEIS COM MAIOR INFLUÊNCIA .	44

CAPÍTULO 5: CONCLUSÕES E RECOMENDAÇÕES	47
5.1. PRINCIPAIS CONCLUSÕES	47
5.2. CONTRIBUTOS PARA A COMUNIDADE CIENTÍFICA E EMPRESARIAL	48
5.3. LIMITAÇÕES DO ESTUDO	48
5.4. PROPOSTAS DE INVESTIGAÇÃO FUTURA	49
REFERÊNCIAS BIBLIOGRÁFICAS	51
APÊNDICES.....	55
APÊNDICE A – QUESTIONÁRIO PARA RECOLHA DE DADOS	55
APÊNDICE B – PREPARAÇÃO DOS DADOS: RESUMO DAS ALTERAÇÕES	79
APÊNDICE C – FLUXOGRAMA MODELO FINAL RAMIFICADO [ESTUDANTES]	80
APÊNDICE D – RESULTADOS DO MODELO FINAL RAMIFICADO PARA PREVISÃO DE NÍVEIS DE ANSIEDADE [ESTUDANTES]	81
APÊNDICE E – FLUXOGRAMA MODELO FINAL RAMIFICADO [ESTUDANTES/PANDEMIA].....	82
APÊNDICE F – RESULTADOS DO MODELO FINAL RAMIFICADO PARA AVALIAÇÃO DO IMPACTO DA PANDEMIA COVID- 19 [ESTUDANTES]	83
APÊNDICE G – FLUXOGRAMA DO MODELO FINAL RAMIFICADO [TRABALHADORES-ESTUDANTE]	84
APÊNDICE H – RESULTADOS DO MODELO FINAL RAMIFICADO PARA PREVISÃO DOS NÍVEIS DE ANSIEDADE [TRABALHADORES-ESTUDANTE]	85
APÊNDICE I – FLUXOGRAMA MODELO FINAL RAMIFICADO [TRABALHADORES-ESTUDANTE/PANDEMIA]	86
APÊNDICE J: RESULTADOS DO MODELO FINAL RAMIFICADO PARA AVALIAÇÃO DO IMPACTO DA PANDEMIA COVID- 19 [TRABALHADORES-ESTUDANTE]	87
APÊNDICE K – FLUXOGRAMA MODELO FINAL RAMIFICADO [TRABALHADORES]	88
APÊNDICE L – RESULTADOS DO MODELO FINAL RAMIFICADO PARA PREVISÃO DOS NÍVEIS DE ANSIEDADE [TRABALHADORES]	89
APÊNDICE M – FLUXOGRAMA MODELO FINAL RAMIFICADO [TRABALHADORES/PANDEMIA]	90
APÊNDICE N – RESULTADOS DO MODELO FINAL RAMIFICADO PARA AVALIAÇÃO DO IMPACTO DA PANDEMIA COVID-19 [TRABALHADORES/PANDEMIA]	91
APÊNDICE O – FLUXOGRAMA MODELO FINAL GERAL	92
APÊNDICE P – FLUXOGRAMA MODELO FINAL GERAL [IMPACTO DA PANDEMIA COVID-19].....	93
APÊNDICE Q – RESULTADOS DO MODELO FINAL GERAL.....	94

Índice de Tabelas

Tabela 2.1: Critérios de inclusão e exclusão.....	8
Tabela 2.2: Artigos incluídos na revisão sistemática da literatura no ano de 2022.....	10
Tabela 2.3: Artigos incluídos na revisão sistemática da literatura no ano de 2021.....	11
Tabela 2.4: Artigos incluídos na revisão sistemática da literatura nos anos 2020 e 2019	12
Tabela 2.5: Critérios de qualidade para avaliação dos artigos incluídos na RSL.....	12
Tabela 2.6: Contextos de aplicação dos artigos incluídos na rsl.....	15
Tabela 2.7: Tipologias de modelos utilizadas na rsl	18
Tabela 3.1: Compreensão das variáveis independentes presentes na base de dados [sociodemográficas]	26
Tabela 3.2: Compreensão das variáveis independentes presentes na base de dados [Estudante]	27
Tabela 3.3: Compreensão das variáveis independentes presentes na base de dados [Trabalhador]..	28
Tabela 3.4: Compreensão das variáveis presentes na base de dados [Saúde mental].....	29
Tabela 3.5: Conjunto final de variáveis independentes	31
Tabela 3.6: Métricas e fórmulas utilizadas para avaliação da performance dos modelos	33
Tabela 3.7: Modelos e parametrizações	35
Tabela 4.1: Resultados dos modelos com melhor qualidade para previsão de níveis de ansiedade em estudantes	38
Tabela 4.2: Resultados dos modelos com melhor qualidade para previsão de níveis de ansiedade em trabalhadores-estudantes	39
Tabela 4.3: Resultados dos modelos com melhor qualidade para previsão de níveis de ansiedade em trabalhadores	40
Tabela 4.4: Resultados dos modelos com melhor qualidade para avaliação do impacto da pandemia em estudantes	42
Tabela 4.5: Resultados dos modelos com melhor qualidade para avaliação do impacto da pandemia em trabalhadores-estudantes	43
Tabela 4.6: Resultados dos modelos com melhor qualidade para avaliação do impacto da pandemia em trabalhadores	44
Tabela 4.7: Perfil de estudantes universitários suscetíveis a agravar o nível de ansiedade.....	44
Tabela 4.8: Perfil de trabalhadores-estudantes suscetíveis a agravar o nível de ansiedade.....	45
Tabela 4.9: Perfil de trabalhadores suscetíveis a agravar o nível de ansiedade.....	45
Tabela B: Preparação dos dados	79

Tabela D: Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Estudantes]	81
Tabela F: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados [Estudantes]	83
Tabela H: Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Trabalhadores-estudante]	85
Tabela J: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados [Trabalhadores-estudante]	87
Tabela L: Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Trabalhadores]	89
Tabela N: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados [Trabalhadores]	91
Tabela Q.1: Resultados do modelo final geral para previsão dos níveis de ansiedade	94
Tabela Q.2: Resultados do modelo final geral para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados	95

Índice de Figuras

Figura 1.1: A metodologia CRISP-DM, segundo Moro (2015)	5
Figura 2.1: Processo de seleção dos artigos na RSL	9
Figura 3.1: Caracterização sociodemográfica da amostra em estudo	22
Figura 3.2: Caracterização da amostra quanto à existência de diagnóstico de ansiedade.....	23
Figura 3.3: Processo de construção do questionário	23
Figura 3.4: Aplicação da metodologia CRISP-DM.....	25
Figura 3.5: Modelos finais (geral e ramificado).....	32
Figura A: Questionário para recolha de dados.....	55
Figura C: Fluxograma modelo final ramificado [Estudantes]	80
Figura E: Fluxograma modelo final ramificado [Estudantes/Pandemia].....	82
Figura G: Fluxograma modelo final ramificado [Trabalhadores-estudante].....	84
Figura I: Fluxograma modelo final ramificado [Trabalhadores-estudante/Pandemia].....	86
Figura K: Fluxograma modelo final ramificado [Trabalhadores].....	88
Figura M: Fluxograma modelo final ramificado [Trabalhadores/Pandemia].....	90
Figura O: Fluxograma modelo final geral para previsão de níveis de ansiedade.....	92
Figura P: Fluxograma modelo final geral para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados	93

Lista de Abreviaturas e Siglas

COVID-19: Doença do Coronavírus de 2019

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

IA: Inteligência Artificial

ML: *Machine Learning*

OMS: Organização Mundial de Saúde

RSL: Revisão Sistemática da Literatura

SNS: Serviço Nacional de Saúde

SPPSM: Sociedade Portuguesa de Psiquiatria e Saúde Mental

WoS: *Web of Science*

Capítulo 1 – Introdução

A presente introdução consiste em apresentar o conceito de saúde mental, mais especificamente, de perturbação de ansiedade, e evidenciar a sua relevância, bem como a identificação do problema, objetivos e contributos subjacentes neste estudo. É abordada também a metodologia adotada e a estrutura da presente investigação.

1.1. Enquadramento e relevância do tema

Não existe saúde sem saúde mental. Como definido pela Organização Mundial de Saúde (OMS) em 2014, a saúde é o “estado de bem-estar físico, mental e social e não apenas a ausência de doença ou enfermidade”.

Assim, sendo a saúde mental parte integrante da condição de ter saúde, esta caracteriza-se como um estado de bem-estar em que cada indivíduo tem capacidade para compreender o seu próprio potencial, conseguir lidar com os desafios normais do dia-a-dia e trabalhar de forma produtiva e frutífera, de modo a contribuir para a comunidade (Organização Mundial de Saúde [OMS], 2014). Contudo, esta capacidade pode ser comprometida pela existência de uma perturbação mental, o que afeta o bem-estar mental e o normal funcionamento de um indivíduo (Caetano, 2019).

Segundo a Sociedade Portuguesa de Psiquiatria e Saúde Mental (SPPSM), uma perturbação mental caracteriza-se, de um modo geral, “por mudanças ao nível do modo de pensamento e das emoções, normalmente com tradução comportamental, ou de uma combinação destes, associados com diminuição (ou perda) da liberdade de agir e/ou deterioração do dia-a-dia do doente, da sua vida social, do emprego ou das atividades familiares” (SPPSM, n.d.).

Existem múltiplas perturbações mentais, que se encontram organizadas em grupos, segundo definido pela *American Psychiatric Association* (2013), no entanto a presente investigação foca-se apenas num grupo: as perturbações de ansiedade. Esta escolha incide no facto de estas serem designadas como “perturbações mentais comuns” (OMS, 2017). A denominação atribuída respeita a perturbações que são muito predominantes na sociedade e que afetam qualquer tipo de indivíduo, sem qualquer diferenciação de género, etnia ou status social (Caetano, 2019).

Desta forma, uma perturbação de ansiedade é definida, pelo Serviço Nacional de Saúde (SNS) de Portugal (2022), como a existência de “um medo grave, desproporcionado, que perdura há pelo menos seis meses e que têm um verdadeiro impacto na vida quotidiana, com interferência com a capacidade de normal funcionamento do indivíduo”. Estas perturbações englobam fobias, ataques de pânico, perturbações de ansiedade generalizada, stress pós-traumático, entre outras, e podem ser classificadas em diferentes níveis - leve, moderado ou grave.

A ansiedade pode ser dividida em ansiedade traço e estado. A ansiedade-traço refere-se a uma propensão geral a ficar ansioso, ou seja, é uma característica individual do traço de personalidade de um indivíduo. Enquanto ansiedade-estado é definida como uma resposta emocional temporária influenciada por uma determinada situação vivenciada (Gruda *et al.*, 2019).

Contudo, é importante salientar que a ansiedade é uma resposta natural do funcionamento normal de um indivíduo perante uma situação de stress (Muhammad *et al.*, 2022). Porém, a ansiedade excessiva ou inadequada pode desencadear uma perturbação mental e, consequentemente, diminuir a qualidade de vida do indivíduo (Ding *et al.*, 2022). Segundo a revisão mundial sobre saúde mental divulgada pela OMS, as perturbações mentais são a principal causa de incapacidade e uma das principais causas de morbilidade na sociedade (OMS, 2022). No caso específico, as perturbações de ansiedade afetam 16,5% da população portuguesa, representando assim um dos países com maiores níveis de prevalência nas doenças mentais no panorama europeu (Almeida *et al.*, 2013).

Assim, segundo as informações providenciadas ao longo deste estudo, evidencia-se a relevância e pertinência da construção de um modelo preditivo para previsão de sinais de ansiedade clinicamente significativa que tem como objetivo não só a previsão, mas também capacitar os profissionais de saúde a realizar um diagnóstico, se possível precoce, e melhorar a qualidade de vida dos doentes.

1.2. Problema e questão de investigação

Nos últimos anos, tem-se vindo a verificar uma tendência global para um aumento de prevalência de doenças mentais. Em 2019, quase um bilião de pessoas sofria com uma perturbação mental, dos quais 14% correspondia a adolescentes de todo o mundo (OMS, 2022).

Nos meses finais de 2019 foi identificada, numa cidade chinesa, pela primeira vez a doença do Coronavírus 2019, designado por COVID-19. (Antunes *et al.*, 2020). Esta doença pode causar infeção respiratória, cujos sintomas são semelhantes aos de uma gripe, como a existência de febre, tosse, cansaço, entre outros. Contudo, esta doença afeta cada pessoa de formas diferentes e, nalguns casos, pode evoluir para uma condição mais grave, como é o caso de uma pneumonia (OMS, 2022).

A propagação da mesma foi rapidamente transmissível por todo o mundo, o que originou uma pandemia global que foi oficialmente declarada como uma emergência de saúde pública pela OMS em março de 2020 (OMS, 2022). De forma a proteger a saúde pública e impedir a propagação do vírus, os países aplicaram diversas “(...) medidas extraordinárias e de carácter urgente de restrição de direitos e liberdades, em especial no que respeita aos direitos de circulação e às liberdades económicas (...)” (Diário da República Eletrónico, 2020).

Assim, a 18 de março de 2020, foi decretado um estado de emergência em Portugal, que implementou medidas como isolamento social e quarentena. Contudo, é importante definir estes

conceitos que são distintos. Quarentena consiste num período em que pessoas saudáveis estiveram expostas a uma doença transmissível, seja por contacto direto com um doente ou por estar numa região com um surto epidémico, têm a sua liberdade limitada. O isolamento físico representa a separação de pessoas infetadas do resto da população para evitar a propagação do vírus (Antunes *et al.*, 2020).

Embora as medidas impostas tenham originado um efeito positivo no que se refere à proteção da saúde física da comunidade e na mitigação da transmissão do vírus, estas indicam um impacto negativo a nível psicológico, e não só (Antunes *et al.*, 2020). A pandemia de COVID-19 afetou a saúde mental de forma global (Carlton *et al.*, 2022), devido a vários fatores, e diversos estudos suportam esta afirmação.

Segundo Alharbi *et al.* 2022, as perturbações de saúde mental, incluindo a ansiedade, estão relacionados com a doença COVID-19 uma vez que a rápida disseminação do vírus e o aumento das taxas de mortalidade no início de 2020 originou o aumento dos níveis de ansiedade, quer em profissionais de saúde que na população em geral. Em termos quantitativos, de acordo com Saeed *et al.* 2022, a incidência de ansiedade aumentou em dobro no ano 2020 em comparação com os anos anteriores.

Contudo, as medidas aplicadas não tiveram apenas impacto a nível psicológico, mas também originaram dificuldades financeiras e fisiológicas uma vez que o quotidiano da população foi alterado de uma forma radical, desde mudanças no estilo de vida, no nível económico – muitas pessoas viram-se impedidas de se deslocar para o seu local de trabalho e, em muitos casos, originou o encerramento de muitas empresas – o medo de ficar infetado ou de perder parentes, a incerteza do futuro, a acessibilidade a alimentos, entre outras alterações e preocupações (Duru *et al.*, 2022; Carlton *et al.*, 2022).

De acordo com o resumo científico divulgado pela OMS a março de 2022, a prevalência global de ansiedade e depressão aumentou em 25% no primeiro ano de pandemia devido aos fatores de stress mencionados, mas também coincidiu com graves interrupções nos serviços hospitalares pois, durante a pandemia, os serviços para condições mentais, neurológicas e de uso de substâncias foram os mais interrompidos entre todos os outros serviços de saúde essenciais (OMS, 2022).

Neste contexto, uma vez que os efeitos da saúde mental a longo prazo ainda não são conhecidos, é importante implementar ferramentas eficazes para detetar possíveis sinais de ansiedade no decorrer da situação pandémica e através da avaliação dos sintomas de forma longitudinal. Para isto, sendo uma área tão vasta e densa de abordar, conclui-se que é necessário alavancar abordagens inovadoras, tais como aplicar *Machine Learning* para a deteção de níveis de ansiedade.

Assim, a presente investigação tem como finalidade gerar conhecimento que permita identificar o perfil de indivíduos que pode, ao longo dos períodos em estudo, desenvolver ou agravar o nível de

ansiedade associado. Neste sentido, este estudo visa responder à seguinte questão de investigação: De que modo pode o *Machine Learning* ser usado como ferramenta de apoio à prevenção na previsão de níveis significativos de ansiedade?

1.3. Objetivos e contributos da investigação

A presente investigação tem como objetivo principal prever níveis de ansiedade clinicamente significativa em alunos universitários e trabalhadores portugueses, com base em técnicas de *Machine Learning*, durante e após a situação pandémica COVID-19. De forma a concretizar o objetivo principal estabelecido, são definidos três objetivos específicos:

- 1) Criar um modelo preditivo para previsão de níveis de ansiedade dos alunos universitários e trabalhadores portugueses;
- 2) Criar um modelo preditivo para avaliar o impacto da pandemia COVID-19 no nível de ansiedade associado dos alunos universitários e trabalhadores portugueses.
- 3) Identificar o perfil de estudantes universitários e trabalhadores portugueses que são mais suscetíveis a agravar o nível de ansiedade associado, bem como os atributos que afetam com maior influência a alteração do nível de ansiedade associado.

A concretização dos objetivos e, consequentemente, a resposta à questão de investigação formulada, vem contribuir a vários níveis, tais como académico, profissional, pessoal e comunitário. No que diz respeito à academia, vem colmatar uma lacuna na literatura, uma vez que os modelos de *Machine Learning* já foram aplicados à previsão de sinais de ansiedade tanto em estudantes universitários como em trabalhadores, mas são poucos os que recolheram dados relativos a estudantes e trabalhadores portugueses após uma situação específica, a pandemia COVID-19. Relativamente ao nível profissional, contribui para a melhoria da atuação dos profissionais de saúde, isto é, vem auxiliar a definir um diagnóstico e um plano de tratamento mais adequado e personalizado para cada caso, o que, consequentemente, vem melhorar, a nível pessoal, a qualidade de vida dos doentes. Por fim, a nível comunitário, vem contribuir para desmitificar e promover a saúde mental em Portugal.

1.4. Abordagem metodológica

O presente trabalho seguirá a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) uma vez que se permite a construção e implementação de um modelo *Data Mining* a ser utilizado num ambiente real (Moro *et al.*, 2011).

Esta fornece uma visão geral do ciclo da vida do projeto e é composta por um conjunto de seis fases que interagem entre si. Para uma melhor compreensão da metodologia, é necessário entender cada uma das fases que a compõe. A figura 1.1 mostra as semelhanças entre as metodologias *Design Science Research* e CRISP-DM em que as fases de cada uma podem ser comparadas, uma vez que apresentam objetivos idênticos.

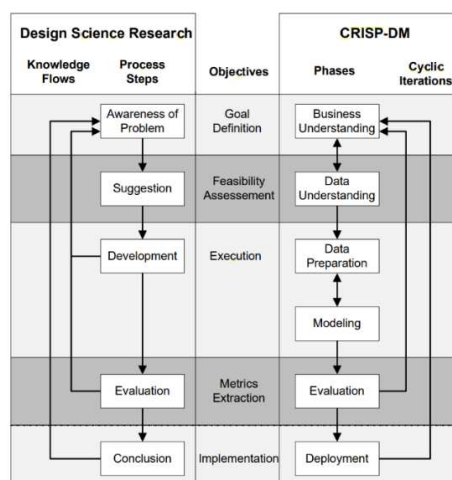


Figura 1.1: A metodologia CRISP-DM, segundo Moro (2015)

Primeiramente, a Compreensão do Negócio (*Business Understanding*) onde é feito o entendimento do trabalho em questão e do seu contexto, posteriormente a Compreensão dos Dados (*Data Understanding*) que diz respeito à recolha e à entendimento dos dados. Na fase seguinte, a Preparação dos Dados (*Data Preparation*) onde são realizadas as transformações necessárias e a sua validação. Segue-se a Modelação (*Modeling*) onde são implementados métodos analíticos para fazer previsões e análises. Na quinta fase, a Avaliação (*Evaluation*), é realizada a avaliação do mesmo tendo por base os critérios de sucesso que foram definidos no início do processo e, por fim, a Implementação (*Deployment*) onde se faz a aplicação prática das fases anteriores.

1.5. Estrutura e organização da dissertação

A elaboração da presente investigação consubstancia um encadeamento lógico, com o objetivo de alcançar os objetivos propostos e responder à questão de pesquisa apresentada. Neste seguimento, o estudo está organizado em cinco capítulos que pretendem refletir as diferentes fases até à sua conclusão.

O primeiro capítulo, que corresponde à introdução, começa por descrever o tema da investigação e a sua relevância, apresentando também a problemática subjacente, a questão de pesquisa e os objetivos propostos, bem como os contributos do estudo e a descrição da metodologia adotada.

O segundo capítulo reflete o enquadramento teórico, designado por Revisão da literatura, onde é essencialmente abordada a temática de *Machine Learning*, com o objetivo principal de entender de que forma os modelos analíticos são utilizados no estudo e previsão de saúde mental.

O terceiro capítulo é dedicado à metodologia utilizada no processo de recolha e tratamento de dados, desde a seleção do público-alvo até à construção do questionário, bem como os métodos de análise utilizados. A análise dos resultados obtidos é apresentada no quarto capítulo, de acordo com a metodologia que se entendeu apropriada, bem como a sua discussão.

Por fim, no quinto capítulo, apresentam-se as conclusões e são identificados os contributos do presente estudo e, ainda, as limitações e pistas futuras de investigação.

Capítulo 2 – Revisão da Literatura

Os modelos de *Machine Learning* (ML) são capazes de prever e oferecer um tratamento diferenciado e personalizado aos pacientes, e, desta forma, têm-se revelado uma ótima ferramenta com potencial para caracterizar, prever e tratar perturbações mentais uma vez que oferecem novas ferramentas para superar desafios para os quais os métodos estatísticos tradicionais não estão adequados (Jiang *et al.*, 2020).

Segundo Paton *et al.* 2022, *Machine Learning*, um dos ramos da Inteligência Artificial (IA), caracteriza-se como uma tecnologia digital que aprende a partir dos dados a identificar padrões e, conseqüentemente, a extrair conhecimento potencialmente útil que suporta a tomada de decisão. As técnicas de ML mais comumente utilizadas no campo psiquiátrico são a aprendizagem supervisionada e aprendizagem não supervisionada (Zhou *et al.*, 2020)

A aprendizagem supervisionada é utilizada quando se sabe qual é a variável de saída, ou seja, tem como objetivo desenvolver um modelo para o qual o resultado pode ser formulado a partir de outras variáveis e, desta forma, fazer uma previsão do resultado futuro aquando aplicadas novas entradas (Zhou *et al.*, 2020). Este tipo de aprendizagem é classificado em problemas de regressão e classificação. No que concerne à aprendizagem não supervisionada, esta é aplicada quando não se sabe qual é a variável de saída, ou seja, é realizada pela descoberta de semelhanças nos dados de entrada sem o auxílio de um resultado rotulado. Deste modo, procura realizar agrupamentos dos exemplos de treino (Jiang *et al.*, 2020). Estes conceitos são suportados por diferentes tipos de algoritmos, bem como diferentes métricas de avaliação, e têm áreas de atuação muito específicas.

Neste sentido, o presente capítulo aborda de que forma são utilizados os modelos de *Machine Learning* na identificação de perturbações de ansiedade. Primeiramente, é descrito o protocolo da Revisão Sistemática da Literatura (RSL) e todo o processo de seleção da mesma, desde a construção da *query* de pesquisa até à seleção dos artigos analisados e incluídos na revisão da literatura. De seguida, tendo por base uma RSL, são abordados os contextos de aplicação dos modelos de *Machine Learning*, a tipologia e, por fim, a avaliação da qualidade dos mesmos.

2.1. Protocolo para a revisão sistemática

Revisão sistemática da literatura (RSL), segundo Caldas 2017, refere-se a um método de pesquisa que é orientado por protocolos específicos. O protocolo deve apresentar, de forma inequívoca, quais as bases de dados bibliográficas consultadas, a *query* de pesquisa utilizada e todo o processo de seleção dos artigos, desde a apresentação dos critérios utilizados para incluir e excluir artigos da análise como também a avaliação da qualidade dos mesmos (Vilelas, 2017).

A revisão de literatura do presente estudo tem como objetivo elaborar o estado da arte relativamente à previsão de sinais de ansiedade clinicamente significativa com recurso a *Machine Learning*. Neste sentido, esta visa responder à seguinte questão: Como se pode aplicar modelos de *Machine Learning* ao estudo da saúde mental, mais especificamente a prever perturbações de ansiedade? Isto é, propõe-se a responder às perguntas seguintes: (1) Em que contextos são aplicados os diferentes modelos analíticos? (2) Quais os modelos analíticos utilizados para deteção de perturbações de ansiedade? (3) Como são avaliados os resultados obtidos?

A base de dados científica utilizada foi a *Web of Science* (WoS), tendo sido selecionada devido à relevância, quer académica como prática, dos *journals* indexados e à vasta gama de publicações na área de interesse (Vilelas, 2017).

O processo de revisão foi dividido em quatro fases: identificação, seleção, elegibilidade e inclusão, como mostra a figura 2.2. Na primeira fase, é realizada uma pesquisa automática na base de dados, que devolve um grande número de referências que são potencialmente elegíveis, através de uma *query* aplicada ao título do artigo. A *query* definida foi: Anxiety AND ("Machine Learning" OR "Data Mining" OR Analytic* OR Algorithm* OR Regression OR Classificat*) (Título), isto é, as palavras-chave utilizadas na *query* resultam da identificação na literatura de palavras/expressões relacionadas com ansiedade e, em simultâneo, a identificação de termos relacionados com *Machine Learning*. Deste processo preliminar de pesquisa são devolvidos 477 resultados, que posteriormente foram sujeitos a uma triagem, a segunda fase do processo.

