



UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Modelagem e Aplicações de Materiais em Biociências
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS EM BIOCÊNCIAS
CAMPUS TOLEDO

Kelly Chapla

**Identificação de possíveis relações entre o uso de
agrotóxicos e os problemas de saúde reportados por
municípios de Santa Tereza do Oeste - PR**

TOLEDO
2023

Kelly Chapla

Identificação de possíveis relações entre o uso de agrotóxicos e os problemas de saúde reportados por munícipes de Santa Tereza do Oeste - PR

Identification of possible relationships between the use of pesticides and health problems reported by residents of Santa Tereza do Oeste - PR

Trabalho de Dissertação apresentado como requisito para obtenção do título de Mestre em Tecnologias em Biociências, do Programa de Pós-Graduação em Tecnologias em Biociências – PPGBio da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Jefferson Gustavo Martins, Dr.

TOLEDO
2023



Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



**Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Toledo**



KELLY CHAPLA

**IDENTIFICAÇÃO DE POSSÍVEIS RELAÇÕES ENTRE O USO DE AGROTÓXICOS E SEU IMPACTO NA
SAÚDE PÚBLICA E AMBIENTAL: ESTUDO DE CASO NA CIDADE DE SANTA TEREZA DO OESTE**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestra Em Tecnologias Em Biociências da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Tecnologias Em Biociências.

Data de aprovação: 01 de Março de 2023

Dr. Jefferson Gustavo Martins, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Alberto Yoshihiro Nakano, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Ana Tereza Bittencourt Guimaraes, Doutorado - Universidade Estadual do Oeste do Paraná (Unioeste)

Dr. Vilson Luiz Dalle Mole, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 11/04/2023.

Dedico esse trabalho à minha família pela compreensão nos momentos de ausência, e em especial a meu esposo que sempre me apoiou e incentivou nessa jornada.

AGRADECIMENTOS

Agradeço ao meu orientador Professor Doutor Jefferson Gustavo Martins pelas orientações, disponibilidade, ensinamentos, sugestões e críticas construtivas, repasse de conhecimento, apoio, compreensão, paciência e dedicação durante todo o desenvolvimento do trabalho.

Agradeço a Professora Doutora Ana Tereza Bittencourt Guimarães pela fundamental parceria, disponibilidade e auxílio sempre que necessário, prontidão e boa vontade e pela gentileza de compartilhar pensamentos, informações e experiências e imensa disseminação de conhecimento, pensando e agindo de forma altruísta.

Aos meus colegas pela ajuda e troca de ideias sempre que possível.

A Secretaria do Curso, e aos coordenadores Prof. Dr. Renato Einseng e atualmente Prof. Dr. Eduardo Vinicius Kuhn pela cooperação e agilidade em cada etapa.

Agradeço a Deus pela luz colocada em nossos caminhos a cada dia! Ao imenso coração de meus pais e suporte que me deram durante essa jornada! Aos meus sogros e cunhada pela disponibilidade em ajudar sempre que precisei. Ao meu esposo que está sempre ao meu lado, me apoiando e incentivando e ao filho Ravi Lucca que desde o meu ventre já me acompanhava nos estudos.

Enfim, a todos que de alguma forma contribuíram para a realização desta pesquisa. Peço desculpas aos que não estão presentes nestas palavras, mas podem estar certos de que fazem parte do meu pensamento e de minha gratidão.

É impossível para um homem aprender aquilo que ele acha que sabe.

Epíteto

RESUMO

Os agrotóxicos são empregados para controle de doenças e pragas provocadas por vetores como insetos, larvas, fungos e carrapatos. Este cenário é caracterizado tanto nos ambientes rurais quanto nos urbanos, mas principalmente decorre das atividades agrícolas com foco no aumento do rendimento na produção de alimentos. O emprego de agrotóxicos tem gerado diversos problemas ambientais e de saúde pública, tema que produziu uma parceria entre a Universidade Estadual do Oeste do Paraná (UNIOESTE) e a Universidade CARDIFF concretizada por um projeto com o objetivo de analisar como a população e os ecossistemas reagem à aplicação de agrotóxicos por longos períodos de tempo. Dentre os resultados do projeto citado, tem-se a criação de uma base de dados coletados por meio de questionários aplicados a uma amostra da população do município de Santa Tereza do Oeste - PR e também pelos resultados de análises de amostras de solo, plantas e animais. O presente estudo emprega técnicas de mineração de dados na base criada pelo projeto citado, por meio da ferramenta computacional WEKA (*Waikato Environment for Knowledge Analysis*), na tentativa de identificar possíveis relações entre o uso de agrotóxicos e problemas de saúde reportados pelos munícipes de Santa Tereza do Oeste - PR. Com base nos resultados obtidos, surgiu um questionamento quanto à forma como é utilizado o questionário do *Self-Reporting Questionnaire 20 items* (SRQ-20), recomendado pela Organização Mundial da Saúde (OMS) para avaliação de desordens psiquiátricas. Em geral, os resultados gerais corroboram com aqueles produzidos por outros estudos relacionados ao tema, ao mesmo tempo em que os resultados individuais de cada técnica de mineração de dados empregadas convergem entre si. Dentre outros, identificou-se a relação do contato com agrotóxicos com problemas de fadiga, neurológicos e gastrointestinais.

Palavras-chave: exposição a agrotóxicos; impactos na saúde e ambiente; população rural e urbana; agricultura.

ABSTRACT

Pesticides are used to control diseases caused by vectors such as insects, larvae, fungi and ticks in rural and urban environments, and especially in agricultural activities to increase income in world food production. However, several environmental and public health problems are associated with the use of pesticides. This case study aimed to understand the public health and environmental scenario of Santa Tereza do Oeste-PR. The information was obtained by means of questionnaires applied to the rural and urban population based on the Protocol for the Assessment of Chronic Poisoning by Pesticides of the Health Department of Paraná and from soil samples were also collected for analysis and evaluation of their characteristics and levels of toxicity, originating from a partial product of a project developed by the CARDIFF University. Based on the analysis of these collections, this study employs data mining techniques, using the computational tool WEKA (*Waikato Environment for Knowledge Analysis*), in an attempt to identify possible relationships between the use of pesticides and health problems reported by residents of Santa Tereza do Oeste - PR. Based on the results obtained, a question arose as to how the Self-Reporting Questionnaire 20 items (SRQ-20) is used, recommended by the World Health Organization (WHO) for the assessment of psychiatric disorders. In general, the general results corroborate those produced by other studies related to the subject, while the individual results of each employed data mining technique converge with each other. Among others, the relationship between contact with pesticides and fatigue, neurological and gastrointestinal problems was identified.

Keywords: exposure to pesticides; impacts on health and environment; rural and urban population; agriculture.

LISTA DE ILUSTRAÇÕES

Figura 1. Representação do processo de KDD (Fayyad, et al. 1996).....	30
Figura 2. Representação da etapa de Data Mining como parte do processo de descoberta de conhecimento em base de dados (KDD).	30
Figura 3. Modelos de análises de Agrupamento. (A) Gráfico de dispersão de dados não categorizados. (B) Agrupamento utilizando o algoritmo K-means com 2 clusters (k=2). (C) Agrupamento utilizando 3 clusters (k=3). (D). Agrupamento utilizando 4 clusters (k=4).	34
Figura 4. Exemplo de Árvore de Decisão.....	35
Figura 5. Mapa de pontos de amostragem do solo	43
Figura 6. Definição dos intervalos utilizados para a definição das classes (rótulos) e a aplicação das técnicas de mineração de dados.	46
Figura 7. Definição dos conjuntos de Treinamento e Teste utilizados nos experimentos.	49
Figura 8. Matriz de Confusão 2X2.....	51

LISTA DE QUADROS

Quadro 1. Exemplo de Árvore de Decisão gerada com a utilização do algoritmo J48 e os dados relacionados ao Saúde (SQR-20).....	37
Quadro 2. Exemplos de Regras de Associação obtidas com a utilização do algoritmo Apriori e os dados relacionados ao Saúde (SQR-20).....	38
Quadro 3. Exemplo de arquivo ARFF do WEKA.	41
Quadro 4. Recorte das questões do Apêndice B relacionadas à Saúde – SQR-20.....	53
Quadro 5. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48: Saúde – SQR-20.	56
Quadro 6. Regras obtidas com a utilização do algoritmo Apriori: Saúde – SQR-20.	57
Quadro 7. Regras obtidas com a utilização do algoritmo Apriori para instâncias da classe “sim”: Saúde – SQR-20.....	58
Quadro 8. Recorte de questões do Apêndice B relacionadas à Saúde: doenças.....	62
Quadro 9. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48 – Execução A: Saúde (doenças).....	65
Quadro 10. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48 – Execução B: Saúde (doenças).....	66
Quadro 11. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48 – Execução C: Saúde (doenças).	66
Quadro 12. Resultados obtidos com a utilização do algoritmo Apriori: Saúde (doenças).	68
Quadro 13. Resultados obtidos com a utilização do algoritmo Apriori considerando apenas instâncias pertencentes à classe “sim”: Saúde e Doenças.	69
Quadro 14. Recorte de questões do Apêndice B relacionadas à SRQ-20 e Saúde.	73
Quadro 15. Resultados obtidos com a utilização do algoritmo Apriori: SRQ-20 e Doenças.	75
Quadro 16. Resultados obtidos com a utilização do algoritmo Apriori considerando apenas instâncias pertencentes à classe “sim”: SRQ-20 e Doenças.	76
Quadro 17. Regras de Árvore de Decisão da execução dos testes obtidos com utilização do algoritmo J48: SRQ-20 e Doenças.	78
Quadro 18. Classificação de itens do Apêndice A relacionadas à Solo.....	82

LISTA DE TABELAS

Tabela 1. Resultados obtidos com o número de clusteres determinado pelo EM: Saúde – SQR-20.....	54
Tabela 2. Resultados médios das execuções realizadas com o Algoritmo J48: Saúde – SQR-20.....	55
Tabela 3. Relação dos atributos selecionados: Saúde – SQR-20.	59
Tabela 4. Resultados obtidos com clusteres determinados pelo próprio algoritmo EM: Saúde (doenças).	63
Tabela 5. Resultados médios das execuções realizadas com o algoritmo J48: Saúde (doenças).....	65
Tabela 6. Atributos selecionados: Saúde (doenças).	70
Tabela 7. Resultados obtidos com clusteres determinados pelo próprio EM: SRQ-20 e Doenças.	74
Tabela 8. Atributos selecionados: SRQ-20 e Doenças.....	77
Tabela 9. Resultados médios das execuções realizadas com o algoritmo J48: SRQ-20 e Doenças.	79
Tabela 10. Atributos selecionados com Diferentes Algoritmos para dados gerais de solo considerando os valores numéricos originais dos atributos.	83
Tabela 11. Atributos selecionados com Diferentes Algoritmos para dados gerais de solo considerando os valores simbólicos definidos por meio de intervalos iguais dos atributos.....	83
Tabela 12. Atributos selecionados com Diferentes Algoritmos para dados de solo (pesticidas).	84
Tabela 13. Atributos selecionados com Diferentes Algoritmos para dados de solo.....	85

LISTA DE ABREVIATURAS E SIGLAS

EPI	Equipamento de Proteção Individual
SINITOX	Sistema nacional de informações tóxico-farmacológicas
CCI	Centro de Controle de Intoxicações de Maringá
PNDA	Programa Nacional de Defensivos Agrícolas
SIAGRO	Sistema de Monitoramento do Comércio e Uso de Agrotóxicos no Paraná
ADAPAR	Agência de Defesa Agropecuária do Paraná
SESA	Secretaria de Saúde do Estado do Paraná
SINAN	Sistema nacional de agravos de notificação
SUREHMA	Superintendência de Recursos Hídricos e Meio Ambiente
INCA	Instituto Nacional de Câncer
OMS	Organização Mundial da Saúde
EM	<i>Expectation-Maximization</i>
NA	Não se Aplica
CSV	<i>Comma-Separated Values</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Objetivo Geral	15
1.2 Objetivos Específicos	16
1.3 Estrutura do Trabalho.....	16
2 REVISÃO BIBLIOGRÁFICA.....	17
2.2 Solo	17
2.2.1 Biodiversidade, qualidade e sustentabilidade do solo	18
2.2.2 Propriedades físicas do solo	19
2.2.3 Propriedades químicas do solo	21
2.2.4 Aplicação de agrotóxicos.....	22
2.2.5 Agrotóxicos nos contextos do Brasil e do Paraná	23
2.2.6 Impactos da utilização dos agrotóxicos ao ambiente	24
2.2.7 Impactos da utilização dos agrotóxicos à saúde	25
2.3 Descoberta de Conhecimentos em Bases de Dados	28
2.3.1 Mineração de Dados	30
2.3.2 Agrupamento.....	31
2.3.3 Árvores de decisão.....	34
2.3.4 Associação	37
2.3.5 Seleção de atributos.....	39
2.3.6 Ferramentas computacionais utilizadas na Mineração de Dados	40
3 MATERIAIS E MÉTODOS.....	42
3.1 Base de Dados.....	42
3.2 Seleção dos Dados	44
3.3 Pré-Processamento dos Dados.....	45
3.4 Transformação dos Dados	46
3.5 Preparação dos Conjuntos de Treinamento e Teste	48
3.6 Algoritmos de Mineração dos Dados Avaliados	49
3.7 Métricas Empregadas para Avaliar os Resultados Obtidos.....	50
3.8 Interpretação dos Resultados.....	52
4 RESULTADOS	53

4.1 Saúde: SRQ-20	53
4.1.1 Clusterização com algoritmo EM: saúde – SQR-20	54
4.1.2 Árvores de secisão com algoritmo J48: saúde - SQR-20	54
4.1.3 Regras de Associação com algoritmo Apriori: saúde – SQR-20	57
4.1.4 Seleção de Atributos com diferentes algoritmos: saúde – SQR-20	59
4.1.5 Resultados obtidos e correlações: saúde – SQR-20.....	60
4.2 Saúde: Doenças.....	62
4.2.1 Clusterização com algoritmo EM: saúde (doenças)	62
4.2.2 Árvores de decisão com algoritmo J48: saúde (doenças).....	63
4.2.3 Regras de associação com algoritmo Apriori: saúde (doenças).....	67
4.2.4 Seleção de atributos com diferentes algoritmos: saúde (doenças)	70
4.2.5 Resultados obtidos e correlações: saúde (doenças).....	70
4.3 Saúde: SRQ-20 e Doenças	72
4.3.1 Clusterização com algoritmo EM: SRQ-20 e doenças	74
4.3.2 Árvores de decisão com algoritmo J48: SRQ-20 e doenças	77
4.3.3 Regras de associação com algoritmo Apriori: SRQ-20 e doenças.....	74
4.3.4 Seleção de atributos com diferentes algoritmos: SRQ-20 e doenças	76
4.3.5 Resultados obtidos e correlações: SRQ-20 e doenças	77
4.4 Solo	81
4.4.1 Resultados obtidos e correlações: solo.....	85
5 CONSIDERAÇÕES FINAIS	87
5.1 Trabalhos Futuros	87
APÊNDICE A - Dados de Elementos e Características do Solo do Município de Santa Tereza do Oeste - PR	112
APÊNDICE B - Dados de Saúde da População da Zona Rural e Urbana do Município de Santa Tereza do Oeste - PR.....	113

1 INTRODUÇÃO

A definição relativa à qualidade de vida é complexa, uma vez que possui influências dos diferentes pontos de vista e valores de cada cultura (AZEVEDO, 2015). Para Pereira et al. (2012), a definição de qualidade de vida está relacionada ao fato do indivíduo se sentir satisfeito, o que inclui o acesso a oportunidades de ser feliz e de alcançar autorrealização, independentemente de sua saúde, condições sociais e financeiras. Tal definição envolve diferentes percepções humanas e abrange muitos significados que levam em consideração os conhecimentos, as experiências e os valores individuais e coletivos. Adiciona-se a estes, as condições econômicas, ambientais, científico-culturais e políticas, coletivamente construídas e postas à disposição dos indivíduos para que estes possam realizar suas potencialidades, incluindo a acessibilidade à produção e ao consumo, aos meios para produzir cultura, ciência e arte. Ao mesmo tempo, inclui-se também a existência de mecanismos de comunicação, de informação, de participação e de influência nos destinos coletivos, por meio da gestão territorial que assegure água e ar limpos, boas condições ambientais, equipamentos coletivos urbanos, alimentos saudáveis e a disponibilidade de espaços naturais urbanos, bem como da preservação de ecossistemas naturais (DANTAS et al., 2003; HERCULANO, 2000).

Com base na grande importância da atividade agrícola para o Brasil e para o mundo, o meio rural brasileiro tem sido estudado para caracterizar sua população e identificar elementos pertinentes a sua qualidade de vida (IBGE, 2017; RIGOTTO et al., 2014). No que diz respeito à qualidade de vida no meio rural, é importante destacar que o Brasil passou por um processo de modernização de sua agricultura. Este processo é denominado Revolução Verde e trouxe avanços da base técnica da agricultura para a formação dos complexos agroindustriais. Após a Revolução Verde, cada vez mais as máquinas ocupam o lugar dos humanos na agricultura, gerando êxodo rural, principalmente nas regiões Sul, Sudeste e Centro-oeste e nas monoculturas voltadas principalmente à exportação com áreas de agricultura moderna convivendo com áreas de agricultura tradicional, o que vem causando impactos socioeconômicos e ambientais (AGRA; SANTOS, 2001; BALSAN, 2006; MATOS, 2010).

As atividades ligadas ao campo ou ao meio rural precisam atender à crescente demanda de frutas, grãos e hortaliças, sendo que a agricultura no Brasil avança a cada ano, o qual figura como um dos principais produtores agrícolas do mundo. A atividade agrícola se caracteriza como indispensável, pois todos dela dependem para obter os alimentos consumidos no dia-a-dia, sejam estes *in natura* ou com algum tipo de processamento. Muitas pessoas dependem da agricultura para gerar renda de forma direta ou indireta, sendo estimulados a utilizar uma grande variedade de insumos para aumentar produtividade e reduzir perdas nas safras, dentre eles, os agrotóxicos (ARAÚJO et al., 2007; MILLER, 2007; CAPORAL; COSTABEBER, 2004; IBGE, 2017; RIGOTTO et al., 2014).

Com o objetivo de atingir melhores resultados de produção, a utilização contínua de agrotóxicos acaba por gerar maior resistência das pragas. Este processo desencadeia um ciclo que se repete. A redução da infestação de grandes populações de pragas causa a diminuição da competição por alimento, espaço e abrigo, o que favorece sua reprodução e o retorno da praga a níveis populacionais maiores em comparação com os níveis antes da aplicação. Além disso, o princípio de ação da maioria dos agrotóxicos não consegue ser seletiva para exterminar apenas o que se pretende e geram danos a várias outras classes de componentes da fauna e da flora existentes no local de sua aplicação (ARRUDA, 2008; FIOCRUZ, 2018; PASCHOAL, 1979).

Uma diversidade de ingredientes ativos e formulações disponíveis no mercado de agrotóxicos que apresentam risco potencial para a saúde humana (FERREIRA et al., 2007). Além da grande variedade de agrotóxicos à disposição da agricultura, a legislação não é aplicada e respeitada. Como consequência, há ainda falta de controle adequado do uso dos agrotóxicos, bem como a exposição do ser humano e do meio ambiente aos danos causados por esta má utilização (VAN DER HOEK; KONRADSEN, 2006; OLIVEIRA; ZAMBRONE, 2006).

Os agrotóxicos representam um grave problema de saúde pública nos países em desenvolvimento e com economia fundamentada no agronegócio, tal como o Brasil, os quais promovem ações governamentais que atuam de forma integrada na promoção da saúde e prevenção dos agravos (MINISTÉRIO DA SAÚDE, 2009; ARAÚJO et al., 2007). A exposição da população ocorre tanto no ambiente rural quanto urbano, como nas fábricas de agrotóxicos, em seu entorno, no combate às endemias, nas proximidades de áreas agrícolas, nas aplicações aéreas e no consumo

dos alimentos contaminados. Devido a persistência dos agrotóxicos no ambiente podem ser causados efeitos indesejáveis à saúde e ao meio ambiente em função dos resíduos liberados para água, que atingem os alimentos e conseqüentemente o ser humano (SANTANA; MACHINSKI, 2004). Outra forma de exposição pode acontecer de forma acidental, na qual o risco da exposição aos agrotóxicos pode estar dentro das residências, devido a utilização com a finalidade de repelir insetos, combater piolhos e outros parasitas, ou ocasionadas devido à forma e ao local incorreto de armazenamento, à reutilização de embalagens dos agrotóxicos, ao derramamento no transporte ou ainda pela ingestão involuntária de água contaminada (RIGOTTO; VASCONSELOS; ROCHA, 2014; WAISSMANN, 2007).

O Brasil foi identificado como o país que mais gastou com agrotóxicos em termos de volume total, com valores que chegam a US\$ 10 bilhões, sendo seguido de Estados Unidos da América, China, Japão e França. Porém, ao se considerar um possível *ranking* com base na relação entre o volume utilizado e a área plantada em cada país, Portugal, Itália, Eslovênia, Espanha, Suíça, Países Baixos e Grécia estariam à frente do Brasil. Já ao se considerar a relação entre a quantidade de agrotóxicos empregada e a produção agrícola alcançada, tendo em vista que a produção brasileira alimenta um quinto da população mundial, o Brasil passa a figurar na décima terceira posição neste mesmo *ranking* (FAO, 2019). Tomando-se o contexto local do estado do Paraná, verifica-se que este figura dentre as primeiras posições como consumidor de agrotóxicos do Brasil. O volume total de agrotóxicos utilizados foi de cerca de 97.715 toneladas no ano de 2014 e 100.123 toneladas em 2015 (SINITOX, 2017; PARANÁ, 2018). De acordo com a ADAPAR (PARANÁ. Secretaria da Agricultura e do Abastecimento), em 2019, foram consumidas 95.286 toneladas de agrotóxicos no estado, representando um aumento de 2,5% em relação ao ano de 2018, dentre os princípios ativos agrotóxicos mais consumidos estão o Glifosato (SESA, 2021).

A intoxicação por agrotóxicos pode ser considerada como um problema de saúde pública mundial que envolve principalmente os países em desenvolvimento (NISHIYAMA, 2003). Algumas pesquisas relatam a falta da utilização dos Equipamentos de Proteção Individual (EPIs) pelos agricultores, facilitando a ocorrência de intoxicação destes agricultores. Em Nova Friburgo (RJ), por exemplo, a prevalência de intoxicação chegou a cerca de 70% dos entrevistados. Fenzke et al., (2018), citam em seu estudo que a intoxicação é o principal agravo de saúde

relacionado à falta e/ou ao mau uso de EPI. O manuseio inadequado dos agrotóxicos pode resultar em intoxicações agudas e em efeitos de longo prazo, causados pela exposição mesmo em baixas concentrações (NISHIYAMA, 2003).

Publicações da Organização Mundial da Saúde (OMS) e da Organização Internacional do Trabalho (OIT) constataram que os agrotóxicos promovem anualmente cerca de 70 mil casos de intoxicações em trabalhadores de países em desenvolvimento. No entanto, supõe-se ainda que esse número pode estar subestimado diante de uma possível elevada subnotificação desse tipo de episódio (FARIA; FASSA; FACHINI, 2007). Os dados citados na edição do Relatório Nacional de Vigilância em Saúde de Populações Expostas a Agrotóxicos, publicado em 2018, traz um compilado de dados de 2007 a 2015 e demonstra que neste período foram notificados 84.206 casos de intoxicação no Brasil em unidades de saúde pública e privada (FAO, 2019).

Os casos de intoxicação no Paraná ocorrem com frequência e são distribuídos em todo o seu território em decorrência da vocação agrícola do estado e têm sido relatados, principalmente entre as pessoas que os manipulam. A presença de resíduos de agrotóxicos na água, no solo e em alimentos é revelada por estudos realizados pela Superintendência de Recursos Hídricos e Meio Ambiente (SUREHMA) e pela Secretaria de Saúde do Estado do Paraná - SESA (SINITOX, 2017; IBGE, 2001).

1.1 Objetivo Geral

Diante da realidade das intoxicações por agrotóxicos no Brasil e principalmente no Paraná, o presente trabalho tem o objetivo de realizar um estudo de caso com uma amostra da população do município de Santa Tereza do Oeste, localizado no Estado do Paraná, para identificar possíveis relações entre o uso de agrotóxicos e os relatos dos munícipes quanto a problemas de saúde.

1.2 Objetivos Específicos

- ✓ Realizar levantamento relativo a agrotóxicos e consequências do uso destes produtos para o meio ambiente e saúde;
- ✓ Realizar levantamento bibliográfico quanto ao processo de descoberta de conhecimento em bases de dados, com ênfase nas técnicas compreendidas pela fase de mineração de dados;
- ✓ Aplicar técnicas de mineração de dados para analisar os dados produzidos pelo projeto "Resistência, adaptação ou homeostase? Como a população e os ecossistemas respondem à aplicação de pesticidas no longo prazo" com o objetivo de identificar possíveis relações do uso de agrotóxicos e os problemas de saúde reportados por municípios de Santa Tereza do Oeste - PR.