Nesta fase os artigos selecionados foram sujeitos aos critérios de inclusão e de exclusão, apresentados na tabela 2.1. Desta forma, resultaram 103 artigos após a seleção com base nestes mesmos critérios.

Tabela 2.1: Critérios de inclusão e exclusão

Critérios de inclusão	Artigos que abordam a utilização de <i>Machine Learning</i> ;
	Artigos que abordam ansiedade.
Critérios de exclusão	Tipo de documento ≠ Artigo;
	Artigos que não estão em língua inglesa;
	Artigos que não estejam totalmente disponíveis;
	Artigos que não tenham fator de impacto.

A aplicação dos critérios definidos obrigou à leitura dos títulos e resumos, os artigos que não cumpriam estes critérios foram excluídos. Desta forma, o número de artigos para análise de elegibilidade é de 29.

Neste seguimento, foram analisados os textos completos dos artigos e excluídos os que não eram elegíveis de acordo com o último critério de exclusão da tabela 2.1, isto é, por não terem fator de impacto. Assim, resultam 20 artigos.

Por fim, a quarta fase diz respeito à inclusão, ou seja, os artigos que foram incluídos e avaliados na revisão da literatura da presente investigação. Assim, resultaram 20 artigos para inclusão na RSL.

Na figura 2.1 está apresentado o resumo do processo de seleção dos artigos descrito.

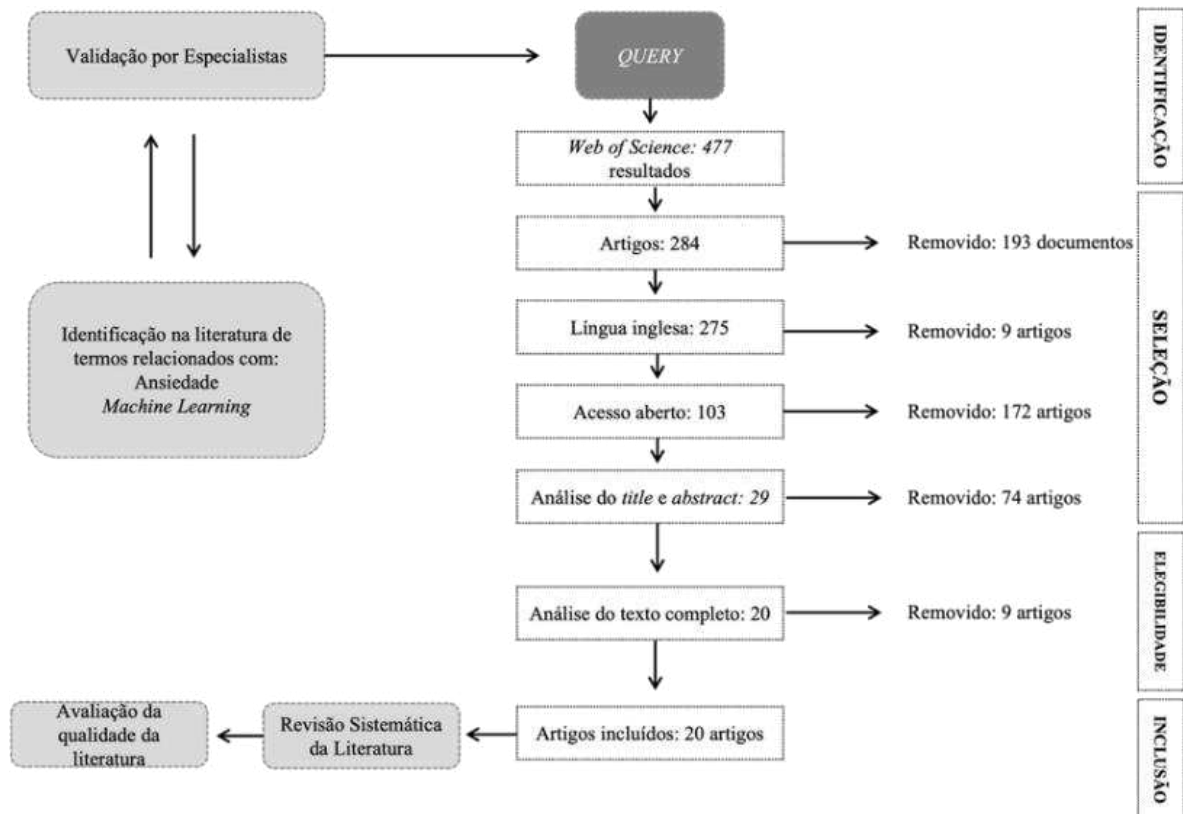


Figura 2.1: Processo de seleção dos artigos na RSL

De seguida, as tabelas 2.2, 2.3 e 2.4 apresentam os artigos incluídos na RSL publicados nos anos de 2022, 2021 e, por fim, 2020 e 2019, respetivamente.

Tabela 2.2: Artigos incluídos na revisão sistemática da literatura no ano de 2022

ID	Ano	Título	Autores	Título da revista
1	2022	Predicting the naturalistic course in anxiety disorders using clinical and biological markers: a machine learning approach	Bokma, W. A.; Zhutovsky, P.; Giltay, E. J.; Schoevers, R. A.; Penninx, J. H.; Van Balkom, J. M.; Batelaan, N. M.; Van Wingen, G. A.	Psychological Medicine
2	2022	Identifying Individuals Who Currently Report Feelings of Anxiety Using Walking Gait and Quiet Balance: An Exploratory Study Using Machine Learning	Stark, M.; Huang, H. K.; Yu, L. F.; Martin, R.; McCarthy, R.; Locke, E.; Yager, C.; Torad, A. A.; Kadry, A. M.; Elwan, M. A.; Smith, M. L.; Bradley, D.; Boolani, A.	Sensors
3	2022	Dynamic Tracking of State Anxiety via Multi-Modal Data and Machine Learning	Ding, Y.; Liu, J. J.; Zhang, X. C.; Yang, Z.	Frontiers in Psychiatry
4	2022	Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown	Anbarasi, L. J.; Jawahar, M.; Ravi, V.; Cherian, S. M.; Shreenidhi, S.; Sharen, H.	Health and Technology
5	2022	Machine Learning Based Depression, Anxiety, and Stress Predictive Model During COVID-19 Crisis	Al-Wesabi, F. N.; Alsolai, H.; Hilal, A. M.; Hamza, M. A.; Al Duhayyim, M.; Negm, N.	Computers Materials & Continua (CMC)
6	2022	Human state anxiety classification framework using EEG signals in response to exposure therapy	Muhammad, F.; Al-Ahmadi, S.	PLOS One
7	2022	Comparison of the Performance of Machine Learning-based Algorithms for Predicting Depression and Anxiety among University Students in Bangladesh: A Result of the First Wave of the COVID-19 Pandemic	Nayan, I. H.; Uddin, S. G.; Hossain, M. I.; Alam, M. M.; Zinnia, M. A.; Haq, I.; Rahman, M. M.; Ria, R.; Methun, I. H.	Asian Journal of Social Health and Behavior

Tabela 2.3: Artigos incluídos na revisão sistemática da literatura no ano de 2021

ID	Ano	Título	Autores	Título da revista
8	2021	Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between Auto-sklearn, Naïve Bayes Classifier, and Traditional Logistic Regression	Van Eeden, W. A.; Luo, C.; Van Hemert, A. M.; Carlier, V. E.; Penninx, B. W.; Wardenaar, K. J.; Hoos, H.; Giltay, E. J.	Psychiatry Research
9	2021	Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders	Richter, T.; Fishbain, B.; Fruchter, E.; Richter-Levin, G.; Okon-Singer, H.	Journal of Psychiatric Research
10	2021	Prediction of state anxiety by machine learning applied to photoplethysmography data	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.	PEERJ
11	2021	Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA).	Wardenaar, K. J.; Riese, H.; Giltay, E. J.; Eikelenboom, M.; Van Hemert, A. J.; Beekman, A. F.; Penninx, J. H.; Schoevers, R. A.	Journal of Affective Disorders
12	2021	Algorithms to Predict Anxiety and Depression Among University Students in China After Analyzing Lifestyles and Sport Habits	Zhang, L. R.; Zhao, S. C.; Lin, Q.; Song, M. M.; Wu, S. R.; Zheng, H.	Neuropsychiatric Disease and Treatment
13	2021	Generalized Anxiety Disorder and Depressive Symptoms among Pakistani Population during the Second Wave of the COVID-19 Pandemic: A Regression Analysis	Yasmin, F.; Dapke, K.; Khalid, M. R.; Naz, H.; Naz, F.; Admani, B.; Asghar, M. S.; Awan, S.; Phadke, R.; Saleem, M.	American Journal of Tropical Medicine and Hygiene
14	2021	Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence	Nemesure, M. D.; Heinz, M. V.; Huang, R.; Jacobson, N. C.	Scientific Reports
15	2021	Machine Learning-Based Predictive Modeling of Anxiety and Depressive Symptoms During 8 Months of the COVID-19 Global Pandemic: Repeated Cross-sectional Survey Study	Hueniken, K.; Some, N. H.; Abdelhack, M.; Taylor, G.; Marshall, T. E.; Wickens, C. M.; Hamilton, H. A.; Wells, S.; Felsky, D.	JMIR Mental Health

Tabela 2.4: Artigos incluídos na revisão sistemática da literatura nos anos 2020 e 2019

ID	Ano	Título	Autores	Título da revista
16	2020	Chinese College Students Have Higher Anxiety in New Semester of Online Learning During COVID-19: A Machine Learning Approach	Wang, C. Y.; Zhao, H.; Zhang, H. R.	Frontiers in Psychology
17	2020	Toward Robust Anxiety Biomarkers: A Machine Learning Approach in a Large-Scale Sample	Boeke, E. A.; Holmes, A. J.; Phelps, E. A.	Biological Psychiatry-Cognitive Neuroscience and Neuroimaging
18	2019	Lifestyles, Depression, Anxiety, and Stress as Risk Factors in Nursing Apprentices: A Logistic Regression Analysis of 1193 Students in Lima, Peru	Diaz-Godino, J.; Fernandez-Henriquez, L.; Pena-Pastor, F.; Alfaro-Flores, P.; Manrique-Borjas, G.; Mayta-Tovalino, F.	Journal of Environmental And Public Health
19	2019	Predicting anxiety from wholebrain activity patterns to emotional faces in young adults: a machine learning approach	Portugal, C. L.; Schrouff, J.; Stiffler, R.; Bertocci, M.; Bebeko, G.; Chase, H.; Lockovitch, J.; Aslam, H.; Graur, S.; Greenberg, T.; Pereira, M.; Oliveira, L.; Phillips, M.; Mourao-Miranda, J.	Neuroimage-Clinical
20	2019	Feeling anxious? Perceiving anxiety in tweets using machine learning	Gruda, D.; Hasan, S.	Computers in Human Behavior

Após a realização da revisão descrita nas tabelas 2.2., 2.3 e 2.4, os artigos selecionados foram submetidos a uma avaliação da sua qualidade através de critérios de qualidade. Estes critérios destinam-se a identificar o contributo dos artigos incluídos para a concretização do objetivo.

Tabela 2.5: Critérios de qualidade para avaliação dos artigos incluídos na RSL

Contexto da aplicação	Q1.1: Descreve o fim para que é aplicado o modelo?
	Q1.2: O modelo é aplicado a uma amostra contextualizada?
	Q1.3: Descreve o modo de recolha dos dados?
	Q1.4: O construto utilizado para medição está devidamente identificado?
Tipologia dos modelos	Q2.1: Descreve as técnicas de análise utilizadas?
	Q2.2: Compara diferentes modelos de <i>Machine Learning</i> ?
Avaliação dos resultados	Q3.1: Realiza técnicas de sensibilidade e robustez?
	Q3.2: Descreve as limitações do estudo?
	Q3.3: Descreve os contributos do estudo?

2.2. Modelos de *Machine Learning*

2.2.1. Contexto de aplicação

Os diferentes estudos analisados na tabela 2.6, publicados entre 2019 e 2022, decorrem em vários contextos, isto é, em diferentes países, populações em estudo, períodos e com finalidades distintas.

Relativamente à população em estudo, a mais mencionada pela literatura são os estudantes universitários (Stark *et al.*, 2022; Ding *et al.*, 2022; Al-Wesabi *et al.*, 2022; Nayan *et al.*, 2022; Zhang *et al.*, 2021; Nemesure *et al.*, 2021; Wang *et al.*, 2020; Diaz-Godino *et al.*, 2019) uma vez que a perturbação mental é considerada um dos problemas cruciais de saúde pública entre os mesmos e, segundo um estudo de 2021, cerca de 24% dos universitários foram identificados com ideação suicida (Zhang *et al.*, 2021), tornando-se evidente a necessidade de desenvolver estratégias para identificar, de forma precoce, a ansiedade em universitários. O tamanho da amostra destes 20 artigos é composto, em média, por mais de 700 participantes (Bokma *et al.*, 2022; Anbarasi *et al.*, 2022; Al-Wesabi *et al.*, 2022; Nayan *et al.*, 2022; Van Eeden *et al.*, 2021; Wardenaar *et al.*, 2021; Zhang *et al.*, 2021; Nemesure *et al.*, 2021; Hueniken *et al.*, 2021; Wang *et al.*, 2020; Boeke *et al.*, 2020; Diaz-Godino *et al.*, 2019; Gruda *et al.*, 2019) e o período de tempo, na grande maioria, corresponde a um período longitudinal, isto é, que não se refere apenas a um momento, mas sim a um período alargado de recolha de dados (Bokma *et al.*, 2022; Van Eeden *et al.*, 2021; Wardenaar *et al.*, 2021; Hueniken *et al.*, 2021; Wang *et al.*, 2020; Gruda *et al.*, 2019).

No que respeita ao método de recolha de dados, o mais utilizado é através de questionários de autorrelato relativos à saúde mental dos participantes (Stark *et al.*, 2022; Ding *et al.*, 2022; Anbarasi *et al.*, 2022; Al-Wesabi *et al.*, 2022; Muhammad *et al.*, 2022; Nayan *et al.*, 2022; Richter *et al.*, 2021; Wardenaar *et al.*, 2021; Zhang *et al.*, 2021; Yasmin *et al.*, 2021; Hueniken *et al.*, 2021; Wang *et al.*, 2020; Diaz-Godino *et al.*, 2019) pois estes representam uma medida confiável para muitos psicólogos e fornecem uma avaliação precisa da ansiedade humana (Muhammad *et al.*, 2022). Apenas 7 dos 20 artigos estudados recorreram a dados secundários disponibilizados em bases de dados, fazendo com que a base de dados mais utilizada seja a NESDA, um Estudo Holandês de Depressão e Ansiedade (Bokma *et al.*, 2022; Van Eeden *et al.*, 2021; Wardenaar *et al.*, 2021).

Na tabela 2.6 é possível constatar diversos construtos de medição como o *Beck Anxiety Inventory* (BAI), o *Mood States-Short Form* (POMS-SF), o *State-Trait Anxiety Inventory* (STAI), o *Generalized Anxiety Disorder* (GAD), o *Depression, Anxiety and Stress Scale*, (DASS), o *Hamilton Anxiety Measure* (HAM-A) e, por fim, o *Self Rating Anxiety Scale* (SAS). Contudo, no caso específico, o construto de medição de ansiedade com mais ocorrência, utilizado em 6 dos 20 artigos, é o *State-Trait Anxiety Inventory* (STAI) de Spielberger (Ding *et al.*, 2022; Richter *et al.*, 2021; Perpetuini *et al.*, 2021; Boeke *et al.*, 2020; Portugal *et al.*, 2019; Gruda *et al.*, 2019). De seguida, 5 do total dos artigos incluídos utilizou

o questionário *General Anxiety Disorder-7* (Anbarasi, *et al.*, 2022; Zhang *et al.*, 2021; Yasmin *et al.*, 2021).

Quanto à finalidade dos estudos, na sua maioria, os artigos identificam a variável dependente como sendo a ansiedade, o que corrobora com o facto de a principal finalidade dos mesmos ser a mensuração (Anbarasi *et al.*, 2022; Al-Wesabi *et al.*, 2022; Muhammad *et al.*, 2022). Contudo, 9 dos 20 artigos acrescem ainda outros objetivos, como a avaliação de qualidade dos modelos de ML (Bokma *et al.*, 2022) e a identificação, isto é, identificar os atributos com maior importância na previsão da ansiedade (Stark *et al.*, 2022; Diaz-Godino *et al.*, 2019).

Por fim, nos casos em que os artigos não especificavam o contexto foi indicado como “Não Especificado” (N.E).

Tabela 2.6: Contextos de aplicação dos artigos incluídos na RSL

ID	População em estudo	País da amostra	Tamanho da amostra	Período	Modo de recolha	Construto de medição	Finalidade do estudo
1	Pacientes com perturbação de Ansiedade	Holanda	887	2004 - 2007	Base de dados: Estudo Holandês de Depressão e Ansiedade (NESDA)	BAI	Mensuração e Qualidade
2	Estudantes com idade entre os 18 e 36 anos	N.E.	133	N.E.	Questionário de autorrelato e testes clínicos	POMS-SF	Identificação
3	Estudantes sem histórico de saúde mental e física	N.E.	25	N.E.	Questionário de autorrelato	STAI	Mensuração e Qualidade
4	População em geral (aleatória)	Índia	704	30 de agosto a 4 de outubro de 2020	Questionário de autorrelato	GAD-7	Mensuração
5	Estudantes Universitários	N.E.	938	Após o término da situação pandémica COVID-19 (sem data concreta)	Questionário de autorrelato	DASS-21	Mensuração
6	Participantes saudáveis e sem histórico de saúde mental e física	N.E.	23	N.E.	Base de dados: DASPS Questionário de autorrelato e testes clínicos	HAM-A	Mensuração
7	Estudantes Universitários	Bangladesh	2121	Agosto a setembro de 2020	Questionários de autorrelato	GAD-7	Mensuração
8	Participantes com perturbação de Ansiedade e/ou Depressão e saudáveis (18 a 64 anos)	Holanda	2 anos: 2596 4 anos: 2402 6 anos: 2256 9 anos: 2068	2004 - 2007	Base de dados: Estudo Holandês de Depressão e Ansiedade (NESDA)	BAI	Mensuração e Qualidade
9	Participantes com perturbação de Ansiedade e/ou Depressão e saudáveis	Israel (cidade Heifa)	101	N.E.	Questionários de autorrelato e testes clínicos	STAI-T	Mensuração
10	Participantes com idade entre os 20 e 70 anos	N.E.	102	N.E.	Medições fotopletismográficas	STAI-Y	Mensuração
11	Participantes com um diagnóstico de depressão e/ou ansiedade de 6 meses	Holanda	1701	2004 - 2007	Base de dados: Estudo Holandês de Depressão e Ansiedade (NESDA) Questionários de autorrelato	BAI	Mensuração

ID	População em estudo	País da amostra	Tamanho da amostra	Período	Modo de recolha	Construto de medição	Finalidade do estudo
12	Estudantes universitários	China	881	Novembro de 2020	Questionários de autorrelato	GAD-7	Mensuração
13	Participantes com mais de 15 anos, residentes nas 4 províncias do Paquistão	Paquistão	550	10 a 25 de janeiro de 2021	Questionários de autorrelato	GAD-7	Mensuração
14	Estudantes universitários da Universidade de Nice Sophia Antipolis	França	4184	N.E.	Base de dados: Dryad	N.E.	Mensuração
15	Participantes adultos	Canadá	6021	Maio a dezembro de 2020	Questionários de autorrelato	GAD-7	Identificação e Mensuração
16	Alunos universitários, entre os 18 e 22 anos, não graduados	China	1172	1) 15 a 17 de fevereiro de 2020; 2) 15 a 17 de março de 2020	Questionário de autorrelato	SAS ^{*7}	Mensuração e Qualidade
17	Participantes com idades entre 18 e 35 anos	Estados Unidos da América	879	N.E.	Base de dados: <i>Brain Genomics Superstruct Project</i>	STAI	Mensuração e Qualidade
18	Estudantes de Enfermagem da Universidade Privada San Juan Bautista	Perú (cidade Lima)	1193	2018	Questionário de autorrelato	DASS-21	Identificação
19	Participantes diagnosticados com ansiedade e saudáveis, entre os 18 e 25 anos	N.E.	154	N.E.	Base de dados: DIAMOND ^{*7}	STAI + HAM-A	Mensuração
20	Users do Twitter	Estados Unidos da América	1º conjunto: 10 510 tweets de 10 386 users Conjunto final: 3.33 milhões tweets de 1418 users	1º conjunto: 2 dias de Dezembro de 2017 Conjunto final: 18 meses	Microblog (Twitter)	STAI	Mensuração e Identificação

2.2.2. Tipologia dos modelos

Com o objetivo de perceber a tipologia dos modelos analíticos utilizados, a tabela 2.7 sistematiza os 20 artigos incluídos na RSL através da análise do problema em estudo, as técnicas de análise utilizadas e o modelo analítico escolhido que, em alguns casos, é identificado como o modelo com melhor desempenho.

Quanto ao problema em estudo, 10 dos 20 artigos identificam um problema de classificação (Bokma *et al.*, 2022; Stark *et al.*, 2022; Anbarasi *et al.*, 2022) e os restantes um problema de regressão (Ding *et al.*, 2022; Perpetuini *et al.*, 2021; Wardenaar *et al.*, 2021). Contudo, independentemente do problema, a técnica de análise mais utilizada, em 18 dos 20 artigos, são algoritmos de *Machine Learning* (Al-Wesabi *et al.*, 2022; Muhammad *et al.*, 2022; Nayan *et al.*, 2022). No entanto, 7 dos artigos complementam a investigação com a utilização de análises estatísticas (Stark *et al.*, 2022; Perpetuini *et al.*, 2021).

Dos 18 artigos que utilizam os algoritmos de Machine Learning, o modelo analítico mais escolhido é o Random Forest (Bokma *et al.*, 2022; Stark *et al.*, 2022; Muhammad *et al.*, 2022) uma vez que este algoritmo demonstra ter um bom desempenho em diferentes problemas de Machine Learning (Bokma *et al.*, 2022). De seguida, o algoritmo *XG Boost - Extreme Gradient Boosting* - é o algoritmo mais mencionado (Nemesure *et al.*, 2021; Hueniken *et al.*, 2021; Wang *et al.*, 2020). Ambos são considerados algoritmos de árvore de decisão, embora sejam utilizados de forma diferente.

Tabela 2.7: Tipologias de modelos utilizadas na RSL

ID	Problema	Técnicas de análise	Comparação de modelos analíticos	Modelo analítico escolhido
1	Classificação	Algoritmos de ML	Não	<i>Random Forest</i>
2	Classificação	Análises Estatísticas + Algoritmos de ML	Sim	<i>Random Forest</i>
3	Regressão	Algoritmos de ML	Sim	LASSO
4	Classificação	Análises Estatísticas + Algoritmos de ML	Sim	<i>Random Forest</i>
5	Classificação	Algoritmos de ML	Sim	IFSSML-DAS ¹
6	Classificação	Algoritmos de ML	Sim	<i>Random Forest</i>
7	Classificação	Algoritmos de ML	Sim	<i>Support Vector Machine (SVM)</i>
8	Classificação	Algoritmos de ML	Sim	<i>Auto-sklearn</i>
9	Classificação	Algoritmo de ML	Sim	<i>Random Forest</i>
10	Regressão	Análises Estatísticas + Algoritmos de ML	Sim	<i>Generalized Linear Model</i>
11	Regressão	Algoritmos de ML	Sim	Super Learner
12	Regressão	Análises Estatísticas	Não	-
13	Regressão	Análises Estatísticas + Algoritmos de ML	Não	Regressão logística
14	Classificação	Algoritmos de ML	Sim	XG Boost
15	Regressão	Análises Estatísticas + Algoritmos de ML	Sim	XG Boost
16	Regressão	Análises Estatísticas + Algoritmo de ML	Sim	XG Boost
17	Regressão	Algoritmos de ML	Sim	Modelo empilhado ²
18	Classificação	Análises Estatísticas	Não	Regressão Logística
19	Regressão	Análises Estatísticas + Algoritmo de ML	Não	<i>Gaussian Process Regression Model</i>
20	Regressão	Algoritmos de ML	Sim	<i>Bayesian Ridge Regression</i>

¹ IFSSML-DAS: Modelo automatizado de previsão DAS² Modelo empilhado: modelo que agrega três modelos de regressão que utilizam diferentes fontes de dados

2.2.3. Avaliação dos resultados

Como forma de avaliação dos resultados, os artigos incluídos na RSL foram alvo de análise, identificação dos testes de sensibilidade e robustez realizados, limitações identificadas e os contributos da pesquisa de cada um dos artigos.

Posto isto, os artigos incluídos na RSL aplicaram diferentes técnicas de robustez. Dos 20 estudos analisados, 15 artigos recorreram à comparação de modelos analíticos, isto é, através da análise dos valores das métricas de sensibilidade identificaram o modelo analítico que melhor se ajusta ao objetivo do estudo (Stark *et al.*, 2022; Al-Wesabi *et al.*, 2022; Perpetuini *et al.*, 2021). De seguida, a alteração de variáveis independentes foi a técnica mais utilizada, onde, por exemplo, foram retiradas ou acrescentadas novas variáveis à análise (Bokma *et al.*, 2022; Ding *et al.*, 2022) e, por fim, a alteração da amostra, no caso específico, Gruda *et al.* 2019 preferiu aplicar as suas técnicas de análise de dados a uma outra amostra, ou seja, a um novo conjunto de dados, mais concretamente, de *tweets* e utilizadores da plataforma social: o *Twitter*.

Relativamente às métricas de avaliação dos modelos que avaliam a sensibilidade dos modelos, estas diferem consoante o problema de classificação ou de regressão. No caso dos problemas de classificação, as métricas mais utilizadas foram a precisão, a sensibilidade (métrica com a capacidade de avaliar o modelo a identificar, com sucesso, os resultados classificados como positivos) e especificidade (identifica a capacidade do modelo avaliar, com sucesso, os resultados classificados como negativos) e AUC (representa a área que se encontra abaixo da curva ROC, curva esta que mostra o quão bom o modelo é a distinguir as duas classes do target). No que respeita ao problema de regressão, as métricas mais recorrentes para avaliação são o R^2 (medida que representa o quão próximos os dados estão da linha de regressão ajustada, isto é, quanto maior o valor, melhor o modelo se ajusta aos seus dados) e métricas para medir o erro: RMSE (Root-Mean-Square Error) e MAE (Mean Absolute Error).

No que diz respeito às limitações mencionadas nos artigos analisado, a mais referida foram os dados, mais especificamente, o tamanho da amostra (Stark *et al.*, 2022; Ding *et al.*, 2022; Anbarasi *et al.*, 2022), o facto dos dados serem autorrelatados pelos participantes, quando o modo de recolha utilizado são questionários de autorrelato (Richter *et al.*, 2021; Zhang *et al.*, 2021), o período de estudo, isto é, a recolha dos dados ser realizada de forma transversal e não longitudinal (Hueniken *et al.*, 2021; Diaz-Godino *et al.*, 2019), o que não permite recolher conhecimento de eventuais mudanças no nível de ansiedade durante o período estudado (Hueniken *et al.*, 2021). No entanto, Wang *et al.* (2020) utiliza dados longitudinais e refere que a principal limitação é a fidelidade de resposta dos participantes. De mencionar que, apesar de ter sido menos mencionada, uma limitação com grande

impacto nos dados é o estigma e a discriminação relativamente a problemas de saúde mental (Zhang *et al.*, 2021; Gruda *et al.*, 2019).

Por fim, os contributos foram categorizados em duas vertentes: acrescentar uma nova dimensão de análise, ou seja, quando a investigação aborda uma nova perspetiva de análise, em relação à bibliográfica existente (Wesabi *et al.*, 2022; Perpetuini *et al.*, 2021) e o aprofundamento do âmbito, quando o estudo procura explorar questões que já foram abordadas na literatura

Capítulo 3 – Metodologia

Neste capítulo são explicitados os procedimentos metodológicos adotados para o desenvolvimento do presente estudo. Primeiramente, é apresentado o posicionamento da investigação, bem como a definição da amostra e todo o processo de seleção e recolha dos dados. Em segundo, é descrita a metodologia adotada e, por fim, as técnicas de análise dos dados utilizadas.

3.1. Posicionamento da investigação

Com base nos objetivos do presente estudo, optou-se por realizar uma metodologia quantitativa através da elaboração de um questionário. Este foi elaborado na plataforma *LimeSurvey* e divulgado *online* em diferentes canais digitais, como o *LinkedIn*.

Importa salientar que esta abordagem foi conduzida através de uma amostra não probabilística por conveniência, tendo em conta que o pretendido era obter uma amostra com estudantes universitários e trabalhadores portugueses que apresentam ou não sinais de ansiedade clinicamente significativa no período estudado. É igualmente importante de referir que este tipo de amostra apresenta limitações como, por exemplo, a representatividade da população, ou seja, não é possível realizar afirmações gerais e rigorosas sobre a população, uma vez que a amostra está mais sujeita a ser influenciada por variáveis que não estão sob o controlo do investigador, como também não é possível estimar o erro.

3.2. Amostra e dados

3.2.1. Processo de seleção da amostra

A população-alvo da presente investigação é constituída por alunos universitários e trabalhadores portugueses, maiores de idade e com domínio da língua portuguesa, quer sejam ou não diagnosticados por um profissional de saúde com alguma perturbação mental.

Esta escolha incide, principalmente, na revisão de literatura apresentada em que os alunos universitários são os mais mencionados por representarem um grupo vulnerável no que diz respeito à saúde mental. Contudo, o presente estudo pretende avaliar não só possíveis sinais de ansiedade clinicamente significativa em estudantes universitários, mas também em trabalhadores portugueses, uma vez que é um grupo distinto e que, devido às especificidades das diferenças de idade, pode haver diferentes impactos nas diferentes faixas etárias, especialmente considerando o impacto da situação pandémica vivenciada (Antunes *et al.*, 2020).

3.2.2. Caracterização da amostra

A amostra selecionada é composta por 597 respostas, o que equivale a 597 inquiridos, recolhidas entre 30 de julho e 30 de setembro de 2022. A ferramenta de recolha foi um questionário de autorrelato, que se encontra no apêndice A, e através das respostas ao mesmo foi possível realizar uma análise relativa aos dados sociodemográficos.

Como se pode observar na figura 3.1, as respostas foram dadas, de semelhante proporção, por homens e mulheres. Contudo, os dados dizem respeito a 57% mulheres e 43% homens. Neste caso, esta variável pode ser interessante analisar para perceber a influência do género na previsão de níveis de ansiedade e de avaliação do impacto da pandemia COVID-19. De seguida, o estado civil mais predominante é ser solteiro/a, com uma percentagem de cerca 80%, o que se torna mais evidente uma vez que a moda de idades da base de dados é de 22 anos, e segue-se o estado civil de casado, com apenas 12%.

Relativamente às condições de vida do agregado familiar, estas são representadas por uma escala de 1 a 5, em que 1 representa menores condições de vida como, por exemplo, menos estabilidade financeira e menos habilitações literárias do agregado e, por outro lado, o 5 representa melhores condições de vida do mesmo. Como se pode verificar na figura 3.1, o somatório dos níveis abaixo e acima do ponto intermédio, isto é, em que a resposta é 3, são bastante distintas, com 20% (7% + 13%) e 60% (27% + 33%), respetivamente. Logo, pode-se constatar que a grande maioria dos inquiridos tem boas condições de vida.

No que concerne à profissão, isto é, ser estudante, trabalhador/a ou trabalhador/a-estudante, pode-se verificar na figura 3.1 que uma grande parte da população alvo já se encontra a trabalhar, representando quase metade da percentagem total, e as restantes categorias estão distribuídas uniformemente.



Figura 3.1: Caracterização sociodemográfica da amostra em estudo

Relativamente à saúde mental dos respondentes, pode-se observar na figura 3.2 que maior parte dos respondentes não é diagnosticado por um/a profissional de saúde com perturbação de ansiedade, representado uma percentagem de aproximadamente 84%.



Figura 3.2: Caracterização da amostra quanto à existência de diagnóstico de ansiedade

3.3. Construção do questionário

Com base na revisão de literatura realizada, a ferramenta mais utilizada para recolha de dados são os questionários de autorrelato que são validados por múltiplos profissionais de saúde. Segundo Vilelas (2009), a aplicação de um questionário apresenta algumas vantagens no que diz respeito aos custos, tratamento de dados e redução da margem de erro.

A construção do questionário é composta por quatro fases, como representado na figura 3.3. Primeiramente, a elaboração de possíveis questões a incluir, de seguida a discussão com profissionais de saúde e especialistas, segue-se o processo de validação e teste por parte das orientadoras de modo a garantir que o documento não detinha qualquer erro e, por fim, a disponibilização do questionário final. De salientar que todo o processo de construção da presente ferramenta de recolha de dados teve como base as melhores práticas e ética definidas pelo Código de Conduta do ISCTE.



Figura 3.3: Processo de construção do questionário

O documento elaborado foi dividido em quatro secções, como se verifica no questionário em apêndice, no apêndice A. Na primeira secção, após ser aceite o consentimento informado, constam as perguntas relativas à caracterização do inquirido como a idade, género, situação económica do agregado familiar, situação profissional atual, entre outras. As secções dois e três dizem respeito a informações relativas a estudantes universitários e/ou trabalhadores/as portugueses/as, respetivamente. Por fim, a última secção diz respeito à saúde mental dos respondentes, isto é, inclui perguntas como a existência de algum diagnóstico por profissionais de saúde de quaisquer perturbações de saúde mental e, especificamente, de perturbações de ansiedade. Estas questões têm como objetivo conhecer o histórico do inquirido relativamente à saúde mental e, posteriormente, através da adaptação do questionário *Beck Anxiety Inventory* (BAI), utilizando escalas tipo *Likert* de quatro níveis, a avaliação do impacto da pandemia COVID-19 no estado de saúde mental.

A escolha da adaptação do questionário BAI decorre da discussão com profissionais de saúde sobre a eficácia da sua utilização na definição de sintomas e níveis de ansiedade.

O principal objetivo deste estudo é avaliar experiências de ansiedade que estudantes universitários/as e trabalhadores/as possam ter vivenciado nos períodos seguintes: durante e períodos subsequentes à pandemia COVID-19.

Assim, as questões elaboradas no presente inquérito têm em consideração três momentos de avaliação:

Momento 1: Período pandémico, isto é, desde que surgiu a Doença do coronavírus (COVID-19) e foram aplicadas medidas de distanciamento social com vista à sua contenção, como o confinamento;

Momento 2: Período pós pandemia COVID-19, ou seja, a fase inicial de desconfinamento com restrições impostas, como o uso de máscara, elaboração de testes PCR como forma de acesso a espaços, por exemplo, entre outras;

Momento 3: Momento atual, isto é, fase em que são retiradas a maior parte das restrições (fase de preenchimento do inquérito).

3.4. Metodologia CRISP-DM

Como definido anteriormente, a metodologia adotada na presente investigação é o CRISP-DM. Esta metodologia é composta por seis fases, contudo neste estudo não é abordada a última etapa, a implementação, uma vez que esta pressupõe a aplicação prática das fases anteriores.

Assim, a figura 3.4 resume o que foi feito em cada uma das fases, no caso concreto nas cinco fases do CRISP-DM. Primeiramente, a compreensão do negócio, em que é realizada uma revisão sistemática da literatura para perceber como é que os modelos de *Machine Learning* são utilizados na

saúde mental. Posteriormente, a compreensão dos dados passa por estudar as variáveis presentes na base de dados, isto é, as variáveis independentes e dependentes.

Posto isto, na terceira fase, segue-se a preparação dos dados onde há a criação, eliminação e transformação de dados, bem como identificação de *outliers* e valores omissos. Após a realização das alterações necessárias, procede-se à verificação do *type* das variáveis e apresentação do conjunto final de dados para utilizar na fase que se segue, a modelação.

De seguida, a modelação começa pela explicação dos modelos criados segundo a questão de investigação e objetivos estabelecidos. Contudo, antes da construção dos modelos é necessário verificar o balanceamento da variável que se pretende prever, o *target*, e, por isso, procedeu-se ao balanceamento das variáveis dependentes, uma vez que são utilizadas duas para diferentes modelos, e à partição dos modelos em duas amostras, de treino e teste. De seguida, são explicadas as métricas e algoritmos utilizados na modelação, e definidas as variáveis *input* a utilizar em cada um dos modelos. Assim, é possível começar a modelar.

Por fim, na avaliação é realizada a comparação de resultados obtidos através das métricas e são identificados os modelos com melhor qualidade.

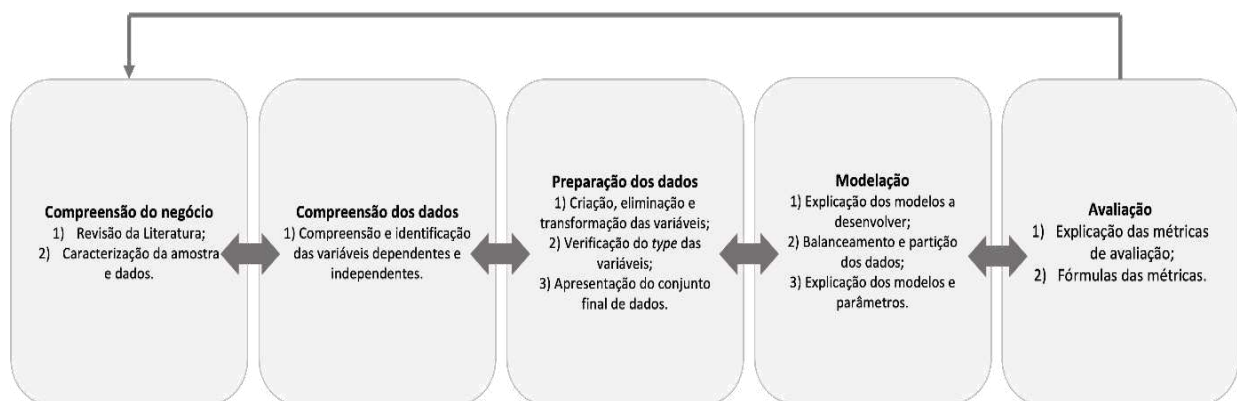


Figura 3.4: Aplicação da metodologia CRISP-DM

3.4.1 Compreensão dos dados

Após a fase da Compreensão do Negócio, a primeira fase da metodologia CRISP-DM, segue-se a Compreensão dos Dados. Nesta etapa é realizada uma breve compreensão das variáveis presentes na base de dados em estudo, desde todo o processo de recolha de dados, à sua compreensão e caracterização.

Primeiramente, os dados recolhidos pelo questionário disponibilizado foram extraídos em formato CSV da plataforma de recolha, o *LimeSurvey*, e foram utilizadas as ferramentas *Microsoft Excel* para visualização e o *SPSS Modeler* para o tratamento inicial dos dados.

Assim, verificou-se que a presente base de dados contém respostas de 1763 inquiridos, em que apenas 639 submissões (36%) estavam completas, isto é, continham resposta à totalidade de perguntas realizadas no decorrer do questionário. No entanto, a pesquisa destina-se a estudantes e trabalhadores/as portugueses/as e, por isso, foram eliminadas 16 respostas em que a nacionalidade é diferente de 'Portuguesa' e 26 respostas de inquiridos que não são estudantes universitários nem trabalhadores/as. Deste modo, a amostra final contém dados relativos a 597 respondentes.

Estas respostas são avaliadas com base em 30 variáveis independentes (*input*). Desta forma, para uma melhor compreensão das variáveis em estudo, segue-se uma explicação mais pormenorizada de cada conjunto de variáveis, como observado nas tabelas 3.1, 3.2 e 3.3.

Tabela 3.1: Compreensão das variáveis independentes presentes na base de dados [Sociodemográficas]

Variável	Descrição	Observações
Características Sociodemográficas		
Idade	Idade do inquirido.	[18,77]
Género	Género com que o indivíduo se identifica	"Homem", "Mulher", "Outro"
Nacionalidade	Nacionalidade do inquirido.	"Portuguesa", "Outra"
Estado_Civil	Estado civil do inquirido.	"Solteiro/a", "Unido/a de facto", "Casado/a", "Divorciado/a", "Viúvo/a"
Dependentes	Dependentes a cargo do inquirido.	"Sim", "Não"
Agregado_Familiar	Agregado com que o indivíduo vive.	"Sozinho/a", "Ambos os pais", "Um dos pais", "Irmãos/as", "Amigos/as", "Outros/as estudantes universitários", "Cônjuge ou companheiro/a", "Filhos/as", "Outro"
Distrito	Local de residência do inquirido.	
Cond_vida_agreg_familiar	Nível de condições de vida do agregado familiar do inquirido	[1,5] ³
Profissão	Profissão do inquirido (Trabalhador)	"Não", "Sim, a tempo parcial", "Sim, a tempo inteiro"
Estudante	Profissão do inquirido (Estudante)	"Sim", "Não"

³condições de vida do agregado familiar com base numa escala de 1 a 5: 1 representa um agregado com menores condições (menos estabilidade financeira, menos estudos) e 5 representa um agregado com melhores condições

Tabela 3.2: Compreensão das variáveis independentes presentes na base de dados [Estudante]

Variável	Descrição	Observações
Características da profissão [Estudante]		
Ensino	Tipo de ensino que o inquirido frequenta.	“Ensino Universitário Público”, “Ensino Universitário Privado”, “Ensino Politécnico”, “Outro”
Area_Estudos	Área de estudos que o inquirido frequenta.	“Artes”, “Línguas e Humanidades”, “Informática”, “Saúde”, “Serviços Sociais”, (...)
Ano_Ingresso	Ano em que ingressou a faculdade que atualmente frequenta.	
Ciclo_Estudos	Ciclo de estudos que frequenta.	“Licenciatura”, “Pós-graduação”, “Mestrado”, “Mestrado Integrado”, “Doutoramento”
Ano_Estudos	Ano curricular que atualmente frequenta.	“1º ano”, “2º ano”, “3º ano”, “4º ano (se aplicável ao curso que frequenta)”, “5º ano (se aplicável ao curso que frequenta)”
Estudante_Finalista	Último ano curricular do inquirido.	“Sim”, “Não”
Estudante_Deslocado	Estatuto atribuído ao inquirido que estuda fora da sua zona de residência.	“Sim”, “Não”
Distância_Casa_Univ	Distância a que o inquirido, designado como estudante deslocado, se encontra da sua habitação (casa).	
Comparação1e2_Resultados	Comparação dos resultados obtidos durante a pandemia COVID-19 e o período inicial de desconfinamento.	“Obtive melhores resultados”, “Mantive os mesmos resultados”, “Pioriei significativamente os resultados”
Comparação2e3_Resultados	Comparação dos resultados obtidos durante o período inicial de desconfinamento e atualmente.	“Obtive melhores resultados”, “Mantive os mesmos resultados”, “Pioriei significativamente os resultados”

Tabela 3.3: Compreensão das variáveis independentes presentes na base de dados [Trabalhador]

Variável	Descrição	Observações
Características da profissão [Trabalhador]		
Area_Trabalho	Área de trabalho que o inquirido frequenta.	“Artes”, “Línguas e Humanidades”, “Informática”, “Saúde”, (...)
Situação_Profissional	Situação profissional do inquirido.	“Trabalhador/a por conta de outrem”, “Trabalhador/a por conta própria ou Empregador/a”, “Trabalhador/a familiar não remunerado”
Anos_Trabalho	Número de anos que o inquirido exerce a função que ocupa atualmente	“Menos de 1 ano”, “Entre 1 e 5 anos”, “Entre 5 a 10 anos”, “Mais de 10 anos”
Vínculo_Trabalhador	Vínculo do inquirido.	“Contrato permanente”, “Contrato a termo certo”, “Contrato a termo incerto”, “Colocado por empresas de trabalho temporário”, “Contrato de prestação de serviços e a recibos verdes”, “Outro vínculo”
Comparação1e2_Work_life_balance	Comparação do equilíbrio entre a família e o trabalho durante o período inicial de desconfinamento e atualmente.	[1, 5] NOTA: 1 representa grande dificuldade em conciliar; 5 representa pouca dificuldade em conciliar
Comparação2e3_Work_life_balance	Comparação do equilíbrio entre a família e o trabalho durante o período inicial de desconfinamento e atualmente.	[1, 5]
Características de Saúde Mental		
Perturbação_Mental	Inquirido diagnosticado com uma perturbação de saúde mental por um/a profissional de saúde.	“Depressão”, “Perturbação de Ansiedade”, “Perturbação Bipolar”, “Perturbação do Comportamento Alimentar”, “Esquizofrenia”, “Perturbação de personalidade de Borderline”, “Não diagnosticado/a”, “Outro diagnóstico”
Perturbação_Ansiedade	Inquirido diagnosticado com ansiedade por um/a profissional de saúde.	“Ansiedade generalizada”, “Fobia social”, “Perturbação de Pânico”, “Fobias”, “Perturbação de Stress Pós-Traumático”, “Perturbação Obsessivo-Compulsiva”, “Não diagnosticado/a”, “Outro diagnóstico”
Autorrelato_ansioso	Autorrelato do inquirido quanto à ansiedade.	“Sim, grande parte do tempo”, “Sim, apenas em situações de stress”, “Não”
Acompanhamento	Procura de acompanhamento durante o período em estudo.	“Sim”, “Não”

Relativamente às variáveis de *input* presentes na base de dados, é necessário determinar um subconjunto de variáveis que melhor explique a variável dependente. Isto é, dentro de todas as variáveis *input* disponíveis, deve ser encontrado um subconjunto de variáveis relevantes para o modelo.

Importa salientar que, no que diz respeito à variável dependente, o “*Target*”, este foi definido posteriormente na Preparação dos Dados, a terceira fase da metodologia. Contudo, as variáveis utilizadas para a definição da mesma foram com base no questionário BAI e, por isso, importa realizar uma análise teórica das mesmas, como representada na tabela 3.4.

Tabela 3.4: Compreensão das variáveis presentes na base de dados [Saúde Mental]

Variável	Descrição	Observações
Características de Saúde Mental [BAI]		
Comparação_BAI_1e2	Comparativamente ao período pandémico, o quanto é que o/a inquirido/a se sentiu incomodado/a com os sintomas de ansiedade no período inicial de desconfinamento.	“(0) De modo algum”, “(1) Ligeiramente, mas não me incomodou muito”, “(2) Moderadamente, por vezes não era agradável, mas pude suportar”, “(3) Severamente, incomodava-me muito e difícil de suportar”
Comparação_BAI_2e3	Comparativamente ao período inicial de desconfinamento, o/a quanto é que o inquirido/a se sentiu incomodado/a com os sintomas de ansiedade atualmente.	“(0) De modo algum”, “(1) Ligeiramente, mas não me incomodou muito”, “(2) Moderadamente, por vezes não era agradável, mas pude suportar”, “(3) Severamente, incomodava-me muito e difícil de suportar”

3.4.2 Preparação dos dados

A terceira fase da metodologia consiste na preparação e validação dos dados para que seja possível fazer uma análise concreta e, posteriormente, retirar conclusões fidedignas. Para isto, foi necessário realizar algumas alterações, nomeadamente recodificações, agrupamentos e criação/eliminação de variáveis. A ferramenta de *software* escolhida para realizar estas atividades foi o *SPSS Modeler*.

Relativamente às variáveis originais, foi verificado através do nó ‘*Type*’ no *SPSS Modeler* que se encontravam corretamente classificadas, no que diz respeito ao seu tipo: *nominal*, *ordinal*, *continuous* ou *flag*.

De seguida, a variável ‘Distrito’ foi recodificada numa nova variável do tipo nominal, designada por ‘Regiões’. Esta está dividida em “Norte”, “Centro”, “Sul” e “Ilhas”.

No passo seguinte foi realizada a recodificação da variável ‘Idade’ para ‘Idade_TILEN’, isto é, foram criados escalões etários com cinco níveis. Desta forma, esta transformação altera o tipo da variável de

continuous para ordinal. Este processo é fulcral para perceber de que forma a idade pode influenciar o *target*, ou seja, se algum escalão específico se diferencia em relação aos restantes por ter uma maior ou menor probabilidade de classificar o *target*.

Relativamente às variáveis que dizem respeito ao questionário adaptado BAI, ou seja, às variáveis relativas à saúde mental, foi realizada a reclassificação das mesmas para “0”, “1”, “2” e “3”. De seguida, procedeu-se à alteração do tipo destas variáveis para inteiro com o objetivo de se poder realizar somatórios e, posteriormente, classificar a pontuação com os diferentes níveis de ansiedade, segundo a pontuação definida pelo BAI. Desta forma, foram criadas as seguintes variáveis ‘Pontuação_BAI_1e2’ e ‘Pontuação_BAI_2e3’ como variáveis do tipo *continuous*. Para uma melhor compreensão futura das mesmas, foi realizada novamente uma nova reclassificação de forma a obter variáveis do tipo nominal, às quais foi acrescentado o sufixo *_classif*. Esta está dividida em “Grau mínimo de ansiedade”, “Ansiedade leve”, “Ansiedade Moderada” e “Ansiedade Severa”. Estas alterações são importantes para definir o *target* do primeiro modelo que tem como finalidade classificar o nível de ansiedade dos respondentes.

Posteriormente, foi criada a variável ‘Variação_BAI’ que concerne à variação das pontuações de ansiedade entre as variáveis ‘Pontuação_BAI_1e2’ e ‘Pontuação_BAI_2e3’. Esta variável é considerada, à partida, útil para a análise uma vez que permite identificar se há uma variação positiva, negativa ou nula, ou seja, permite identificar se a pandemia COVID-19 teve influência no nível de ansiedade sentido pelos inquiridos nos períodos em estudo. De seguida, a diferença entre as variáveis foi classificada como “Melhoria”, “Agravamento” e “Constante”, e foi acrescentado o sufixo *_classif* à variável das pontuações.

Por fim, foram criadas duas variáveis designadas como ‘Diagnostico_P_Mental’ e ‘Diagnostico_P_Ansiedade’ que identificam se o inquirido é diagnosticado com pelo menos uma das perturbações associadas, isto é, corresponde ao histórico de saúde mental do respondente, qualquer que seja o diagnóstico associado.

Relativamente à eliminação de variáveis, optou-se por excluir seis variáveis: ‘Nacionalidade’, ‘Ano de Ingresso’, ‘Distância_Casa_Univ’, ‘Perturbação_Mental’, ‘Perturbação_Ansiedade’ e ‘Autorrelato_Ansioso’, ‘Acompanhamento’ uma vez que não são úteis para a análise. No caso da terceira e da última variáveis referidas, isto é ‘Distância_Casa_Univ’ e ‘Acompanhamento’, por apresentarem maioritariamente não respostas que podem induzir a interpretações ambíguas e a variável de autorrelato uma vez que está relacionada com a variável dependente, o que pode enviesar o estudo. Todas as alterações anteriormente mencionadas podem ser consultadas no apêndice B.