1.3 Estrutura do Trabalho

A presente dissertação está organizada em cinco capítulos. O primeiro compreende sua motivação, os objetivos gerais e específicos da pesquisa. O Capítulo 2 contempla a revisão bibliográfica, os conceitos e os ferramentais empregados. Neste sentido, tem-se uma explanação quanto à biodiversidade, propriedades físicas e químicas do solo, aplicação e utilização de agrotóxicos e seus impactos, descoberta de conhecimento em base de dados, mineração de dados, e técnicas de agrupamento, árvores de decisão, associação, seleção de atributos e ferramentas computacionais. O Capítulo 3 compreende a descrição da metodologia utilizada durante o desenvolvimento do trabalho, englobando a base de dados e a preparação dos dados, bem como a aplicação dos algoritmos e as métricas utilizadas para a avaliação dos resultados. O Capítulo 4 contém os resultados alcançados e sua discussão. As considerações finais são apresentadas no Capítulo 5.

2 REVISÃO BIBLIOGRÁFICA

O conceito de meio ambiente de acordo com a Lei da Política Nacional do Meio Ambiente, art. 3º, I, é definido como o conjunto de condições, leis, influências e interações de ordem física, química e biológica, que permitem, abrigam e regem a vida em todas as suas formas. É um conceito amplo, composto da interação entre os seres vivos e o meio em que vivem, constituído por solo, água, ar atmosférico, os elementos da biosfera (hidrosfera, atmosfera e litosfera), fauna e flora (MACHADO, 2013; SILVA, 2019; FIORILLO; RODRIGUES, 1997). Quanto aos componentes do meio ambiente, nos próximos tópicos, serão abordados assuntos com ênfase no solo, que é um dos temas deste trabalho.

2.2 Solo

O solo é um componente essencial dos ecossistemas (PEREIRA et al., 2018), sendo caracterizado como um *habitat* com grande biodiversidade, tanto interna quanto externamente a ele, o que compreende protozoários, fungos, invertebrados, vertebrados e plantas (PAVAO-ZUCKERMAN, 2008). Esses seres vivos interagem nas teias alimentares do solo, que são responsáveis pelos ciclos biogeoquímicos relacionados ao carbono, ao nitrogênio e ao fósforo (CARVALHAIS et al., 2012; CAVICCHIOLI et al., 2019; OULAS et al., 2015).

Os microrganismos absorvem a energia solar, transformando a luz em biomassa por meio da fotossíntese, contribuindo para a produção primária global. Diversos microorganismos produzem e consomem a maioria dos gases de efeito estufa (dióxido de carbono, óxido nitroso e metano) e são responsáveis por metade do oxigênio produzido na Terra (CAVICCHIOLI et al., 2019).

O solo regula a maioria de seus processos e serve como alicerce para várias atividades humanas. Além disso, os impactos antrópicos sobre o solo são cada vez mais negativos. Por ser um recurso não renovável, a qualidade do solo deve ser mantida para garantir a sustentabilidade das funções do ecossistema por ele fornecidos (DUARTE et al., 2018; FAGERIA; STONE, 2006).

O solo é um componente crucial para quase todos os aspectos da vida, apresentando-se como um componente fundamental do ecossistema terrestre e para

os seres humanos, por armazenar e filtrar água, atuar na regulação da temperatura e na emissão de gases do efeito de estufa e promover a decomposição da matéria orgânica, a reciclagem de nutrientes e o controle de pragas e doenças de animais e plantas, além da desintoxicação do meio ambiente, que podem ser prejudicados pelo uso de agrotóxicos (ARSHAD; MARTIN, 2002; LINHARES; RODRIGUES, 2021; BREVIK et al., 2015; WALL; KNOX, 2014).

Diante da possibilidade de influenciarem na interação do solo com agrotóxicos, sua contaminação e também na intoxicação dos seres humanos, as seções seguintes apresentam brevemente um descritivo de algumas características do solo.

2.2.1 Biodiversidade, qualidade e sustentabilidade do solo

A qualidade do solo é a capacidade do solo de funcionar como um sistema vivo dentro dos limites do ecossistema e do uso da terra para sustentar a produtividade biológica, manter a qualidade ambiental e promover a saúde vegetal e animal (LINHARES; RODRIGUES, 2021; DORAN; PARKIN, 1994). É possível avaliar a qualidade do solo em agro ecossistemas tendo a produtividade como o principal veículo de resposta. (BÜNEMANN et al., 2018).

Além dos fatores anteriores, a qualidade do solo pode ser avaliada por meio de um conjunto de atributos sensíveis do solo que refletem sua capacidade de funcionamento de forma sustentável (KARLEN; DITZLER; ANDREWS, 2003). O teor de carbono orgânico do solo (SOC) é um indicador de qualidade do solo muito importante, em função da dinâmica da biota do solo e por sua influência quanto à fertilidade, disponibilidade de água no solo e estabilidade de agregados em áreas de cultivo (BOUMA; MCBRATNEY, 2013; LAL, 2002). Tem-se, ainda, a atividade e a composição das comunidades microbianas do solo. Estas são muito sensíveis às mudanças do solo e às condições ambientais e devem ser consideradas em uma identificação mais abrangente da qualidade do solo (BLECKER et al., 2012).

A base para o uso e manejo sustentável do solo é a promoção e o monitoramento de sua qualidade e conta com indicadores capazes de integrar seus atributos biológicos, químicos e físicos. Identificar e parametrizar os dados de qualidade do solo, além de determinar quais destes parâmetros são melhores para avaliar o impacto do uso do solo, constituem os maiores desafios para a ciência do

solo e exigem pesquisas com colaboração de profissionais de diversas áreas (químicos, físicos, biólogos, estatísticos e agrônomos) (ABOIM et al., 2008).

2.2.2 Propriedades físicas do solo

A qualidade física do solo é avaliada por meio de algumas de suas características, tais como textura, densidade, umidade, densidade de partículas e porosidade. Estas características permitem manter o suporte à vida vegetal e animal e podem indicar processos dinâmicos relacionados ao fluxo de água e ar no solo (PULIDO MONCADA et al., 2014). Além disso, tais características servem para avaliar a qualidade e função física do solo em diferentes manejos, assim como sua relação com o funcionamento do solo e suas características químicas e biológicas, interferindo em processos como impedimentos mecânicos na emergência de plântulas e no desenvolvimento das raízes (KARLEN; DITZLER; ANDREWS, 2003).

As atividades agrícolas podem reduzir o volume e aumentar a densidade do solo, prejudicando sua qualidade física e comprometendo sua produtividade (HAMZA; ANDERSON, 2005; REICHERT et al., 2016). As propriedades mais afetadas pelas atividades são a densidade, porosidade e umidade. A densidade do solo representa a relação entre a massa sólida seca do solo e seu volume total, considerando tanto o volume dos sólidos quanto o volume dos poros e podendo ser utilizada para avaliar alterações de estrutura e porosidade do solo (AL-SHAMMARY et al., 2018). A estrutura diz respeito ao agrupamento das unidades de partículas do solo que formam os agregados e a textura se refere à distribuição das partículas de acordo com seu tamanho e proporções de silte, areia e argila. Solos arenosos suportam maiores valores de densidade do que solos argilosos, sendo que a compactação pode induzir a mudanças nos ciclos de nitrogênio e carbono, causando a deformação de raízes e diminuição da biodiversidade do solo (NAWAZ; BOURRIÉ; TROLARD et al., 2013).

A porosidade é essencial para o crescimento das raízes, movimento do ar, água e solutos pelo solo. Esta consiste no espaço do solo não ocupado por sólidos, mas sim ar e água. A textura e estrutura dos solos se deve ao tamanho, quantidade e continuidade dos poros. Esta propriedade física do solo pode ser classificada em microporosidade e macroporosidade, sendo que os microporos são responsáveis pela

retenção e armazenamento de água e os macroporos permitem a aeração e a infiltração de água no solo (BOUMA et al., 1977).

A proporção dos poros varia de acordo com o nível de compactação e a classe textural do solo, sendo que solos arenosos possuem maior proporção de macroporos e solos argilosos possuem maior proporção de microporos. A aeração atende a demanda de respiração biológica do solo, pois existe uma necessidade contínua de troca de CO₂ e oxigênio entre o solo e a atmosfera (MORALES-OLMEDO; ORTIZ; SELLÉS, 2015).

A compactação do solo causa a saturação de suas camadas superficiais e sua capacidade de infiltração, levando ao escoamento das chuvas e possível perda de nutrientes. Por outro lado, solos com alta capacidade de infiltração são definidos por sua estabilidade estrutural e capacidade de manter poros contínuos, o que os permitem limitar o escoamento (RAMPAZZO et al., 2014). Além disso, o nível de compactação do solo também afeta sua temperatura e interfere na germinação, crescimento das raízes e absorção de água e nutrientes pelas plantas e na atividade microbiana (RICHART et al., 2005).

Como já identificado anteriormente, outro fator importante que controla a composição física do solo é a umidade. O fenômeno de degradação do solo causado pela agricultura intensiva é agravado quando o teor de água no solo é muito alto ou muito baixo. O teor de água do solo regula a vulnerabilidade à compactação do solo, resistência à penetração, capacidade de suporte de carga e a mineralização da matéria orgânica. As interferências nos parâmetros físicos do solo determinam o efeito da compactação do solo nas propriedades físicas, e afetam as propriedades químicas e a biota do solo (BEYLICH et al., 2010).

A degradação física do solo afeta negativamente a biomassa microbiana, e pode influenciar nas atividades enzimáticas (BUCK; LANGMAACK; SCHRADER, 2000), principalmente aquelas relacionadas à compactação (UDAWATTA; GANTZER; JOSE, 2017). Diversas comunidades microbianas que crescem nas raízes metabolizam o poluente pela atividade enzimática, produzindo compostos menos prejudiciais (LIN et al., 2011).

As propriedades físicas do solo podem afetar as demais propriedades importantes do solo e são impactadas pelo sistema de produção agrícola utilizado. As práticas agrícolas atuais causam grande estresse a solo, água e biodiversidade. Tendo em vista a proteção do solo e suas funções, é essencial a identificação de sua

degradação física, bem como sua relação com seu uso e suas propriedades químicas, sua biota e o processo biológico (BONMATIN et al., 2014; OLIVEIRA FILHO, BARETTA, 2016; PORTUGAL et al., 2010).

2.2.3 Propriedades químicas do solo

A principal função do solo no tocante a suas propriedades químicas é fornecer nutrientes para o crescimento das plantas. No entanto, os aspectos considerados na determinação da composição química do solo englobam a disponibilidade de nutrientes, a concentração de metais e a mobilidade de pesticidas. Estes mesmos aspectos são afetados pelos seguintes fatores: capacidade de troca catiônica (CEC), pH, conteúdo de matéria orgânica e mineralogia. Devido à aplicação de pesticidas, fertilizantes, esterco e corretivos ao solo, sua capacidade de imobilizar pesticidas e imobilizar e desintoxicar metais pesados está frequentemente relacionada à dinâmica de seus aspectos químicos (HEIL; SPOSITO, 1997; KAUR et al., 2017).

Dentre os aspectos da qualidade do solo, o nível de fertilidade é considerado crucial para o controle da disponibilidade de nutrientes essenciais às plantas, pois o crescimento e o desenvolvimento delas dependem da combinação e concentração de macronutrientes e micronutrientes minerais no solo. A disponibilidade de nutrientes do solo pode sofrer interferências devido a mudanças no uso da terra e no ecossistema, podendo ter sérios efeitos na produção agrícola em decorrência dos produtos utilizados (ANDREU; PICÓ, 2004; HEIL; SPOSITO, 1997; KAUR et al., 2017; LAUBER et al., 2009).

A matéria orgânica do solo é uma fonte de nutrientes que é liberada por meio da decomposição de microorganismos, tal como o húmus que contribui para a fertilidade do solo, imobilização de poluentes e redução da lixiviação dos produtos químicos para as águas subterrâneas (CANELLAS; FAÇANHA, 2004). Os principais fatores que afetam a fertilidade do solo são a disponibilidade de matéria orgânica e metais, sendo que a agricultura intensiva pode ser causa de efeitos deletérios na microflora e diminuição de nutrientes no solo (KAUR et al., 2017; WOŁEJKO et al., 2020).

2.2.4 Aplicação de agrotóxicos

O termo agrotóxico passou a ser adotado no Brasil a partir da Lei Federal nº 7.802, de 1989, regulamentada pelo Decreto nº 4.074, de 2002, e traz o seguinte conceito: compostos de substâncias químicas destinadas ao controle, destruição ou prevenção, direta ou indiretamente, de agentes patogênicos para plantas e animais úteis e às pessoas (SANTANA; MOURA; NOGUEIRA, 2013; DOU, 1989). No contexto da agricultura, os pesticidas são muito utilizados contra diversas pragas e doenças que infestam as culturas agrícolas. Dentre os vetores que causam danos à produção, foram identificadas cerca de 9.000 espécies de insetos e ácaros, 50.000 fito patógenos e 8.000 ervas daninhas. A redução das perdas de rendimento decorrentes de danos causados por pragas e doenças é imprescindível para garantir a produção mundial de alimentos (ZHANG; JIANG; OU, 2011). Tais perdas na produção de frutas, vegetais e cereais poderiam chegar a 78%, 54% e 32%, respectivamente (AKTAR; SENGUPTA; CHOWDHURY, 2009).

Os agrotóxicos são usados para eliminar insetos, larvas, fungos e carrapatos, com o intuito de controlar as doenças provocadas por esses vetores e regular o crescimento da vegetação nos ambientes rural e urbano. Os agrotóxicos têm seu uso tanto em atividades agrícolas como não agrícolas. Os agrícolas são relacionados ao setor de produção, seja na limpeza do terreno e preparação do solo, na etapa de acompanhamento da lavoura, no armazenamento e no beneficiamento de produtos agrícolas, nas pastagens e nas florestas plantadas. Já o uso não agrícola ocorre em florestas e ecossistemas, como lagos e açudes (INCA, 2021).

A Agência Nacional de Vigilância Sanitária (ANVISA) foi criada pela Lei nº. 9.782, de 26 de janeiro de 1999, e é caracterizada como órgão de normatização, controle, regulação e fiscalização da vigilância sanitária brasileira. Esta classifica os agrotóxicos segundo a praga que se deseja controlar, grupo químico, toxicidade aguda, carcinogênese (processo de formação do câncer) e tipo de formulação (BRASÍLIA, 1999; FRARES, 1997).

Considerando o primeiro critério, isto é, a praga que se deseja controlar, os agrotóxicos são classificados em três grupos principais (além de algumas exceções), inseticidas, herbicidas, fungicidas, os quais são destinados ao controle de insetos, ervas daninhas e fungos, respectivamente. Quanto ao grupo químico, tem-se os

agrupamentos organofosforados, carbamatos, organoclorados, pirertóides, biperidílicos, mercuriais, dentre outros (FRARES, 1997).

Tendo-se como parâmetro o grupo químico, a classificação compreende os grupos carcinogênese de acordo com classificação IARC (INTERNATIONAL AGENCY FOR RESEARCH ON CANCER), tem-se grupo 1 (cancerígeno para humanos), 2A (Provavelmente cancerígeno para humanos), 2B (Possivelmente cancerígeno para humanos) e 3 (não classificados como cancerígenos para humanos) (FRARES, 1997; IARC, 2021).

A toxicidade da maioria dos agrotóxicos é expressa em valores referentes à Dose Média Letal (DL50), por via oral e representada por miligramas do ingrediente ativo do produto por quilograma de peso vivo, necessários para matar 50% da população de ratos ou de outro animal teste. A DL50 é usada para estabelecer as medidas de segurança a serem seguidas para reduzir os riscos que o produto pode apresentar à saúde humana. Os agrotóxicos são classificados quanto à sua toxicidade aguda em quatro classes, e de acordo com a DL50, os produtos enquadrados nas Classes I e II são considerados como “extremamente tóxicos” e “altamente tóxicos”, ambos identificados com tarja vermelha e podem levar à morte se ingeridos ou entrarem em contato com a pele e os olhos. Já os produtos classificados como Classe III são definidos como “moderadamente tóxicos”, sendo identificados com tarja amarela. Por fim, os produtos compreendidos na Classe IV são “pouco tóxicos” ou “improváveis de causar dano agudo”, sendo identificados com tarja azul e podendo causar intoxicação, mas sem possibilidade de morte (ANVISA, 2019). No entanto, a classificação se relaciona à intoxicação aguda e não indica os riscos de doenças de evolução prolongada, de acometimento crônico, como, por exemplo, câncer, neuropatias, hepatopatias e problemas respiratórios crônicos (WHO, 1990).

Quanto ao tipo de formulação, a classificação abrange solventes, aderentes, umectantes, entre outros. Já a apresentação pode ocorrer sob as formas de líquido, pó e granulado (FRARES, 1997).

2.2.5 Agrotóxicos nos contextos do Brasil e do Paraná

Em países tropicais como o Brasil, o calor e a umidade ao longo da safra tornam a agricultura mais suscetível a epidemias de pragas (GALLUP; SACHS; 2000) e

requerem o uso intensivo de agrotóxicos para garantir a produção (PIGNATI et al., 2017). A utilização em massa de agrotóxicos na agricultura foi iniciada no país na década de 1960 com a implantação do Programa Nacional de Defensivos Agrícolas (PNDA). O programa buscava a modernização da agricultura e o aumento da produtividade e, ao mesmo tempo, vinculava a utilização de agrotóxicos à concessão de créditos agrícolas. Desta forma, o próprio Estado se caracterizou como um dos principais incentivadores do uso de tais produtos (SIQUEIRA, 2013; JOBIM et al., 2010). Atualmente, ainda há políticas públicas que fomentam o uso e o comércio de agrotóxicos. Este setor da economia teve uma expansão de 190% na última década e é caracterizado pelo fato de mais de 70% estar sob o domínio de apenas dez empresas (GONÇALVES et al., 2012; SOARES; PORTO, 2012).

O Brasil também se destaca no agronegócio com ênfase no comércio global e na produção de *commodities* agrícolas, pela disponibilidade de áreas agricultáveis e seus altos níveis de produção (PORTO, 2013). Esse processo é responsável pelo crescente uso de agrotóxicos, sendo que a região Sul é responsável por, aproximadamente, 30% desse consumo (CREMONESE, 2012). Já o Estado do Paraná se destaca no uso de agrotóxicos entre os estados brasileiros, sendo o terceiro maior consumidor de agrotóxicos do Brasil, responsável pela utilização de 12 quilos de agrotóxico por hectare ao ano, enquanto a média brasileira de consumo é de 4 quilos por hectare ao ano (NEVES; BELLINI, 2013). Devido ao consumo elevado, os casos de intoxicação no Paraná ocorrem com frequência e são distribuídos em todo o seu território em decorrência da vocação agrícola do estado. Em estudo realizado no estado do Paraná, detectou-se que os agrotóxicos mais utilizados na cultura do algodão são as classes de maior toxicidade (GABOARDI; CANDIOTTO; RAMOS, 2019; IBGE, 2001; IBGE 2012).

2.2.6 Impactos da utilização dos agrotóxicos ao ambiente

Diversos problemas ambientais estão associados ao uso intenso de agrotóxicos, principalmente a contaminação de ecossistemas terrestres e alimentares (CALDAS, 2016). . O desenvolvimento de pragas mais resistentes aos organoclorados estimula de forma contínua sua substituição por novos compostos mais eficientes

contra pragas e doenças, dentre eles, destacam-se os organofosforados, os carbamatos e os piretróides (HALLMANN et al., 2014; PARK et al., 2015; SOARES, 2010).

Diversas pesquisas foram desenvolvidas ao longo dos anos para entender o comportamento desses compostos químicos no meio ambiente, incluindo seu ciclo, destino e toxicidade para a biota e verificou-se que agrotóxicos organoclorados são metabolizados por bactérias e podem permanecer no solo como um composto tóxico por dias, semanas e anos (CARVALHO, 2017; ELLIOTT et al., 1997). Além disso, existem evidências de que os organofosforados são altamente tóxicos para os artrópodes em geral, como camarões, caranguejos e até vertebrados e muitos outros pesticidas, como os herbicidas, têm demonstrado efeitos no sistema neural e excretor de mamíferos (CASIDA; QUISTAD, 2005; SANTOS et al., 2007).

Uma série de evidências científicas sobre os impactos do uso de agrotóxicos e seus resíduos no solo foram obtidas em artigos e compiladas no Dossiê da Associação Brasileira de Saúde Coletiva (CARNEIRO et al. 2002; CARNEIRO et al. 2015) e no livro organizado por Ferment et al. (2015).

2.2.7 Impactos da utilização dos agrotóxicos à saúde

Os agrotóxicos não afetam apenas as culturas em que são usados (THOMAZ; JÚNIOR, 2014). A extensiva utilização de agrotóxicos nas lavouras apresenta repercussões na saúde dos trabalhadores rurais que os utilizam diretamente, na comunidade circunvizinha exposta a eles e aos consumidores dos alimentos (CONSEA, 2014). Dentre os impactos socioambientais, há a redução da diversidade biológica e de ecossistemas, a contaminação das águas, do solo e de seres humanos (CARNEIRO et al., 2015).

Tais efeitos podem ser decorrentes de intoxicações agudas e crônicas por agrotóxicos. A intoxicação aguda decorre de um único contato (dose única) ou múltiplos contatos (doses repetidas) com um agrotóxico (ou mistura de agrotóxicos) em um período de 24 horas. Os efeitos podem surgir de imediato ou no decorrer de alguns dias, no máximo duas semanas, dependendo do princípio ativo. A depender da quantidade de produto absorvida, a intoxicação pode ocorrer de forma leve, moderada ou grave. Por outro lado, a intoxicação crônica é caracterizada pelo

surgimento tardio, após meses ou anos de exposição, acarretando danos muitas vezes irreversíveis. Os sintomas são subjetivos, o diagnóstico e nexos causais são difíceis de serem estabelecidos. Ao contrário do que acontece com as intoxicações agudas, é muito difícil estabelecer as relações de causa e efeito para os envenenamentos crônicos, pois no segundo caso as manifestações mórbidas que surgem meses ou anos após a exposição contínua de pequenas doses de pesticidas (SECRETARIA DA SAÚDE DO PARANÁ, 2018).

As intoxicações agudas por agrotóxicos são as mais conhecidas e afetam, principalmente, as pessoas expostas em seu ambiente de trabalho (exposição ocupacional). Dentre outras, tal como apresentado no Apêndice B, as intoxicações são caracterizadas por efeitos como irritação da pele e olhos, coceira, cólicas, vômitos, diarreias, espasmos, dificuldades respiratórias, convulsões e morte.

Já as intoxicações crônicas podem afetar toda a população, pois são decorrentes da exposição múltipla aos agrotóxicos, ou seja, da presença de resíduos de agrotóxicos em alimentos e no ambiente, que geralmente estão em doses baixas. Os efeitos adversos decorrentes da exposição crônica aos agrotóxicos podem aparecer muito tempo após a exposição, dificultando a correlação com o agente. Dentre os efeitos associados à exposição crônica a ingredientes ativos de agrotóxicos podem ser citados infertilidade, impotência, abortos, malformações, neurotoxicidade, desregulação hormonal, efeitos sobre o sistema imunológico e câncer (INCA, 2021).

Além das implicações sobre a saúde geral e a qualidade de vida dos agricultores expostos a agrotóxicos, alguns produtos químicos presentes no processo produtivo podem ser nocivos à audição. A perda auditiva pode representar um sinal precoce de intoxicação por agrotóxicos (BIELEFELD, 2009), a qual é caracterizada pela perda do tipo sensorio-neural bilateral simétrica, de caráter irreversível seletiva em altas frequências (3kHz a 6kHz) de configuração em entalhe por lesão principalmente em células ciliadas da cóclea (MORATA, 2002).

A sintomatologia apresentada nos casos de intoxicação por agrotóxicos pode ser confundida com outros agravos, pois os trabalhadores não associam sintomas como astenia, náuseas, vômitos, cefaleia, dificuldade respiratória e dores abdominais à exposição a este agente de risco (BRASIL, 2006; MANJABOSCO et al., 2004). A dimensão das intoxicações por agrotóxicos não é totalmente conhecida e o controle dos problemas decorrentes desta exposição ainda é dificultado pelos dados discrepantes entre as notificações do Sistema Nacional de Agravos de Notificação

(SINAN) e do Sistema Nacional de Informações Tóxico-Farmacológicas (SINITOX) que alimentam a base de dados da ANVISA. A insuficiência de dados sobre a utilização destes produtos é recorrente no Brasil e, conforme afirma Pignati et al. (2017), tem favorecido o ocultamento e a invisibilidade desse importante problema de saúde pública e ambiental (ARAÚJO et al., 2007; MINISTÉRIO DA SAÚDE, 2006; SOARES, 2009).

Com a implantação, a partir dos anos 1980, dos Centros de Controle de Intoxicações em vários estados brasileiros, as notificações dos agravos causados pelos agrotóxicos passaram a ser mais sistematizadas, constituindo-se um Sistema Nacional de Informações Tóxico-Farmacológicas (SINITOX, 2006). Os dados gerados nos diversos estados do país são consolidados pela Fundação Oswaldo Cruz, do Ministério da Saúde, a qual publica anualmente as estatísticas de casos de intoxicação registrados pelos Centros de Informação Científica e Tecnológicas (FIOCRUZ/CICT, 2000).