Após todas as alterações nas variáveis *input*, importa identificar o *target* da presente investigação. Com base na questão de investigação e objetivos definidos, são definidas duas variáveis dependentes:

(1) 'Pontuação_BAI_1e2_classif' → *Target* (type: categórico); (2) 'Variação_BAI_classif' → *Target* (type: categórico)

A primeira variável dependente tem como objetivo prever os níveis de ansiedade nos alunos universitários e trabalhadores portugueses, através da definição do nível de ansiedade. No que diz respeito à segunda variável dependente, esta tem como propósito avaliar o impacto da pandemia COVID-19 na saúde mental dos mesmos. Para isto, foi desenvolvido mais do que um modelo, como poderemos observar na figura 3.5, na próxima etapa do CRISP-DM.

Assim, após a limpeza e tratamento dos dados, obteve-se um conjunto de dados composto por 25 variáveis independentes, como se pode observar na tabela 3.5.

Tabela 3.5: Conjunto final de variáveis independentes

Variável	Descrição	Observações
Características Demográficas		
Idade_TILEN	Escalões etários	1: [18, 22[2: [22, 23[3: [23, 25[4: [25, 35[5: [35, 77]
Género	Género com que o indivíduo se identifica	
Estado_Civil	Estado civil do inquirido.	
Dependentes	Dependentes a cargo do inquirido.	
Agregado_Familiar	Agregado com que o indivíduo vive.	
Regiões	Local de residência do inquirido por regiões.	"Norte", "Centro", "Sul"
Cond_vida_agreg_familiar	Nível de condições de vida do agregado familiar do inquirido	
Profissão	Profissão do inquirido (Trabalhador)	
Estudante	Profissão do inquirido (Estudante)	
Características da profissão [Estudante]		
Ensino	Tipo de ensino que o inquirido frequenta.	
Area_Estudos	Área de estudos que o inquirido frequenta.	
Ciclo_Estudos	Ciclo de estudos que frequenta.	
Ano_Estudos	Ano curricular que atualmente frequenta.	
Estudante_Finalista	Último ano curricular do inquirido.	
Estudante_Deslocado	Estatuto atribuído ao inquirido que estuda fora da sua zona de residência.	
Comparação1e2_Resultados	Comparação dos resultados obtidos durante a pandemia COVID-19 e o período inicial de desconfinamento.	
Comparação2e3_Resultados	Comparação dos resultados obtidos durante o período inicial de desconfinamento e atualmente.	

Variável	Descrição	Observações
Características da profissão [Trabalhador]		
Area_Trabalho	Área de trabalho que o inquirido frequenta.	
Situação_Profissional	Situação profissional do inquirido.	
Anos_Trabalho	Número de anos que o inquirido exerce a função que ocupa atualmente	
Vínculo_Trabalhador	Vínculo do inquirido.	
Comparação1e2_Work_life_balance	Comparação do equilíbrio entre a família e o trabalho durante o período inicial de desconfinamento e atualmente.	
Comparação2e3_Work_life_balance	Comparação do equilíbrio entre a família e o trabalho durante o período inicial de desconfinamento e atualmente.	
Características de Saúde Mental		
Diagnostico_P_Mental	Inquirido diagnosticado com pelo menos uma perturbação de saúde mental por um/a profissional de saúde.	1: ≥ 1 diagnóstico 0: não diagnosticado
Diagnostico_P_Ansiedade	Inquirido diagnosticado com pelo menos uma perturbação de ansiedade por um/a profissional de saúde.	1: ≥ 1 diagnóstico 0: não diagnosticado

3.4.3 Modelação e avaliação

Concluídas as fases anteriores da metodologia CRISP-DM, a presente etapa consiste na seleção de vários algoritmos de *Machine Learning*. No caso concreto, na seleção de algoritmos de classificação, dos quais as árvores de decisão, redes neurais e regressão logística são exemplos.

Como mencionado na fase anterior e ilustrado na figura 3.5, foi desenvolvido mais do que um modelo final. O objetivo da criação de dois modelos de classificação é perceber qual apresenta melhores resultados na previsão de níveis de ansiedade e avaliação do impacto da pandemia na saúde mental da população-alvo.

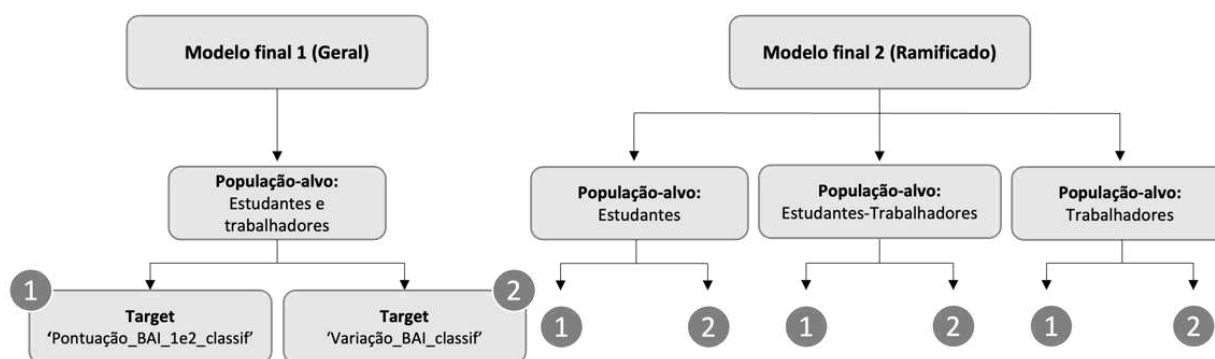


Figura 3.5: Modelos Finais (Geral e ramificado)

Assim, em primeiro lugar, é necessário perceber se as variáveis “*target*” estão balanceadas. Para isto, foi utilizado o nó *Distribution*, do separador *Graphs*, no *SPSS Modeler*. Verificou-se nesta etapa inicial que as variáveis dependentes não se encontram balanceadas, pelo contrário, existe uma grande discrepância entre as categorias. Deste modo, optou-se por fazer o balanceamento de forma que todas as categorias representem, aproximadamente, 25% (*target*: ‘Pontuação_BAI_1e2_classif’) e 33,33% (*target*: ‘Variação_BAI_classif’). Para isso, foi utilizada a opção *Balance Node (boost)* que leva ao aumento da ocorrência de categorias pouco frequentes até que seja corrigido o desequilíbrio nos dados.

De seguida, procedeu-se à partição dos dados através do nó *Partition*, ou seja, a amostra original é dividida em duas. Neste caso, 70% para treino (constituído por dados com os quais o modelo será construído) e 30% para teste (constituído por dados com os quais o modelo será testado).

A avaliação da qualidade dos modelos teve em consideração diversas métricas, sendo que duas das mencionadas são baseadas na matriz de classificação. No presente estudo recorreu-se à percentagem de casos corretamente classificados, designado por PCCC, à sensibilidade, que traduz a capacidade do modelo identificar, com sucesso, os resultados classificados como positivos, e à especificidade, que identifica a capacidade do modelo identificar, com sucesso, os resultados classificados como negativos. A tabela 3.6 resume as fórmulas utilizadas para o cálculo das mesmas.

Tabela 3.6: Métricas e fórmulas utilizadas para avaliação da performance dos modelos

Métrica	Fórmula
PCCC	$PCCC = \frac{VP^4 + VN^5}{VP + VN + FP^6 + FN^7}$
Sensibilidade	$Sensibilidade = \frac{VP}{VP + FP}$
Especificidade	$Especificidade = \frac{VN}{VN + FP}$

Para a criação de todos os modelos, foi realizado um estudo dos resultados das métricas para prever cada uma das categorias da variável dependente, o *Target*. No caso concreto, sendo que este é categórico, as métricas foram calculadas por categoria e não para o modelo no geral, exceto o valor

⁴ Verdadeiro Positivo

⁵ Verdadeiro Negativo

⁶ Falso Positivo

⁷ Falso Negativo

de PCCC. Os resultados das mesmas acabam por ser mais elevados por serem mais específicos, isto é, calculados por categorias.

O cálculo da diferença entre as métricas Sensibilidade e Especificidade explicita a estabilidade entre as mesmas, isto é, quanto menor a diferença, melhor a estabilidade.

Árvores de decisão

O presente estudo recorreu apenas a árvores de decisão uma vez que não possuem pressupostos, são de rápida resposta e facilmente interpretáveis e, simultaneamente, por serem os que apresentam melhores resultados e serem os mais mencionados segundo a Revisão Sistemática da Literatura realizada. Há diversos exemplos de árvores de decisão, contudo neste trabalho são destacados os seguintes: *Random Tree*, C5.0 e CHAID.

O processo de criação dos modelos passou por, numa primeira fase, construir diversos modelos para cada algoritmo, mantendo as variáveis independentes. De seguida, após observar os resultados das métricas, foram realizadas alterações nos parâmetros dos mesmos. Por fim, foi selecionado o modelo com maior qualidade para prever os níveis de ansiedade e avaliar o impacto da pandemia na saúde mental em estudantes universitários e trabalhadores/as portugueses/as.

Para todos os modelos (estudantes, estudantes-trabalhadores, trabalhadores/as) foram utilizados os mesmos algoritmos, mantendo todas as alterações nas parametrizações, para que fosse mais fácil interpretar e comparar os resultados dos mesmos. Assim, a tabela 3.7 resume os modelos e parametrizações realizadas.

Os modelos obtidos através de algoritmos podem ser construídos com recurso a uma única árvore (*single tree*), como se encontra por *default* no *SPSS Modeler*. Contudo, de forma a melhorar o desempenho dos mesmos são realizadas alterações nas parametrizações. Com recurso à técnica *bagging* é possível melhorar a capacidade de generalização dos modelos através do aumento da estabilidade, ou com recurso à técnica *boosting* que pretende melhorar a precisão do modelo na fase de treino. Para além disso, procura-se fixar limitações ao crescimento das árvores de forma a evitar o sobrejustamento dos dados de treino dos modelos, através dos parâmetros número mínimo de casos por Nó Pai e número mínimo de casos por Nó filho.

Tabela 3.7: Modelos e parametrizações

Modelos	Algoritmos		
	Random Trees	C5.0	CHAID
A	Default (10 depth ⁸ MCNS* ⁹ 5)	-	-
B	15 depth MCNS 1	-	-
C	20 depth MCNS 1	-	-
D	10 depth MCNS 1	-	-
E	-	Default (Cross validate =10)	-
F	-	Ensemble: Boosting	-
G	-	-	Default (depth* ¹ = 5)
H	-	-	Depth* ¹ : 10
I	-	-	Ensemble: Boosting
J	-	-	Ensemble: Bagging

⁸ Maximum tree depth⁹ Minimum child node size

Capítulo 4 – Resultados e Discussão

Este capítulo visa dar sequência à metodologia abordada no capítulo anterior, evidenciando os resultados obtidos segundo as análises efetuadas e a discussão dos mesmos, tendo em consideração os objetivos estabelecidos.

De seguida, são apresentados os resultados do modelo 2, o modelo ramificado, e a discussão dos mesmos. A escolha do modelo incide no facto de se verificar que o modelo ramificado é o que apresenta melhores métricas de um modo geral. Por isto, optou-se por apresentar de forma mais pormenorizada este modelo e os resultados do modelo 1 encontram-se em apêndice, nos apêndices O e P (fluxogramas do modelo) e Q (os resultados do modelo). De salientar que se prevê que haja uma maior explicabilidade do modelo ramificado uma vez que as variáveis independentes são mais específicas a cada grupo do que no modelo geral.

Em apêndices, do apêndice C a N, encontram-se os fluxogramas e resultados do modelo ramificado para cada um dos grupos.

4.1. Modelo preditivo para previsão de níveis de ansiedade

4.1.1. Modelo final 2 [Ramificado]: Estudantes

Primeiramente, começou-se por analisar quais as variáveis independentes que poderão acrescentar mais valor ao modelo e, para isso, foram construídos vários gráficos de forma a perceber empiricamente a influência que cada variável pode ter ou não no *target*, como se pode observar no fluxograma no apêndice C.

Assim, conclui-se que as variáveis *input* consideradas para análise do modelo relativo apenas a estudantes universitários são: 'Estado_Civil', 'Dependentes', 'Cond_Vida_agreg_familiar', 'Ciclo_Estudos', 'Estudante_Deslocado', 'Regiões', 'Género', 'Agregado_familiar', 'Idade_TILEN', 'Ensino' e 'Area_Estudos'.

Como se pode observar na tabela no apêndice D, como na tabela 4.1, os modelos não apresentam problemas de sobre ajustamento, isto é, acontece quando o modelo faz muitas boas previsões para a amostra de treino e, para a amostra de teste apresenta uma descida considerável na sua *performance*. Contudo, alguns modelos mostram menor estabilidade entre as métricas.

Na generalidade, os modelos apresentam bons resultados para prever os níveis de ansiedade em estudantes universitários, contudo concluiu-se que os modelos que apresentam maior qualidade são:

1) Modelo D (RT 10 depth MCNS 1), uma vez que apresenta não só os melhores valores de PCCC, como também uma melhor diferença entre a previsão para a amostra de treino e teste, e uma boa estabilidade entre as métricas (S-E);

2) Modelo F (C5.0 *Boosting*), pois apresenta melhores métricas em termos de PCCC e estabilidade para todas as categorias;

3) Modelo I (CHAID *Boosting*) por apresentar melhores resultados em todas as métricas, comparativamente aos restantes.

Na tabela 4.1 estão representados, de uma forma resumida, os resultados dos modelos anteriormente selecionados e apura-se o modelo com maior qualidade para prever os níveis de ansiedade dos estudantes.

Tabela 4.1: Resultados dos modelos com melhor qualidade para previsão de níveis de ansiedade em estudantes

Melhores Modelos															
				Grau mínimo			Ansiedade leve			Ansiedade moderada			Ansiedade severa		
PCCC				S	E	S - E	S	E	S - E	S	E	S - E	S	E	S - E
Random Trees	D	Treino	93,5%	94,2%	94,3%	-0,1%	94,2%	98,1%	-3,9%	89,3%	100%	-10,7%	96,9%	99,1%	-2,1%
		Teste	90,2%	90%	93,8%	-3,8%	91,2%	98%	-6,8%	84,4%	100%	-15,6%	97,1%	96%	1,1%
C5.0	F	Treino	96,7%	96,9%	96,7%	0,2%	95,8%	100%	-4,2%	94,4%	99%	-4,6%	100%	100%	0%
		Teste	92,4%	92,6%	92,4%	0,2%	87,2%	100%	-12,8%	96,8%	98%	-1,2%	94,3%	100%	-5,7%
CHAID	I	Treino	92,9%	86,8%	95,1%	-8,3%	97,1%	96,2%	0,8%	93,1%	99%	-6%	95,4%	100%	-4,6%
		Teste	92,5%	83,3%	94,5%	-11,2%	93,3%	99%	-5,7%	89,7%	96,8%	-7,1%	100%	100%	0%

Os três modelos apresentam métricas muito semelhantes, sendo que o modelo F é o que apresenta melhores resultados de PCCC. Contudo, é importante perceber qual é que prevê de forma mais equilibrada os quatro níveis de ansiedade.

Assim, avaliando os quatro níveis, descarta-se de imediato o algoritmo D uma vez que apresenta uma discrepância muito elevada entre as métricas sensibilidade e especificidade na previsão do nível de “Ansiedade moderada” (16% aproximadamente).

Relativamente aos restantes modelos, o modelo F apresenta melhores resultados na previsão de todos os níveis. Assim, este é o modelo escolhido para prever os níveis de ansiedade em estudantes universitários.

4.1.2. Modelo final 2 [Ramificado]: Trabalhadores-estudantes

Analisado o primeiro modelo que respeita aos estudantes/as portugueses/as, segue-se o estudo relativo a trabalhadores-estudantes. Para esta análise conclui-se que as variáveis independentes consideradas são: ‘Estado_Civil’, ‘Dependentes’, ‘Cond_Vida_agreg_familiar’, ‘Regiões’, ‘Género’,

‘Agregado_familiar’, ‘Idade_TILEN’, ‘Ciclo_Estudos’, ‘Estudante_Deslocado’, ‘Ensino’, ‘Area_Estudos’, ‘Anos_trabalho’, ‘Area_trabalho’, ‘Situação_profissional’, ‘Vínculo_trabalhador’ e ‘Profissão’.

Como se pode observar na tabela presente no apêndice H e na tabela 4.2, nenhum dos modelos se encontra em sobre ajustamento.

Na generalidade, os modelos apresentam ótimos resultados para prever os níveis de ansiedade em trabalhadores-estudantes. Assim, concluiu-se que os modelos que apresentam maior qualidade são:

1) Modelo C (RT 20 depth MCNS 1): apesar de não ser o modelo que apresenta melhor valor de PCCC, é o modelo que tem uma menor diferença entre o PCCC de treino e teste, e que, simultaneamente, apresenta bons valores em termos de estabilidade do modelo;

2) Modelo F (C5.0 *Boosting*): é o modelo que apresenta melhores resultados em todas as métricas.

3) Modelo I (CHAID *Boosting*): o modelo apresenta as melhores métricas em termos de PCCC, sensibilidade, especificidade e estabilidade.

Tabela 4.2: Resultados dos modelos com melhor qualidade para previsão de níveis de ansiedade em trabalhadores-estudantes

Melhores Modelos															
				Grau mínimo			Ansiedade leve			Ansiedade moderada			Ansiedade severa		
PCCC				S	E	S - E	S	E	S - E	S	E	S - E	S	E	S - E
Random Trees	C	Treino	96,5%	100%	95,5%	4,5%	96%	100%	-4%	97,2%	100%	-2,8%	93,5%	100%	-6,5%
		Teste	95,6%	100%	94,6%	5,4%	88,6%	100%	-11,4%	97,4%	100%	-2,6%	97,3%	100%	-2,7%
C5.0	F	Treino	97,6%	100%	97%	3%	97,6%	100%	-2,4%	97,2%	100%	-2,8%	96%	100%	-4%
		Teste	94,9%	100%	93,4%	6,6%	86,2%	100%	-13,8%	92,7%	100%	-7,3%	100%	100%	0%
CHAID	I	Treino	96,9%	100%	96%	4%	93,7%	100%	-6,3%	94,4%	100%	-5,6%	100%	100%	0%
		Teste	95,7%	100%	94,7%	5,3%	97,3%	100%	-2,7%	95%	100%	-5%	91,7%	100%	-8,3%

Primeiramente, opta-se por excluir o modelo F por ser o que apresenta menor estabilidade na previsão de três categorias: “Grau mínimo de ansiedade” com cerca de 7% de diferença, “Ansiedade leve” com aproximadamente 14% e “Ansiedade moderada” com 7,3%.

Os restantes modelos foram devidamente analisados e conclui-se que o melhor modelo para a previsão de níveis de ansiedade em trabalhadores-estudantes é o modelo C. A PCCC é relativamente equivalente ao modelo I, contudo o modelo C apresenta melhor estabilidade em todos os níveis de

ansiedade o que é uma mais-valia uma vez que se quer optar por um modelo que seja capaz de classificar cada uma das categorias com os menores erros possíveis.

4.1.3. Modelo final 2 [Ramificado]: Trabalhadores/as

Por fim, é analisado o modelo relativo aos trabalhadores/as portugueses/as. Segundo a análise realizada, as variáveis consideradas para análise do modelo relativo apenas a trabalhadores/as são: 'Estado_Civil', 'Dependentes', 'Cond_Vida_agreg_familiar', 'Regiões', 'Género', 'Agregado_familiar', 'Idade_TILEN', 'Anos_trabalho', 'Area_trabalho', 'Situação_profissional' e 'Vínculo_trabalhador'.

Como consta na tabela no apêndice L e na tabela 4.3, nenhum dos modelos de previsão de níveis de ansiedade em trabalhadores/as apresenta sobre ajustamento, pelo contrário. Relativamente aos modelos, os que apresentam melhor qualidade são:

1) Modelo C (RT 20 *depth* MCNS 1): é o modelo que apresenta os melhores valores de PCCC, em simultâneo tem uma boa estabilidade entre as métricas e ótimos resultados na previsão de todas as categorias do *target*;

2) Modelo F (C5.0 *Boosting*): apesar do modelo apresentar uma maior instabilidade na previsão do “Grau mínimo de ansiedade”, os valores das métricas, sensibilidade e especificidade, são superiores ao outro modelo C5.0. O modelo escolhido apresenta melhores métricas em termos de PCCC e as restantes métricas;

3) Modelo I (CHAID *Boosting*): apresenta os melhores resultados, quer em termos de PCCC como de estabilidade do modelo.

Tabela 4.3: Resultados dos modelos com melhor qualidade para previsão de níveis de ansiedade em trabalhadores

Melhores Modelos															
				Grau mínimo			Ansiedade leve			Ansiedade moderada			Ansiedade severa		
PCCC				S	E	S - E	S	E	S - E	S	E	S - E	S	E	S - E
Random Trees	C	Treino	96,3%	100%	95,3%	4,7%	92,6%	100%	-7,4%	97,6%	100%	-2,4%	96,3%	100%	-3,7%
		Teste	94,8%	100%	93,3%	6,7%	90,9%	100%	-9,1%	94,1%	100%	-5,9%	94,2%	100%	-5,8%
C5.0	F	Treino	98,1%	100%	97,5%	2,5%	97,1%	100%	-2,9%	97,6%	100%	-2,4%	97,7%	100%	-2,3%
		Teste	94,4%	100%	93%	7%	92,5%	100%	-7,5%	90,8%	100%	-9,2%	96,3%	100%	-3,7%
CHAID	I	Treino	95,5%	94,7%	95,8%	-1,1%	94,9%	100%	-5,1%	95,2%	98,5%	-3,3%	97,1%	100%	-2,9%
		Teste	93,9%	94,1%	93,8%	0,3%	94,2%	100%	-5,8%	91,5%	98,1%	-6,5%	96,1%	100%	-3,9%

Após a análise da tabela 4.3, conclui-se que o modelo que melhor prevê os níveis de ansiedade em trabalhadores/as é o modelo I pois, apesar de ser o que tem um menor resultado de PCCC, é o que tem melhor estabilidade, isto é, regista menores diferenças entre as métricas sensibilidade e especificidade, e os resultados obtidos em teste são melhores relativamente aos restantes modelos.

4.2. Modelo preditivo de avaliação do impacto da pandemia COVID-19 no nível de ansiedade

As variáveis independentes utilizadas para a avaliação do impacto da pandemia COVID -19 no nível de ansiedade são, igualmente, as variáveis escolhidas para a criação do modelo anterior, o modelo preditivo para previsão de níveis de ansiedade, de forma a obter uma interpretação fácil e comparativa.

4.2.1. Modelo final 2 [Ramificado]: Estudantes universitários

Tal como o modelo anterior, de previsão de níveis de ansiedade, nenhum dos modelos para avaliação do impacto da pandemia no nível de ansiedade dos alunos/as universitários/as apresenta problemas de sobre ajustamento, como observado na tabela 4.4 e no apêndice F, uma vez que os valores de PCCC do treino e do teste são muito semelhantes.

Relativamente às *Random Trees*, o modelo que apresenta melhor qualidade é o modelo B, ou seja, cujo nível de profundidade máximo é 15 e o número mínimo de nós filho é 1, uma vez que, apesar de não ser o modelo com o maior valor de PCCC, é aquele que apresenta uma menor diferença entre o PCCC de treino e teste, e é também aquele que apresenta bons valores de estabilidade do modelo.

Relativamente ao C5.0, o modelo F (*Boosting*) é o que apresenta melhores métricas quer em termos de PCC, como de estabilidade em todas as categorias.

Por fim, o modelo que indica melhor qualidade, utilizando o CHAID, é o I (*Boosting*) pois os modelos G e H apresentam resultados baixos, logo são excluídos da análise, e entre modelo I e o J é o que apresenta melhores resultados em todas as métricas, até porque na previsão de “Agravamento” o modelo J regista uma instabilidade entre as métricas de cerca 20%, o que é um valor relativamente elevado, como observado na tabela no apêndice F.

Tabela 4.4: Resultados dos modelos com melhor qualidade para avaliação do impacto da pandemia em estudantes

Melhores Modelos												
				Agravamento			Constante			Melhoria		
PCCC				S	E	S - E	S	E	S - E	S	E	S - E
Random Trees	B	Treino	92,9%	88,4%	100%	-11,6%	91,9%	98,2%	-6,4%	100%	91,7%	8,3%
		Teste	92,1%	89,4%	100%	-10,6%	89,8%	98,7%	-8,9%	100%	90,6%	9,4%
C5.0	F	Treino	93,1%	93,4%	96,4%	-3%	87,5%	100%	-12,5%	100%	93,6%	6,4%
		Teste	92,9%	90,7%	100%	-9,3%	89,8%	100%	-10,2%	100%	90,2%	9,8%
CHAID	I	Treino	95,7%	93,5%	100%	-6,5%	94,1%	98,2%	-4,1%	100%	95,5%	4,5%
		Teste	88,8%	87,2%	100%	-12,8%	83,7%	98,7%	-15%	100%	86,5%	13,5%

Na tabela 4.4 pode-se verificar que o modelo com melhor qualidade é o F devido aos resultados de PCCC e às restantes métricas que apresentam melhor estabilidade neste modelo, exceto na previsão da categoria “Constante”, mas que não invalida o mesmo.

Desta forma, este é o modelo utilizado para caracterizar tanto o perfil de estudantes universitários que são mais suscetíveis a desenvolver ou a agravar o nível de ansiedade associado devido à pandemia COVID-19, como para identificar quais os melhores preditores.

4.2.2. Modelo final 2 [Ramificado]: Trabalhadores-estudantes

Considerando os resultados obtidos expostos na tabela no apêndice J e na tabela 4.5, verifica-se que o modelo A tem uma descida considerável na performance, visto que a diferença entre a PCCC de treino e teste é de 13,3%. Assim, este modelo é excluído da análise do impacto da pandemia nos níveis de ansiedade dos trabalhadores-estudantes. Desta forma, os modelos com melhores resultados e, consequentemente, com melhor qualidade são:

1) Modelo C (RT 20 *depth* MCNS 1): não é o modelo que apresenta melhores valores de PCCC, contudo é o que apresenta um maior equilíbrio, no que concerne às diferenças das métricas. Apesar de ter resultados semelhantes ao modelo B, este regista menor estabilidade em todos os níveis;

2) Modelo F (C5.0 *Boosting*): o modelo é o que apresenta melhores resultados em termos de PCCC, mesmo se verificando que tem uma maior diferença entre os valores de treino e teste, contudo, em termos de estabilidade é melhor, dando ênfase ao valor registado pelo modelo E na categoria “Constante” de cerca 19% de desequilíbrio entre sensibilidade e especificidade;

3) Modelo I (CHAID *Boosting*): é o modelo que apresenta os melhores resultados em todas as métricas.

Tabela 4.5: Resultados dos modelos com melhor qualidade para avaliação do impacto da pandemia em trabalhadores-estudantes

Melhores Modelos												
				Agravamento			Constante			Melhoria		
PCCC				S	E	S - E	S	E	S - E	S	E	S - E
Random Trees	C	Treino	95,8%	96,2%	100%	-3,8%	91,7%	100%	-8,3%	100%	94%	6%
		Teste	85,9%	87,9%	100%	-12,1%	83,7%	94,6%	-10,9%	87%	85,5%	1,4%
C5.0	F	Treino	95,8%	98,7%	100%	-1,3%	92,4%	98,5%	-6,1%	96,6%	95,5%	1,1%
		Teste	88,8%	92,9%	100%	-7,1%	84,4%	93,9%	-9,6%	89,5%	88,3%	1,2%
CHAID	I	Treino	98,2%	97,5%	100%	-2,5%	97,3%	100%	-2,7%	100%	97,4%	2,6%
		Teste	91,9%	88,9%	100%	-11,1%	94,4%	95,2%	-0,8%	91,7%	92,1%	-0,4%

Concluindo, o modelo que tem melhores resultados para avaliar o impacto da pandemia no nível de ansiedade de trabalhadores-estudantes é o F. Apesar de este não apresentar os melhores valores em termos de PCCC, na categoria “Agravamento” é o modelo que tem uma maior sensibilidade e também equilíbrio das métricas, o que acaba por reforçar esta escolha, pois é o modelo que tem maior capacidade de identificar, com sucesso, os resultados classificados como positivos, ou seja, de identificar o agravamento do nível de ansiedade em trabalhadores-estudantes.

4.2.3. Modelo final 2 [Ramificado]: Trabalhadores/as

No que concerne à avaliação do impacto da pandemia, os modelos que relatam melhores resultados, segundo a tabela 4.6, bem como no apêndice N, são:

1) Modelo B (15 depth MCNS 1): é o modelo que apresenta maior PCCC e também melhores valores em termos de estabilidade. Em termos de identificação das categorias, os resultados são superiores aos dos outros modelos, principalmente na avaliação de caso de “Agravamento”;

2) Modelo E (C5.0 *default*): o modelo apresenta uma PCCC inferior ao modelo F, contudo optou-se por abdicar um pouco da percentagem de casos corretos, em prol de uma concordância entre a percentagem de casos corretamente classificados como positivos e a percentagem de casos corretamente classificados como negativos. De notar que o modelo F apresenta valores na amostra de teste inferiores a 70%;

3) Modelo I (CHAID *Boosting*): é o modelo que apresenta melhores valores quer em termos de PCCC como em termos de estabilidade.

Tabela 4.6: Resultados dos modelos com melhor qualidade para avaliação do impacto da pandemia em trabalhadores

Melhores Modelos												
				Agravamento			Constante			Melhoria		
PCCC				S	E	S - E	S	E	S - E	S	E	S - E
Random Trees	B	Treino	87,6%	93,5%	100%	-6,5%	84,5%	91%	-6,5%	84,6%	90,5%	-5,9%
		Teste	88,4%	94,4%	100%	-5,6%	89,1%	90%	-0,9%	80,4%	93%	-12,6%
C5.0	E	Treino	79,1%	79,7%	100%	-20,3%	79,4%	86,9%	-7,4%	77,6%	83,5%	-5,8%
		Teste	82,2%	92,7%	100%	-7,3%	71,7%	90,3%	-18,6%	81,6%	84,3%	-2,7%
CHAID	I	Treino	89,4%	96,4%	100%	-3,6%	85,7%	93,5%	-7,8%	85,2%	91%	-5,8%
		Teste	77,4%	93,9%	100%	-6,1%	65,9%	84,8%	-18,9%	71,4%	81,1%	-9,7%

Assim, o modelo escolhido para prever os níveis de ansiedade em trabalhadores/as é o B uma vez que é o que tem melhores resultados nas métricas de sensibilidade, especificidade, PCC e, também, mais estabilidade.

4.3. Suscetibilidade ao agravamento do nível de ansiedade e variáveis com maior influência

Após a análise dos resultados obtidos no modelo para avaliação do impacto da pandemia COVID-19 no nível de ansiedade associado a estudantes e trabalhadores/as portugueses/as, são identificados os perfis de estudantes-universitários que são mais suscetíveis a agravar o nível de ansiedade associado, como as variáveis preditoras que têm maior relevância nesta alteração.

Assim, no modelo relativo a estudantes portugueses, os preditores com mais importância são 'Area_Estudos', com 15%, 'Agregado_familiar' e 'Idade_TILEN' com a igual percentagem de 11%.

Tabela 4.7: Perfil de estudantes universitários suscetíveis a agravar o nível de ansiedade

	Perfil	Regra
'Agravamento'	Estudantes universitários com a seguinte caracterização são classificados com um agravamento no nível de ansiedade associado com 92,3% de confiança e com um suporte de 13 casos: 1. Alunos universitários nas áreas e Informação e Jornalismo; 2. Alunos universitários da área de Artes.	Nó 1 Area_Estudos = Informação e Jornalismo Area_Estudos = Artes

Relativamente ao modelo dos Trabalhadores-estudantes, a variável preditora, segundo este modelo, com maior importância é 'Anos_trabalho' com 8%.

Tabela 4.8: Perfil de trabalhadores-estudantes suscetíveis a agravar o nível de ansiedade

	Perfil	Regra
'Agravamento'	Trabalhadores-estudantes com a seguinte caracterização são classificados com um agravamento no nível de ansiedade associado com 71,1% de confiança e com um suporte de 76 casos: 1. Trabalhar há menos de 1 ano, e 2. Frequentarem a Licenciatura.	Nó 23 Anos_trabalho = "Menos de 1 ano" Ciclo_Estudos = "Licenciatura"

Por fim, as variáveis preditoras do modelo relativo aos trabalhadores são 'Vínculo_Trabalhador', 'Area_Trabalho' e 'Anos_Trabalho' com uma igual importância de com 13%.

Tabela 4.9: Perfil de trabalhadores suscetíveis a agravar o nível de ansiedade

	Perfil	Regra
'Agravamento'	Trabalhadores com a seguinte caracterização são classificados com um agravamento no nível de ansiedade associado com 82,7% de confiança. 1. Ter idade igual ou inferior a 34 anos; 2. Trabalhar nas áreas das Ciências Empresariais ou Serviços Pessoais; 3. Ser solteiro ou unido de facto; 4. Ser homem, e 5. Ter dependentes a cargo.	Idade_TILEN <= 4 Area_trabalho = Ciências Empresariais, Serviços Pessoais Estado_Civil = Solteiro/a, Unido/a de facto Género = Homem Dependentes = Sim

Capítulo 5: Conclusões e recomendações

5.1. Principais conclusões

Em Portugal, desde 2020, uma pandemia de trato respiratório veio modificar a realidade e o quotidiano da população. Esta teve um grave impacto na saúde mental dos portugueses/as e afetou também os serviços hospitalares neste campo. De forma a aumentar a eficácia e eficiência da previsão de níveis de ansiedade e a mitigar os resultados futuros da pandemia, é imperativo apoiar os profissionais de saúde no diagnóstico dos seus pacientes. Nesta sequência, é necessário fornecer ferramentas capazes de avaliar a saúde mental, de forma a combater possíveis consequências irreversíveis, como também de modo a proporcionar uma melhor qualidade de vida para diagnosticados/as com perturbações de foro mental.

Perante este problema, o presente estudo procura desenvolver um modelo que seja capaz de identificar os níveis de ansiedade em alunos e trabalhadores/as portugueses/as e avaliar o impacto que a pandemia COVID-19 possa ter causado na saúde mental dos mesmos. Uma vez que esta se trata de uma área complexa, recorreu-se a técnicas de *Machine Learning* para a criação de modelos preditivos com capacidade para determinar perfis de estudantes e trabalhadores/as portugueses/as que sejam mais suscetíveis a agravar o diagnóstico associado.

Com base nos resultados obtidos, os objetivos estabelecidos foram cumpridos. De facto, através de *Machine Learning* foi possível construir dois modelos preditivos capazes de identificar tanto os níveis de ansiedade na população-alvo, como também avaliar o impacto da pandemia na saúde mental da mesma, em que foi possível constatar se houve um agravamento, indiferença ou melhoria da saúde mental. Como mencionado, os modelos revelaram uma ótima qualidade de previsão e avaliação e é possível verificar através dos resultados obtidos pelas métricas utilizadas: PCCC, sensibilidade, especificidade e a diferença entre sensibilidade e especificidade, que proporciona a identificação da estabilidade do modelo.

Deste modo, o estudo permite responder à questão de investigação, uma vez que se constatou que os modelos de *Machine Learning* constituem uma boa ferramenta de apoio à prevenção de perturbações de ansiedade.

5.2. Contributos para a comunidade científica e empresarial

Perante o problema exposto, o estudo vem contribuir tanto para a academia, como para os profissionais da área de aplicação deste estudo. Por um lado, a RSL efetuada sobre a utilização de técnicas de *Machine Learning* na saúde mental contribui para ajudar futuros/as investigadores/as nas áreas de Dados e de Saúde, na medida em que é disponibilizado um resumo dos principais artigos publicados em revistas científicas e são identificadas as utilizações destas técnicas no estudo da saúde mental em quatro dimensões de análise (contextos, tipologias e avaliação dos resultados), ou seja, esta investigação torna-se um bom ponto de partida.

Adicionalmente, a recolha de dados de estudantes e trabalhadores/as portugueses/as e longitudinais também vem contribuir não só para a literatura, como, em simultâneo, para a comunidade estudantil e para profissionais, uma vez que é realizada uma análise mais vocacionada e pouco mencionada na literatura existente.

Para os/as investigadores/as mais focados em técnicas de *Machine Learning*, o estudo revela uma forte capacidade preditiva dos modelos de classificação baseados em árvores de decisão. É de salientar, que os resultados obtidos, através da modelação e aplicados à saúde mental, foram superiores aos obtidos por estudos anteriores e, também, que a avaliação de grupos demonstra uma maior qualidade de predição, em detrimento de uma avaliação geral e menos pormenorizada.

Por fim, para os profissionais de saúde, apresenta-se uma ferramenta a explorar que pode fazer diferença na área de saúde mental, ou seja, uma boa técnica de suporte à previsão de diagnósticos e prevenção de sinais de ansiedade.

5.3. Limitações do estudo

No que respeita às limitações, é importante salientar que a falta de informação e a difícil recolha de dados sobre uma área, relativa à Saúde Mental, que, ainda, tem um longo trabalho de desmitificação a ser feito, afetou o estudo.

Primeiramente, por não haver um fácil acesso a dados de foro psicológico, principalmente dados relativos a portugueses/as, por se tratar de dados de autorrelato, isto é, que derivam da perceção dos/as inquiridos/as e devido ao tamanho da amostra que pode enviesar a análise.

5.4. Propostas de investigação futura

Para futuros trabalhos e para colmatar possíveis falhas no presente estudo apresentam-se as seguintes propostas:

- 1) Inclusão de outras variáveis independentes nos modelos (dados hospitalares, hábitos e estilos de vida, entre outras);
- 2) Aumento da dimensão da amostra;
- 3) Alargar o estudo, no que respeita à população-alvo;
- 4) Alargar o estudo à previsão de outras perturbações mentais;
- 5) Aplicar outros algoritmos de *Machine Learning*, como redes neuronais e regressão logística.

Referências bibliográficas

- Almeida, J., Xavier, M., Cardoso, G., Pereira, M. (2013), Estudo Epidemiológico Nacional De Saúde Mental: 1º Relatório, Lisboa, Faculdade de Ciências Médicas, Universidade Nova de Lisboa
- Alharbi, A. A., Alqassim, A. Y., Muaddi, M. A., Makeen, A. M., Hagawi, A. M., Shaabi, L. A., Alhweity, S. M., Alharbi, M. A., Sultan, M. A., Alshamakhi, A. H., Aburasain, N. S., & Horner, R. D. (2022). Generalized anxiety disorder symptoms during COVID-19 pandemic in Jazan, Saudi Arabia. *Heliyon*, 8(5), e09424. <https://doi.org/10.1016/J.HELIYON.2022.E09424>
- Al-Wesabi, F. N., Alsolai, H., Hilal, A. M., Hamza, M. A., Duhayyim, M. al, & Negm, N. (2022). Machine Learning Based Depression, Anxiety, and Stress Predictive Model During COVID-19 Crisis. *CMC-Comput. Mat. Contin.*, 70(3), 5803–5820. <https://doi.org/10.32604/CMC.2022.021195>
- American Psychiatric Association, (2013), *Diagnostic And Statistical Manual Of Mental Disorders*, Arlington, American Psychiatric Publishing (5o Edição) (Edição original, 1952)
- Anbarasi, L. J., Jawahar, M., Ravi, V., Cherian, S. M., Shreenidhi, S., & Sharen, H. (2022). Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown. *Health and Technology*, 12(4), 825–838. <https://doi.org/10.1007/S12553-022-00674-7/TABLES/10>
- Antunes, R., Frontini, R., Amaro, N., Salvador, R., Matos, R., Morouço, P., & Rebelo-Gonçalves, R. (2020). Exploring Lifestyle Habits, Physical Activity, Anxiety and Basic Psychological Needs in a Sample of Portuguese Adults during COVID-19. *International Journal of Environmental Research and Public Health*, 17(12), 1–13. <https://doi.org/10.3390/IJERPH17124360>
- Boeke, E. A., Holmes, A. J., & Phelps, E. A. (2020). Towards robust anxiety biomarkers: A machine learning approach in a large-scale sample. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 5(8), 799. <https://doi.org/10.1016/J.BPSC.2019.05.018>
- Bokma, W. A., Zhutovsky, P., Giltay, E. J., Schoevers, R. A., Penninx, B. W. J. H., van Balkom, A. L. J. M., Batelaan, N. M., & van Wingen, G. A. (2022). Predicting the naturalistic course in anxiety disorders using clinical and biological markers: a machine learning approach. *Psychological Medicine*, 52(1), 57–67. <https://doi.org/10.1017/S0033291720001658>
- Caetano, B. L. (2019). *Calma Mente: decodificar a Ansiedade, uma campanha para a Saúde Mental*. https://repositorio.iscte-iul.pt/bitstream/10071/19480/4/master_beatriz_leitao_caetano.pdf
- Caldas, R., Mundt, M., Potthast W., Neto, F., & Markert, B. (2017). A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. *Gait & Posture*, 57: 204-210
- Carlton, C. N., Garcia, K. M., Andino, M. V., Ollendick, T. H., & Richey, J. A. (2022). Social anxiety disorder is Associated with Vaccination attitude, stress, and coping responses during COVID-19. *Cognitive Therapy and Research*, 46(5), 916–926. <https://doi.org/10.1007/S10608-022-10310-3/TABLES/2>
- Diária da República Eletrónico* (2020). *Decreto-Lei n.º 10-A/2020*. <https://dre.pt/dre/detalhe/decreto-lei/10-a-2020-130243053>
- Díaz-Godiño, J., Fernández-Henríquez, L., Peña-Pastor, F., Alfaro-Flores, P., Manrique-Borjas, G., & Mayta-Tovalino, F. (2019). Lifestyles, Depression, Anxiety, and Stress as Risk Factors in Nursing Apprentices: A Logistic Regression Analysis of 1193 Students in Lima, Peru. *Journal of Environmental and Public Health*, 2019. <https://doi.org/10.1155/2019/7395784>
- Ding, Y., Liu, J., Zhang, X., & Yang, Z. (2022). Dynamic Tracking of State Anxiety via Multi-Modal Data and Machine Learning. *Frontiers in Psychiatry*, 13, 757961. <https://doi.org/10.3389/FPSYT.2022.757961/FULL>
- Duru, Y. B., Gunal, V., Agaoglu, C. Y., & Tatli, C. (2022). The role of covid-19 anxiety and intolerance of uncertainty in predicting resilience. *Scandinavian Journal of Psychology*, 63(5), 522–529. <https://doi.org/10.1111/SJOP.12820>
- Gruda, D., & Hasan, S. (2019). Feeling anxious? Perceiving anxiety in tweets using machine learning. *Computers in Human Behavior*, 98, 245–255. <https://doi.org/10.1016/J.CHB.2019.04.020>
- Hueniken, K., Somé, N. H., Abdelhack, M., Taylor, G., Marshall, T. E., Wickens, C. M., Hamilton, H. A., Wells, S., & Felsky, D. (2021). Machine Learning–Based Predictive Modeling of Anxiety and

- Depressive Symptoms During 8 Months of the COVID-19 Global Pandemic: Repeated Cross-sectional Survey Study. *JMIR Ment Health* 2021;8(11):E32876 <https://Mental.Jmir.Org/2021/11/E32876>, 8(11), e32876. <https://doi.org/10.2196/32876>
- Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, 51(5), 675–687. <https://doi.org/10.1016/J.BETH.2020.05.002>
- Moro, S., Laureano, R., & Cortez, P. (2011). Using data mining for bank direct marketing: An application of the CRISP-DM methodology. *ESM 2011 - 2011 European Simulation and Modelling Conference: Modelling and Simulation 2011, Figure 1*, 117–121.
- Muhammad, F., & Al-Ahmadi, S. (2022). Human state anxiety classification framework using EEG signals in response to exposure therapy. *PLOS ONE*, 17(3), e0265679. <https://doi.org/10.1371/JOURNAL.PONE.0265679>
- Nayan, M., Uddin, M., Hossain, M., Alam, M., Zinnia, M., Haq, I., Rahman, M., Ria, R., & Haq Methun, M. (2022). Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among University Students in Bangladesh: A result of the first wave of the COVID-19 pandemic. *Asian Journal of Social Health and Behavior*, 5(2), 75. https://doi.org/10.4103/SHB.SHB_38_22
- Nemesure, M. D., Heinz, M. v., Huang, R., & Jacobson, N. C. (2021). Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Scientific Reports* 2021 11:1, 11(1), 1–9. <https://doi.org/10.1038/s41598-021-81368-4>
- OMS (2014). “Mental health: a state of well-being”. Consultado a 1 de setembro de 2022. Disponível em http://www.who.int/features/factfiles/mental_health/en/
- OMS (2017). *Depression and Other Common Mental Disorders: Global Health Estimates*, Genebra.
- OMS (2022). “OMS divulga Informe Mundial de Saúde Mental: transformar a saúde mental para todos”. Consultado a 2 de setembro de 2022. Disponível em <https://bvsms.saude.gov.br/oms-divulga-informe-mundial-de-saude-mental-transformar-a-saude-mental-para-todos/>
- Paton, L. W., & Tiffin, P. A. (2022). Technology Matters: Machine learning approaches to personalised child and adolescent mental health care. *Child and Adolescent Mental Health*, 27(3), 307–308. <https://doi.org/10.1111/CAMH.12546>
- 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. <https://doi.org/10.7717/PEERJ.10448/SUPP-2>
- Portugal, L., Schrouff, J., Stiffler, R., Bertocci, M., Bebko, G., Chase, H., Lockovitch, J., Aslam, H., Graur, S., Greenberg, T., Pereira, M., Oliveira, L., Phillips, M., & Mourão-Miranda, J. (2019). Predicting anxiety from wholebrain activity patterns to emotional faces in young adults: a machine learning approach. *NeuroImage: Clinical*, 23, 101813. <https://doi.org/10.1016/J.NICL.2019.101813>
- Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G., & Okon-Singer, H. (2021). Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, 141, 199–205. <https://doi.org/10.1016/J.JPSYCHIRES.2021.06.044>
- Saeed, H., Eslami, A., Nassif, N. T., Simpson, A. M., & Lal, S. (2022). Anxiety Linked to COVID-19: A Systematic Review Comparing Anxiety Rates in Different Populations. *International Journal of Environmental Research and Public Health*, 19(4). <https://doi.org/10.3390/IJERPH19042189>
- SNS (2022). “Saúde Mental: Ansiedade”. Consultado a 2 de setembro de 2022. Disponível em <https://www.sns24.gov.pt/tema/saude-mental/ansiedade/>
- SPPSM (n.d.), “Termos em Saúde/Perturbação Mental”. Consultado em 2 de setembro de 2022. Disponível em <https://www.sppsm.org/informemente/termos-em-saudeperturbacao-mental-2/>
- Stark, M., Huang, H., Yu, L. F., Martin, R., McCarthy, R., Locke, E., Yager, C., Torad, A. A., Kadry, A. M., Elwan, M. A., Smith, M. L., Bradley, D., & Boolani, A. (2022). Identifying Individuals Who Currently

- Report Feelings of Anxiety Using Walking Gait and Quiet Balance: An Exploratory Study Using Machine Learning. *Sensors*, 22(9), 3163. <https://doi.org/10.3390/S22093163/S1>
- Van Eeden, W. A., Luo, C., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W., Wardenaar, K. J., Hoos, H., & Giltay, E. J. (2021). Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression. *Psychiatry Research*, 299, 113823. <https://doi.org/10.1016/J.PSYCHRES.2021.113823>
- Vilelas, J. (2009). *Investigação: o Processo de Construção do Conhecimento*. Lisboa: Edições Sílabo.
- Vilelas, J. (2017). *Investigação: O Processo de Construção do Conhecimento* (2a ed.). Lisboa: Edições Sílabo
- Wang, C., Zhao, H., & Zhang, H. (2020). Chinese College Students Have Higher Anxiety in New Semester of Online Learning During COVID-19: A Machine Learning Approach. *Frontiers in Psychology*, 11, 3465. <https://doi.org/10.3389/FPSYG.2020.587413/BIBTEX>
- Wardenaar, K. J., Riese, H., Giltay, E. J., Eikelenboom, M., van Hemert, A. J., Beekman, A. F., Penninx, B. W. J. H., & Schoevers, R. A. (2021). Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders*, 293, 295–304. <https://doi.org/10.1016/J.JAD.2021.06.029>
- World Health Organization (2014), “Mental health: a state of well-being” (online). Consultado a 27 de agosto de 2022. Disponível em http://www.who.int/features/factfiles/mental_health/en/
- World Health Organization (2017), *Depression and Other Common Mental Disorders: Global Health Estimates*, Geneva, WHO Document Production Services
- Yasmin, F., Dapke, K., Khalid, M. R., Naz, H., Naz, F., Admani, B., Asghar, M. S., Awan, S., Phadke, R., & Saleem, M. (2021). Generalized Anxiety Disorder and Depressive Symptoms among Pakistani Population during the Second Wave of the COVID-19 Pandemic: A Regression Analysis. *The American Journal of Tropical Medicine and Hygiene*, 105(4), 915. <https://doi.org/10.4269/AJTMH.21-0380>
- Zhang, L., Zhao, S., Lin, Q., Song, M., Wu, S., & Zheng, H. (2021). Algorithms to Predict Anxiety and Depression Among University Students in China After Analyzing Lifestyles and Sport Habits. *Neuropsychiatric Disease and Treatment*, 17, 2011. <https://doi.org/10.2147/NDT.S315148>
- Zhou, Z., Wu, T. C., Wang, B., Wang, H., Tu, X. M., & Feng, C. (2020). Machine learning methods in psychiatry: a brief introduction. *General Psychiatry*, 33(1), 100171. <https://doi.org/10.1136/GPSYCH-2019-100171>

Apêndices

Apêndice A – Questionário para recolha de dados

Machine Learning como ferramenta de apoio à prevenção: identificação de sinais de ansiedade

Existe(m) 33 questão(ões) neste questionário.

Machine Learning como ferramenta de apoio à prevenção: identificação de sinais de ansiedade

Consentimento Informado

No âmbito de uma dissertação de Mestrado em *Business Analytics* do **Iscte – Instituto Universitário de Lisboa**, estou a desenvolver um estudo que tem por objetivo avaliar as experiências de ansiedade, de estudantes universitários/as e trabalhadores/as portugueses/as durante e após o confinamento decorrente da pandemia COVID-19.

A sua participação consiste no preenchimento de um questionário online, com uma duração aproximada de 10 minutos, no qual pedimos que responda a algumas perguntas sobre as suas experiências de ansiedade e de bem-estar psicológico.

A participação neste estudo é estritamente **voluntária**, isto é, pode escolher participar ou não participar. No caso de participar, pode interromper a sua participação em qualquer momento sem ter de prestar qualquer justificação. Esta é também **anónima e confidencial**, ou seja, os dados recolhidos destinam-se apenas para fins académicos. Em nenhum momento do estudo precisa de se identificar. O presente estudo destina-se a respondentes **maiores de idade** (idade superior a 18 anos) e com **domínio da língua portuguesa**.