O número de casos de intoxicação se elevou em alguns estados, como por exemplo, o estado de Goiás que teve registros de 116 casos de intoxicação por agrotóxicos de uso agrícola no ano de 1999 e em 2012, esse número subiu para 547 casos. No Brasil, o número anual de óbitos ocasionados pelo uso de agrotóxicos é superior a uma centena, e as curas não confirmadas, no período entre 1999 e 2013 totalizam 348 casos (SINITOX, 2017).

A toxicidade de agrotóxicos como o Glifosato, que é amplamente utilizado no Brasil, pode causar sintomas agudos como dificuldade respiratória, hipertermia, convulsões, perda de apetite, enjoo, vômitos, sangramento nasal, desmaios, entre outros sintomas. Em estado crônico pode causar diferentes tipos de câncer, fibrose pulmonar, lesões hepáticas e dermatites de contato (OMS, 1996).

Incidentes com a produção e uso de agrotóxicos ocorrem em todo o país, no entanto alguns casos se de maior relevância merecem destaque: o caso de intoxicação de mães lactantes em Lucas do Rio Verde, no Mato Grosso e a pulverização de crianças e adolescentes em escola no município de Rio Verde, em Goiás. Em Lucas do Rio Verde foram realizadas pesquisas em elementos como água, animais e humanos, por pesquisadores da Universidade Federal do Mato Grosso (UFMT). Em todos os casos foi detectada contaminação, inclusive foram detectados diferentes tipos de agrotóxicos no leite materno de mulheres residentes na cidade que estavam amamentando da segunda à oitava semana após o parto (CARNEIRO et al.

2015). Em uma das mães que participou da pesquisa foram detectados seis tipos diferentes de agrotóxico no leite materno. De acordo com Palma (2011), o leite contaminado quando consumido pelos recém-nascidos provoca agravos à saúde, pois são mais vulneráveis à exposição a agentes químicos principalmente, porque se alimentam exclusivamente ou quase exclusivamente do leite materno, pelo menos até os seis meses de idade. A ocorrência de contaminação das mães pode ter ocorrido de forma ocupacional, mas existe também o fator ambiental que deve ser levado em consideração. Estudo realizado por Pignati e Machado (2011) sobre a agricultura do município de Lucas do Rio Verde demonstrou dados de exposição da população a 136 litros de agrotóxicos por habitante na safra de 2010.

Além disso, em todo o país tem aumentado o número de casos de câncer relacionados à fatores ambientais e ao uso intensivo de agrotóxicos. Estudos têm indicado a influência dessas substâncias químicas no desenvolvimento de neoplasias que de acordo com Instituto Nacional de Câncer (INCA) é um tumor que ocorre pelo crescimento anormal do número de células. Esses dados demonstram a importância de uma avaliação mais apurada e detalhada (DUTRA; SOUZA, 2017).

O Estado do Paraná conta com o Centro de Controle de Intoxicações de Maringá (CCI – Maringá). O CCI – Maringá é um órgão de assessoria e consultoria na área de urgências toxicológicas que desenvolve as seguintes atividades: fornecimento de informações toxicológicas a profissionais de saúde e população, vigilância epidemiológica das intoxicações (toxicovigilância), acompanhamento ambulatorial toxicológico, acompanhamento de internações hospitalares, análises toxicológicas, divulgação e ação educativa, atividades científicas e manutenção de banco de soros e antídotos. Este atende em regime de plantão permanente, 24 horas/dia, e está vinculado à Universidade Estadual de Maringá, à Secretaria Estadual de Saúde do Paraná/15ª Regional de Saúde, ao Ministério da Saúde, ANVISA, FIOCRUZ e à Organização Mundial da Saúde (NEVES; BELLINI, 2013).

2.3 Descoberta de Conhecimentos em Bases de Dados

Segundo Fayyad (1996), o modelo tradicional para transformação dos dados em informação (conhecimento) consiste em um processamento manual dos primeiros por especialistas que produzem relatórios a serem analisados. Porém, na maioria das

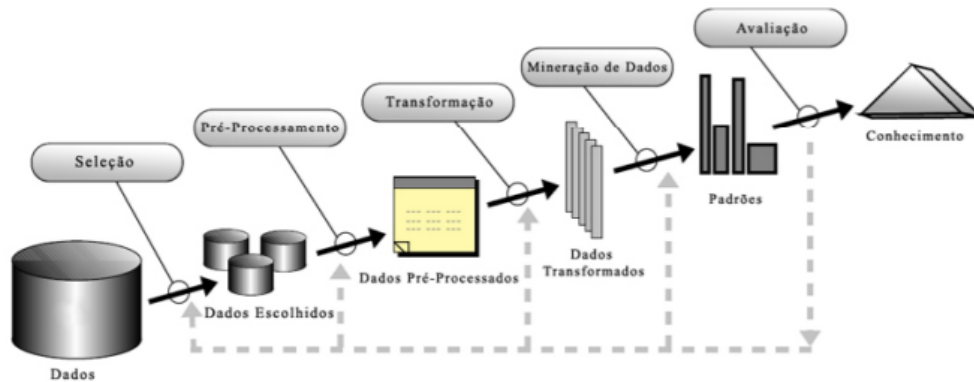
situações, devido ao grande volume de dados, a execução de forma manual desse processo se torna impraticável. As tecnologias voltadas para coleta, armazenamento e disponibilização de dados vêm evoluindo e disponibilizando técnicas, métodos e ferramentas computacionais automatizadas capazes de auxiliar na extração de informações úteis contidas em dados complexos (CIOS et al., 2007; COSTA, 2004).

O processo conhecido como Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases - KDD*), ilustrado na Figura 1, possui as seguintes etapas: seleção de dados, pré-processamento, transformação, mineração de dados e interpretação (avaliação). Os conceitos de cada etapa estão apresentados a seguir:

- a) A Seleção de Dados possui o objetivo de identificar o conjunto de dados que será utilizado para a análise, isto é, os elementos (observações, casos ou registros) e os atributos (características ou variáveis) que os representarão.
- b) O Pré-processamento é utilizado com intuito de eliminar dados redundantes, discrepantes (ambiguidade) e inconsistentes (ruídos). A presença de dados com tais características dentre aqueles previamente selecionados pode influenciar o processo e prejudicar seus resultados (STOROPOLI, 2016; SOUZA, 2013).
- c) A Transformação aplica tratamentos sobre os dados para melhor adequá-los à extração de padrões para se chegar ao formato e armazenamento ideal para posterior mineração de dados.
- d) A Mineração de Dados (*Data Mining - DM*) utiliza algoritmos para identificar padrões e regras (até então desconhecidos) nos dados analisados pelo sistema e pelo analista.
- e) A Interpretação (Avaliação) analisa a relevância e realiza o refinamento do conhecimento descoberto para o domínio em questão para identificação de tendências e construção de informação (STOROPOLI, 2016; SOUZA, 2013).

Diante da grande variedade de técnicas presentes na etapa de Mineração de Dados e do enfoque deste trabalho, aqui nós apresentamos algumas definições e técnicas relacionadas a esta etapa.

Figura 1. Representação do processo de KDD (Fayyad, et al. 1996).

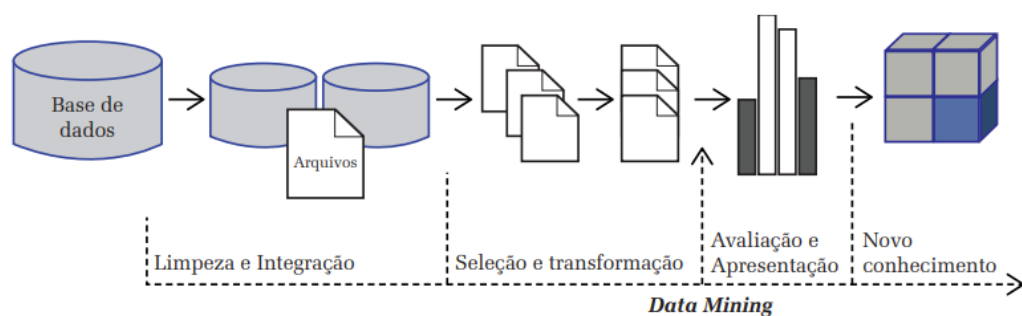


Fonte: Fayyad, et al. 1996.

2.3.1 Mineração de Dados

A expressão Mineração de Dados (MD) surgiu no começo da década de 1990 em comunidades de bases de dados e se refere à etapa de análise de dados do KDD, conforme ilustrado na Figura 2. As definições sobre Mineração de Dados variam de acordo com o campo de atuação. Esta é caracterizada como multidisciplinar, sendo que as áreas consideradas como de maior expressão dentro da Mineração de Dados são Estatística, Aprendizado de Máquina e Banco de Dados (SOBRAL, 2015; CARDOSO; MACHADO, 2008; CIOS et al., 2007; COSTA, 2004; MCCUE, 2007; QUONIAM et al., 2001).

Figura 2. Representação da etapa de Data Mining como parte do processo de descoberta de conhecimento em base de dados (KDD).



Fonte: FERNANDES; FILHO, 2019.

A Mineração de Dados emprega um conjunto de técnicas, procedimentos e algoritmos para a extração de um determinado tipo de conhecimento dos dados. É utilizada para auxiliar na tomada de decisão com maior grau de confiança, pois auxilia na descoberta de relações e padrões ocultos, o que dificilmente seria possível a partir de análises realizadas por especialistas humanos sem o auxílio de ferramentas computacionais. Tais descobertas permitem a geração de regras para correlacionar dados, revelando e descrevendo padrões e propriedades presentes nos dados analisados, bem como realizar previsões quanto ao comportamento de dados futuros e identificar anomalias (SOBRAL, 2015; CARDOSO; MACHADO, 2008; CIOS et al., 2007; COSTA, 2004; QUONIAM et al., 2001).

Por meio da aplicação de diferentes técnicas, a Mineração de Dados permite que os especialistas concentrem esforços apenas nos itens mais significativos dos dados (ZHOU, 2003; CAMILO; SILVA, 2009; LAROSE, 2005). Tais técnicas podem ter diferentes objetivos, tais como classificação por árvore de decisão, agrupamento, associação e sumarização. A Classificação procura pela identificação de modelos a partir de um conjunto de dados (elementos) para distinguir os diferentes tipos (classes) existentes neste conjunto, tal que estes modelos possam ser utilizados futuramente para prever o tipo (classe) de um novo elemento no futuro. O Agrupamento ou *Clustering* em inglês, agrupa elementos semelhantes. A Associação busca por relacionamentos entre os dados por meio da definição de Regras de Associação (*Association Rules*). Por fim, a Sumarização foca na sintetização dos dados e a Detecção de Desvios procura por mudanças no comportamento dos dados (CAMILO; SILVA, 2009; LAROSE, 2005; RIBEIRO, 2008). Berry, Linoff (2000) consideram fundamentais também as tarefas de Regressão e Predição. Neste trabalho, serão abordadas técnicas inerentes a Agrupamento, Regras de Associação e Árvore de Decisão, as quais serão melhor descritas nas seções seguintes.

2.3.2 Agrupamento

A análise de *agrupamento* (*Clustering*) destina-se a segmentar indivíduos de uma amostra de modo a formar conjuntos (agrupamentos ou *clusters*) mutuamente excludentes que apresentem similaridades entre seus elementos. As medidas de

similaridade são fundamentais para a análise de *cluster*, pois permitem avaliar o grau de semelhança entre os indivíduos que compõem uma amostra, subsidiando um posterior processo de particionamento (ou agrupamento). São amplas as aplicações das tarefas de agrupamento, tais como pesquisa de mercado, reconhecimento de padrões, processamento de imagens, análise de dados, taxonomia de plantas e animais, pesquisas geográficas, classificação de documentos da *Web* e detecção de comportamentos atípicos (fraudes). Geralmente, a tarefa de agrupamento é combinada com outras tarefas, além de serem usadas na fase de preparação dos dados (OLIVEIRA; CARVALHO, 2008; ALBUQUERQUE et al., 2006; VALLI, 2012).

De maneira geral, o agrupamento é constituído pelos seguintes passos: seleção das variáveis de agrupamento e medida de distância; seleção do algoritmo de agrupamento; análise de agrupamento; e interpretação dos grupos formados. Durante a seleção das variáveis de agrupamento, os atributos devem ser adequadamente selecionados de forma a codificar a maior quantidade possível de informações relacionada à tarefa de interesse. Além disso, a medida de distância define a quantificação de similaridade (ou dissimilaridade) entre os elementos (vetores com as medidas registradas para cada atributo). Neste ponto, é importante destacar que um atributo não pode ser dominante sobre o outro, isto é, todos devem passar por um processo de transformação para que apresentem uma mesma escala (intervalo de valores) final (LINDEN, 2009; HAIR et al., 2005; CUNHA et al., 2008).

Os algoritmos de agrupamentos podem ser classificados em hierárquicos e não-hierárquicos. Algoritmos hierárquicos agrupam os objetos por meio da construção de uma hierarquia entre os elementos ou estrutura em forma de árvore (dendrograma) e usam uma matriz de distâncias para definir a similaridade entre os elementos. A árvore de grupos apresenta a estrutura hierárquica entre os grupos, permitindo decidir qual nível de agrupamentos é mais apropriado para cada aplicação (MATHWORKS, 2002).

Existem dois tipos de métodos no procedimento hierárquico: aglomerativo e divisivo. No método aglomerativo (*bottom-up*), cada elemento inicialmente representa um grupo separado. Em cada iteração, dois grupos (ou elementos) que apresentam maior similaridade são combinados formando um novo grupo, e esse processo é repetido até que todos os objetos formem um único grupo. No método divisivo (*top-down*), todos os objetos iniciam-se em um único grupo, o qual é dividido em cada iteração em dois grupos de objetos com a maior dissimilaridade. Esse processo continua até que cada elemento forme um único grupo e exige grande capacidade

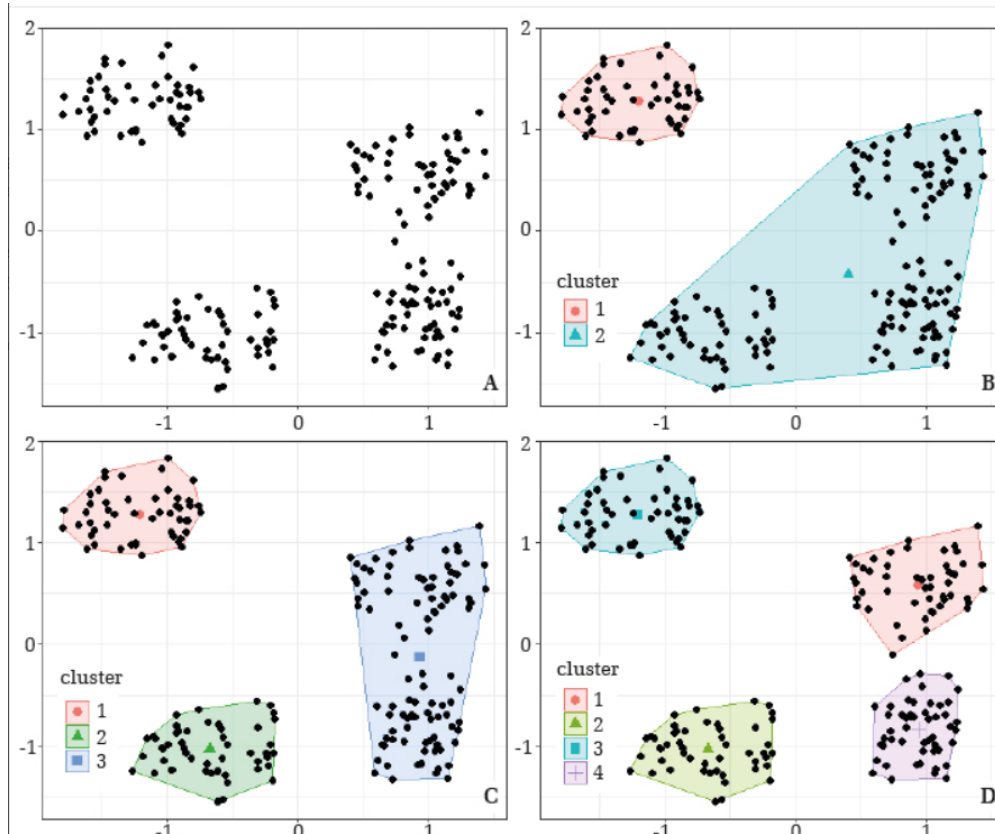
computacional (HAIR et al., 2005; KAUFMAN; ROUSSEEUW, 1990 (MANNING E SCHUTZE, 2003; ROCHA, FALCÃO, 2009; RICHARD et al., 2001).

O procedimento não-hierárquico ou não supervisionado produz apenas uma solução de agrupamento para um conjunto de sementes iniciais, seu conceito está baseado na escolha de uma solução inicial de grupos e na alteração dos elementos destes grupos para se obter a melhor configuração de grupos. Neste método, ao contrário do hierárquico, deve-se primeiramente especificar o número de grupos desejados. O método amplamente conhecido é o K médias (do inglês *K-means*), o qual inicia com a seleção aleatória dos centroides dos agrupamentos do conjunto de dados. *K-means* segue, a cada interação, atribuindo cada elemento ao centroide mais próximo considerando a menor distância quadrática Euclidiana. Os centroides dos agrupamentos são também atualizados a partir do cálculo da nova de cada cluster. O algoritmo termina quando não há mudança significativa dos centroides (MATHWORKS, 2002; ELMAGARMID; IPEIROTIS; VERYKIOS, 2016; AI-ANAZI; ALMAHMOUD; AI-TURAIKI, 2016), conforme demonstrado na Figura 3.

O algoritmo *Expectation-Maximization* (EM) deriva do Modelo de Mistura Gaussiana (*Gaussian Mixture Model*, GMM), sendo um método de estimação que apresenta solução para diversos problemas, no contexto de dados incompletos. Os conceitos fundamentais deste método foram introduzidos por Dempster, Laird e Rubin (1977) e sua ideia central é otimizar os parâmetros de uma função de distribuição de probabilidades, de forma que esta represente os dados da maneira mais fidedigna possível. EM se baseia em cálculos estatísticos que computam a estimativa de parâmetros de máxima semelhança nos casos em que as equações não podem ser resolvidas diretamente onde o modelo depende de variáveis latentes não observados (TORRES et al. 2017; WU; STEINBAUER; KUO, 2005).

EM e K-means são semelhantes no tocante ao processo iterativo de refinamento do modelo para encontrar a melhor configuração dos agrupamentos. No entanto, K-means utiliza a distância Euclidiana em suas avaliações, enquanto EM usa métodos estatísticos e a probabilidade de associação a um cluster. Ou seja, EM não atribui as amostras a um cluster, mas o faz de forma parcial e proporcionalmente a diferentes clusters. O algoritmo EM é frequentemente usado para fornecer as funções de forma mais eficaz (JUNG; KANG; HEO, 2014). Diante do exposto e da caracterização dos da base de dados, optou-se por utilizar o EM neste trabalho.

Figura 3. Modelos de análises de Agrupamento. (A) Gráfico de dispersão de dados não categorizados. (B) Agrupamento utilizando o algoritmo K-means com 2 clusters ($k=2$). (C) Agrupamento utilizando 3 clusters ($k=3$). (D). Agrupamento utilizando 4 clusters ($k=4$).



Fonte: FERNANDES; FILHO, 2019.

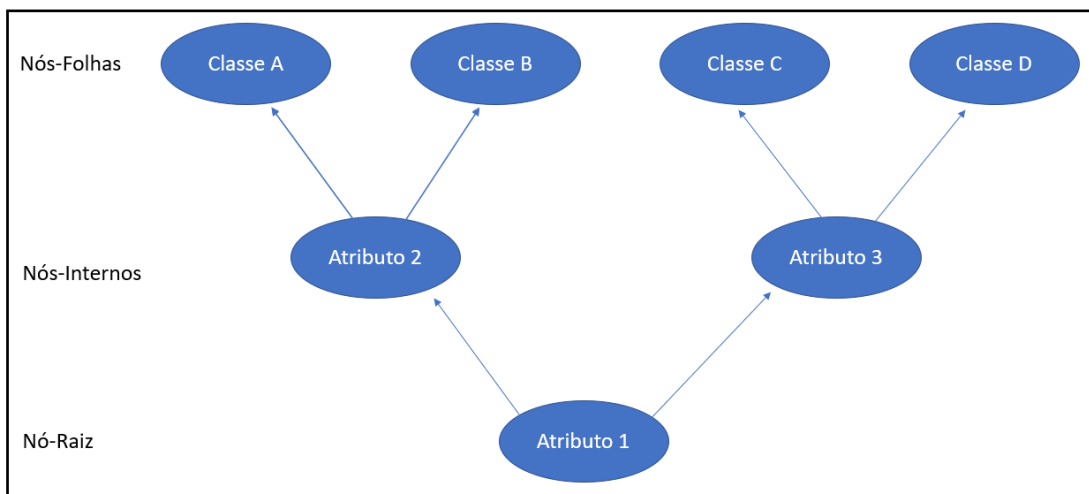
Nos passos finais, a Análise de Agrupamento deve verificar se os conjuntos gerados são condizentes com a representação dos dados. Já a Interpretação dos grupos formados os integra com outras evidências experimentais e análises afim de ter as conclusões corretas (WIVES, 1999).

2.3.3 Árvores de decisão

O método de classificação por Árvore de Decisão (*Decision Trees*) é uma representação simples e eficiente de encontrar o atributo que gera a melhor divisão dos dados, subconjunto com maior pureza. Este método basicamente funciona como um fluxograma em forma de árvore, tal que cada nó interno indica uma condição, cujo teste sobre cada elemento determina se o mesmo pertence ao conjunto de elementos

da esquerda ou da direita. Uma sequência de nós internos define as condições para que um elemento seja classificado para cada possível classe, sendo que as classes são representadas pelos nós-folhas. Para classificar um elemento, percorre-se a árvore a partir de sua raiz, aplicando os testes das condições dos nós internos até que se alcance um nó-folha. Ao atingir um nó folha, a classe que o rotula é atribuída àquele elemento, conforme exemplo na Figura 4 (HAN, KAMBER, 2000; CARVALHO 2005; RIBEIRO, 2008).

Figura 4. Exemplo de Árvore de Decisão.



Fonte: SANTOS; FRANCISCO; ALMEIDA (2015). Adaptado pela Autora, 2023.

O sucesso das árvores de decisão se deve ao fato de ser uma técnica que possui um bom grau de assertividade, desde que seja realizada uma análise detalhada dos dados que serão usados. Além disso, os modelos preditivos de apoio à tomada de decisão construídos são estruturas gráficas hierárquicas de fácil entendimento e aplicação, caracterizadas por segmentar dados heterogêneos de acordo com suas similaridades de maneira que se tornem mais homogêneos em relação à variável alvo. A relevância de tais árvores para a tomada de decisão se dá por sua capacidade preditiva, ou seja, a capacidade do modelo em prever no presente as interações que ocorrerão no futuro com um nível de certeza, auxiliando na resolução de problemas em diferentes áreas (CERVANTES et al., 2015; RAMYA et al., 2015; EVANGELINE; SUDHASINI, 2018).

Para a construção da árvore, o conjunto de elementos de treinamento são divididos recursivamente tendo como base os valores dos atributos selecionados até

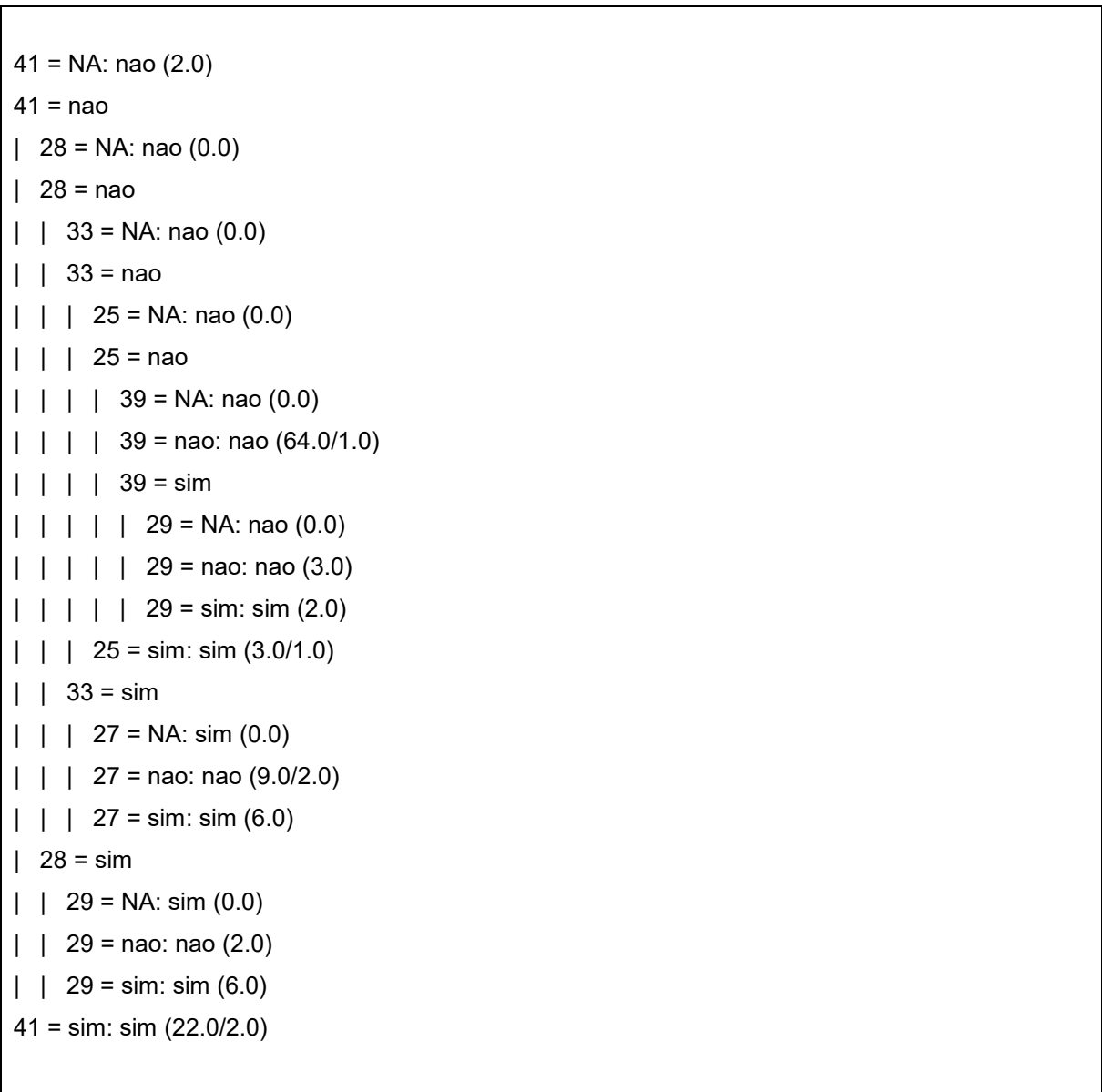
que seja satisfeita uma das seguintes condições de parada: todos os dados de um mesmo nó pertencem a uma mesma classe; não há mais atributos sobrando para o particionamento; não há mais dados de treinamento (HAN; KAMBER, 2000; CARVALHO 2005; RIBEIRO, 2008).