Não existem respostas certas ou erradas. É expectável que responda a todas as questões da forma mais sincera possível. Nenhuma resposta será analisada ou reportada individualmente.

Face a estas informações, por favor indique se aceita participar neste estudo e clique no botão do canto inferior direito para continuar na página seguinte. O preenchimento do questionário presume que compreendeu e que aceita as condições do presente estudo, consentindo participar.

Se pretender algum esclarecimento sobre este estudo, por favor contacte: Carolina Marques pelo e-mail cmdms@iscte-iul.pt (<mailto:cmdms@iscte-iul.pt>).

Figura A: Questionário para recolha de dados

Aceita participar voluntariamente neste estudo? *

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Aceito
- ☐ Não aceito

Qual é a sua idade? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

❗ A sua resposta deve ser no mínimo 18

❗ Neste campo apenas pode ser introduzido um valor inteiro.

Por favor, escreva aqui a sua resposta:

Como se identifica? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Homem
- ☐ Mulher

☐ Outro

Qual a sua nacionalidade? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

❗ Se seleccionar a opção 'Outra:', por favor especifique a sua escolha utilizando o respetivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

☐ Portuguesa

☐ Outra:

Qual é o seu estado civil? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **apenas uma** das seguintes opções:

☐ Solteiro/a

☐ Unido/a de facto

☐ Casado/a

☐ Divorciado/a

☐ Viúvo/a

Tem dependentes a cargo? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Sim
- ☐ Não

Com quem vive? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **todas** as que se aplicam:

- ☐ Sozinho/a
- ☐ Ambos os pais
- ☐ Um dos pais
- ☐ Irmãos/ãs
- ☐ Amigos/as
- ☐ Outros/as estudantes universitários/as
- ☐ Cônjuge ou companheiro/a
- ☐ Filhos/as
- ☐ Outros. Quem?:

Qual é o seu local de residência atual? (onde vive maior parte do tempo) *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

❗ Se selecionar a opção 'Outro. Qual?', por favor especifique a sua escolha utilizando o respetivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Aveiro
- ☐ Beja
- ☐ Braga
- ☐ Bragança
- ☐ Castelo Branco
- ☐ Coimbra
- ☐ Évora
- ☐ Faro
- ☐ Guarda
- ☐ Leiria
- ☐ Lisboa
- ☐ Portalegre
- ☐ Porto
- ☐ Santarém
- ☐ Setúbal
- ☐ Viana do Castelo
- ☐ Vila Real
- ☐ Viseu
- ☐ Região Autónoma dos Açores
- ☐ Região Autónoma da Madeira
- ☐ Outro. Qual?

Em que medida (1 a 5) quantifica o nível de condições de vida do seu agregado familiar?

*

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

Obs.: 1 representa menores condições de vida (por exemplo, menos estabilidade financeira, menos estudos, entre outros fatores); 5 representa melhores condições de vida (como, mais estabilidade financeira, mais qualificações, entre outros).

Tem algum tipo de atividade profissional? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Não
- ☐ Sim, a tempo parcial
- ☐ Sim, a tempo inteiro

É estudante? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **apenas uma** das seguintes opções:

☐ Sim

☐ Não

O principal objetivo deste estudo é avaliar experiências de ansiedade que estudantes universitários/as e trabalhadores/as possam ter vivenciado nos determinados períodos de tempo: durante a pandemia COVID-19 e períodos subsequentes.

Assim, as questões elaboradas no presente inquérito têm em consideração três momentos de avaliação:

Momento 1: Período pandémico (isto é, desde que surgiu a Doença do coronavírus - COVID-19 - e foram aplicadas medidas de distanciamento social com vista à sua contenção, como o confinamento);

Momento 2: Período pós pandemia COVID-19 (fase inicial de desconfinamento com restrições impostas, como o uso de máscara, elaboração de testes PCR como forma de acesso a espaços, por exemplo, entre outras);

Momento 3: Momento atual (fase em que são retiradas maior parte das restrições).

Qual o tipo de ensino que frequenta? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

❗ Se selecionar a opção 'Outro: Qual?', por favor especifique a sua escolha utilizando o respetivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

☐ Ensino Universitário Público

☐ Ensino Universitário Privado

☐ Ensino Politécnico

☐ Outro: Qual?

Qual é a área de estudos que frequenta? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

❗ Se selecionar a opção 'Outra. Qual?', por favor especifique a sua escolha utilizando o respetivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

☐ Artes (Exemplo: Cinema, Dança, Design, Música, Fotografia)

☐ Línguas e Humanidades (Exemplo: Arqueologia, História, Línguas, Tradução)

☐ Ciências Sociais e do Comportamento (Exemplo: Antropologia, Sociologia, Criminologia, Economia)

☐ Informação e Jornalismo (Exemplo: Jornalismo, Relações Públicas)

☐ Ciências Empresariais (Exemplo: Contabilidade, Gestão, Finanças, Marketing, Comércio, Secretariado)

☐ Direito (Exemplo: Direito, Solicitadoria, Administração Público-Privada)

☐ Ciências da Vida (Exemplo: Biologia, Bioquímica, Biorrecursos, Ecologia e Ambiente)

- ☐ Ciências Físicas (Exemplo: Física e Química, Geologia, Proteção Civil, Ciências do Mar)
- ☐ Matemática e Estatística (Exemplo: Matemática Aplicada)
- ☐ Informática (Exemplo: Ciência de Dados, Computação, Informática de Gestão)
- ☐ Engenharia e Técnicas Afins (Exemplo: Biomecânica, Ciências de Engenharia Aeronáutica, Engenharia Eletrotécnica)
- ☐ Arquitetura e Construção (Exemplo: Engenharia Civil, Engenharia Geoespacial)
- ☐ Agricultura, Silvicultura e Pescas (Exemplo: Ciências Agrárias, Zootecnia, Engenharia Agro-Pecuária)
- ☐ Saúde (Exemplo: Medicina, Enfermagem, Farmácia, Terapia da Fala, Dietética e Nutrição, Enfermagem/Medicina Veterinária)
- ☐ Serviços Sociais (Exemplo: Animação Sociocultural, Educação Social, Serviço Social)
- ☐ Serviços Pessoais (Exemplo: Desporto, Gastronomia, Gestão Hoteleira, Turismo e Lazer)
- ☐ Outra. Qual?

Em que ano ingressou na faculdade? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

❗ A resposta deve estar entre 1900 e 2021

❗ Neste campo apenas pode ser introduzido um valor inteiro.

Por favor, escreva aqui a sua resposta:

Qual o ciclo de estudos que frequenta? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Licenciatura
- ☐ Pós-graduação
- ☐ Mestrado
- ☐ Mestrado Integrado
- ☐ Doutoramento

Qual o ano curricular que frequenta? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ 1º ano
- ☐ 2º ano
- ☐ 3º ano
- ☐ 4º ano (se aplicável ao curso que frequenta)
- ☐ 5º ou mais anos (se aplicável ao curso que frequenta)

Selecione o ano curricular que frequenta independentemente do número de anos que repete algum dos anos curriculares (isto é, caso esteja há 3 anos na faculdade e, neste momento, está a frequentar, em grande parte ou total, o 2º ano curricular, então deve seleccionar a opção: "2ºano")

Frequenta, neste momento, o último ano curricular? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?) e Resposta foi maior do que '2º ano' na pergunta '16 [AnoCurricular]' (Qual o ano curricular que frequenta?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Sim
- ☐ Não

Selecione se é ou não finalista.

É estudante deslocado? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Sim
- ☐ Não

Qual a distância a que se encontra da sua habitação permanente (casa)? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?) e A resposta for 'Sim' na pergunta '18 [EstudanteDeslocado]' (É estudante deslocado?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ até 50 km
- ☐ Entre 51 e 100 km
- ☐ Entre 101 e 150 km
- ☐ Entre 151 e 200 km
- ☐ Mais de 201 km

Comparativamente aos resultados de avaliação obtidos durante a situação pandémica (COVID-19) , como classifica os resultados de avaliação obtidos no periodo pós confinamento (fase inicial de desconfinamento com restrições impostas)? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Obtive melhores resultados
- ☐ Mantive os mesmos resultados
- ☐ Piorei significativamente os resultados

Comparativamente aos resultados de avaliação obtidos no período após a situação pandémica (fase inicial de desconfinamento com restrições impostas), como classifica os resultados de avaliação obtidos atualmente? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Sim' na pergunta '11 [Estudante]' (É estudante?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Obtive melhores resultados
- ☐ Mantive os mesmos resultados
- ☐ Pirei significativamente os resultados

O principal objetivo deste estudo é avaliar experiências de ansiedade que estudantes universitários/as e trabalhadores/as possam ter vivenciado nos determinados períodos de tempo: durante a pandemia COVID-19 e períodos subsequentes.

Assim, as questões elaboradas no presente inquérito têm em consideração três momentos de avaliação:

Momento 1: Período pandémico (isto é, desde que surgiu a Doença do coronavírus - COVID-19 - e foram aplicadas medidas de distanciamento social com vista à sua contenção, como o confinamento);

Momento 2: Período pós pandemia COVID-19 (fase inicial de desconfinamento com restrições impostas, como o uso de máscara, elaboração de testes PCR como forma de acesso a espaços, por exemplo, entre outras);

Momento 3: Momento atual (fase em que são retiradas maior parte das restrições).

Qual a área que trabalha atualmente? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta não foi 'Não' na pergunta '10 [Profissao]' (Tem algum tipo de

atividade profissional?)

❗ Se selecionar a opção 'Outra. Qual?', por favor especifique a sua escolha utilizando o respectivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Artes (Exemplo: Cinema, Dança, Design, Música, Fotografia)
- ☐ Línguas e Humanidades (Exemplo: Arqueologia, História, Línguas, Tradução)
- ☐ Ciências Sociais e do Comportamento (Exemplo: Antropologia, Sociologia, Criminologia, Economia)
- ☐ Informação e Jornalismo (Exemplo: Jornalismo, Relações Públicas)
- ☐ Ciências Empresariais (Exemplo: Contabilidade, Gestão, Finanças, Marketing, Comércio, Secretariado)
- ☐ Direito (Exemplo: Direito, Solicitoria, Administração Público-Privada)
- ☐ Ciências da Vida (Exemplo: Biologia, Bioquímica, Biorrecursos, Ecologia e Ambiente)
- ☐ Ciências Físicas (Exemplo: Física e Química, Geologia, Proteção Civil, Ciências do Mar)
- ☐ Matemática e Estatística (Exemplo: Matemática Aplicada)
- ☐ Informática (Exemplo: Ciência de Dados, Computação, Informática de Gestão)
- ☐ Engenharia e Técnicas Afins (Exemplo: Biomecânica, Ciências de Engenharia Aeronáutica, Engenharia Eletrotécnica)
- ☐ Arquitetura e Construção (Exemplo: Engenharia Civil, Engenharia Geoespacial)
- ☐ Agricultura, Silvicultura e Pescas (Exemplo: Ciências Agrárias, Zootecnia, Engenharia Agro-Pecuária)
- ☐ Saúde (Exemplo: Medicina, Enfermagem, Farmácia, Terapia da Fala, Dietética e Nutrição, Enfermagem/Medicina Veterinária)
- ☐ Serviços Sociais (Exemplo: Animação Sociocultural, Educação Social, Serviço Social)
- ☐ Serviços Pessoais (Exemplo: Desporto, Gastronomia, Gestão Hoteleira, Turismo e Lazer)
- ☐ Outra. Qual?

Obs.: No caso de trabalhar em mais do que uma área de negócio, deve indicar (1) Aquela que lhe ocupa mais tempo ou (2) Aquela que lhe aufera maiores rendimentos

Qual é a sua situação profissional? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta não foi 'Não' na pergunta '10 [Profissao]' (Tem algum tipo de atividade profissional?)

❗ Se selecionar a opção 'Outra situação. Qual?', por favor especifique a sua escolha utilizando o respetivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Trabalhador/a por contra de outrem
- ☐ Trabalhador/a por conta própria ou Empregador/a
- ☐ Trabalhador/a familiar não remunerado (Exemplo: Cuidadores/as)
- ☐ Outra situação. Qual?

Há quanto tempo exerce a função que ocupa atualmente (número de anos)? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta não foi 'Não' na pergunta '10 [Profissao]' (Tem algum tipo de atividade profissional?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Menos de 1 ano
- ☐ Entre 1 e 5 anos
- ☐ Entre 5 e 10 anos
- ☐ Mais de 10 anos

Qual é o seu tipo de vínculo? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta não foi 'Não' na pergunta '10 [Profissao]' (Tem algum tipo de atividade profissional?)

❗ Se selecionar a opção 'Outro vínculo. Qual?', por favor especifique a sua escolha utilizando o respetivo campo de texto.

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Contrato permanente
- ☐ Contrato a termo certo
- ☐ Contrato a termo incerto
- ☐ Colocado por empresas de trabalho temporário
- ☐ Contrato de prestação de serviços e a recibos verdes
- ☐ Outro vínculo. Qual?

Comparativamente ao período pandémico COVID-19 vivenciado, em que medida quantifica a conciliação da vida laboral com a vida familiar no período pós confinamento (fase inicial de desconfinamento com restrições impostas)? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta não foi 'Não' na pergunta '10 [Profissao]' (Tem algum tipo de atividade profissional?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

Obs.: 1 representa uma grande dificuldade em conciliar a vida laboral com a familiar; 5 representa pouca dificuldade em conciliar a vida laboral com a familiar.

Comparativamente ao período pós situação pandémica (fase inicial de desconfinamento com restrições impostas), em que medida quantifica a conciliação da vida laboral com a vida familiar atualmente? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta não foi 'Não' na pergunta '10 [Profissao]' (Tem algum tipo de atividade profissional?)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ 1
☐ 2
☐ 3
☐ 4
☐ 5

Obs.: 1 representa uma grande dificuldade em conciliar a vida laboral com a familiar; 5 representa pouca dificuldade em conciliar a vida laboral com a familiar.

Saúde Mental

O presente inquérito tem em consideração três momentos de avaliação:

Momento 1: Período pandémico (isto é, desde que surgiu a Doença do coronavírus - COVID-19 - e foram aplicadas medidas de distanciamento social com vista à sua contenção, como o confinamento);

Momento 2: Período pós pandemia COVID-19 (fase inicial de desconfinamento com restrições impostas, como o uso de máscara, elaboração de testes PCR como forma de acesso a espaços, por exemplo, entre outras);

Momento 3: Momento atual (fase em que são retiradas maior parte das restrições).

Caso seja diagnosticado/a com uma perturbação de saúde mental por um/a profissional de saúde, por favor indique qual ou quais: *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione **todas** as que se aplicam:

- ☐ Depressão
- ☐ Perturbação de Ansiedade
- ☐ Perturbação Bipolar
- ☐ Perturbação do Comportamento Alimentar
- ☐ Esquizofrenia
- ☐ Perturbação de personalidade de Borderline
- ☐ Não diagnosticado/a

☐ Outro diagnóstico. Qual?:

Caso seja diagnosticado/a com ansiedade por um/a profissional de saúde, por favor indique qual ou quais as perturbações mentais diagnosticadas: *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Perturbação de Ansiedade' na pergunta '28 [PerturbacaoMental]' (Caso seja diagnosticado/a com uma perturbação de saúde mental por um/a profissional de saúde, por favor indique qual ou quais:)

Por favor, selecione **todas** as que se aplicam:

- ☐ Ansiedade generalizada
- ☐ Fobia social
- ☐ Perturbação de Pânico
- ☐ Fobias
- ☐ Perturbação de Stress Pós-traumático
- ☐ Perturbação Obsessivo-Compulsiva
- ☐ Não diagnosticado/a

☐ Outro diagnóstico. Qual?:

Considera-se uma pessoa ansiosa? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Não diagnosticado/a' na pergunta '28 [PerturbacaoMental]' (Caso seja diagnosticado/a com uma perturbação de saúde mental por um/a profissional de saúde, por favor indique qual ou quais:)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Sim, grande parte do tempo
- ☐ Sim, apenas em situações de stress
- ☐ Não

Procurou acompanhamento psicológico devido a sentimentos de ansiedade em algum dos momentos mencionados (durante e/ou após a situação pandémica COVID-19)? *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?) e A resposta for 'Não diagnosticado/a' na pergunta '28 [PerturbacaoMental]' (Caso seja diagnosticado/a com uma perturbação de saúde mental por um/a profissional de saúde, por favor indique qual ou quais:)

Por favor, selecione **apenas uma** das seguintes opções:

- ☐ Sim
- ☐ Não

Comparativamente ao período da pandemia e às medidas de distanciamento social aplicadas com vista à sua contenção (confinamento), identifique o quanto foi incomodado/a por cada um dos seguintes sintomas no período pós confinamento (fase inicial de desconfinamento com restrições impostas): *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione a posição apropriada para cada elemento:

		(1) Ligeiramente, mas não me incomodou muito	(2) Moderadamente, por vezes não era agradável, mas pude suportar	(3) Severamente, incomodava- me muito e difícil de suportar
	(0) De modo algum			

Dormência ou formigueiro	☐	☐	☐	☐
Sensação de calor	☐	☐	☐	☐
Tremores nas pernas	☐	☐	☐	☐
Incapaz de relaxar	☐	☐	☐	☐
Medo que aconteça o pior	☐	☐	☐	☐
Atordoado/a ou tonto/a	☐	☐	☐	☐
Palpitação ou aceleração do coração	☐	☐	☐	☐
Sem equilíbrio	☐	☐	☐	☐
Aterrorizado/a	☐	☐	☐	☐
Nervoso/a	☐	☐	☐	☐
Sensação de sufoco	☐	☐	☐	☐
Tremores nas mãos	☐	☐	☐	☐
Trémulo/a	☐	☐	☐	☐
Medo de perder o controle	☐	☐	☐	☐
Dificuldade de respirar	☐	☐	☐	☐
Medo de morrer	☐	☐	☐	☐

Assustado/a	☐	☐	☐	☐
Indigestão ou desconforto no abdômen	☐	☐	☐	☐
Sensação de desmaio	☐	☐	☐	☐
Rosto afogueado	☐	☐	☐	☐
Suor (não devido ao calor)	☐	☐	☐	☐

Comparativamente ao período pós confinamento (fase inicial de desconfinamento com restrições impostas), identifique o quanto foi incomodado por cada um dos seguintes sintomas atualmente: *

Responda a esta pergunta apenas se as seguintes condições são verdadeiras:

A resposta for 'Aceito' na pergunta '1 [Consentimento]' (Aceita participar voluntariamente neste estudo?)

Por favor, selecione a posição apropriada para cada elemento:

	(0) De modo algum	(1) Ligeiramente, mas não me incomodou muito	(2) Moderadamente, por vezes não era agradável, mas pude suportar	(3) Severamente, incomodava-me muito e difícil de suportar
Dormência ou formigueiro	☐	☐	☐	☐
Sensação de calor	☐	☐	☐	☐

Tremores nas pernas	☐	☐	☐	☐
Incapaz de relaxar	☐	☐	☐	☐
Medo que aconteça o pior	☐	☐	☐	☐
Atordoado/a ou tonto/a	☐	☐	☐	☐
Palpitação ou aceleração do coração	☐	☐	☐	☐
Sem equilíbrio	☐	☐	☐	☐
Aterrorizado/a	☐	☐	☐	☐
Nervoso/a	☐	☐	☐	☐
Sensação de sufoco	☐	☐	☐	☐
Tremores nas mãos	☐	☐	☐	☐
Trémulo/a	☐	☐	☐	☐
Medo de perder o controle	☐	☐	☐	☐
Dificuldade de respirar	☐	☐	☐	☐
Medo de morrer	☐	☐	☐	☐
Assustado/a	☐	☐	☐	☐
Indigestão ou desconforto no	☐	☐	☐	☐

abdómen				
Sensação de desmaio	☐	☐	☐	☐
Rosto afogueado	☐	☐	☐	☐
Suor (não devido ao calor)	☐	☐	☐	☐

**Agradecemos a sua participação neste inquérito.
A sua resposta foi registada.**

Carolina Marques
Professora Doutora Catarina Ferreira da Silva
Professora Doutora Carla Moleiro

Apêndice B – Preparação dos dados: resumo das alterações

Tabela B: Preparação dos dados

Ação	Tratamento	Variável	Classificação	Fórmula
1	Reclassificação	Regiões	<i>Nominal</i>	“Norte”, “Centro”, “Sul”
2	Reclassificação	Idade_TILEN	<i>Ordinal</i>	1: [18, 22[2: [22, 23[3: [23, 25[4: [25, 35[5: [35, 77]
3	Reclassificação e Alteração do tipo da variável	Comparação_BAI_1e2	<i>Integer</i>	0
		Comparação_BAI_1e2		1
	Criação	Pontuação_BAI_1e2	<i>Continuous</i>	2
		Pontuação_BAI_2e3		3
	Classificação	Pontuação_BAI_1e2_classif	<i>Ordinal</i>	Somatório dos valores atribuídos a cada um dos sintomas.
		Pontuação_BAI_2e3_classif		0 ≤ Pontuação ≤ 10: “Grau mínimo de Ansiedade” 11 ≤ Pontuação ≤ 19: “Ansiedade leve” 20 ≤ Pontuação ≤ 30: “Ansiedade moderada” Pontuação > 30: “Ansiedade severa”
4	Criação	Variação_BAI	<i>Continuous</i>	Diferença entre a ‘Pontuação_BAI_2e3’ e a ‘Pontuação_BAI_1e2’
	Classificação	Variação_BAI_classif	<i>Nominal</i>	‘Variação_BAI’ = 0: “Constante” ‘Variação_BAI’ > 0: “Agravamento” ‘Variação_BAI’ < 0: “Melhoria”
5	Reclassificação e Alteração do tipo da variável	Perturbação_Mental	<i>Integer</i>	“Sim”: 1
		Perturbação_Ansiedade		“Não”: 0
	Criação	Diagnostico_P_Mental	<i>Continuous</i>	“N/A”: 0
		Diagnostico_P_Ansiedade		‘Perturbação_Mental’ ≥ 1: 1 ‘Perturbação_Mental’ = 0: 0
				‘Perturbação_Ansiedade’ ≥ 1: 1 ‘Perturbação_Ansiedade’ = 0: 0

Apêndice C – Fluxograma modelo final ramificado [Estudantes]

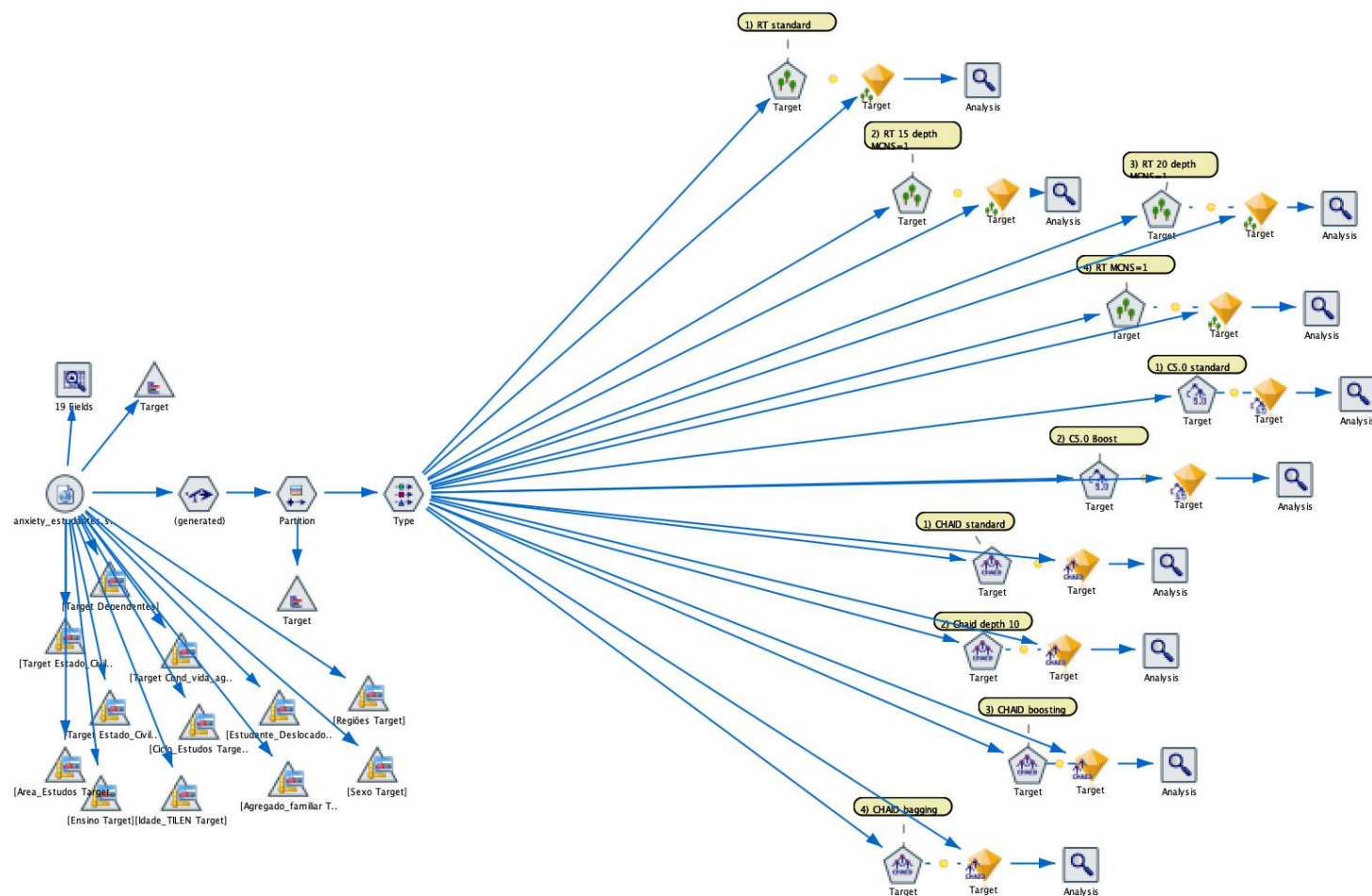


Figura C: Fluxograma modelo final ramificado [Estudiantes]