O passo principal de um algoritmo que constrói uma árvore de decisão é a escolha de um atributo para rotular o nó atual da árvore. Deve-se escolher o atributo que tenha o maior poder de discriminação entre as classes para os exemplos no nó atual. Para isso, deve-se utilizar uma medida de poder de discriminação de classes (CARVALHO 2005; HAN; KAMBER, 2000; CARVALHO 2005; RIBEIRO, 2008).

Um dos mais conhecidos algoritmos de indução de árvore de decisão é o C4.5, criado por Ross Quinlan na década de 1990. Este foi implementado na linguagem JAVA sob o nome J48, utilizado neste trabalho e que vem sendo popularmente utilizado por apresentar um determinado padrão de comportamento em conjuntos de dados de diferentes formas de representação, não tendo a exigência de uma distribuição de probabilidade específica. As principais vantagens desta técnica é a capacidade de processar valores em falta ou dados com ruídos, bem como gerar resultados de alto desempenho com baixo custo computacional (CHAUHAN; CHAUHAN, 2013; LIN; CHEN 2012; CERVANTES et al., 2015; BHARGAVA et al., 2013). O J48 é um algoritmo de código aberto, implementado pelo software WEKA (WITTEN et al., 2016).

O Quadro 1 apresenta um exemplo das árvores de decisão geradas durante os experimentos. Nesta representação, os níveis das árvores são caracterizados pelos espaçamentos horizontais e verticais, sendo que ao final são apresentadas a classe ('sim' e 'não') e também a quantidade de acertos e erros produzidos pela aplicação das condições definidas pelas ramificações da árvore. Para o caso apresentado, o primeiro atributo empregado na construção da árvore foi o 41, sendo que há ramificações para cada um dos possíveis valores que ele assume ('NA', 'sim' e 'não'). Quando o atributo 41 continha o valor 'NA' (primeira linha), houveram 2 acerto para a classe 'nao'. Já para as ocorrências do valor 'sim', houveram 22 acertos e 2 erros para a classe 'sim'. Para os casos em que o atributo 41 continha o valor 'não', houveram outros atributos utilizados para a definição da classe. Pela combinação do atributo 41 ('nao'), 28 ('nao'), 33 ('nao'), 25 ('nao') e 39 ('não') houveram 64 acertos e apenas um erro produzido.

Quadro 1. Exemplo de Árvore de Decisão gerada com a utilização do algoritmo J48 e os dados relacionados ao Saúde (SQR-20).



Fonte: A autora, 2023.

2.3.4 Associação

As técnicas de associação (*Association*), também conhecidas na área de *marketing* como grupos de afinidade, permitem identificar relações entre elementos de uma base de dados. Estas relações são caracterizadas pela ocorrência de combinações de atributos e valores, os quais formam as Regras de Associação

(*Association Rules*) (HIPPI et al., 2000; HAN, KAMBER, 2001; VASCONSELOS; CARVALHO, 2004).

O Quadro 2 apresenta 2 exemplos de regras de associação obtidas neste trabalho por meio da ferramenta WEKA. No primeiro exemplo, quando o atributo 24 e 29 tinham o valor 'não' (condição ou antecedente), a classificação do SQR-20 foi definida como 'sim' (rótulo ou consequente). Já no segundo exemplo, quando o atributo 24 e 28 tinham o valor 'não' e o atributo 31 tinha o valor 'sim', a classificação do SQR-20 foi definida como 'não'. Neste ponto, cabe destacar importantes conceitos nesta notação. O símbolo '==>' separa a condição à esquerda do rótulo à direita. Tanto a condição quanto o rótulo possuem uma métrica definida como 'suporte', a qual é dada pela frequência em que um conjunto (atributo ou rótulo, e respectivo valor) é encontrado no conjunto de dados. Neste exemplo, tem-se suportes 3 (1,59%) e 3 (1,59%) para a primeira regra, bem como 41 (21,69%) e 21 (11,11%) para a segunda regra. Já quanto à regra como um todo, há o conceito definido como 'confiança'. Este é dado pela frequência (razão) com que o antecedente e o consequente (atributos e respectivos valores) aparecem juntos na mesma regra. Para os exemplos em questão, tem-se 100% (ou 1, dado por 3/3) e 51% (ou 0,51, dado por 21/41).

Quadro 2. Exemplos de Regras de Associação obtidas com a utilização do algoritmo Apriori e os dados relacionados ao Saúde (SQR-20).

<p>24=nao 29=nao 3 ==> SQR-20=sim 3 conf:(1) 24=nao 28=nao 31=sim 41 ==> iqr20rotulo=nao 21 conf:(0.51)</p>

Fonte: A autora, 2023.

Quando este conceito foi introduzido pela primeira vez por Agrawal, Imielinski e Swami (1993), foi apresentado igualmente um algoritmo designado AIS para a descoberta deste tipo de regras. Este algoritmo produz regras cujos consequentes possuem apenas um item. No entanto, a estratégia do AIS resulta na geração de *itemsets* (conjuntos de atributos) desnecessários e regras com valores baixos para suporte e confiança. Com o objetivo de otimizar a geração de candidatos do AIS, surgiram os algoritmos *Apriori* e *AprioriTID*. Partindo desta plataforma de trabalho, muitos outros autores propuseram alternativas para obter desempenhos superiores

aos algoritmos *Apriori* e *AprioriTID* na tarefa da descoberta de regras de associação (AGRAWAL et al., 1994; VELOSO, 2003).

Um dos algoritmos mais utilizados para mineração por regras de associação é o Apriori. O algoritmo emprega busca em profundidade e gera conjuntos de itens candidatos (padrões) de k elementos a partir de conjuntos de itens de $k - 1$ elementos. Os padrões não frequentes são eliminados. Toda a base de dados é rastreada e os conjuntos de itens frequentes são obtidos a partir dos conjuntos de itens candidatos (VASCONSELOS; CARVALHO, 2004).

2.3.5 Seleção de atributos

A seleção de atributos consiste em identificar o subconjunto de atributos mais relevantes para previsão do objeto em estudo. Desta forma, é necessário eliminar os atributos que sejam pouco significativos (redundantes ou que causem ruídos) para que se obtenha uma melhora no desempenho produtivo e redução do tempo de treinamento e do espaço de armazenamento dos dados. Existem várias formas de selecionar atributos, mas todas visam a minimização de um conjunto de dados, mantendo os atributos com maior relevância para o modelo (BASTIANI, 2018).

A ferramenta WEKA disponibiliza um conjunto de diferentes algoritmos de seleção de atributos, dentre os quais estão aqueles utilizados neste estudo: *CfsSubsetEval-GreedyStepwise* e *CfsSubsetEval-BestFisrt*, *ClassifierAttributeEval*, *InfoGainAttributeEval* e *GainRatioAttributeEval* (SANTOS, 2018).

O *CfsSubsetEval* busca por um subconjunto de atributos que possua alta correlação com as classes da base de dados. Para isso, calcula as correlações de subconjuntos de atributos com as classes e avalia a relevância de cada subconjunto com base na habilidade individual de predição de cada atributo e o grau de redundância entre aqueles atributos que formam o conjunto. Subconjuntos de atributos que apresentem alta correlação e classes que apresentem baixa intercorrelação são desejáveis. (BASTIANI, 2018; NETTO, 2013).

O *ClassifierAttributeEval* algoritmo permite que um classificador seja utilizado como parâmetro e com base nos melhores resultados, os melhores atributos são selecionados (SANTOS, 2018).

O algoritmo *GainRatioAttributeEval* usa uma métrica para ordenar os atributos de uma base de dados calculando o ganho de informação de um atributo com relação ao número de saídas que um teste com aquele atributo pode resultar. O algoritmo é uma versão ponderada do algoritmo *InfoGainAttributeEval* e ranqueia os melhores atributos por ganho de informação e redução da entropia (TAKAMOTO; TOMAZELA, 2011; LIMA, 2016). O algoritmo *InfoGainAttributeEval* segue as mesmas especificações do *Gain Ratio* no sentido de que seleciona os atributos através de ganho de informação em relação à classe, calculado da seguinte forma: $GanhoInfo(Classe, Atributo) = Classe - Classe | Atributo$ onde a Classe é a entropia a priori, calculada sem considerar o corte com o atributo em questão, e Classe | Atributo é a entropia a posteriori, calculada considerando a poda com valores do atributo corrente (WITTEN; FRANK, 2005; MEDEIROS, 2018).

2.3.6 Ferramentas computacionais utilizadas na Mineração de Dados

Diversas ferramentas foram desenvolvidas no intuito de tornar a aplicação da Mineração de Dados menos técnica e possibilitar que profissionais de outras áreas pudessem utilizá-la. Neste sentido, o mercado de ferramentas de mineração de dados tem se tornado bastante atraente. Empresas como a Microsoft e a Oracle, além de várias outras, estão atentas a estas movimentações do mercado e apresentam soluções neste domínio. Academicamente existem também alguns casos de soluções para Data Mining, tal como o WEKA (*Waikato Environment for Knowledge Analysis*) (HAN et al., 2000).

WEKA é considerada uma das melhores ferramentas livres por possuir uma série de algoritmos para as tarefas de mineração, além de fornecer as funcionalidades para pré-processamento e visualização. Estes podem ser aplicados diretamente a ferramenta ou utilizados por meio de programação em Java. Cabe ressaltar que a ferramenta disponibiliza diferentes alternativas para a realização das avaliações e que estas variam de acordo com a técnica a ser utilizada. Destaca-se que além das alternativas genéricas identificadas a seguir para classificação e agrupamento, cada algoritmo possui parâmetros próprios (WAIKATO, 2021; CAMILO; SILVA, 2009; WITTEN; FRANK, 2005).

Para os algoritmos de Agrupamentos, podem ser empregadas 4 alternativas. A opção padrão é “*Use training set*”, sendo que esta utiliza o mesmo conjunto de dados para treinar e testar o modelo de classificação. “*Supplied test set*” permite que sejam fornecidos 2 conjuntos de dados, sendo um utilizado para o treinamento e o outro para os testes. A opção “*Percentage split*” separa uma parte dos dados para treinamento e outra para teste considerando o percentual informado pelo usuário. Por fim, opção “*Classes to clusters evaluation*” permite avaliar o quão aderentes os dados são em relação a clusters previamente identificados.

No tocante às técnicas relacionadas a Classificação, as três primeiras opções anteriores são mantidas, e uma nova opção “*Cross-validation*” realiza validação cruzada, sendo que o conjunto de dados fornecido é dividido em partições (*folds*) iguais. A cada execução, uma partição é usada para compor conjunto de teste e as demais para compor o conjunto de treinamento. O número de partições, e consequentemente o número de execuções, deve ser indicado pelo usuário.

WEKA emprega o formato ARFF (*Attribute-Relation File Format*). Conforme ilustrado no Quadro 3, este é um padrão para representar um conjunto de elementos por meio de seus atributos e cujo resultado final é um arquivo texto. Os arquivos ARFF foram desenvolvidos pelo Projeto de Aprendizado de Máquina do Departamento de Ciência da Computação da Universidade de Waikato para uso com o *software* de aprendizado de máquina WEKA - *Waikato Environment for Knowledge Analysis*.

Quadro 3. Exemplo de arquivo ARFF do WEKA.