Apêndice D – Resultados do modelo final ramificado para previsão de níveis de ansiedade [Estudantes]

Tabela. D: Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Estudantes]

		Random Trees								C5.0				CHAID							
		A		B		C		D		E		F		G		H		I		J	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		86%	80,5%	94,2%	88%	95,3%	88,7%	93,5%	90,2%	89,5%	82,7%	96,7%	92,4%	57,1%	59,9%	75,7%	64,4%	92,9%	92,5%	88,7%	84,2%
Grau mínimo	S	80,8%	55,6%	92,2%	83,9%	95%	88,5%	94,2%	90%	92,7%	57,1%	96,9%	92,6%	57,7%	48%	60,8%	50%	86,8%	83,3%	85,7%	100%
	E	92,2%	92,8%	96,2%	91,2%	95,4%	88,8%	94,3%	93,8%	88,7%	87,5%	96,7%	92,4%	86,1%	89%	89,1%	85,9%	95,1%	94,5%	90%	83,5%
	S - E	-11,4%	-37,2%	-4%	-7,3%	-0,4%	-0,3%	-0,1%	-3,8%	4%	-30,4%	0,2%	0,2%	-28,4%	41%	-28,4%	-35,9%	-8,3%	-11,2%	-4,2%	16,5%
Ansiedade leve	S	85,5%	88,5%	94,2%	87,5%	93,4%	80%	94,2%	91,2%	87,7%	77,1%	95,8%	87,2%	59,6%	52,2%	88,2%	59,1%	97,1%	93,3%	83,6%	84,6%
	E	94,8%	91,6%	97,1%	95%	99,5%	99%	98,1%	98%	98,5%	93,9%	100%	100%	83,8%	82,6%	90,5%	81,8%	96,2%	99%	95,5%	98,9%
	S - E	-9,3%	-3,1%	-2,9%	-7,5%	-6,1%	-19%	-3,9%	-6,8%	-10,9%	-16,7%	-4,2%	-12,8%	-24,1%	-30,4%	-2,3%	-22,7%	0,8%	-5,7%	-12%	-14,3%
Ansiedade moderada	S	89,1%	83,3%	93,2%	93,3%	95,9%	93,8%	89,3%	84,4%	89,3%	87,2%	94,4%	96,8%	75,6%	70,6%	78,3%	79,2%	93,1%	89,7%	89,9%	71,8%
	E	94,4%	90,3%	99%	98,1%	99%	98%	100%	100%	99,5%	96,8%	99%	98%	81%	86,1%	88,7%	87%	99%	96,8%	100%	100%
	S - E	-5,4%	-7%	-5,8%	-4,7%	-3,1%	-4,3%	-10,7%	-15,6%	-10,2%	-9,6%	-4,6%	-1,2%	-5,4%	-15,5%	-10,3%	-7,9%	-6%	-7,1%	-10,1%	-28,2%
Ansiedade severa	S	89,0%	95,1%	97%	87,5%	97%	92,5%	96,9%	97,1%	89,2%	97,4%	100%	94,3%	48,1%	73,8%	80,5%	71,7%	95,4%	100%	95,5%	90,7%
	E	100%	100%	100%	100%	100%	100%	99,1%	96%	100%	100%	100%	100%	94,1%	90%	100%	100%	100%	100%	100%	100%
	S - E	-11%	-4,9%	3%	-12,5%	3%	-7,5%	-2,1%	1,1%	-10,8%	-2,6%	0%	-5,7%	46%	-16,2%	-19,5%	-28,3%	-4,6%	0%	-4,5%	-9,3%

Apêndice E – Fluxograma modelo final ramificado [Estudantes/Pandemia]

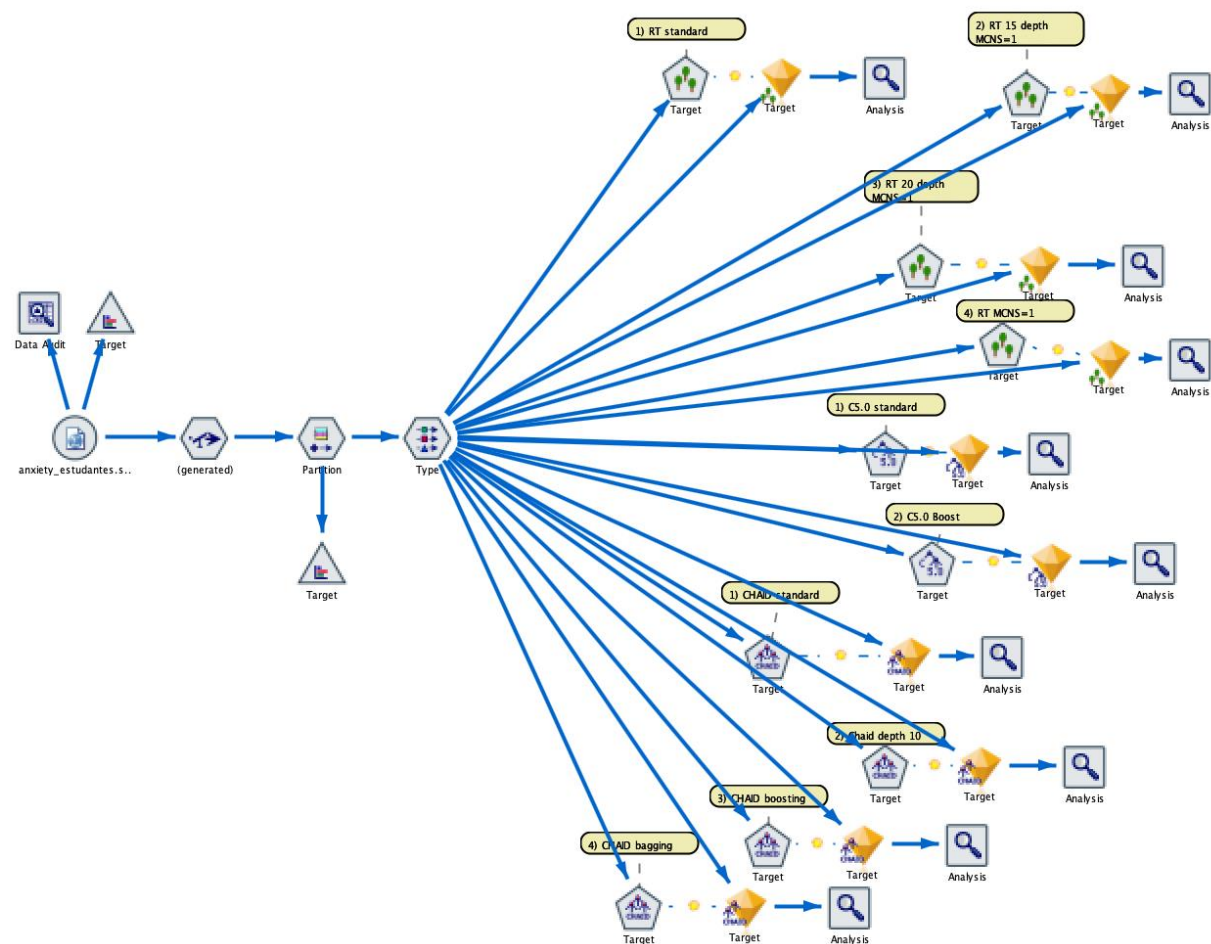


Figura E: Fluxograma modelo final ramificado [Estudantes/Pandemia]

Apêndice F – Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 [Estudantes]

Tabela F: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados [Estudantes]

Random Trees										C5.0				CHAID							
A				B		C		D		E		F		G		H		I		J	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		94,5%	90,4%	92,9%	92,1%	95,3%	87,3%	94,4%	86,5%	88,9%	83,1%	93,1%	92,9%	69%	73,8%	82,4%	78,6%	95,7%	88,8%	88,5%	83,7%
Agravamento	S	92,5%	84%	88,4%	89,4%	91,6%	87,5%	88,7%	84,8%	84,5%	90%	93,4%	90,7%	67,6%	76,5%	78%	72,5%	93,5%	87,2%	87,6%	78,8%
	E	100%	100%	100%	100%	99,4%	94,2%	100%	100%	99,4%	94%	96,4%	100%	90,2%	96%	98,6%	94,7%	100%	100%	96,9%	98,6%
	S - E	-7,5%	-16%	-11,6%	-10,6%	-7,8%	-6,7%	-11,3%	-15,2%	-14,8%	-4%	-3%	-9,3%	-22,6%	-19,5%	-20,7%	-22,1%	-6,5%	-12,8%	-9,3%	-19,7%
Constante	S	91,8%	91,3%	91,9%	89,8%	96,3%	88,2%	96,5%	84,8%	88,2%	81%	87,5%	89,8%	62,9%	66,7%	81,4%	81,6%	94,1%	83,7%	84,5%	81,6%
	E	97,6%	100%	98,2%	98,7%	97,7%	98,7%	97,1%	95%	96,3%	95,1%	100%	100%	96,3%	93,7%	98,1%	98,7%	98,2%	98,7%	96,8%	95,3%
	S - E	-5,9%	-8,7%	-6,4%	-8,9%	-1,4%	-10,4%	-0,6%	-10,2%	-8,1%	-14,2%	-12,5%	-10,2%	-33,4%	-27%	-16,7%	-17,1%	-4,1%	-15%	-12,2%	-13,7%
Melhoria	S	100%	100%	100%	100%	98,8%	85,7%	98,6%	91,2%	96,8%	78,6%	100%	100%	100%	100%	95,6%	84,6%	100%	100%	95,4%	93,9%
	E	94,4%	87,5%	91,7%	90,6%	96%	89%	94,5%	85,9%	88,9%	85,4%	93,6%	90,2%	73,8%	77,2%	81%	78%	95,5%	86,5%	89,8%	83,3%
	S - E	5,6%	12,5%	8,3%	9,4%	2,7%	-3,3%	4,1%	5,3%	7,9%	-6,8%	6,4%	9,8%	26,2%	22,8%	14,5%	6,6%	4,5%	13,5%	5,6%	10,6%

Apêndice G – Fluxograma do modelo final ramificado [Trabalhadores-Estudante]

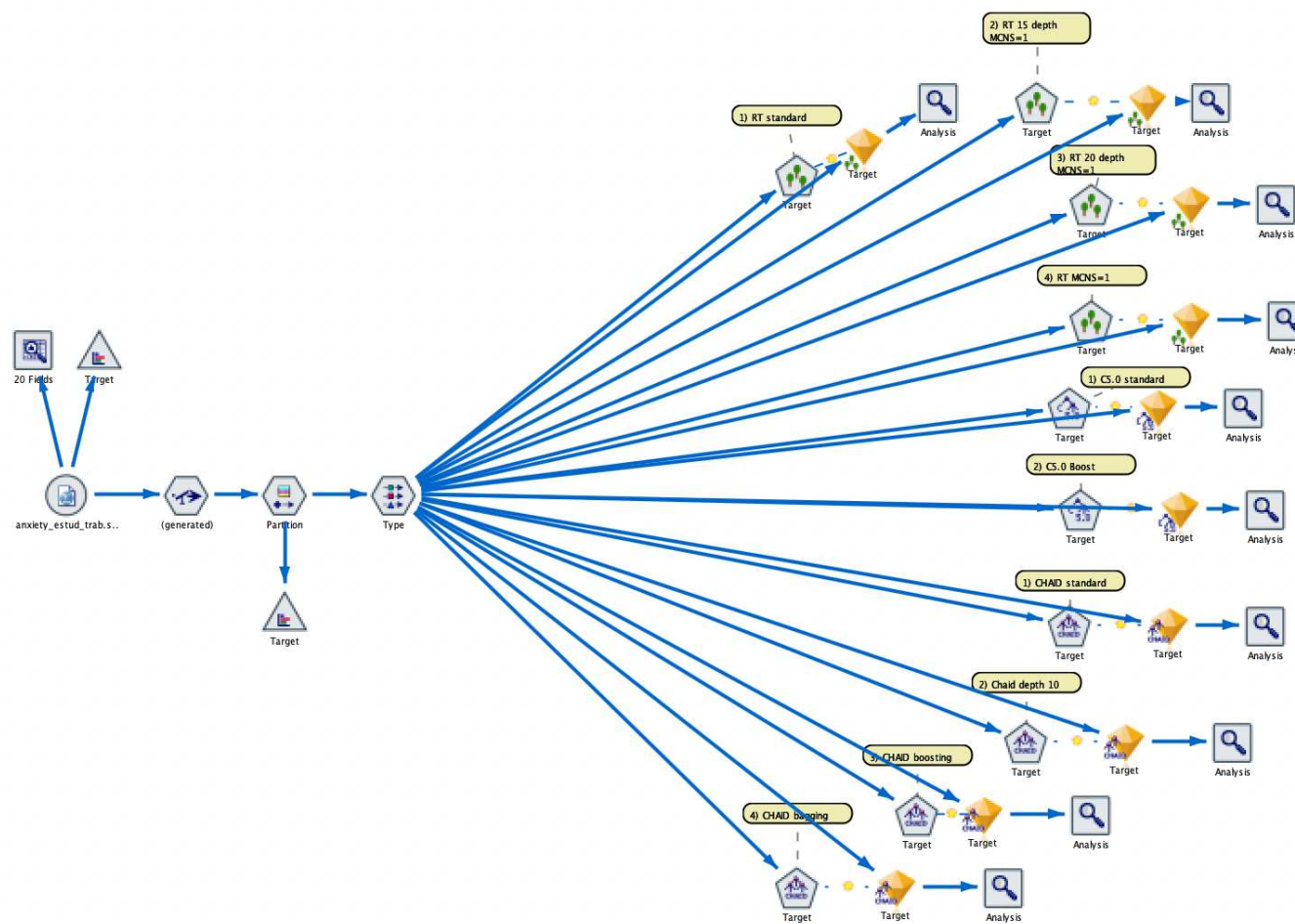


Figura G: Fluxograma modelo final ramificado [Trabalhadores-Estudante]

Apêndice H – Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Trabalhadores-Estudante]

Tabela H: Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Trabalhadores-Estudante]

		Random Trees								C5.0				CHAID							
		A		B		C		D		E		F		G		H		I		J	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		95,3%	94,2%	94,2%	97,1%	96,5%	95,6%	98,9%	91,3%	89,5%	92%	97,6%	94,9%	87,1%	85,6%	91,5%	87,7%	96,9%	95,7%	92,1%	93,5%
Grau mínimo	S	100%	100%	100%	100%	100%	100%	100%	87,1%	96,2%	87,5%	100%	100%	95%	88,9%	96,2%	87%	100%	100%	100%	100%
	E	94,1%	92,8%	92,9%	96,3%	95,5%	94,6%	98,7%	92,5%	88%	93%	97%	93,4%	85,8%	85,1%	90,5%	87,8%	96%	94,7%	90,3%	92,3%
	S - E	5,9%	7,2%	7,1%	3,7%	4,5%	5,4%	1,3%	-5,4%	8,1%	-5,5%	3%	6,6%	9,2%	3,8%	5,8%	-0,9%	4%	5,3%	9,7%	7,7%
Ansiedade leve	S	89,9%	90,3%	87%	96,6%	96%	88,6%	97,3%	84,8%	79,3%	90,6%	97,6%	86,2%	83%	73,5%	87,9%	74,2%	93,7%	97,3%	81,9%	97,1%
	E	100%	100%	100%	100%	100%	100%	100%	96,2%	99%	97,2%	100%	100%	99%	98,1%	99%	97,2%	100%	100%	100%	100%
	S - E	-10,1%	-9,7%	-13%	-3,4%	-4%	-11,4%	-2,7%	-11,3%	-19,7%	-6,5%	-2,4%	-13,8%	-16%	-24,6%	-11,1%	-23%	-6,3%	-2,7%	-18,1%	-2,9%
Ansiedade moderada	S	95,8%	95,2%	97,2%	95,1%	97,2%	97,4%	100%	92,5%	95,8%	100%	97,2%	92,7%	88,2%	87,5%	88,3%	91,3%	94,4%	95%	95,8%	88,9%
	E	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	S - E	-4,2%	-4,8%	-2,8%	-4,9%	-2,8%	-2,6%	0%	-7,5%	-4,2%	0%	-2,8%	-7,3%	-11,8%	-12,5%	-11,7%	-8,7%	-5,6%	-5%	-4,2%	-11,1%
Ansiedade severa	S	97,4%	92,1%	96%	97,3%	93,5%	97,3%	98,7%	100%	91%	88,4%	96%	100%	86,9%	92,3%	95,9%	94,7%	100%	91,7%	93,9%	91,9%
	E	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	S - E	-2,6%	-7,9%	-4%	-2,7%	-6,5%	-2,7%	-1,3%	0%	-9%	-11,6%	-4%	0%	-13,1%	-7,7%	-4,1%	-5,3%	0%	-8,3%	-6,1%	-8,1%

Apêndice I – Fluxograma modelo final ramificado [Trabalhadores-estudante/Pandemia]

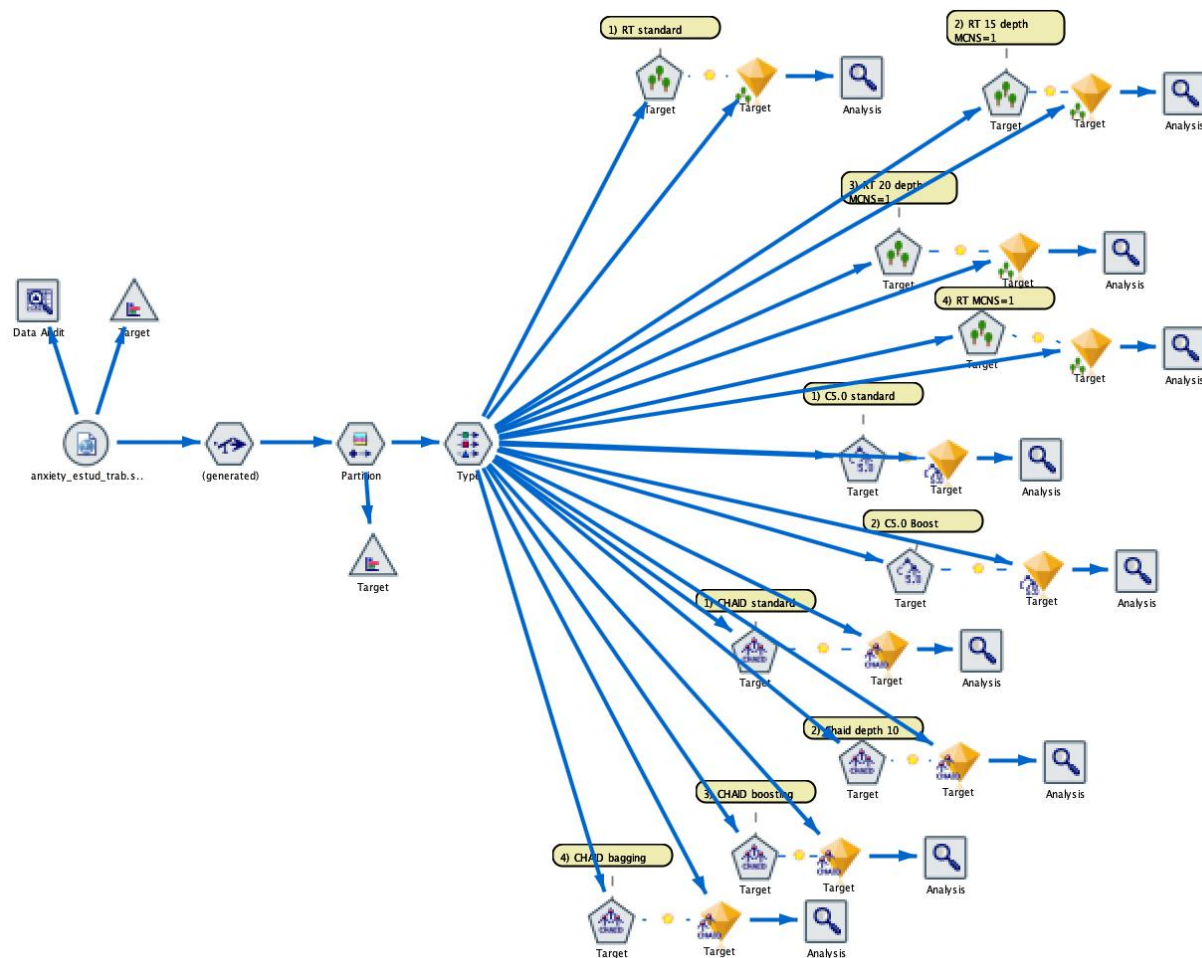


Figura .1: Fluxograma modelo final ramificado [Trabalhadores-Estudante/Pandemia]

Apêndice J: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 [Trabalhadores-Estudante]

Tabela J: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados [Trabalhadores-estudante]

Random Trees										C5.0				CHAID							
A				B		C		D		E		F		G		H		I		J	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		93,1%	79,8%	95,8%	86,3%	95,8%	85,9%	97,2%	84,2%	86,5%	84,7%	95,8%	88,8%	84,2%	75,5%	84,7%	75,5%	98,2%	91,9%	88,9%	80%
Agravamento	S	98,7	82,9%	96,3%	85,3%	96,2%	87,9%	97,4%	88,2%	92,2%	96,9%	98,7%	92,9%	91,1%	85,7%	91,9%	80%	97,5%	88,9%	94%	85,3%
	E	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	S - E	-1,3%	-17,1%	-3,7%	-14,7%	-3,8%	-12,1%	-2,6%	-11,8%	-7,8%	-3,1%	-1,3%	-7,1%	-8,9%	-14,3%	-8,1%	-20%	-2,5%	-11,1%	-6%	-14,7%
Constante	S	86,1%	77,8%	94,7%	82,9%	91,7%	83,7%	95,9%	76,7%	81,2%	73,5%	92,4%	84,4%	79,7%	71%	76,2%	70%	97,3%	94,4%	84%	77,1%
	E	97,8%	90,5%	98,6%	95,5%	100%	94,6%	99,3%	96,6%	94,6%	92,2%	98,5%	93,9%	91,9%	85,1%	94,7%	88,7%	100%	95,2%	95%	89,2%
	S - E	-11,7%	-12,7%	-3,8%	-12,7	-8,3%	-10,9%	-3,4%	-19,8%	-13,4%	-18,7%	-6,1%	-9,6%	-12,2%	-14,1%	-18,5%	-18,7%	-2,7%	-0,8%	-11%	-12,1%
Melhoria	S	95%	78,6%	96,6%	90,9%	100%	87%	98,5%	91,7%	86,8%	84,4%	96,6%	89,5%	80,7%	68,8%	87%	77,8%	100%	91,7%	87,9%	77,4%
	E	92,3%	80,3%	95,5%	84,1%	94%	85,5%	96,7%	81,8%	86,4%	84,8%	95,5%	88,3%	85,4%	78,8%	84,7%	76%	97,4%	92,1%	89,2%	81,2%
	S - E	2,7%	-1,7%	1,1%	6,9%	6%	1,4%	1,8%	9,8%	0,4%	-0,5%	1,1%	1,2%	-4,7%	-10%	2,3%	1,8%	2,6%	-0,4%	-1,3%	-3,7%

Apêndice K – Fluxograma modelo final ramificado [Trabalhadores]

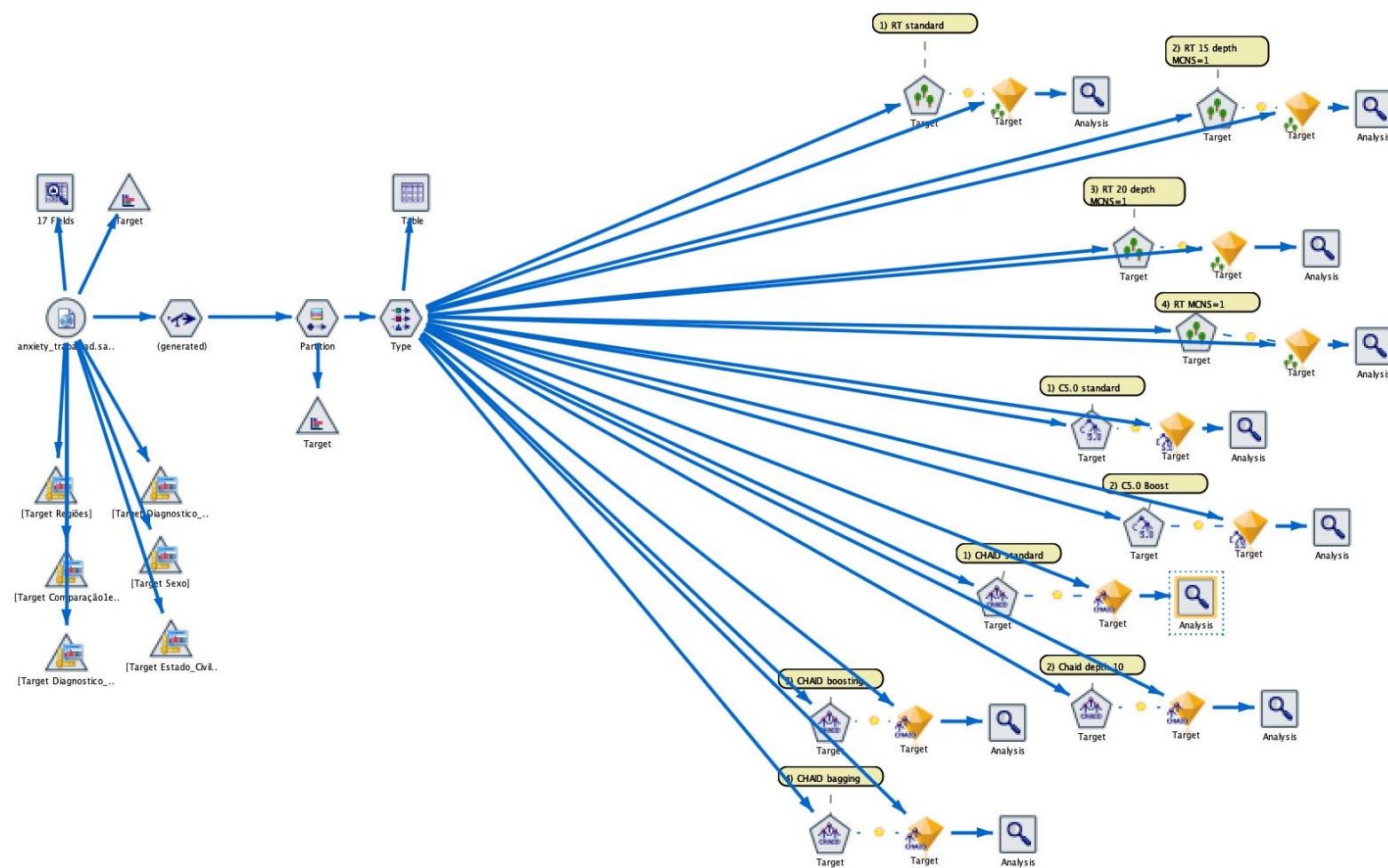


Figura K: Fluxograma modelo final ramificado [Trabalhadores]

Apêndice L – Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Trabalhadores]

Tabela L: Resultados do modelo final ramificado para previsão dos níveis de ansiedade [Trabalhadores]

Random Trees										C5.0				CHAID								
A				B		C		D		E		F		G		H		I		J		
		Treino		Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		94,7%	89,7%		95,4%	94,4%	96,3%	94,8%	95,7%	95,8%	93,4%	87,3%	98,1%	94,4%	76,6%	68,1%	77,4%	70%	95,5%	93,9%	84,7%	82,6%
Grau mínimo	S	95,8%	100%		100%	100%	100%	100%	100%	100%	94,8%	90,5%	100%	100%	88,1%	72%	78,3%	83,3%	94,7%	94,1%	84%	90,2%
	E	94,5%	86,9%		94,1%	93,1%	95,3%	93,3%	94,5%	95%	93%	86,5%	97,5%	93%	81,5%	78,7%	84,9%	82,5%	95,8%	93,8%	87,6%	83,7%
	S - E	1,4%	13,1%		5,9%	6,9%	4,7%	6,7%	5,5%	5%	1,8%	3,9%	2,5%	7%	6,6%	-6,7%	-6,7%	0,8%	-1,1%	0,3%	-3,6%	6,5%
Ansiedade leve	S	93,5%	78,6%		91%	88,1%	92,6%	90,9%	94,1%	96,5%	90,1%	84,6%	97,1%	92,5%	74,7%	58%	76,5%	56,3%	94,9%	94,2%	76,3%	74,1%
	E	98,9%	100%		100%	100%	100%	100%	100%	100%	98,7%	97,5%	100%	100%	94,1%	90,2%	91,1%	87,9%	100%	100%	94,8%	96,2%
	S - E	-5,4%	-21,4%		-9%	-11,9%	-7,4%	-9,1%	-5,9%	-3,5%	-8,6%	-12,9%	-2,9%	-7,5%	-19,4%	-32,2%	-14,6%	-31,6%	-5,1%	-5,8%	-18,6%	-22,2%
Ansiedade moderada	S	92,2%	87,8%		96,1%	95,3%	97,6%	94,1%	92,2%	95,7%	93,4%	85,7%	97,6%	90,8%	75%	68,2%	75%	74,2%	95,2%	91,5%	89,1%	83,9%
	E	100%	100%		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	96,7%	91,8%	97%	95,9%	98,5%	98,1%	98,4%	98%
	S - E	-7,8%	-12,2%		-3,9%	-4,7%	-2,4%	-5,9%	-7,8%	-4,3%	-6,6%	-14,3%	-2,4%	-9,2%	-21,7%	-23,7%	-22%	-21,7%	-3,3%	-6,5%	-9,4%	-14,1%
Ansiedade severa	S	97,8%	92,3%		95,7%	96,2%	96,3%	94,2%	97%	92,5%	95,7%	89,3%	97,7%	96,3%	77%	73,6%	80,1%	68,3%	97,1%	96,1%	90,4%	83,9%
	E	100%	100%		100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	98,9%	95,3%	100%	100%	100%	100%
	S - E	-2,2%	-7,7%		-4,3%	-3,8%	-3,7%	-5,8%	-3%	-7,5%	-4,3%	-10,7%	-2,3%	-3,7%	-23%	-26,4%	-18,7%	-27,1%	-2,9%	-3,9%	-9,6%	-16,1%

Apêndice M – Fluxograma modelo final ramificado [Trabalhadores/Pandemia]

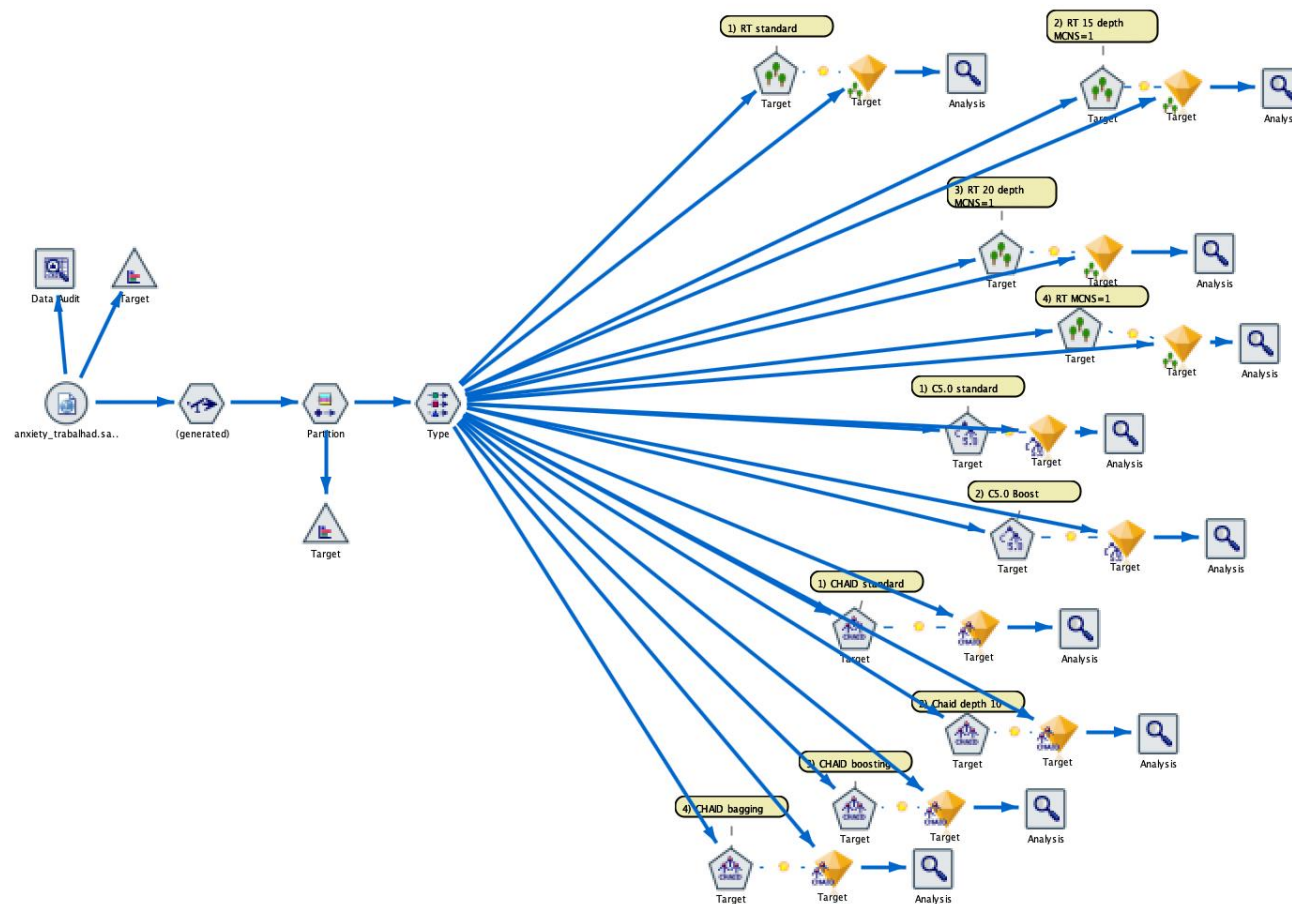


Figura M: Fluxograma modelo final ramificado [Trabalhadores/Pandemia]

Apêndice N – Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 [Trabalhadores/Pandemia]

Tabela N: Resultados do modelo final ramificado para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados [Trabalhadores/Pandemia]

Random Trees										C5.0				CHAID							
A				B		C		D		E		F		G		H		I		J	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		85,3%	84%	87,6%	88,4%	87,3%	84,3%	87%	84,1%	79,1%	82,2%	88,6%	76%	66,7%	68,3%	72,9%	68,5%	89,4%	77,4%	78,7%	77,2%
Agravamento	S	88,7%	84,6%	93,5%	94,4%	95,6%	91,8%	91,9%	89,7%	79,7%	92,7%	87,7%	89,7%	72,7%	85,4%	84,2%	84,8%	96,4%	93,9%	85,1%	84,7%
	E	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	94,9%	94,8%	97,4%	91%	100%	100%	100%	100%
	S - E	-11,3%	-15,4%	-6,5%	-5,6%	-4,4%	-8,2%	-8,1%	-10,3%	-20,3%	-7,3%	-12,3%	-10,3%	-22,2%	-9,4%	-13,3%	-6,2%	-3,6%	-6,1%	-14,9%	-15,3%
Constante	S	79,6%	79,6%	84,5%	89,1%	83,5%	74,1%	79,6%	81,6%	79,4%	71,7%	92,2%	64,4%	58,3%	66%	61,7%	58,6%	85,7%	65,9%	69,1%	74,4%
	E	91,9%	88,4%	91%	90%	91,3%	94,3%	93,7%	87,5%	86,9%	90,3%	90,6%	86,1%	81,6%	78,9%	92,1%	89,5%	93,5%	84,8%	87%	84,3%
	S - E	-12,3%	-8,8%	-6,5%	-0,9%	-7,8%	-20,2%	-14%	-5,9%	-7,4%	-18,6%	1,6%	-21,7%	-23,3%	-12,9%	-30,4%	-30,9%	-7,8%	-18,9%	-17,9%	-9,9%
Melhoria	S	87,1%	88,4%	84,6%	80,4%	81,9%	89,7%	89,4%	78,9%	77,6%	81,6%	85,9%	69,8%	67,6%	53,2%	80%	66,7%	85,2%	71,4%	82,3%	69,8%
	E	87,4%	88,1%	90,5%	93%	90,1%	83,2%	87,4%	89,7%	83,5%	84,3%	92,6%	79,6%	76,2%	78,6%	75,1%	75%	91%	81,1%	82,7%	83,3%
	S - E	-0,3%	0,3%	-5,9%	-12,6%	-8,2%	6,6%	2%	-10,8%	-5,8%	-2,7%	-6,8%	-9,8%	-8,6%	-25,4%	4,9%	-8,3%	-5,8%	-9,7%	-0,4%	-13,6%

Apêndice O – Fluxograma modelo final geral

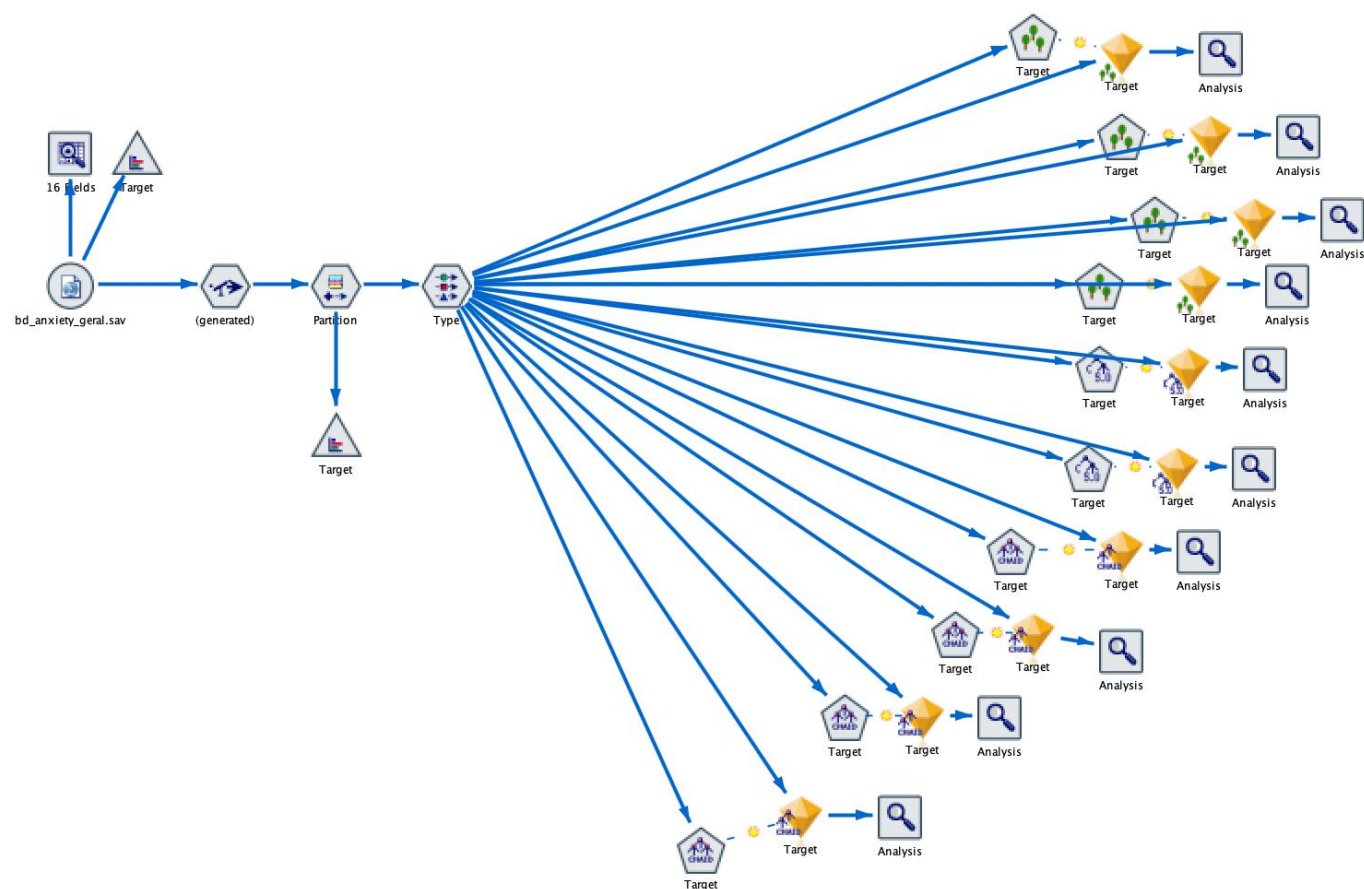


Figura O: Fluxograma modelo final geral para previsão de níveis de ansiedade

Apêndice P – Fluxograma modelo final geral [Impacto da pandemia COVID-19]

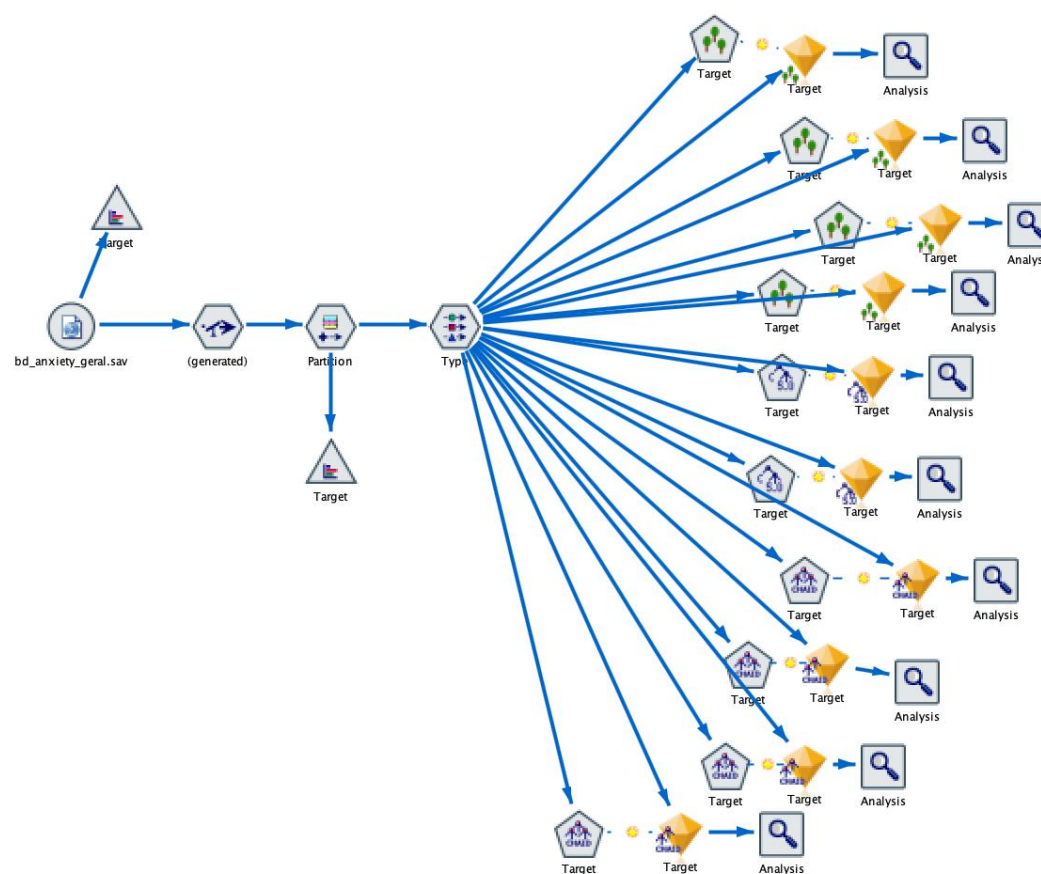


Figura P: Fluxograma modelo final geral para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados

Apêndice Q – Resultados do modelo final geral

Tabela Q.1: Resultados do modelo final geral para previsão dos níveis de ansiedade

Random Trees										C5.0				CHAID							
A				B		C		D		E		F		G		H		I		J	
Treino			Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		79,6%	75,2%	83,2%	80,41	93,2%	90,7%	77,7%	73,8%	77,5%	77,8%	82%	76,5%	58,7%	51,8%	59%	53,3%	69,6%	66,3%	66,2%	62,9%
Grau mínimo	S	91,1%	81,1%	95,5%	88,5%	93,3%	84,3%	92,5%	81,8%	85,2%	86,8%	89,3%	82,9%	59%	58,6%	63,1%	55,7%	61%	53,3%	62,4%	55,2%
	E	87,5%	86,3%	88,7%	87,5%	94,8%	93%	87,4%	86,8%	86,4%	83,6%	89%	87,8%	80,2%	79,4%	81%	81,1%	85,6%	85,1%	83,6%	86,4%
	S - E	3,6%	-5,3%	6,8%	1%	-1,5%	-8,6%	5,1%	-4,9%	-1,2%	3,1%	0,3%	-4,9%	-21,2%	-20,8%	-17,9%	-25,4%	-24,6%	-31,8%	-21,1%	-31,2%
Ansiedade leve	S	75,7%	69,6%	79,7%	70,6%	92,2%	84,8%	72%	71,4%	71,4%	70,5%	80,7%	74,8%	55,6%	45,4%	52,6%	48,1%	65,7%	63%	56,3%	63,4%
	E	93%	92%	94,5%	93,5%	96,7%	96,1%	92,5%	91,2%	90,1%	93,1%	92,2%	90,7%	86,7%	85,3%	85,8%	81%	87,1%	85,2%	84,1%	84,2%
	S - E	-17,3%	-22,3%	-14,9%	-22,9%	-4,5%	-11,3%	-20,5%	-19,8%	-18,7%	-22,5%	-11,5%	-15,9%	-31,1%	-39,9%	-33,2%	-32,9%	-21,4%	-22,1%	-27,9%	-20,8%
Ansiedade moderada	S	73,3%	75,9%	82,7%	80,6%	93,4%	94,5%	70,6%	67,3%	78,1%	80%	84,1%	78,1%	57,9%	49,7%	59,1%	51,2%	71,8%	70%	64,3%	63,2%
	E	96,2%	93,3%	96,7%	95,2%	99,6%	98,9%	95,5%	93,5%	96,5%	97,1%	96,1%	93%	87,5%	86,9%	87,1%	86,4%	93,5%	92,2%	89,5%	87,4%
	S - E	-22,9%	-17,4%	-14%	-14,6%	-6,3%	-4,4%	-24,9%	-26,1%	-18,4%	-17,1%	-12%	-14,9%	-29,6%	-37,2%	-28%	-35,2%	-21,7%	-22,2%	-25,2%	-24,2%
Ansiedade severa	S	83%	76,9%	79,9%	82,5%	94%	97,7%	82,1%	78,9%	78,4%	77,7%	76,8%	73,5%	62,5%	57,1%	62,5%	58,5%	77,9%	78,5%	66,9%	67,7%
	E	97,1%	96,3%	98,6%	98,5%	100%	100%	95,8%	94,4%	98,4%	98,5%	99,6%	98,5%	92%	85,4%	92,9%	90%	93,7%	92,7%	95,1%	92,9%
	S - E	-14,2%	-19,4%	-18,7%	-16%	-6%	-2,3%	-13,7%	-15,5%	-20%	-20,8%	-22,8%	-25%	-29,5%	-28,3%	-30,4%	-31,5%	-15,8%	-14,2%	-28,2%	-25,2%

Tabela Q.2: Resultados do modelo final geral para avaliação do impacto da pandemia COVID-19 nos níveis de ansiedade associados

Random Trees										C5.0				CHAID							
A				B		C		D		E		F		G		H		I		J	
		Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste	Treino	Teste
PCCC		70,4%	74,5%	78,6%	76,8%	76,2%	76,2%	79,3%	78,6%	73,2%	77,3%	81,8%	72,4%	49,8%	46,8%	49,1%	51%	69,3%	69,3%	65,4%	62,6%
Agravamento	S	67,2%	78,3%	76,4%	79,4%	73,6%	75%	77,1%	82,5%	72,9%	81,2%	80,1%	70,4%	60,3%	56,1%	42,6%	53,8%	82%	85,3%	64%	59,3%
	E	95,9%	92,7%	97,6%	96,3%	97,9%	98,3%	98,8%	93,8%	95,5%	97%	97,9%	95,4%	75,2%	70,5%	82,7%	71,3%	89,1%	86,3%	94,9%	92,7%
	S - E	-28,7%	-14,4%	-21,2%	-16,9%	-24,3%	-23,3%	-21,7%	-11,3%	-22,6%	-15,8%	-17,8%	-25%	-14,9%	-14,4%	-40,1%	-17,5%	-7,1%	-0,9%	-30,9%	-33,4%
Constante	S	74,3%	74,5%	78%	72,8%	77%	78,4%	77,4%	73%	77,3%	76,2%	82,1%	69,4%	57,4%	62,5%	67,7%	59,1%	69,2%	69,9%	64,9%	73,2%
	E	83,9%	88,7%	90,1%	91,1%	86,4%	89,1%	90,4%	91,5%	83,8%	89,1%	91,1%	89,3%	74,3%	78,6%	76,3%	81,3%	81,4%	84%	82,6%	83,2%
	S - E	-9,6%	-14,2%	-12,1%	-18,3%	-9,5%	-10,6%	-13%	-18,6%	-6,5%	-12,9%	-9,1%	-19,9%	-16,9%	-16,1%	-8,6%	-22,2%	-12,2%	-14,1%	-17,7%	-10%
Melhoria	S	71,1%	66,7%	83,2%	77,3%	79,3%	76%	85,3%	80%	69,1%	72,3%	84,4%	81,9%	41,7%	35,9%	44,6%	35,8%	56,8%	57,1%	69,9%	56,7%
	E	77,9%	82,2%	81,9%	80,7%	81,8%	80,7%	81,4%	84,4%	81,6%	81,9%	85,1%	77,2%	75,3%	69,8%	67,6%	72,7%	81,6%	83,7%	74,9%	73,1%
	S - E	-6,8%	-15,5%	1,2%	-3,3%	-2,4%	-4,7%	3,9%	-4,4%	-12,6%	-9,5%	-0,8%	4,8%	-33,7%	-33,9%	-23%	-36,9%	-24,8%	-26,5%	-4,9%	-16,4%