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
% a) Creator: R.A. Fisher
% b) Donor: Michael Marshall (MARSHALL@PLU@io.arc.nasa.gov)
% c) Date: July, 1988
%
@RELATION iris
@ATTRIBUTE sepalength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class{Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
```

Fonte: WEKA (Waikato Environment for Knowledge Analysis), 2022.

3 MATERIAIS E MÉTODOS

Este capítulo apresenta a metodologia adotada neste projeto. Além da base de dados e das métricas adotadas para avaliar os resultados obtidos, basicamente são descritos os procedimentos adotados para as etapas previamente identificadas na revisão da literatura, isto é, seleção de dados, pré-processamento, transformação, mineração de dados e interpretação (avaliação).

3.1 Base de Dados

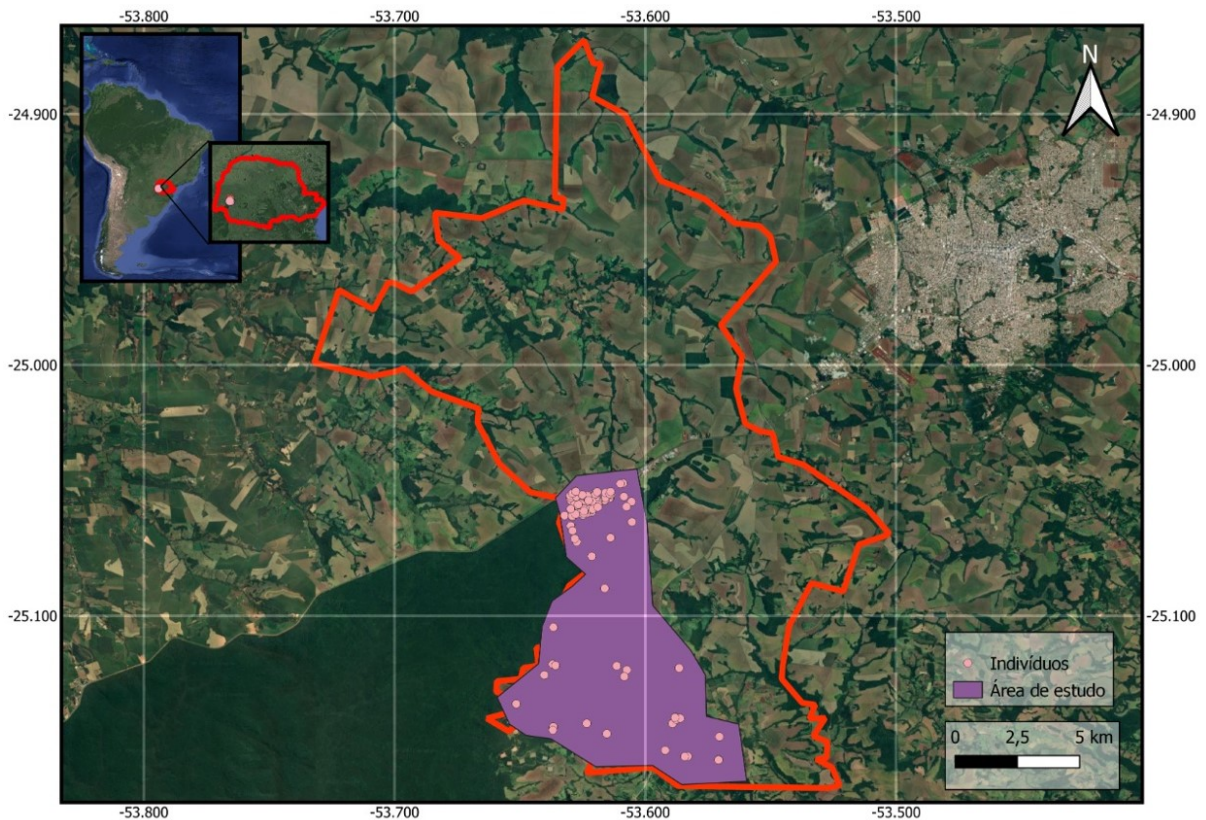
Os dados utilizados nesta pesquisa têm sua origem em um produto parcial de um projeto desenvolvido pela Universidade CARDIFF em parceria com Universidade Estadual do Oeste do Paraná (UNIOESTE). O protocolo de levantamento destes dados está baseado no Protocolo de Avaliação das Intoxicações Crônicas por Agrotóxicos da Secretaria de Estado da Saúde do Paraná, sendo que houve uma coleta parcial em 2019 e outra coleta em 2021. Tal projeto teve sua aprovação junto ao Comitê de Ética em Pesquisa (CEP) segundo o Certificado de Apresentação de Apreciação Ética (CAEE) N. 07558819.0.0000.8142 e Parecer N. 3.414.975, de 26 de junho de 2019.

O conjunto completo de dados contém informações, relativas às zonas urbanas e rurais do município de Santa Tereza do Oeste-PR, quanto à intoxicação do meio ambiente e de pessoas decorrentes de utilização e exposição a agrotóxicos. Quanto ao meio ambiente, foram coletadas amostras de solo e realizadas análises, conforme projeto citado anteriormente para se determinar possíveis níveis de toxicidade em diferentes pontos do município, conforme mapa da Figura 5. Os itens avaliados podem ser visualizados no Apêndice A.

Os dados levantados englobam informações sobre análises de solo e sua composição, conforme conduzido no projeto citado anteriormente. Foram consideradas características e propriedades do solo, dentre elas pode-se citar densidade, micro e macro porosidade, pH e análise para identificação de presença de agrotóxicos específicos. Quanto à composição química, temos os seguintes elementos: cálcio, potássio, cromo, cobre e ferro.

Com relação à intervenção com as pessoas, os dados que compõem a base foram coletados por meio de entrevistas de indivíduos da população rural e urbana de Santa Tereza do Oeste-PR. Neste processo, os entrevistadores interagem com os pesquisadores fazendo-lhes os questionamentos e anotando suas respostas, o que incluía também informações relativas à localização de sua residência. As perguntas do questionário podem ser visualizadas no Apêndice B.

Figura 5. Mapa de pontos de amostragem do solo



Fonte: Projeto desenvolvido pela Universidade CARDIFF, 2021

Os dados obtidos consistem em informações pessoais dos entrevistados como data de nascimento, sexo, escolaridade, local onde mora e número de pessoas na família além de itens relacionados a saúde envolvendo questionamentos sobre dores, tremores, qualidade do sono, digestão, apetite, cansaço, entre outros.

As questões relacionadas à saúde mental do participante da pesquisa foram definidas pelo Questionário de Auto-Avaliação (*Self-Reporting Questionnaire*, SRQ-20). Este foi desenvolvido por Harding et al. (1980) como colaboração para a Organização Mundial da Saúde (OMS) na triagem das desordens psiquiátricas mais

comuns na atenção básica à saúde (WHO, 1994). O SQR-20 é composto por 20 questões elaboradas para a triagem dos atualmente chamados Transtornos Mentais Comuns ou Morbidades Psiquiátricas Menores. O SRQ-20 foi validado em diversos países e pode sofrer alterações na ordem ou formato das perguntas para se adequar a realidades culturais diferentes. No Brasil, o questionário foi validado por Mari e Willians (1986), sendo que se o participante responder “sim” para 7 ou mais questões, este teria maior probabilidade de ter um transtorno mental comum.

Os questionários se referem a dados sobre a aplicação de agrotóxicos, relacionados às condições ambientais e uso de equipamento de proteção individual. Tanto para os agricultores, quanto para os moradores da zona urbana constavam perguntas sobre os principais problemas de saúde, agudos ou crônicos, que poderiam estar presentes na população de Santa Tereza de Oeste. Ao total foram entrevistadas 179 pessoas, 53 em 2019 e 126 em 2021, dentre os quais tem-se agricultores familiares e moradores da cidade, com faixa etária de 18 a 90 anos.

Cabe destacar ainda o fato de que as coletas de amostras para algumas análises, tais como as de solo, foram realizadas em pontos específicos do município de Santa Tereza do Oeste - PR. Por outro lado, também houveram coletas de dados realizadas em outros pontos geográficos, dentre as quais tem-se as entrevistas com uma amostra da população do referido município. Desta forma, para que fosse possível definir algum tipo de relação entre os dados oriundos das análises de coletas e dos dados oriundos das entrevistas, foi empregada uma técnica de krigagem ou krigeagem (do inglês *kriging*). Esta técnica se caracteriza como um método de regressão e é amplamente empregada em aplicações de geoestatística para aproximar ou interpolar dados (SANTOS, 2010). A krigagem tem capacidade de produzir estimativas de qualidade em termos de interpolação, porque está embasada em duas premissas: não-tendenciosidade do estimador e variância mínima das estimativas, reduzindo erros aleatórios pelo controle de parcela deste erro associada à dependência espacial (VIEIRA, 2000).

3.2 Seleção dos Dados

Os dados que compõem a base compreendem um conjunto diverso de informações, conforme descrito no Apêndice B. Do Campo_01 ao Campo_05 têm-se

dados pessoais dos entrevistados, tais como data de nascimento, sexo, escolaridade, número de pessoas na família e local onde mora, o que inclui também a localização geográfica e classificação como sendo zona rural ou zona urbana.

Na sequência, do Campo_06 ao Campo_73 representa-se os resultados das perguntas relacionadas à saúde dos entrevistados. Campo_74 ao Campo_110 compreendem os resultados das análises de solo, o que envolve elementos metálicos (cálcio, potássio, cromo, cobre, ferro, etc.), além de características físicas e presença/quantificação de agrotóxicos.

Foram removidos alguns atributos da base de dados original para aplicação das técnicas de mineração de dados, restando aqueles apresentados nos Apêndices I e II, cujos recortes utilizados nos experimentos são apresentados nas seções do Capítulo 4. Dentre estes, tem-se campos com informações de nomes, e-mails, data e horário da entrevista, telefone, coordenadas geográficas, endereço, bairro e cidade, forma e local de armazenamento de agrotóxicos, se houve instruções sobre utilização de agrotóxicos, nome de medicamentos utilizados e colunas com todas as células sem resposta (em branco).

3.3 Pré-Processamento dos Dados

Dentre as tarefas a serem desenvolvidas nesta etapa, tem-se a padronização quanto à representação dos dados. A título de exemplo, foi definido um número específico de casas decimais para valores numéricos. Houve ainda a uniformização quanto a grafia de palavras e expressões. Para os campos com valores em branco para um determinado elemento (questão, por exemplo), foi definido um valor padrão como “SemResposta”, ao passo que para as questões numéricas com valores em branco foi usado o valor 0 (zero).

Os procedimentos realizados neste passo permitem que padrões sejam identificados com maior facilidade (e com maior frequência) em decorrência da uniformização de valores. Ao mesmo tempo, o fato de empregar valores fora do domínio do campo (atributo), permite análises sem a interferência dos elementos (indivíduos/atributos) com caracterizações não significativas.

3.4 Transformação dos Dados

Esta tarefa compreende a realização de adequações nos dados de acordo com as necessidades do problema e das técnicas que serão utilizadas. A título de exemplo, o algoritmo *Apriori* (compreendido no grupo de Associação) aceita apenas valores simbólicos, isto é, não podem ser utilizados valores numéricos (inteiros ou reais, dentre outros). Neste caso, foi necessário criar símbolos (classes) considerando os possíveis valores de cada atributo numérico, tal como definir classes para o atributo idade com base em faixas etárias.

A Figura 6 apresenta um exemplo hipotético com 10 indivíduos para os dois tipos de rotulagem testados. À esquerda tem-se os indivíduos ilustrados com valores que variam entre 1 e 5. Ao centro, tem-se a representação da divisão dos indivíduos em 5 classes considerando intervalos iguais, o que em geral produziu grupos com diferentes quantidades de componentes. Por fim, à direita tem-se a representação da divisão dos indivíduos em 5 classes considerando quintis, sendo que neste caso todos os grupos tinham a mesma quantidade de componentes.

Figura 6. Definição dos intervalos utilizados para a definição das classes (rótulos) e a aplicação das técnicas de mineração de dados.

Valor Real	Valor Real	Intervalos Iguais	Valor Real	Quintis
1	1	A	1	A
1	1		1	
1	1		1	B
2	2	2		
3	3	C	3	C
3	3		3	
4	4	D	4	D
4	4		4	
4	4		4	E
5	5	5		

Fonte: A autora, 2023.

Para o primeiro tipo de rotulagem (Figura 6, centro), foram realizados testes com a definição de valores diferentes para o número de classes N a serem utilizadas. Em seguida, a partir dos valores mínimo e máximo de cada atributo, foram determinados os valores iniciais e finais de cada um dos N intervalos para cada

atributo. Por fim, cada valor real foi substituído pela identificação do intervalo (classe) em que estava enquadrado.

No decorrer do processo de avaliação dos valores para N, notou-se que à medida que se diminuía o valor de N, aumentava o desvio padrão internamente a cada intervalo definido, devido a uma maior heterogeneidade dos indivíduos que compunham aquele grupo (classe). Por outro lado, à medida que se aumentava o valor de N, reduzia-se a heterogeneidade do conjunto, mas também se aumentava o número de intervalos (classes) vazios, isto é, sem a presença de pelo menos um indivíduo. Ao final desta avaliação, adotou-se o valor 5 para N em decorrência deste ter sido identificado como a melhor dentre as alternativas testadas, reduzindo a heterogeneidade interna aos grupos e o número de agrupamentos sem elementos.

Tendo em vista o exposto anteriormente quanto a definição de N intervalos fixos, a variação dos tamanhos dos agrupamentos gerados e os resultados obtidos, repetiu-se os todos os experimentos com novos agrupamentos. Nesta segunda bateria de testes, os N agrupamentos foram definidos sempre com a mesma quantidade de instâncias ($179 / N$), tendo como base os percentis (Figura 6, direita). Ou seja, os indivíduos foram ordenados de acordo com os valores do atributo e cada valor real do atributo foi substituído pela identificação do intervalo (classe) em que estava enquadrado. Aqui também foram avaliados valores diferentes para o número de classes N a serem utilizadas, tendo havido a mesma relação previamente descrita entre o valor de N e a heterogeneidade dos grupos, mas sem a variação no número de elementos por grupo.

Como etapa final da transformação, os dados em formato de arquivo Comma-Separated Values (CSV) foram convertidos para o formato *Attribute-Relation File Format* (ARFF), o qual é padrão utilizado pela ferramenta WEKA. Para isso, foi utilizado o sítio *web* <https://ikuz.eu/csv2arff/>, sendo que o processo consistiu em realizar o *upload* do arquivo CSV, definir o caráter delimitador (separador) de campos (valores) e salvar o arquivo final em formato ARFF antes de aplicar as técnicas de mineração de dados.

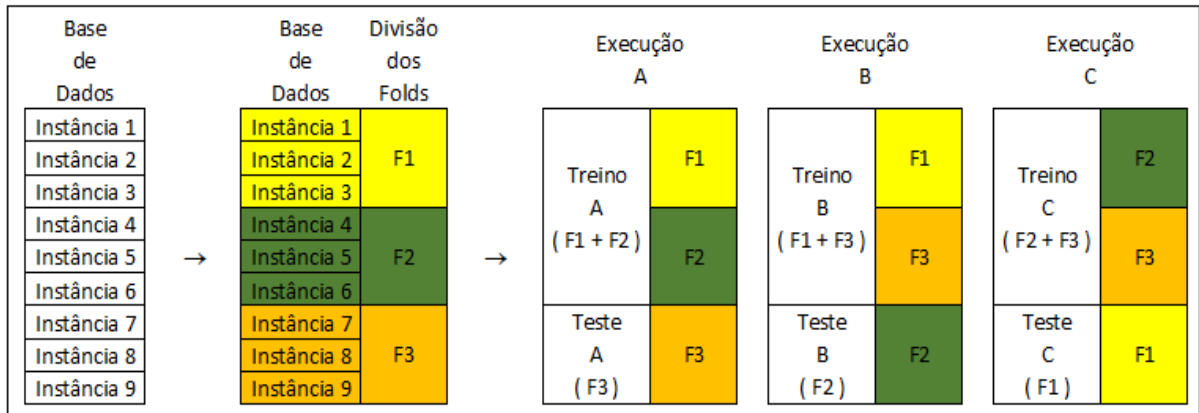
3.5 Preparação dos Conjuntos de Treinamento e Teste

Os dados empregados neste estudo, e que representavam cada um dos indivíduos participantes, foram previamente descritos. Os dados de cada indivíduo foram concatenados em linhas, classificados e rotulados com base nos critérios do SRQ-20. Ou seja, os indivíduos que responderam “sim” para 7 ou mais das 20 questões compuseram a primeira classe, enquanto os demais indivíduos compuseram a segunda classe. Assim, o conjunto de dados foi composto por 179 linhas, cujas colunas representavam os dados relativos a cada um dos indivíduos sob diferentes contextos. Tais contextos (dados) com base/indicação podem ser visualizados nos Apêndices I e II e referem-se a componentes químicos e físicos, granulometria e enzimas do solo, presença de agrotóxicos e dados de saúde. Foram realizados diferentes conjuntos de experimentos considerando os dados de cada um dos contextos de forma individual e também a combinação entre eles. Neste processo, realizou-se variações de valores para os possíveis parâmetros de cada um dos algoritmos de mineração de dados empregado, conforme descrito na Seção 3.6.

Antes da aplicação das técnicas de mineração, os dados foram organizados para compor os conjuntos de treinamento e teste, sendo que ambos os conjuntos mantiveram a representatividade das classes pertencentes ao problema em questão. A Figura 7 ilustra o referido processo considerando um exemplo de uma base de dados com 9 instâncias. Inicialmente, os dados foram divididos em três partes (*folds* - F1, F2 e F3), a partir dos quais foi empregada validação cruzada com três execuções (A, B e C). Em cada execução, uma parte era usada como conjunto de teste e as outras duas eram utilizadas como conjunto de treinamento. Os resultados finais são expressos pela média e desvio padrão (dp) das taxas obtidas para as três execuções, sendo estas taxas de avaliação a acurácia, a precisão, recall e f1 score, as quais são descritas na Seção 3.7.

Destaca-se ainda que a divisão da base de dados em 3 *folds* produziu subconjuntos com 60, 60 e 59 elementos. Consequentemente, a combinação para a execução das 3 execuções considerou conjuntos de treinamento com 119 (60 + 59) ou 120 (60 + 60) elementos, enquanto que o conjunto de teste continha os demais elementos, isto é, 60 ou 59, respectivamente.

Figura 7. Definição dos conjuntos de Treinamento e Teste utilizados nos experimentos.



Fonte: A autora, 2023.

3.6 Algoritmos de Mineração dos Dados Avaliados

Conforme já apresentado, para a aplicação das técnicas de Mineração de Dados, foi utilizada a ferramenta computacional WEKA, tendo em vista que esta possui um conjunto diversificado de algoritmos que implementam tais técnicas. Para a realização dos experimentos relacionados a classificação e agrupamento foi utilizada a opção “*Supplied test set*”, bem como a definição dos conjuntos de treinamento e teste previamente apresentada. Por outro lado, os algoritmos compreendidos nas técnicas de Associação e Seleção de Atributos não permitem o uso destes 2 conjuntos, tendo sido empregados todos os 179 indivíduos que compunham a base como entrada, isto é, a união dos conjuntos de treinamento e teste utilizados anteriormente.

Diferentes valores para os parâmetros de cada algoritmo foram testados. Porém, tal variação não se reverteu em ganhos significativos. Diante disso, de forma geral, a maior parte dos parâmetros foi mantida com seus valores padrão. Para o algoritmo de classificação J48, os melhores resultados foram alcançados a partir do uso de todos os atributos com seus valores padrão. Com relação à Seleção de Atributos, para os algoritmos *CfsSubsetEval*, *InfoGainAttributeEval*, *ClassifierAttributeEval* e *GainRatioAttributeEval* também foram utilizados os valores padrão.

Para o algoritmo de agrupamento EM, os valores testados para o número de clusters foram 2 e -1. No primeiro caso, tem-se a quantidade de classes presentes no problema, enquanto que no segundo caso o próprio algoritmo determina quantos agrupamentos serão criados com base em seus próprios critérios (regras). Os melhores resultados foram alcançados pelo uso dos parâmetros número de execuções (ciclos) executadas do k-means com valor 1.000 e número máximo de iterações executadas com valor 10.000.

Para o algoritmo de associação Apriori, foi aplicada a mineração de regras para a determinação das classes ao invés das clássicas regras de associação. Isso significa que todas as regras tinham como consequente apenas o atributo que definia o rótulo relativo ao IQR20 e seus possíveis valores “sim” ou “não”. O valor 0.5 foi empregado como ponto de corte (valor mínimo) para os atributos (métricas) suporte e confiança de cada regra gerada.

3.7 Métricas Empregadas para Avaliar os Resultados Obtidos

Considerando que há um problema de 2 classes, as medidas anteriores podem ser calculadas a partir de uma Matriz de Confusão 2x2 (Figura 8), em que a diagonal principal representa os acertos e a diagonal secundária representa os erros do classificador. Dentre os indicadores obtidos a partir desta Matriz de Confusão sob um contexto de avaliação mais geral, destaca-se a totalização de acertos e erros considerados verdadeiros positivos (TP), falsos negativos (FN), verdadeiros negativos (TN) e falsos positivos (FP).

Essa representação é útil ao analisar a frequência de erros cometidos em cada classe ou par de classes em termos de erros Tipo I (FP) e Tipo II (FN). TP compreende as amostras positivas classificadas como positivas. FN compreende as amostras positivas classificadas como negativas. TN compreende as amostras negativas classificadas como negativas. FP compreende as amostras negativas classificadas como positivas. TP, FN, TN e FP são expressos em valores absolutos e permitem a derivação de taxas em percentuais, facilitando a avaliação da performance dos classificadores (MARTINS et al. 2021).

Figura 8. Matriz de Confusão 2X2.

		Classe Sugerida pelo Classificador	
		Positiva	Negativa
Classe Real	Positiva	Verdadeiro Positivo	Falso Negativo
	Negativa	Falso Positivo	Verdadeiro Negativo

Fonte: MARTINS et al. 2021.

As Equações 1-4 foram obtidas em MARTINS et al. (2021) e apresentam os cálculos para as métricas Acurácia, Precisão, Recall e F1 Score. Acurácia (Equação 1) considera quantas instâncias positivas reais foram rotuladas como positivas (TP) e quantas instâncias negativas reais foram rotuladas como negativas (TN). Em outras palavras, contabiliza as instâncias que foram classificadas corretamente, ou seja, de acordo com a classificação do SRQ-20, contabiliza-se quantos indivíduos rotulados como “sim” e “não” verdadeiramente foram classificadas como “sim” e “não”. Cabe lembrar que o indivíduo rotulado como “sim” respondeu 7 ou mais perguntas com “sim”, o que o lhe atribui maior probabilidade de ter um transtorno mental comum. Precisão (Equação 2) avalia quantas respostas previstas como positivas eram realmente positivas, ou seja, quantas instâncias classificadas como “sim” pertenciam à classe positiva. Esta é uma boa medida para ser usada em casos onde o custo do FP é alto, o que para este caso seria classificar “não” como “sim”. Recall (Equação 3) calcula quantas instâncias positivas reais foram rotuladas como positivas (TP) por um modelo, ou seja, quantas instâncias classificadas como “sim” eram realmente “sim”. Portanto, Recall seria uma métrica a ser utilizada para selecionar o melhor modelo quando há um alto custo associado à FN, que para este caso seria classificar os indivíduos rotulados com “sim” como “não”. Por fim, o F1 Score (Equação 4) é útil quando se busca um equilíbrio entre precisão e recall, inclusive quando há uma distribuição de classes desigual.

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

$$Precisão = \frac{TP}{TP + FP} \times 100 \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$F1 - Score = 2 * \frac{Precisão \times Recall}{Precisão + Recall} \quad (4)$$

Destaca-se que, diante do presente contexto, adotou-se a premissa de que seria melhor classificar um indivíduo cujo rótulo real seja “não” como “sim”, sendo este encaminhado para realizar um possível acompanhamento e tratamento. Ainda nesta linha, a adoção de uma premissa contrária implicaria na negação de acompanhamento e tratamento para alguém que realmente necessitasse e, por consequência, possíveis danos irreversíveis ao indivíduo, a seus familiares e àqueles que com ele convive.

3.8 Interpretação dos Resultados

A última etapa do processo KDD consiste na Interpretação (avaliação) dos resultados. Nesta etapa procurou-se identificar as relações entre o uso de agrotóxicos com os relatos de saúde e impactos no solo, tendo como base os dados apresentados na Seção 3.1 Base de Dados.

Foram realizadas análises dos contextos individuais, conforme os Apêndices I e II. Estes contextos se referem a fertilidade do solo, propriedades físicas do solo, granulometria do solo e presença de agrotóxicos no solo, além de dados de saúde dos entrevistados relativos a uma autoavaliação dos mesmos. Além dos contextos individuais, também foram avaliadas diferentes alternativas quanto a suas combinações.

Na tentativa de validar os resultados obtidos pelos diferentes algoritmos e conjuntos de dados, realizou-se uma comparação dos mesmos, a qual envolveu a análise das taxas alcançadas por cada um, bem como dos atributos e valores envolvidos.

4 RESULTADOS

Este capítulo apresenta os resultados obtidos a partir do desenvolvimento da metodologia proposta e do uso do ferramental descrito nos capítulos anteriores.

4.1 Saúde: SRQ-20

Foram utilizadas as questões da saúde descritas no Quadro 4, que compreendem um recorte do Apêndice B. Os resultados obtidos pelos testes realizados por meio das diferentes técnicas são apresentados nas subseções seguintes.

Quadro 4. Recorte das questões do Apêndice B relacionadas à Saúde – SQR-20.

Identificador	Questões do Apêndice B relacionadas à Saúde
24	Você tem dores de cabeça frequentes?
25	Você tem falta de apetite?
26	Você dorme mal?
27	Você se assusta com facilidade?
28	Você tem tremores musculares?
29	Você se sente nervoso, tenso ou preocupado?
30	Você tem má digestão?
31	Você sente que suas ideias ficam embaralhadas de vez em quando?
32	Você tem se sentido triste ultimamente?
33	Você tem chorado mais do que de costume?
34	Você encontra dificuldade para realizar com satisfação suas atividades diárias?
35	Você tem dificuldade para tomar decisões?
36	Você acha que seu trabalho é penoso e lhe causa sofrimento?
37	Você se acha incapaz de desempenhar um papel útil na sua vida?
38	Você tem perdido o interesse pelas coisas?
39	Você se sente uma pessoa sem valor?
40	Alguma vez você pensou em acabar com sua vida?
41	Você se sente cansado o tempo todo?
42	Você sente alguma coisa desagradável no estômago?
43	Você se cansa com facilidade?

Fonte: A autora, 2023.

4.1.1 Clusterização com algoritmo EM: saúde – SQR-20

Os resultados obtidos a partir do algoritmo de clusterização *Expectation Maximization* (EM) estão expostos na Tabela 1. Nessa execução o próprio EM determinou a quantidade de agrupamentos seguindo seus critérios internos.

Tabela 1. Resultados obtidos com o número de clusters determinado pelo EM: Saúde – SQR-20.

Clusters	Execuções		
	A	B	C
0	1	1	2
1	29	17	16
2	6	7	-
3	23	35	42

Fonte: A autora, 2023.

Durante as análises, percebeu-se que não é possível criar agrupamentos que separem as instâncias das classes “sim” e “não” em ambos os conjuntos de treinamento e teste. Foram identificados agrupamentos que representaram subconjuntos das instâncias com classificação “sim”; agrupamentos que representaram subconjuntos das instâncias com classificação “não”; além de agrupamentos com instâncias pertencentes a ambas as classificações. Embora não seja apresentado aqui, os resultados para a execução do EM com a definição de apenas 2 clusters, considerado previamente a existências de 2 classes, também não permitiu a separação das classes.

4.1.2 Árvores de decisão com algoritmo J48: saúde - SQR-20

As regras que definiram as Árvores de Decisão geradas com a aplicação do Algoritmo J48 ao conjunto de dados de treinamento são apresentadas no Quadro 5. Ao realizar uma análise sob a perspectiva dos atributos envolvidos na geração das regras (Quadro 5), identificou-se dois atributos em comum (28 e 41) nas regras produzidas em todas as execuções. Além disso, ao se avaliar a relevância dos atributos, observou-se o 41 figura nos primeiros níveis na primeira e terceira

execuções. A segunda e terceira execuções apresentaram mais atributos em comum entre si (41, 25, 28). O atributo 31 só apareceu na Execução A do teste, enquanto os atributos 37, 36, 26, 43 e 30 só apareceram na segunda execução. E os atributos 33, 39, 27 e 29 só estavam presentes na Execução C.

A partir da aplicação das regras apresentadas no Quadro 5 aos conjuntos de teste, foram calculados média e desvio padrão apresentados na Tabela 2. A acurácia atingiu 82,1% ($\sigma=2,1$) e as demais métricas obtiveram valores próximos, o que indica que a taxas gerais de erros são semelhantes para ambas as classes (“sim” e “não”), sendo que as instâncias com valores NA em seus atributos estão compreendidos entre as classes.

Tabela 2. Resultados médios das execuções realizadas com o Algoritmo J48: Saúde – SQR-20.

Acurácia		Precisão		Recall		F-Measure	
%	σ	%	σ	%	σ	%	σ
82,1	2,1	73,7	0,7	71,4	8,7	72,0	5,1

Fonte: A autora, 2023.

Em relação aos erros de classificação, como esperado, foram identificadas instâncias da classe “não” preditas como “sim” e vice-versa. A título de exemplo, Execução A do Quadro 5, uma regra gerada com o uso dos atributos 32 (não) e 41 (não) apresentou 67 acertos e apenas 2 erros. Nesta mesma execução houveram 10 acertos e 2 erros a partir de uma regra gerada com o uso dos atributos 32 (sim), 31 (não), 42 (não) e 28 (não). Em ambos os casos, os erros se referem a instâncias pertencentes à classe “sim” que foram classificadas como “não”. Houveram também casos em que instâncias pertencentes à classe “não” foram classificadas como “sim”. Uma regra gerada a partir apenas do atributo 41 (sim) produziu 22 acertos e 2 erros (Execução C do Quadro 5).

Mediante os resultados obtidos com a aplicação do Algoritmo J48 e a análise dos atributos utilizados, pode-se realizar o questionamento sobre a forma como o questionário SQR-20 considera potencial de problemas de saúde, mais especificamente psiquiátricos. Ou seja, talvez a quantidade de respostas “sim” não seja um indicativo de transtorno mental comum. Neste sentido, talvez poderia ser mais interessante considerar quais as perguntas tiveram “sim” como resposta para que seja possível ter mais assertividade no diagnóstico.

Quadro 5. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48: Saúde – SQR-20.

Execução A	Execução B	Execução C
32 = NA: nao (3.0)	25 = NA: nao (3.0)	41 = NA: nao (2.0)
32 = nao	25 = nao	41 = nao
41 = NA: nao (0.0)	37 = NA: nao (0.0)	28 = NA: nao (0.0)
41 = nao: nao (67.0/2.0)	37 = PrefereNaoResponder: nao (0.0)	28 = nao
41 = sim	37 = nao	33 = NA: nao (0.0)
31 = NA: sim (0.0)	36 = NA: nao (1.0)	33 = nao
31 = nao: nao (3.0)	36 = PrefereNaoResponder: sim (1.0)	25 = NA: nao (0.0)
31 = sim: sim (8.0)	36 = nao	25 = nao
32 = sim	36 = nao	39 = NA: nao (0.0)
31 = NA: sim (0.0)	28 = NA: nao (0.0)	39 = nao: nao (64.0/1.0)
31 = nao	28 = nao	39 = sim
42 = NA: nao (0.0)	41 = NA: nao (0.0)	29 = NA: nao (0.0)
42 = nao	41 = nao: nao (66.0/1.0)	29 = nao: nao (3.0)
28 = NA: nao (0.0)	41 = sim	29 = sim: sim (2.0)
28 = nao: nao (10.0/2.0)	30 = NA: nao (0.0)	25 = sim: sim (3.0/1.0)
28 = sim: sim (2.0)	30 = nao: nao (2.0)	33 = sim
42 = sim: sim (3.0)	30 = sim: sim (2.0)	27 = NA: sim (0.0)
31 = sim: sim (24.0/1.0)	28 = sim	27 = nao: nao (9.0/2.0)
	32 = NA: sim (0.0)	27 = sim: sim (6.0)
	32 = nao: nao (5.0/1.0)	28 = sim
	32 = sim: sim (6.0)	29 = NA: sim (0.0)
	36 = sim	29 = nao: nao (2.0)
	26 = NA: sim (0.0)	29 = sim: sim (6.0)
	26 = nao	41 = sim: sim (22.0/2.0)
	43 = NA: nao (0.0)	
	43 = nao: nao (4.0)	
	43 = sim: sim (2.0)	
	26 = sim: sim (5.0)	
	37 = sim: sim (10.0/1.0)	
	25 = sim: sim (12.0)	

Fonte: A autora, 2023.

Baseado nos atributos utilizados e suas combinações que culminaram nas regras geradas, por meio da aplicação do Algoritmo J48, ressalta-se que erros referentes a instâncias pertencentes à classe “sim” que foram classificadas como “não” e vice versa, assim como verificou-se que o atributo comum em todos os grupos foi o 41, corroborando com o questionamento sobre o uso de uma simples contagem de respostas e a importância/peso de cada uma das perguntas.

4.1.3 Regras de associação com algoritmo apriori: saúde – SQR-20

Após o processo de mineração de dados efetuado pelo algoritmo Apriori pela ferramenta WEKA na base de dados, obteve-se uma série de regras de associação, das quais algumas são listadas no Quadro 6. Das 3 milhões de regras inicialmente geradas, a maioria representava elementos da classe “não”, cujas respostas às questões de saúde do SQR-20 tinham logicamente a resposta “não” em sua maioria.

Quadro 6. Regras obtidas com a utilização do algoritmo Apriori: Saúde – SQR-20.

```

25=nao 28=nao 33=nao 38=nao 39=nao 41=nao 88 ==> iqr20rotulo=nao 88 conf:(1)
25=nao 28=nao 33=nao 36=nao 37=nao 39=nao 41=nao 86 ==> iqr20rotulo=nao 86 conf:(1)
28=nao 32=nao 37=nao 39=nao 41=nao 85 ==> iqr20rotulo=nao 85 conf:(1)
25=nao 26=nao 28=nao 32=nao 33=nao 37=nao 38=nao 39=nao 42=nao 43=nao 45 ==>
iqr20rotulo=nao 45 conf:(1)
27=nao 28=nao 29=nao 32=nao 36=nao 38=nao 39=nao 41=nao 42=nao 32 ==> iqr20rotulo=nao 32
conf:(1)
24=nao 27=nao 29=sim 32=nao 33=nao 34=nao 39=nao 40=nao 41=nao 19 ==> iqr20rotulo=nao 19
conf:(1)
33=nao 43=sim 38 ==> iqr20rotulo=nao 19 conf:(0.5)

```

Fonte: A autora, 2023.

Diante dos resultados apresentados no Quadro 6, o corrente experimento foi repetido, porém considerando apenas as instâncias pertencentes à classe “Sim”. Algumas das regras geradas são apresentadas no Quadro 7, podendo-se verificar que a maioria dos atributos empregados na composição das regras possuem o valor “sim”. Assim, a caracterização apresentada no Quadro 6 e Quadro 7 corrobora com o

exposto anteriormente quanto à possibilidade de haver questões chaves no conjunto de perguntas do SQR-20. Destaca-se que em ambos, as regras apresentadas em geral têm um alto valor para a confiança (1 ou 100%), com a particularidade de que há valores maiores para os suportes das regras apresentadas no Quadro 6 do que no Quadro 7, em decorrência da proporcionalidade com que as classes “sim” e “não” estão representadas na base de dados.

Quadro 7. Regras obtidas com a utilização do algoritmo Apriori para instâncias da classe “sim”: Saúde – SQR-20.

```

24=sim 25=sim 26=sim 29=sim 31=sim 41=sim 43=sim 3 ==> iqr20rotulo=sim 3   conf:(1)
24=sim 25=sim 26=sim 29=sim 31=sim 37=nao 41=sim 3 ==> iqr20rotulo=sim 3   conf:(1)
24=sim 25=sim 26=sim 29=sim 31=sim 39=nao 41=sim 3 ==> iqr20rotulo=sim 3   conf:(1)
27=sim 31=sim 33=sim 18 ==> iqr20rotulo=sim 18 conf:(1)
26=sim 29=sim 32=sim 41=sim 18 ==> iqr20rotulo=sim 18 conf:(1)
27=sim 29=sim 31=sim 32=sim 18 ==> iqr20rotulo=sim 18 conf:(1)
29=sim 31=sim 41=sim 29 ==> iqr20rotulo=sim 28 conf:(0.97)
31=sim 41=sim 30 ==> iqr20rotulo=sim 29 conf:(0.97)
26=sim 31=sim 41=sim 23 ==> iqr20rotulo=sim 22 conf:(0.96)
27=sim 33=sim 22 ==> iqr20rotulo=sim 21 conf:(0.95)
31=sim 39=nao 41=sim 21 ==> iqr20rotulo=sim 20 conf:(0.95)
28=nao 31=sim 41=sim 21 ==> iqr20rotulo=sim 20 conf:(0.95)

```

Fonte: A autora, 2023.

Com a repetição do experimento, considerando apenas as instâncias pertencentes à classe “Sim” (conforme algumas das regras geradas exibidas no Quadro 7), foi verificado que o atributo 41 (Você se sente cansado o tempo todo?) é exibido com maior frequência, seguidos dos atributos 31 (Você sente que suas ideias ficam embaralhadas de vez em quando?), 27 (Você se assusta com facilidade?) e 29 (Você se sente nervoso, tenso ou preocupado?). Desta forma, destaca-se a importância do atributo 41 para o estudo em questão. Além disso, notou-se que de forma geral, houve associação entre os atributos 41 e 31 nas regras geradas, indicando que há relação entre eles.

4.1.4 Seleção de atributos com diferentes algoritmos: saúde – SQR-20

A Tabela 3 apresenta os resultados obtidos por diferentes algoritmos implementados na ferramenta WEKA e cujo objetivo compreende a seleção ou a ordenação de atributos tendo em vista diferentes critérios quanto a sua influência na distinção das classes presentes na base de dados.

Tabela 3. Relação dos atributos selecionados: Saúde – SQR-20.

Algoritmo	(Ordem dos) Atributos Selecionados
<i>GainRatioAttributeEval</i>	41, 32, 25, 33, 37 43, 38, 31, 35, 28, 34, 36, 29, 26, 40, 39, 30, 27, 42, 24
<i>ClassifierAttributeEval</i>	43, 30, 31, 42, 29, 28, 27, 26, 25, 32, 33, 34, 39, 41, 40, 38, 35, 37, 36, 24
<i>InfoGainAttributeEval</i>	32, 41, 33, 31, 43, 35, 37, 29, 25, 34, 26, 38, 36, 40, 28, 27, 30, 39, 42, 24
<i>CfsSubsetEval-BestFisrt</i>	25, 31, 32, 33, 41, 43
<i>CfsSubsetEval-GreedyStepwise</i>	41, 32, 33, 43, 31, 25, 35, 36, 38, 29, 26, 37, 34, 28, 30, 40, 27, 39, 24, 42

Fonte: A autora, 2023.

Na análise de Seleção de Atributos com diferentes algoritmos foi possível identificar alguns padrões quanto a sua relevância e que corroboram com os resultados encontrados anteriormente. Verificou-se que o atributo 41 estava presente em todos os algoritmos utilizados e sua ordem de aparição em cada um deles foi: 1º lugar com *GainRatioAttributeEval*, 2º lugar com *InfoGainAttributeEval*, 5º lugar com *CfsSubsetEval* e 14º lugar com *ClassifierAttributeEval*, evidenciando a importância desse atributo para o estudo em questão.

Além do atributo 41, os atributos comuns entre todos os algoritmos utilizados foram: 32 (em 1º lugar com *InfoGainAttributeEval*, 2º lugar com *GainRatioAttributeEval*, 3º lugar com *CfsSubsetEval* e 10º lugar com *ClassifierAttributeEval*); 25 (em 1º lugar com *CfsSubsetEval*, 3º lugar com *GainRatioAttributeEval* e 9º lugar com *ClassifierAttributeEval* e *InfoGainAttributeEval*); 33 (em 3º lugar com *InfoGainAttributeEval*, 4º lugar com *GainRatioAttributeEval* e *CfsSubsetEval* e 11º lugar com *ClassifierAttributeEval*); 31 (em 2º lugar com

CfsSubsetEval, 3º lugar com *ClassifierAttributeEval*, 4º lugar com *InfoGainAttributeEval* e 8º lugar com *GainRatioAttributeEval*).

Todos os atributos selecionados pelo algoritmo *CfsSubsetEval-BestFisrt* (25, 31, 32, 33, 41, 43) também foram selecionados dentre as primeiras opções pelos demais algoritmos (*ClassifierAttributeEval*, *GainRatioAttributeEval* e *InfoGainAttributeEval*). Tais resultados confirmam a tese de que há atributos mais influentes que outros na distinção das classes presentes na base de dados. Em contrapartida, detectou-se que os atributos 24, 27 e 39 ficaram nas últimas posições dentre aqueles ordenados pelos algoritmos *ClassifierAttributeEval*, *GainRatioAttributeEval*, *InfoGainAttributeEval* e *CfsSubsetEval-GreedyStepwise*.

4.1.5 Resultados obtidos e correlações: saúde – SQR-20

A utilização dos agrotóxicos tem trazido uma série de consequências, que são agravadas por diversas determinantes de ordens cultural, social e econômica (PERES et al., 2005). De acordo com a Carneiro et al. 2012 mesmo que alguns dos ingredientes ativos dos agrotóxicos possam ser classificados como medianamente ou pouco tóxicos, pelos seus efeitos agudos, não se pode deixar de considerar os efeitos crônicos, que podem ocorrer meses, anos ou até décadas após a exposição, manifestando-se em várias doenças. Diante deste amplo e diverso contexto, bem como dos resultados obtidos, buscou-se por referências para tentar justificar tais resultados.

Diversos estudos mostram o potencial tóxico de agrotóxicos sobre a saúde da população urbana e também rural. Se quem está longe dos campos agrícolas está sujeito aos agrotóxicos, a situação pode se agravar para aqueles que, além de consumirem o alimento contaminado pelo excesso de resíduos, manipulam tais produtos, durante a aplicação (CARNEIRO et al. 2012; PERES et al., 2003).

Dentre as constatações, a comunidade científica tem detectado a presença de agrotóxicos diversos em amostras de sangue humano, urina e leite materno (BELO et al. 2012; PIGNATI et al. 2012). Desse modo, são grandes as possibilidades de ocorrência de anomalias congênitas, câncer, disfunções na reprodução humana, bem como distúrbios endócrinos, neurológicos e mentais (ARMAS et al. 2007; SIQUEIRA; KRUSE, 2008).

A exposição tóxica aos organofosforados pode ocorrer por inalação, via transdérmica ou ingestão. O principal alvo para a ação desses agentes neurotóxicos é a enzima acetilcolinesterase (AChE), a qual controla as ações centrais e periféricas do neurotransmissor acetilcolina (ACh). Os organofosforados inibem irreversivelmente a acetilcolinesterase, que deixa de hidrolisar a ACh, levando a um acúmulo desta nas sinapses centrais e periféricas, promovendo hiperestimulação colinérgica, que resulta em broncorreia, fasciculação muscular e cardíaca, convulsões, depressão respiratória e doenças neurológicas como Parkinson e Alzheimer (EDDLESTON et al. 2008; KOUREAS, 2012).

Existem estudos na literatura que comprovam que os organofosforados afetam o sistema nervoso dos organismos, inibindo a atividade da acetilcolinesterase (AChE). Richendrfer e Creton (2015) estudaram os efeitos de organofosforados sobre o comportamento e a atividade da AChE de larvas do *Danio rerio* (peixe-zebra) e constataram que houve mudanças na velocidade de natação (hipoatividade e hiperatividade), descanso e tigmotatismo. Em outro estudo, Crosby et al. (2015), também identificaram que as larvas de *Danio rerio* (peixe-zebra) reduziram significativamente a atividade de natação quando expostos a agrotóxicos.

No presente estudo, verificou-se que existem associações para os seguintes sintomas: dores de cabeça frequentes, falta de apetite, noites mal dormidas, sentimentos de nervosismo, tensão e preocupação, incapacidade de desempenhar um papel útil e desvalorização de si mesmo, assim como possuir ideias embaralhadas. Tais distúrbios psiquiátricos são preocupantes, pois se não forem tratados de maneira adequada, podem ser potenciais fatores de desencadeamento de ideação suicida.

Estudos apontam que a exposição a agrotóxicos é um fator de risco para suicídio e existe um risco maior de morte por suicídio entre trabalhadores rurais que residem em áreas de maior exposição a agrotóxicos. Precedendo o suicídio ou a tentativa de suicídio, existe a ideação suicida, que é a presença de pensamentos passivos ou ativos sobre o fim da vida. Trata-se de um tema pouco explorado na população de áreas rurais no Brasil, contando com uma prevalência de 5%. Diversos estudos indicam a associação entre intoxicação aguda por agrotóxicos e distúrbios psiquiátricos. Entretanto, existe ainda uma lacuna na literatura quanto à associação entre distúrbios psiquiátricos e a exposição crônica, isto é, a exposição prolongada e cumulativa a baixas doses de agrotóxicos (FREIRE; KOIFMAN, 2013; KIM; KO; LEE 2013; FARIA, 2014; GONZAGA; BALDO; CALDEIRA, 2021).

4.2 SAÚDE: DOENÇAS

Foram utilizadas as questões da saúde descritas no Quadro 8, as perguntas compreendem um recorte do Apêndice B. Os resultados obtidos pelos testes realizados por meio das diferentes técnicas são apresentados nas subseções seguintes.

Quadro 8. Recorte de questões do Apêndice B relacionadas à Saúde: doenças.

Identificador	Questões do Apêndice B relacionadas à Saúde
44	Contato com agrotóxicos
45	Problemas ortopédicos
46	Dor muscular
47	Fadiga
48	Doença respiratória
49	Doenças cardiovasculares
50	Alterações endócrinas
51	Doença hepática
52	Doença renal urinária
53	Problemas neurológicos
54	Problemas gastrointestinais
55	Problemas de pele
56	Tabagista
57	Consumo de álcool

Fonte: A autora, 2023.

4.2.1 Clusterização com algoritmo EM: saúde (doenças)

Os resultados obtidos a partir do algoritmo de clusterização EM aplicados ao conjunto de dados anteriormente descrito estão expostos na Tabela 4. Nesta execução o próprio EM determinou a quantidade de agrupamentos seguindo seus critérios internos.

Tabela 4. Resultados obtidos com clusteres determinados pelo próprio algoritmo EM: Saúde (doenças).

Clusteres	Execuções		
	A	B	C
0	6	7	5
1	30	29	28
2	23	24	27

Fonte: A autora, 2023.

Conforme representado na tabela anterior, não foi possível distinguir os agrupamentos das instâncias classificadas como “sim” e como “não”. Os resultados da Tabela 4, indicam que as demais instâncias, independentemente da classe à qual pertenciam, ficaram distribuídas principalmente em dois agrupamentos, sendo que um terceiro agrupamento (zero – ‘0’) era composto basicamente por instâncias em que a maior parte dos atributos continham a resposta ‘NA’.

Novamente, nos 2 cenários testados, com a definição de apenas 2 agrupamentos e sem a definição do número de agrupamentos, não foi possível separar as instâncias das classes ‘sim’ e ‘nao’.

4.2.2 Árvores de decisão com algoritmo J48: saúde (doenças)

Para os testes de Árvores de Decisão, com a aplicação do Algoritmo J48, foram geradas as regras apresentadas no Quadro 9, Quadro 10 e Quadro 11, as quais produziram os resultados expostos na Tabela 5. Verificou-se que novamente, os erros de classificação compreenderam instâncias da classe “não” preditas como “sim” e vice-versa. A título de exemplo, na Execução A do Quadro 9, uma regra gerada com o uso dos atributos referentes a problemas neurológicos (não) e alterações endócrinas (não) apresentou 58 acertos e 9 erros. Nesse caso, os erros se referem a instâncias pertencentes à classe “sim” que foram classificadas como “não”. Nesta mesma execução houveram 17 acertos e 4 erros a partir de uma regra gerada com o uso dos atributos referentes a problemas neurológicos (sim) e fadiga (sim), indicando que instâncias pertencentes à classe “não” foram classificadas como “sim”.

Na execução B do Quadro 10, uma regra gerada com o uso do atributo relacionado a problemas neurológicos (não) apresentou 74 acertos e 17 erros. Nesse

caso, os erros se referem a instâncias pertencentes à classe “sim” que foram classificadas como “não”. Ainda, na mesma execução houveram 27 acertos e 9 erros a partir de uma regra gerada com o uso do atributo relativo a dor muscular, indicando que instâncias pertencentes à classe “não” foram classificadas como “sim”.

Na execução C do Quadro 11, uma regra gerada com o uso dos atributos problemas gastrointestinais (“não”) e 47 Fadiga (não) apresentou 68 acertos e 7 erros. Os erros são referentes a instâncias pertencentes à classe “sim” que foram classificadas como “não”. Além disso, na mesma execução houveram 15 acertos e apenas 1 erro a partir de uma regra gerada com o uso dos atributos referentes a problemas gastrointestinais (“sim”) e problemas ortopédicos (“sim”), indicando que instâncias pertencentes à classe “não” foram classificadas como “sim”.

A média e desvio padrão definidos para a avaliação dos resultados foram calculados com base nos valores obtidos nas 3 execuções e podem ser visualizados na Tabela 5. A acurácia atingiu 69,2% ($\sigma=7,3$) e as demais métricas obtiveram valores inferiores a 54%, inclusive com valores mais elevados para o desvio padrão (σ), o que indica que tal recorte por si só não tem capacidade de diferenciar as classes “sim” e “não”.

Quadro 9. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48 – Execução A: Saúde (doenças).

```

problemas_neurologicos = NA: nao (12.0/5.0)
problemas_neurologicos = nao
| alteracoes_endocrinas = NA: nao (0.0)
| alteracoes_endocrinas = nao: nao (58.0/9.0)
| alteracoes_endocrinas = sim
| | doenca_respiratoria = NA: nao (0.0)
| | doenca_respiratoria = nao
| | | contato_com_agrotoxicos = NA: nao (3.0)
| | | contato_com_agrotoxicos = nao: sim (1.0)
| | | contato_com_agrotoxicos = sim
| | | | doencas_cardiovasculares = NA: nao (0.0)
| | | | doencas_cardiovasculares = nao: sim (2.0)
| | | | doencas_cardiovasculares = sim: nao (2.0)
| | doenca_respiratoria = sim: sim (2.0)
problemas_neurologicos = sim
| fadiga = NA: sim (0.0)
| fadiga = nao
| | doenca_renal_urinaria = NA: nao (0.0)
| | doenca_renal_urinaria = nao
| | | problemas_pele = NA: nao (0.0)
| | | problemas_pele = nao
| | | | problemas_gastrointestinais = NA: nao (0.0)
| | | | problemas_gastrointestinais = nao: nao (7.0/1.0)
| | | | problemas_gastrointestinais = sim: sim (3.0)
| | | problemas_pele = sim: nao (9.0/1.0)
| | doenca_renal_urinaria = sim: sim (4.0/1.0)
| fadiga = sim: sim (17.0/4.0)

```

Fonte: A autora, 2023.

Tabela 5. Resultados médios das execuções realizadas com o algoritmo J48: Saúde (doenças).

Acurácia		Precisão		Recall		F-Measure	
%	σ	%	σ	%	σ	%	σ
69,2	7,3	53,8	12,2	43,2	9,5	47,8	10,7

Fonte: A autora, 2023.

Quadro 10. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48 – Execução B: Saúde (doenças).

```

problemas_neurologicos = NA: nao (11.0/3.0)
problemas_neurologicos = nao: nao (74.0/17.0)
problemas_neurologicos = sim
| dor_muscular = NA: sim (0.0)
| dor_muscular = nao: nao (7.0/1.0)
| dor_muscular = sim: sim (27.0/9.0)

```

Fonte: A autora, 2023.

Quadro 11. Regras geradas para a definição das Árvores de Decisão com utilização do algoritmo J48 – Execução C: Saúde (doenças).

```

problemas_gastrointestinais = NA: nao (15.0/7.0)
problemas_gastrointestinais = nao
| fadiga = NA: nao (0.0)
| fadiga = nao: nao (68.0/7.0)
| fadiga = sim
| | tabagista = NA: sim (0.0)
| | tabagista = exfumante: nao (2.0)
| | tabagista = nao: sim (9.0/2.0)
| | tabagista = sim: sim (0.0)
problemas_gastrointestinais = sim
| tabagista = NA: sim (0.0)
| tabagista = exfumante: nao (4.0/1.0)
| tabagista = nao
| | problemas_ortopedicos = nao: nao (3.0/1.0)
| | problemas_ortopedicos = sim: sim (15.0/1.0)
| tabagista = sim: sim (3.0/1.0)

```

Fonte: A autora, 2023.

Mediante os resultados obtidos e a análise dos atributos utilizados, verificou-se que os atributos 44 (Teve contato com agrotóxicos) e 47 (Você tem fadiga?) foram comuns na primeira e última execução (A e C), sendo que o atributo 53 (Você tem

problemas neurológicos?) foi comum entre a primeira e a segunda execução. Não houve atributos comuns entre as 3 execuções dos testes.

As regras geradas levantam questionamentos quanto às relações das doenças listadas na base de dados e os problemas de saúde ocasionados em função da exposição a agrotóxicos. Embora os problemas ortopédicos não estejam diretamente ligados a intoxicação, estes podem ser consequências do acúmulo de algum componente químico no corpo dos indivíduos durante tempo prolongados de exposição. Em contrapartida, na execução A e B, o atributo problemas neurológicos apareceu no topo da árvore. Já na execução C, no topo da árvore apareceu o atributo problemas gastrointestinais, indicando que ambos os problemas de saúde estão diretamente relacionados com a exposição a agrotóxicos.

4.2.3 Regras de associação com algoritmo apriori: saúde (doenças)

Os resultados encontrados por meio da utilização das regras de associação com algoritmo Apriori, mostraram a existência de associações entre os atributos descritos a seguir. Verificou-se, em vários casos, problemas ortopédicos associados a dor muscular e fadiga. Já para alguns casos em que a resposta para o atributo contato com agrotóxicos foi “sim”, notou-se a associação com problemas ortopédicos e dor muscular. Houve ainda relações entre respostas “sim” para problemas ortopédicos e fadiga e doenças cardiovasculares, dentre as quais muitas ainda estavam atrelados também a ocorrência de problemas neurológicos. Além disso, notou-se associação entre doenças cardiovasculares com problemas de pele. Por fim, verificou-se que houve associação entre dor muscular, problemas neurológicos e problemas gastrointestinais (Quadro 12).

Das 3 milhões de regras geradas, a maioria apresentou a resposta “não”, o que reforça que o mais importante é a pergunta e não o número de perguntas, que gerará um resultado da somatória. A partir disto, o experimento foi repetido considerando apenas as instâncias pertencentes à classe “Sim”, sendo algumas das regras geradas apresentadas no Quadro 13. O Quadro 12 e o Quadro 13 são caracterizados por regras com alto valor para a confiança (acima de 0,88 ou 88%), mas com baixíssimos valores para suporte, o que representa o fato de haver poucas ocorrências na base de dados que atendam tais regras.

Quadro 12. Resultados obtidos com a utilização do algoritmo Apriori: Saúde (doenças).

```
doenca_respiratoria=nao      alteracoes_endocrinas=nao      doenca_renal_urinaria=nao
problemas_neurologicos=nao   consumo_de_alcool=1_2_vezes_por_semana  25  ==>
iqr20rotulo=nao 23  conf:(0.92)

fadiga=nao doencas_cardiovasculares=sim alteracoes_endocrinas=nao doenca_renal_urinaria=nao
problemas_gastrointestinais=nao 23 ==> iqr20rotulo=nao 21  conf:(0.91)

contato_com_agrotoxicos=sim      problemas_ortopedicos=nao      fadiga=nao
problemas_neurologicos=nao 30 ==> iqr20rotulo=nao 27  conf:(0.9)

contato_com_agrotoxicos=sim dor_muscular=nao fadiga=nao doencas_cardiovasculares=nao
doenca_renal_urinaria=nao problemas_neurologicos=nao problemas_gastrointestinais=nao 26 ==>
iqr20rotulo=nao 23  conf:(0.88)

problemas_gastrointestinais=nao tabagista=nao consumo_de_alcool=1_2_vezes_por_semana 25
==> iqr20rotulo=nao 22  conf:(0.88)

problemas_ortopedicos=sim      doenca_renal_urinaria=nao      problemas_neurologicos=nao
problemas_gastrointestinais=nao 24 ==> iqr20rotulo=nao 21  conf:(0.88)
```

Fonte: A autora, 2023.

Nos resultados obtidos com a análise de Regras de Associação, utilizando algoritmo Apriori, verificou-se que houve relação entre contato com agrotóxicos, problemas ortopédicos, dor muscular, fadiga, doença respiratória, doenças cardiovasculares, doença hepática, doença renal urinaria, problemas neurológicos e problemas gastrointestinais.

Quadro 13. Resultados obtidos com a utilização do algoritmo Apriori considerando apenas instâncias pertencentes à classe “sim”: Saúde e Doenças.

contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	doenca_hepatica=sim	
doenca_renal_urinaria=sim	problemas_neurologicos=sim	problemas_gastrointestinais=sim	1 ==>
iqr20rotulo=sim 1	conf:(1)		
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	alteracoes_endocrinas=nao	
doenca_hepatica=sim	doenca_renal_urinaria=sim	problemas_neurologicos=sim	
problemas_gastrointestinais=sim	1 ==>	iqr20rotulo=sim 1	conf:(1)
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	doenca_hepatica=sim	
doenca_renal_urinaria=sim	problemas_neurologicos=sim	problemas_gastrointestinais=sim	
problemas_pele=nao	1 ==>	iqr20rotulo=sim 1	conf:(1)
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	doenca_hepatica=sim	
doenca_renal_urinaria=sim	problemas_neurologicos=sim	problemas_gastrointestinais=sim	
tabagista=nao	1 ==>	iqr20rotulo=sim 1	conf:(1)
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	doenca_hepatica=sim	
doenca_renal_urinaria=sim	problemas_neurologicos=sim	problemas_gastrointestinais=sim	
consumo_de_alcool=raramente	1 ==>	iqr20rotulo=sim 1	conf:(1)
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	alteracoes_endocrinas=nao	
doenca_hepatica=sim	doenca_renal_urinaria=sim	problemas_neurologicos=sim	
problemas_gastrointestinais=sim	problemas_pele=nao	1 ==>	iqr20rotulo=sim 1 conf:(1)
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	alteracoes_endocrinas=nao	
doenca_hepatica=sim	doenca_renal_urinaria=sim	problemas_neurologicos=sim	
problemas_gastrointestinais=sim	tabagista=nao	1 ==>	iqr20rotulo=sim 1 conf:(1)
contato_com_agrotoxicos=sim	problemas_ortopedicos=sim	dor_muscular=sim	fadiga=sim
doenca_respiratoria=sim	doencas_cardiovasculares=sim	doenca_hepatica=sim	
doenca_renal_urinaria=sim	problemas_neurologicos=sim	problemas_gastrointestinais=sim	
problemas_pele=nao	tabagista=nao	1 ==>	iqr20rotulo=sim 1 conf:(1)

Fonte: A autora, 2023.

4.2.4 Seleção de atributos com diferentes algoritmos: saúde (doenças)

A Tabela 6 apresenta os resultados obtidos por diferentes algoritmos implementados na ferramenta WEKA e cujo objetivo compreende a seleção ou a ordenação de atributos tendo em vista diferentes critérios quanto a sua influência na distinção das classes presentes na base de dados.

Tabela 6. Atributos selecionados: Saúde (doenças).

Algoritmo	(Ordem dos) Atributos Selecionados
<i>GainRatioAttributeEval</i>	54, 53, 47, 46, 46, 50, 49, 57, 51, 52, 56, 44, 48, 55
<i>ClassifierAttributeEval</i>	57, 47, 48, 46, 56, 45, 49, 50, 51, 52, 55, 54, 53, 44
<i>InfoGainAttributeEval</i>	53, 54, 47, 46, 45, 57, 49, 50, 51, 56, 52, 44, 48, 55
<i>CfsSubsetEval</i>	47, 53, 54

Fonte: A autora, 2023.

Os atributos fadiga (47), problemas neurológicos (53) e problemas gastrointestinais (54), constam em todas as execuções com os diferentes algoritmos utilizados, e são os únicos que foram selecionados na análise com algoritmo *CfsSubsetEval*. O atributo 47 apareceu com a seguinte ordenação: 1º lugar (*CfsSubsetEval*), 2º lugar (*ClassifierAttributeEval*), 3º lugar (*InfoGainAttributeEval*). Considerando os algoritmos *GainRatioAttributeEval* e *CfsSubsetEval* o atributo 53 apareceu em 2º lugar, e em 1º lugar com o algoritmo *InfoGainAttributeEval*, já com ao algoritmo *ClassifierAttributeEval* apareceu em penúltimo lugar. Para o atributo 54 a ordenação obtida foi: 1º lugar (*GainRatioAttributeEval*), 2º lugar (*InfoGainAttributeEval*), 3º lugar (*CfsSubsetEval*) e 12º lugar.

4.2.5 Resultados obtidos e correlações: saúde (doenças)

Para os testes de Árvores de Decisão, verificou-se que foram geradas regras com os atributos referentes a problemas neurológicos e alterações endócrinas, problemas neurológicos e fadiga. Neste sentido, diversos estudos mostram o potencial tóxico de agrotóxicos sobre a saúde da população urbana e também rural.

Os efeitos da exposição aos agrotóxicos são diversos, dependendo das características do produto e tempo de exposição. A longo prazo poderão haver distúrbios neurológicos, endócrinos, mentais, intoxicação e câncer, seja por exposição direta ao agrotóxico ou consumo de alimentos contaminados com grandes quantidades desse produto (ARMAS et al. 2007; BUTINOF et al. 2015; DELLAVECHIA et al. 2016; SIQUEIRA; KRUSE, 2008). Algumas destas possibilidades foram identificadas com a execução do algoritmo J48 para Árvores de Decisão.

Em revisão de artigos sobre agrotóxicos, exposição humana e toxicidade, resultados apresentados por Mostafalou e Abdollahi (2017) corroboram com os apontamentos anteriores e apontam que os tumores cerebrais e cânceres (próstata, mama, colorretal, pancreático e pulmão) estão fortemente associados à exposição a agrotóxico. Além destes, também se tem as doenças dos sistemas musculoesquelético, circulatório e da pele, tais como Alzheimer, Parkinson, esclerose lateral amiotrófica, asma, bronquite, defeitos congênitos, entre outras (MOSTAFALOU; ABDOLLAHI, 2017).

Um outro estudo foi realizado por Simas (2020) para investigar o estado de saúde dos trabalhadores e compreender as relações existentes entre a saúde e o trabalho na bananicultura. Este foi realizado em uma região do Vale do Ribeira, tendo sido verificada a presença de transtornos mentais comuns em 25% dos indivíduos. Destaca-se que todos os participantes estiveram expostos aos agrotóxicos, sendo que quanto maior o tempo de exposição, maior foi a presença de distúrbios. Os trabalhadores apresentaram alterações do estado de saúde, sendo evidenciados distúrbios osteomusculares, distúrbios respiratórios, transtornos mentais comuns e fadiga muscular.

Conforme evidenciado no item 4.1.5 os organofosforados afetam o sistema nervoso dos organismos, inibindo a atividade da acetilcolinesterase (AChE). Estudos apontam a correlação entre o herbicida e o aparecimento ou o agravamento de algumas patologias como as doenças celíaca e a de Parkinson, a osteoporose, o hipotireoidismo, o glaucoma entre outras. Isso acontece, pois o glifosato é derivado do aminoácido glicina, que facilmente se insere numa variedade de vias metabólicas (NASCIMENTO et al. 2019; MEFTAUL et al. 2020).

Em estudo realizado por Oliveira (2017) com trabalhadores rurais expostos e não expostos a agrotóxicos, verificou-se que para os sintomas de problemas gastrointestinais, ficou evidente a inter-relação com os agrotóxicos para os

trabalhadores expostos de maneira ocupacional com 3,92 vezes maior para o desenvolvimento de problemas gastrointestinais, sendo a dor no estômago e a epigastralgia os sintomas mais comuns, relacionados com o uso dos agroquímicos, com prevalência no grupo exposto.

Com base nos resultados obtidos com a utilização do algoritmo Apriori, foi verificado que em vários casos, problemas ortopédicos estavam associados a dor muscular e fadiga. Já para alguns casos que a resposta para o atributo contato com agrotóxicos foi “sim”, notou-se a associação com problemas ortopédicos e dor muscular. Houve ainda relações entre respostas “sim” para problemas ortopédicos e fadiga e doenças cardiovasculares, dentre as quais muitas ainda estavam atrelados também a ocorrência de problemas neurológicos. Além disso, houve associação entre doenças cardiovasculares com problemas de pele. E verificou-se que houve associação entre dor muscular, problemas neurológicos e problemas gastrointestinais. Ainda, notou-se que houve relação entre contato com agrotóxicos, doença respiratória, doenças cardiovasculares, doença hepática, doença renal urinária. Diferentes estudos relacionados aos efeitos da intoxicação (aguda ou crônica) corroboram com as relações identificadas no presente trabalho, dentre eles destacam-se ARMSTRONG; GUO, 2019; BACON; BIAZON, 2014; BRONDANI, et al. 2020; CAVALCANTI et al. 2016; CHIELLE et al. 2015; DOULL, 2000; EDDLESTON et al. 2008; HILLIPS, 2004; JACOBSON et al. 2021; KOUREAS, 2012; LIONETTO et al., 2013; LOPES; ALBUQUERQUE, 2021; MESA, 2020; MESNAGE et al. 2017; MILLS et al. 2020; MURAKAMI et al. 2017; NAGARAJU et al. 2020; NEVES et al. 2020; OLIVEIRA, 201; PANDEY et al. 2019; PORTIER, 2020; PORTO; SOARES, 2012; PRADO et al. 2021; RAMAI et al. 2021; ROMÃO, 2004; SILVA; WESSELING, 2017; SILVÉRIO et al. 2020; SOUDANI et al. 2019; STORM; ROZMAN; VASCONCELLOS et al. 2019; VERVAET et al. 2020; WAHLANG et al. 2019.

4.3 Saúde: SRQ-20 e Doenças

Foram utilizadas as questões do SRQ-20 e de saúde descritas no Quadro 14, as perguntas compreendem um recorte do Apêndice B. Os resultados obtidos pelos testes realizados por meio das diferentes técnicas são apresentados nas subseções seguintes.

Quadro 14. Recorte de questões do Apêndice B relacionadas à SRQ-20 e Saúde.

Identificador	Questões do Apêndice B relacionadas à Saúde
24	Você tem dores de cabeça frequentes?
25	Você tem falta de apetite?
26	Você dorme mal?
27	Você se assusta com facilidade?
28	Você tem tremores musculares?
29	Você se sente nervoso, tenso ou preocupado?
30	Você tem má digestão?
31	Você sente que suas ideias ficam embaralhadas de vez em quando?
32	Você tem se sentido triste ultimamente?
33	Você tem chorado mais do que de costume?
34	Você encontra dificuldade para realizar com satisfação suas atividades diárias?
35	Você tem dificuldade para tomar decisões?
36	Você acha que seu trabalho é penoso e lhe causa sofrimento?
37	Você se acha incapaz de desempenhar um papel útil na sua vida?
38	Você tem perdido o interesse pelas coisas?
39	Você se sente uma pessoa sem valor?
40	Alguma vez você pensou em acabar com sua vida?
41	Você se sente cansado o tempo todo?
42	Você sente alguma coisa desagradável no estômago?
43	Você se cansa com facilidade?
44	Contato com agrotóxicos
45	Problemas ortopédicos
46	Dor muscular
47	Fadiga
48	Doença respiratória
49	Doenças cardiovasculares
50	Alterações endócrinas
51	Doença hepática
52	Doença renal urinária
53	Problemas neurológicos
54	Problemas gastrointestinais
55	Problemas de pele
56	Tabagista
57	Consumo de álcool

Fonte: A autora, 2023.

4.3.1 Clusterização com algoritmo EM: SRQ-20 e doenças

Os resultados das médias e desvio padrão obtidos a partir do algoritmo de clusterização EM estão expostos na Tabela 7. Nesta execução o próprio EM determinou a quantidade de agrupamentos ideal de 5 conjuntos seguindo seus critérios internos. Os resultados obtidos para a execução do EM com predefinição da criação de apenas 2 clusteres foi aqui omitida por não terem sido relevantes.

Tabela 7. Resultados obtidos com clusteres determinados pelo próprio EM: SRQ-20 e Doenças.

<u>Clusteres</u>	<u>Execuções</u>		
	A	B	C
0	8	22	5
1	21	1	17
2	5	23	38
3	28	1	-
4	5	7	-
5	1	6	-

Fonte: A autora, 2023.

Durante as análises, percebeu-se que “não” é possível criar agrupamentos que separem as instâncias com classificação “sim” daquelas com classificação “não” em ambos os conjuntos de treinamento e teste. Foram identificados agrupamentos que representaram subconjuntos das instâncias com classificação “sim”; agrupamentos que representaram subconjuntos das instâncias com classificação “não”; além de agrupamentos com instâncias pertencentes a ambas as classificações. Houveram agrupamentos gerados em consequência das situações em que “NA” foi dada como respostas, inclusive, em uma das três execuções realizadas, foram criados apenas três agrupamentos, conforme identificado na Tabela 7.

4.3.2 Regras de associação com algoritmo apriori: SRQ-20 e doenças

O emprego do algoritmo Apriori da ferramenta WEKA gerou uma série de regras de associação, das quais algumas serão listadas no Quadro 15. Pode-se observar que não houve associação entre os atributos testados em relação à classe “sim”, sendo observado apenas valores “sim” isolados como resposta de uma ou outra

questão. Por exemplo, para o atributo 29 (você se sente nervoso, tenso ou preocupado?) na última regra do Quadro 15.

Quadro 15. Resultados obtidos com a utilização do algoritmo Apriori: SRQ-20 e Doenças.

```

25=nao 28=nao 30=nao 37=nao 38=nao 42=nao doencas_cardiovasculares=nao
doenca_hepatica=nao 53 ==> iqr20rotulo=nao 52 conf:(0.98)
25=nao 28=nao 31=nao 33=nao 38=nao 40=nao fadiga=nao alteracoes_endocrinas=nao 53 ==>
iqr20rotulo=nao 52 conf:(0.98)
26=nao 28=nao 30=nao 33=nao 38=nao 39=nao fadiga=nao doenca_hepatica=nao
doenca_renal_urinaria=nao 50 ==> iqr20rotulo=nao 49 conf:(0.98)
28=nao 33=nao 37=nao 38=nao 39=nao 41=nao fadiga=nao alteracoes_endocrinas=nao
doenca_hepatica=nao doenca_renal_urinaria=nao problemas_neurologicos=nao
problemas_gastrointestinais=nao 49 ==> iqr20rotulo=nao 48 conf:(0.98)
25=nao 27=nao 33=nao 34=nao 35=nao 36=nao 37=nao 38=nao 39=nao 41=nao fadiga=nao
doenca_hepatica=nao doenca_renal_urinaria=nao 49 ==> iqr20rotulo=nao 48 conf:(0.98)
25=nao 28=nao 29=sim 33=nao 39=nao 41=nao 48 ==> iqr20rotulo=nao 47 conf:(0.98)

```

Fonte: A autora, 2023.

Das 3 milhões de regras geradas, a grande maioria referenciava a classe “não”. Diante disso, o experimento foi repetido apenas com as instâncias pertencentes à classe “Sim”, sendo algumas das regras geradas apresentadas no Quadro 16. Novamente, as regras apresentam alto valor para a confiança (acima de 0,98 ou 98%), mas com baixos valores para suporte (principalmente para o segundo caso), o que representa o fato de haver poucas ocorrências na base de dados que atendam tais regras.

Quadro 16. Resultados obtidos com a utilização do algoritmo Apriori considerando apenas instâncias pertencentes à classe “sim”: SRQ-20 e Doenças.

```

24=sim 27=sim 29=sim 41=sim 7 ==> iqr20rotulo=sim 7  conf:(1)
24=sim 26=sim 29=sim 41=sim 9 ==> iqr20rotulo=sim 9  conf:(1)
24=sim 29=sim 41=sim problemas_neurologicos=sim 8 ==> iqr20rotulo=sim 8  conf:(1)
24=sim 29=sim 41=sim problemas_ortopedicos=sim 7 ==> iqr20rotulo=sim 7  conf:(1)
24=sim 29=sim 41=sim dor_muscular=sim 7 ==> iqr20rotulo=sim 7  conf:(1)
24=sim 29=sim 41=sim problemas_gastrointestinais=sim 7 ==> iqr20rotulo=sim 7  conf:(1)
24=sim 26=sim 29=sim 31=sim 41=sim 8 ==> iqr20rotulo=sim 8  conf:(1)
26=sim 29=sim 31=sim 32=sim 38=sim 40=sim 41=sim 7 ==> iqr20rotulo=sim 7  conf:(1)
26=sim 29=sim 31=sim 32=sim 33=sim 40=sim 41=sim 7 ==> iqr20rotulo=sim 7  conf:(1)

```

Fonte: A autora, 2023.

Foram verificadas dentre as relações existentes que os sintomas dores de cabeça estão associados com cansaço, problemas ortopédicos, neurológicos, gastrointestinais dores musculares. Ainda, quando a resposta foi positiva para a pergunta “você dorme mal”, houve associação com problemas psicológicos (nervosismo, tensão, preocupação, ideias embaralhadas, tristeza, choro, cansaço e pensamentos de suicídio).

4.3.3 Seleção de atributos com diferentes algoritmos: SRQ-20 e doenças

A Tabela 8 apresenta os resultados obtidos por diferentes algoritmos implementados na ferramenta WEKA e cujo objetivo compreende a seleção ou a ordenação de atributos tendo em vista diferentes critérios quanto a sua influência na distinção das classes presentes na base de dados.

Tabela 8. Atributos selecionados: SRQ-20 e Doenças.

Algoritmo	(Ordem dos) Atributos Selecionados
<i>GainRatioAttributeEval</i>	41, 32, 25, 33, 37, 43, 38, 31, 35, 28, 34, 36, 29, 26, 40, 39, 30, 27, 54, 53, 47, 42, 24, 46, 45, 50, 49, 57, 51, 52, 56, 44, 48, 55
<i>ClassifierAttributeEval</i>	57, 32, 34, 35, 36, 37, 38, 33, 31, 56, 30, 25, 26, 27, 28, 29, 39, 40, 41, 42, 51, 52, 53, 54, 55, 50, 49, 48, 44, 43, 45, 47, 46, 24
<i>InfoGainAttributeEval</i>	32, 41, 33, 31, 43, 35, 37, 29, 48, 34, 26, 38, 36, 40, 28, 27, 30, 53, 54, 47, 39, 46, 42, 24, 45, 57, 49, 50, 51, 56, 52, 44, 48, 55
<i>CfsSubsetEval</i>	25, 31, 32, 33, 41, 43

Fonte: A autora, 2023.

Os atributos 41 (Você se sente cansado o tempo todo?), 32 (Você tem se sentido triste ultimamente?) e 33 (Você tem chorado mais do que de costume?), constam em todas as execuções com os diferentes algoritmos utilizados. O atributo 41 apareceu com a seguinte ordenação: 1º lugar com *GainRatioAttributeEval*, 2º lugar com *InfoGainAttributeEval*, 5º lugar com *CfsSubsetEval* e 19º lugar *ClassifierAttributeEval*.

O atributo 32 foi ordenado em 1º lugar com algoritmo *InfoGainAttributeEval*, 2º lugar com os algoritmos *GainRatioAttributeEval* e *ClassifierAttributeEval* e 3º lugar com o algoritmo *CfsSubsetEval*.

O atributo 25 (Você tem falta de apetite?) foi listado em 1º lugar com o algoritmo *CfsSubsetEval*, 3º lugar com *GainRatioAttributeEval*, em 12º lugar com *ClassifierAttributeEval* e não foi selecionado com o *InfoGainAttributeEval*.

O algoritmo *CfsSubsetEval* fez a seleção e ordenação apenas de atributos referentes ao *SQRQ-20*. Pode-se inferir que tais atributos são mais relevantes em comparação com os demais, podendo garantir mais assertividade no diagnóstico.

4.3.4 Árvores de decisão com algoritmo J48: SRQ-20 e doenças

Para os testes de Árvores de Decisão, com a aplicação do Algoritmo J48, foram geradas as regras apresentadas no Quadro 17, as quais produziram os resultados expostos na Tabela 9. Observou-se que os atributos 32, 25 e 41 apareceram no topo da árvore (nó inicial) na primeira, segunda e terceira execução,

respectivamente. Ressalta-se que na primeira execução o atributo 41 apareceu logo na sequencia após o topo da árvore.

Quadro 17. Regras de Árvore de Decisão da execução dos testes obtidos com utilização do algoritmo J48: SRQ-20 e Doenças.

Regras de Árvore de Decisão dos Grupos A	Regras de Árvore de Decisão dos Grupos B	Regras de Árvore de Decisão dos Grupos C
32 = NA: nao (3.0)	25 = NA: nao (3.0)	41 = NA: nao (2.0)
32 = nao	25 = nao	41 = nao
41 = NA: nao (0.0)	37 = NA: nao (0.0)	28 = NA: nao (0.0)
41 = nao: nao (67.0/2.0)	37 = PrefereNaoResponder: nao (0.0)	28 = nao
41 = sim	37 = nao	25 = NA: nao (0.0)
31 = NA: sim (0.0)	36 = NA: nao (1.0)	25 = nao
31 = nao: nao (3.0)	36 = PrefereNaoResponder: sim (1.0)	33 = NA: nao (0.0)
31 = sim: sim (8.0)	36 = nao	33 = nao
32 = sim	28 = NA: nao (0.0)	39 = NA: nao (0.0)
31 = NA: sim (0.0)	28 = nao	39 = nao: nao (64.0/1.0)
31 = nao	41 = NA: nao (0.0)	39 = sim
42 = NA: nao (0.0)	41 = nao: nao (66.0/1.0)	29 = NA: nao (0.0)
42 = nao	41 = sim	29 = nao: nao (3.0)
28 = NA: nao (0.0)	30 = NA: nao (0.0)	29 = sim: sim (2.0)
28 = nao	30 = nao: nao (2.0)	29 = sim: sim (2.0)
30 = NA: nao (0.0)	30 = sim: sim (2.0)	33 = sim
contato_com_agrotoxicos = NA: sim (2.0)	28 = sim	27 = NA: nao (0.0)
32 = NA: sim (0.0)	32 = NA: sim (0.0)	27 = nao: nao (8.0/1.0)
contato_com_agrotoxicos = nao: nao (1.0)	32 = nao: nao (5.0/1.0)	27 = sim: sim (6.0)
32 = sim: sim (6.0)	36 = sim	25 = sim: sim (4.0/1.0)
contato_com_agrotoxicos = sim: nao (7.0)	26 = NA: sim (0.0)	28 = sim
28 = sim: sim (2.0)	26 = nao	29 = NA: sim (0.0)
42 = sim: sim (3.0)	43 = NA: nao (0.0)	29 = nao: nao (2.0)
31 = sim: sim (24.0/1.0)	43 = nao: nao (4.0)	29 = sim: sim (6.0)
	43 = sim: sim (2.0)	41 = sim: sim (22.0/2.0)
	26 = sim: sim (5.0)	
	37 = sim: sim (10.0/1.0)	
	25 = sim: sim (12.0)	

Fonte: A autora, 2023.

A média e desvio padrão definidos para a avaliação dos resultados foram calculados com base nos valores obtidos nas 3 execuções e podem ser visualizados na Tabela 9. A acurácia atingiu 82,1% ($\sigma=2,7$) e as demais métricas obtiveram valores próximos, o que indica que a taxas gerais de erros são semelhantes para ambas as classes (“sim” e “não”).

Tabela 9. Resultados médios das execuções realizadas com o algoritmo J48: SRQ-20 e Doenças.

<u>Acurácia</u>		<u>Precisão</u>		<u>Recall</u>		<u>F-Measure</u>	
%	σ	%	σ	%	σ	%	σ
82,1	2,7	73,0	1,4	73,1	10,5	72,4	6,2

Fonte: A autora, 2023.

Observou-se que quanto aos erros de classificação, como esperado, foram identificadas instâncias da classe “não” preditas como “sim” e vice-versa. A título de exemplo, Execução A do Quadro 17, uma regra gerada a partir dos atributos 32 (não) e 41 (não) apresentou 67 acertos e apenas 2 erros. Nesta mesma execução houveram 24 acertos e 1 erro a partir de uma regra gerada com o uso dos atributos 32 (sim) e 31 (sim). Em ambos os casos, os erros se referem a instâncias pertencentes à classe “sim” foram classificadas como “não”. Houveram também casos em que instâncias pertencentes à classe “não” foram classificadas como “sim”. Uma regra gerada a partir dos atributos 25 (não), 37 (não), 36 (não), 28 (não) e 41 (não) produziu 66 acertos e 1 erro para o atributo 41 (Execução B do Quadro 17). Já para uma regra gerada para a classe “sim”, para o mesmo atributo (41 = sim) houveram 22 acertos e 2 erros (Execução C do Quadro 17). Ainda na Execução C (Quadro 17), uma regra considerando os atributos 41 (não), 28 (não), 25 (não), 33 (não) e 39 (não) gerou 64 acertos e 1 erro.

4.3.5 Resultados obtidos e correlações: SRQ-20 e doenças

Os resultados obtidos nesta seção e a análise dos atributos utilizados corroboram os resultados produzidos por estudos anteriores. A exposição múltipla a agrotóxicos e efeitos à saúde em agricultores foi investigada por Araújo et al. (2007) e verificou-se que os resultados das atividades da enzima Acetilcolinesterase (AChE)

são sugestivas de episódios sistemáticos de quadros leves a moderados de intoxicação, com elevada prevalência de sinais e sintomas relacionados à exposição aos agrotóxicos, especialmente cefaleia, visão turva, vertigem, fadiga, fraqueza, câibras, parestesias e distúrbios cognitivos (dificuldade de concentração, esquecimento e confusão mental). A inibição da enzima gera o acúmulo do neurotransmissor acetilcolina (ACh) nas junções sinápticas do sistema nervoso periférico e central ocasionando a síndrome colinérgica, a qual envolve uma superestimulação das estruturas inervadas pelas fibras colinérgicas (coração, glândulas, músculos lisos), causando fraqueza muscular e cansaço (EDDLESTON; HILLIPS, 2004; CAVALCANTI et al. 2016).

Foram verificadas dentre as relações existentes que os sintomas dores de cabeça estão associados com cansaço, problemas ortopédicos, neurológicos, gastrointestinais e dores musculares. Ainda, quando a resposta foi positiva para a pergunta “você dorme mal”, houve associação com problemas psicológicos (nervosismo, tensão, preocupação, ideias embaralhadas, tristeza, choro, cansaço e pensamentos de suicídio).

Conforme discutido nos itens 4.15 e 4.2.5 a exposição a agrotóxicos afetam o sistema nervoso, gerando distúrbios neurológicos, psicológicos, endócrinos, mentais, problemas gastrointestinais e doenças dos sistemas musculoesquelético, circulatório, da pele e fadiga muscular, conforme evidenciado em trabalhos disponíveis na literatura (ARMAS et al. 2007; BUTINOF et al. 2015; CAVALCANTI et al. 2016; DELLAVECHIA et al. 2016; EDDLESTON; HILLIPS, 2004; KOUREAS, et al. 2012; MOSTAFALOU; ABDOLLAHI, 2017; OLIVEIRA, 2017; SIQUEIRA; KRUSE, 2008; STORM; ROZMAN; DOULL, 2000).

Gonzaga, Baldo, Caldeira (2021) identificaram a ocorrência de ideação suicida entre camponeses vivendo sob exposição ambiental a agrotóxicos. Lopes e Albuquerque (2021) constataram que agricultores expostos constantemente a agentes químicos podem apresentar transtornos mentais, alteração nas atividades do telômeros, danos ao DNA e nos mecanismos de defesa das células, náuseas, dor de cabeça, anomalias no tecido hematopoiético e fígado, alteração nos níveis hormonais, depressão, mialgia, ansiedade e irritabilidade e maior chance de cometer suicídio. Outro estudo realizado na Serra Gaúcha por Faria (2005) obteve uma clara associação entre intoxicações por agrotóxicos e a ocorrência de problemas de saúde

de evolução prolongada, destacando-se os transtornos psiquiátricos menores e os sintomas respiratórios.

Novamente, enfatiza-se que é válido questionar a forma como o SQR-20 considera potencial de problemas de saúde, pois a quantidade de respostas “sim” possivelmente não seja um indicativo de transtorno mental comum, tornando-se mais importante levar em consideração quais as perguntas tiveram “sim” como resposta, para obter mais assertividade no diagnóstico.

4.4 Solo

Depois das plantas, o solo é o principal receptor das pulverizações de agrotóxicos. Neste ponto, a interação pesticida-solo depende principalmente das características físico-químicas de ambos e pode influenciar na população da macro e microfauna, como é o caso do herbicida glifosato (MARTINS, 2006; MORAES; ROSSI, 2010; ZILLI et al., 2008; BASTOS et al., 2007). Neste ponto, foram avaliadas as diferentes técnicas descritas na Seção 2.3, seguindo as definições da Seção 3.4 e os parâmetros definidos na Seção 3.6. O Quadro 18 apresenta um recorte do Apêndice B, sendo que os testes foram realizados para todo o conjunto e também para cada um dos contextos (granulometria, fertilidade, componentes físicos e pesticidas) em separado.

Quadro 18. Classificação de itens do Apêndice A relacionadas à Solo.

Categoria	Identificador	Questões do Apêndice A relacionadas à Solo
Granulometria	1	Areia
	2	Argila
	3	Silte
Componentes de Fertilidade	4	Alumínio
	5	Bário
	6	Calcio
	7	Cabono
	8	Chumbo
	9	Cobre
	10	Cromo
	11	Ferro
	12	Fósforo
	13	Íon Alumínio
	14	Magnésio
	15	Manganês
	16	Niquel
	17	Nitrogênio
	18	Potássio
	19	Sódio
	20	Zinco
Componentes Físico	21	Matéria orgânica
	22	pH
	23	SB
	24	Tefetiva
Pesticidas	25	Ampa
	26	Glifosato
	27	Organoclorado

Fonte: A autora, 2023.

Exceto pelos algoritmos implementados na ferramenta WEKA com o objetivo de seleção ou ordenação dos atributos, a aplicação das demais técnicas não alcançou resultados relevantes e não serão aqui apresentados. A Tabela 10 e a Tabela 11 apresentam os resultados obtidos com o emprego de todos os atributos apresentados no Quadro 18, respectivamente com o emprego dos valores numéricos originais dos atributos e com o emprego dos rótulos definidos por meio de intervalos iguais.

Tabela 10. Atributos selecionados com Diferentes Algoritmos para dados gerais de solo considerando os valores numéricos originais dos atributos.

Algoritmo	(Ordem dos) Atributos Selecionados
<i>GainRatioAttributeEval</i>	24, 4, 6, 7, 8, 5, 3, 10, 2, 26, 27, 1, 9, 11, 23, 18, 20, 21, 22, 19, 17, 12, 16, 13, 14, 15, 25
<i>ClassifierAttributeEval</i>	24, 4, 6, 7, 8, 5, 3, 10, 2, 26, 27, 1, 9, 11, 23, 18, 20, 21, 22, 19, 17, 12, 16, 13, 14, 15, 25
<i>InfoGainAttributeEval</i>	24, 4, 6, 7, 8, 5, 3, 10, 2, 26, 27, 1, 9, 11, 23, 18, 20, 21, 22, 19, 17, 12, 16, 13, 14, 15, 25
<i>CfsSubsetEval</i> – <i>GreedyStepwise</i>	25, 26, 27, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24
<i>CfsSubsetEval - BestFirst</i>	25

Fonte: A autora, 2023.

Tabela 11. Atributos selecionados com Diferentes Algoritmos para dados gerais de solo considerando os valores simbólicos definidos por meio de intervalos iguais dos atributos.

Algoritmo	(Ordem dos) Atributos Selecionados
<i>ClassifierAttributeEval</i>	24, 4, 6, 7, 8, 5, 3, 10, 2, 26, 27, 1, 9, 11, 23, 18, 20, 21, 22, 19, 17, 12, 16, 13, 14, 15, 25
<i>InfoGainAttributeEval</i>	1, 4, 3, 24, 19, 20, 6, 23, 10, 14, 13, 11, 16, 2, 9, 18, 26, 27, 15, 17, 22, 12, 21, 7, 5, 8, 25
<i>GainRatioAttributeEval</i>	19, 1, 4, 26, 20, 27, 3, 24, 11, 12, 10, 18, 15, 9, 23, 6, 13, 14, 2, 16, 17, 25, 7, 21, 22, 8, 5
<i>CfsSubsetEval</i> – <i>GreedyStepwise</i>	19, 1, 4, 26, 3, 12, 27, 20, 24, 2, 18, 10, 13, 11, 25, 16, 6, 8, 9, 23, 17, 15, 5, 14, 7, 22, 21
<i>CfsSubsetEval - BestFirst</i>	1, 4, 19

Fonte: A autora, 2023.

A partir da análise da Tabela 10, verificou-se que foram selecionados os mesmos atributos e ordenados da mesma forma pelos algoritmos *GainRatioAttributeEval*, *ClassifierAttributeEval* e *InfoGainAttributeEval*. Diante disso, é possível destacar a relevância desses atributos para o estudo.

Observou-se que a ordenação do atributo 25 (AMPA) foi em último para os algoritmos *GainRatioAttributeEval*, *ClassifierAttributeEval* e *InfoGainAttributeEval* e em primeiro lugar com a utilização dos algoritmos *CfsSubsetEval* – *GreedyStepwise*

e *CfsSubsetEval - BestFirst*. Quanto ao algoritmo *CfsSubsetEval - BestFirst*, verificou-se que o mesmo selecionou apenas o atributo 25, o que tornou tal atributo como sendo o único comum para todos os algoritmos utilizados.

Ao se considerar a análise dos dados da Tabela 11, identificou-se que foram selecionados 3 atributos em comum 1 (areia), 4 (Alumínio) e 19 (sódio), alternando apenas a ordenação da seleção dos atributos para cada algoritmo. Comparando os resultados obtidos na análise de seleção de atributos com a árvore de decisão, observou-se que o atributo 4 consta em ambas análises, o que reforça sua relevância para o presente estudo.

A Tabela 12 e a Tabela 13 apresentam os resultados obtidos com o emprego dos atributos definidos pelo recorte identificado como “Pesticidas” no Quadro 18, respectivamente com o emprego dos valores numéricos originais dos atributos e com o emprego dos rótulos definidos por meio de intervalos iguais. Para ambos os casos, o número limitado de atributos acabou por limitar as possibilidades quanto aos atributos selecionados e também a ordem com que isso ocorreu. Destaca-se que basicamente houve a mesma ordenação (27, 26, 25) para os algoritmos *ClassifierAttributeEval*, *GainRatioAttributeEval* e *InfoGainAttributeEval*, além da inversão dos atributos selecionados nas 2 últimas execuções destas tabelas, houve uma inversão, com maior relevância e a seleção apenas do atributo 25.

Tabela 12. Atributos selecionados com Diferentes Algoritmos para dados de solo (pesticidas).

Algoritmo	(Ordem dos) Atributos Selecionados
<i>ClassifierAttributeEval</i>	27, 26, 25
<i>InfoGainAttributeEval</i>	27, 26, 25
<i>GainRatioAttributeEval</i>	27, 26, 25
<i>CfsSubsetEval - GreedyStepwise</i>	25
<i>CfsSubsetEval - BestFirst</i>	25

Fonte: A autora, 2023.

Tabela 13. Atributos selecionados com Diferentes Algoritmos para dados de solo.

Algoritmo	(Ordem dos) Atributos Selecionados
<i>ClassifierAttributeEval</i>	27, 26, 25
<i>InfoGainAttributeEval</i>	26, 27, 25
<i>GainRatioAttributeEval</i>	26, 27, 25
<i>CfsSubsetEval - GreedyStepwise</i>	26, 27
<i>CfsSubsetEval - BestFirst</i>	26, 27

Fonte: A autora, 2023.

4.4.1 Resultados obtidos e correlações: solo

O uso intensivo de agrotóxicos pode contribuir para processos de contaminação do solo e da água. O uso inadequado, ausência de treinamento, informação e fiscalização pode intensificar esses processos de contaminação ambiental. Na saúde humana, os efeitos são diversos, dependendo das características do produto e tempo de exposição (ARMAS et al. 2007; BUTINOF et al. 2015; DELLAVECHIA et al. 2016; SIQUEIRA; KRUSE, 2008).

Por vezes, pulverizações desnecessárias, ou com dosagens acima das recomendadas, são realizadas. Na análise de Seleção de Atributos com diferentes algoritmos para o teste simbólico de intervalos iguais referentes a Pesticidas para as amostras de coletas do solo, notou-se que foram selecionados os atributos: Glifosato e Organoclorados em todos os algoritmos utilizados. Além da utilização do herbicida glifosato de forma isolada, é comum a mistura de agrotóxicos manipulada pelo próprio agricultor (BASTOS et al. 2007; MARTINS, 2006; MORAES; ROSSI, 2010; ZILLI et al. 2008).

Segundo Moreira et al. (2002), os principais meios de contaminação humana por agrotóxicos são os ambientais, os ocupacionais e os alimentares. Neste sentido, estudos foram realizados envolvendo ambientes aquáticos (BELCHIOR et al. 2017), com efeitos aos sistemas nervosos dos organismos e inibição da atividade da acetilcolinesterase (WANG et al. 2015; ZHAO et al. 2015), redução significativa no tamanho dos organismos juvenis e gerando efeitos sobre a fertilidade (CUHRA et al. 2013).

Parte dos agrotóxicos que são utilizados nas plantações afetam todo o ecossistema e a cadeia alimentar, tendo em vista que podem ocorrer desvios dos

alvos por meio do vento, da chuva, entre outros fatores. Esta parcela acaba por contaminar o solo, o lençol freático e as águas fluviais, sem contar que os produtos que possuem mais persistência no ambiente são acumulados biologicamente na cadeia alimentar e nos seres humanos, gerando impactos na qualidade de vida (PIGNATI, 2007; CONSEA, 2014). No entanto, é comum que as investigações não identifiquem cenários de contaminação no ambiente agrícola, detectando somente a presença dos agrotóxicos e nitrato em baixas concentrações. Muitos fatores concorrem para a dissipação de agrotóxicos no ambiente, levando à mineralização completa da molécula ou gerando outros produtos como metabólitos e produtos de degradação (GOMES; BARIZON, 2014).

O conceito abrangente de qualidade de vida preconizado pela Organização Mundial de Saúde (OMS) incorpora de uma forma complexa, a saúde física, o estado psicológico, o nível de dependência, as relações sociais, as crenças pessoais e o relacionamento com características que se destacam no ambiente (WHO, 1990). Os efeitos deletérios à saúde humana são descritos por alterações nos sistemas nervoso, cardiovascular, respiratório, na pele, nos olhos, além de alterações hematológicas e reações alérgicas aos agrotóxicos (FIGUEIREDO, 2009; SIQUEIRA, L. S.; KRUSE, 2008).

Ressalta-se a importância de prevenção e controle do uso de agrotóxicos, já que o uso indiscriminado causa danos tanto para o meio ambiente como para trabalhadores e indivíduos envolvidos nesse contexto. Em função da gravidade deste tema, é importante identificar aspectos que envolvam a atuação dos profissionais de saúde e programas de conscientização para a melhoria da qualidade de vida, a fim de causar o mínimo impacto possível à saúde dos trabalhadores, dos animais, dos vegetais, da água e da terra. Outra alternativa interessante é o estímulo aos movimentos pelo desenvolvimento sustentável que busquem controlar e eliminar o “uso e abuso” de agrotóxico na agropecuária, substituindo os por outras técnicas de controle de pragas, assim como a implantação de Sistema de Monitoramento de Resíduos de Agrotóxicos em águas de córregos, rios, lagos e pântanos em áreas de monoculturas e implantação do Monitoramento ou Programa de Análise de Resíduos de Agrotóxicos em Alimentos.

5 CONSIDERAÇÕES FINAIS

Um conjunto diverso de estudos indicam a influência de substâncias químicas no desenvolvimento de problemas de saúde. No entanto, quando se trata de diagnóstico de intoxicações e problemas psiquiátricos mediante exposição a agrotóxicos, existem dificuldades de se obter uma conclusão e realizar afirmações sobre a relação entre eles. Devido à complexidade do assunto, a apresentação de sintomas “sutis” e a falta de sensibilidade aguçada no momento do diagnóstico por profissionais de saúde, em muitos casos não é possível realizar essa associação.

Os resultados das análises de solo presentes na base de dados utilizada indicam a presença de agrotóxicos nas amostras coletadas, isto é, Ampa, Glifosato e Organoclorado. Já as análises dos dados referentes à saúde e solo indicam a existência de relacionamentos quanto ao uso de agrotóxicos e os problemas de saúde relatados por munícipes de Santa Tereza do Oeste - PR. Dentre outros, identificou-se problemas de fadiga, neurológicos e gastrointestinais, os quais foram elencados nas regras de associação ou nas árvores de decisão. Tais atributos também estavam contidos na análise de seleção de atributos, para todos os algoritmos testados. Adicionalmente, outro atributo que apareceu com frequência foi o 41, o qual representa a questão “Você se sente cansado o tempo todo”. Foram identificadas, ainda, relações entre a exposição de agrotóxicos e sofrimento mental e distúrbios psiquiátricos na amostra entrevistada, revelando inclusive o pensamento de suicídio.

De forma geral, os diferentes algoritmos empregados apresentaram resultados que corroboram entre si. Ou seja, os mesmos atributos selecionados são utilizados na construção das árvores de decisão e na composição das regras de associação. Além do exposto, também foram identificados um conjunto diverso de pesquisas que confirmam os achados do presente trabalho.

5.1 Trabalhos Futuros

Ao analisar os resultados obtidos durante o desenvolvimento deste trabalho, no que concerne à saúde, identificou-se uma possível necessidade de se repensar a forma com que é utilizado o questionário do *Self-Reporting Questionnaire 20 Items* (SRQ-20). Atualmente o SQR-20 considera apenas a contagem das questões que

obtiveram respostas “sim”, mas ao analisar os atributos empregados pelos algoritmos de mineração de dados identificou-se maior importância para alguns deles em detrimento dos demais. Neste sentido, poderia ser mais interessante considerar quais as perguntas tiveram “sim” como resposta para que seja possível ter mais assertividade no diagnóstico, o que poderia envolver pesos diferentes e o cálculo de uma média ponderada com base nestes pesos.

Diante da relevância do tema dos agrotóxicos e do contingente da população que está exposta, direta ou indiretamente, identifica-se na necessidade de ampliação dos estudos nesta área. A identificação de indícios de sofrimento mental, distúrbios psiquiátricos e a tendência de pensamentos suicidas indicam a necessidade de novos estudos. Investigações individualizadas poderiam permitir que aqueles que realmente precisem tenham os tratamentos adequados. Ao mesmo tempo, investigações considerando toda a região permitiriam um melhor mapeamento da população.

REFERÊNCIAS

- ABOIM, Marcela C.R.; COUTINHO, Heitor L.C.; PEIXOTO, Raquel S.; et al. Soil bacterial community structure and soil quality in a slash-and-burn cultivation system in Southeastern Brazil. **Applied Soil Ecology**, [s.l.], v. 38, n. 2, p. 100-108, fev. 2008. Elsevier BV. <http://dx.doi.org/10.1016/j.apsoil.2007.09.004>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0929139307001151?via%3Di> hub. Acesso em: 01 mar. 2023.
- AGRA, N. G.; SANTOS, R. F. D. Agricultura brasileira: situação atual e perspectivas de desenvolvimento. *In*: CONGRESSO DA SOCIEDADE BRASILEIRA DE ECONOMIA E SOCIOLOGIA RURAL, 39, **Anais**, [s.l.]. Recife, 2001.
- AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules. *In* PROCEEDINGS OF THE 20TH INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, v. 94. 1994, San Francisco, USA. **Anais** [s.l.]: Morgan Kaufmann Publishers, 1994. p. 487- 499.
- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A.. **Database mining**: a performance perspective. *Ieee Transactions On Knowledge And Data Engineering*, [S.L.], v. 5, n. 6, p. 914-925, dez. 1993. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/69.250074>. Disponível em: <https://ieeexplore.ieee.org/document/250074>. Acesso em: 21 abr. 2023.
- AI-ANAZI, S.; ALMAHMOUD, H.; AI-TURAIKI, I. Finding similar documents using different clustering techniques. **Procedia Comput Science**. Saudi Arabia, versão 82: p. 28-34. 2016.
- AL-SHAMMARY, A. A. G.; KOUZANI, A. Z.; KAYNAK, A.; KHOO, S. Y. et al. Soil Bulk Density Estimation Methods: A Review. **Pedosphere**, China, v. 28, n. 4, p. 581-596, 2018.
- ALBUQUERQUE, M. A. A.; FERREIRA, R. L. C.; SILVA, J. A. A.; Santos, E. S.; Stosic, B.; Souza, A. L. Estabilidade em análise de agrupamento: estudo de caso em ciência florestal. **Revista Árvore**, Minas Gerais, v. 30, p. 257-265. 2006.
- ANDREU, V.; PICÓ, Y. Determination of pesticides and their degradation products in soil: critical review and comparison of methods. **TrAC Trends in Analytical Chemistry**, Canada, v. 23, n. 10, p. 772-789, 2004.
- AKTAR, M. W.; SENGUPTA, D.; CHOWDHURY, A. Impact of pesticides use in agriculture: their benefits and hazards. **Interdisciplinary toxicology**, India, v. 2, n. 1, p. 1-12, 2009.

ANVISA, Agência Nacional de Vigilância Sanitária. **Reclassificação Toxicológica dos Agrotóxicos**. ANVISA, [2019]. Disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2019/publicada-reclassificacao-toxicologica-de-agrotoxicos#:~:text=Categoria%201%20%E2%80%93%20Produto%20Extremamente%20T%C3%B3xico,T%C3%B3xico%20%E2%80%93%20faixa%20azul%3A%20599.> Acesso em: 27 jun. 2022.

BRASÍLIA. **Lei nº 9.782, de 26 de janeiro de 1999**. Lei de Criação da ANVISA. 1999. BRASÍLIA: ANVISA, [1999]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9782.htm#:~:text=LEI%20N%C2%BA%209.782%2C%20DE%2026%20DE%20JANEIRO%20DE%201999.&text=Define%20o%20Sistema%20Nacional%20de,Sanit%C3%A1ria%2C%20e%20d%C3%A1%20outras%20provid%C3%A1ncias. Acesso em: 05 abr. 2022.

ARAÚJO A. J.; LIMA, J. S.; MOREIRA, J. C. et al. Exposição múltipla a agrotóxicos e efeitos à saúde: estudo transversal em amostra de 102 trabalhadores rurais, nova friburgo, rj. **Ciência & Saúde Coletiva**, [S.L.], v. 12, n. 1, p. 115-130, mar. 2007. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1413-81232007000100015>. Disponível em: <https://www.scielo.br/j/csc/a/WVKwzDvn9PhNQYqW3Z7pMbD/abstract/?lang=pt>. Acesso em: 21 mar. 2023.

ARMAS, E. D.; MONTEIRO, R. T. R.; ANTUNES, P. M. et al. Diagnóstico espaço-temporal da ocorrência de herbicidas nas águas superficiais e sedimentos do Rio Corumbataí e principais afluentes. **Química Nova**, [s.l.], v. 30, n. 5, p. 1119-1127, out. 2007. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0100-40422007000500013>. Disponível em: <https://www.scielo.br/j/qn/a/RccQSbxDybF3R35QdSsLCRg/?lang=pt>. Acesso em: 10 fev. 2023.

ARRUDA, T. F. **Agrotóxicos e Sistema Penal**. Dissertação (Programa de Pós-Graduação em Ciências Agrárias) - Universidade Estadual de Maringá - UEM, Maringá. Paraná, 2008.

ARSHAD, M. A.; MARTIN, S. Identifying critical limits for soil quality indicators in agroecosystems. **Agriculture, Ecosystems & Environment**, Cachan Cedex, France, v. 88, n. 2, p. 153-160, 2002.

ARMSTRONG, L. E.; GUO, G. L. **Understanding environmental contaminants' direct effects on non-alcoholic fatty liver disease progression**. *Current Environmental Health Reports*, 6(3), 95–104. 2019.

AZEVEDO, M. S. A. **O Envelhecimento Ativo e a Qualidade de Vida: Uma Revisão Integrativa**. Dissertação de Mestrado (Curso de Mestrado em Enfermagem Comunitária) Escola Superior de Enfermagem do Porto - Porto, 2015. Disponível em: <https://comum.rcaap.pt/bitstream/10400.26/10776/1/marta%2020%20de%20abril%20-%20tese%20final%20-%20pdf.pdf>. Acesso em: 21/02/2023.

BACON, B. G.; BIAZON, A. C. B. Exposição a agrotóxicos: perfil dos trabalhadores rurais de uma comunidade rural de Campo Mourão-PR. **Rev Saúde Biol (SaBios)**. Campo Mourão - PR, 9 (2):13-9. 2014.

BALSAN, R. Impactos decorrentes da modernização da agricultura brasileira. **Campo-Território**: revista de geografia agrária, Uberlândia, v. 1, n. 2, p. 123-151, 2006.

BASTIANI, M. **Aplicação de Controle Estatístico de Processos e Algoritmos de Mineração de Dados na Gestão da Produção de Frangos de Corte**. Dissertação (Ciências Exatas e da Tecnologia da Computação) - Universidade Tecnológica Federal do Paraná – UTFPR. Medianeira, 2018, Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/3449>. Acesso em: 01/02/2023.

BASTOS, C. S.; DANIELLE, J.; MARIA, R. Seletividade de pesticidas à *Chrysoperla externa* (Hagen) (Neuroptera: Chrysopidae). Brasília, DF: **Ministério da Agricultura, Pecuária e Abastecimento**, 2007. (Comunicado técnico, 346).

BELCHIOR, D. C. V.; SARAIVA, A. S.; LÓPEZ, A. M. C.; SCHEIDT, G. N. Impactos de Agrotóxicos sobre o Meio Ambiente e a Saúde Humana. **Cadernos de Ciência & Tecnologia**, Brasília, v. 34, n. 1, p. 135-151, 2017

BELO, M. S. S. P.; PIGNATI, W.; DORES, E. G. de C.; et al. Uso de agrotóxicos na produção de soja do estado do Mato Grosso: um estudo preliminar de riscos ocupacionais e ambientais. **Revista Brasileira de Saúde Ocupacional**, [s.l.], v. 37, n. 125, p. 78-88, jun. 2012. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0303-76572012000100011>. Disponível em: <https://www.scielo.br/j/rbso/a/6WpPZxTdH4GdPPCh4TwndHc/?lang=pt>. Acesso em: 23 mar. 2023.

BEYLICH, A.; OBERHOLZER, H.-R.; SCHRADER, S.; HÖPER, H. et al.. Evaluation of soil compaction effects on soil biota and soil biological processes in soils. **Soil And Tillage Research**, [s.l.], v. 109, n. 2, p. 133-143, ago. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.still.2010.05.010>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167198710001017>. Acesso em: 16 fev. 2023.

BERNARDI, A. P. A. **Exposição ocupacional a ruído e solventes e alterações auditivas periféricas e centrais**. Tese. Universidade de São Paulo. São Paulo, 2007. Disponível em: https://www.teses.usp.br/teses/disponiveis/6/6132/tde-27042021-094706/publico/DR_891_Bernardi_2007.pdf. Acesso em: 04/02/2023.

BERRY, M. J. A.; LINOFF, G. S. **Mastering Data Mining**, New York: John Wiley & Sons, Inc. 2000.

BHARGAVA, N.; GIRJA, S.; RITU, D. B.; MANISHA, M. Decision tree analysis on J48 algorithm for data mining, **International Journal of Advanced Research in Computer Science and Software Engineering (JARCSSE)**. 2013.

BIELEFELD, E. C.; HU, B. H.; HARRIS, C.; HENDERSON, D. Damage and threshold shift resulting from cochlear exposure to Paraquat-generated superoxide. **J Occup Med Toxicol**, 2009; 207(1-2):35-42.

BLECKER, S. W.; STILLINGS, L. L.; AMACHER, M. C.; IPPOLITO, J. A. et al. Development of vegetation based soil quality indices for mineralized terrane in arid and semiarid regions. **Ecological Indicators**, 20, p. 65-74, 2012/09/01/ 2012.

BONMATIN, J. et al. (2014). Environmental fate and exposure; neonicotinoids and fipronil. **Environmental Science and Pollution Research**, [s.l.], 22 (1): 35-67.

BORGELT, C. **An Implementation of the Fp-growth algorithm**. Otto-Von-Guericke-University of Magdeburg. Germany, 2005.

BOUMA, J.; JONGERIUS, A.; BOERSMA, O.; JAGER, A. et al. The Function of Different Types of Macropores During Saturated Flow through Four Swelling Soil Horizons. **Soil Science Society of America Journal**, Amsterdam, 41, n. 5, p. 945-950, 1977.

BOUMA, J.; MCBRATNEY, A. Framing soils as an actor when dealing with wicked environmental problems. **Geoderma**, 200-201, p. 130-139, 2013/06/01, 2013.

BRASIL - Ministério da Saúde (MS). **Plano Integrado de Vigilância em Saúde de Populações Expostas a Agrotóxicos**. Brasília: Editora do Ministério da Saúde; 2009.

BRASIL - Ministério da Saúde (MS). **Protocolo de Atenção à Saúde dos Trabalhadores Expostos a Agrotóxicos**. Brasília: Editora do Ministério da Saúde; 2006.

BRASIL - Ministério da Saúde (MS). **Plano Integrado de Vigilância em Saúde de Populações Expostas a Agrotóxicos**. Brasília: Editora do Ministério da Saúde; 2009.

BRAZ, A. M.; OLIVEIRA, I. J.; CAVALCANTI, L. C. S.; ALMEIDA, A. C.; CHÁVEZ, E. S. Análise de Agrupamento (Cluster) para Tipologia de Paisagens. **Mercator**, Fortaleza, v.19, e19011, 2020.

BREVIK, E. C.; CERDÀ, A.; MATAIX-SOLERA, J.; PEREG, L. et al. **The interdisciplinary nature of Soil**. **SOIL**, 1, n. 1, p. 117-129, 2015.

BRONDANI, V. F.; SCHIMITH, M. D.; PUHL, G. S.; et al. Agrotóxicos e saúde de trabalhadores rurais: tendências da produção científica no brasil. **Research, Society And Development**, [s.l.], v. 9, n. 9, p. 1-23, 14 set. 2020. Research, Society and Development. <http://dx.doi.org/10.33448/rsd-v9i9.8258>. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/8258>. Acesso em: 17 jan. 2023.

BÜNEMANN, E. K.; BONGIORNO, G.; BAI, Z.; CREAMER, R. E. et al. Soil quality – A critical review. **Soil Biology and Biochemistry**, [s.l.] 120, p. 105-125, 2018/05/01, 2018.

BUCK, C.; LANGMAACK, M.; SCHRADER, S. Influence of mulch and soil compaction on earthworm cast properties. **Applied Soil Ecology**, Oklahoma 14, n. 3, p. 223-229, 2000.

BUTINOF, M.; FERNANDEZ, A. R.; STIMOLO, I. M.; et al. Pesticide exposure and health conditions of terrestrial pesticide applicators in Córdoba Province, Argentina. *Cad. Saúde Pública*, Rio de Janeiro, 31 (3), Mar, 2015.

CALDAS, E. D. Pesticide Poisoning in Brazil. College of Health Science, University of Brasilia In: **Reference Module in Earth Systems and Environmental Sciences**, London, Elsevier, 2016.

CAMILO, C. O.; SILVA, J. C. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Technical Report - RT-INF_001-09 - Relatório Técnico August - 2009 – Agosto. Instituto de Informática Universidade Federal de Goiás. 2009.

CANELLAS, L. P.; FAÇANHA, A. R. Chemical nature of soil humified fractions and their bioactivity. **Pesquisa Agropecuária Brasileira**, Brasília – DF, 39, p. 233-240, 2004.

CAPORAL, F. R.; COSTABEBER, J. A. **Agroecologia e Extensão Rural Contribuições Para a Promoção do Desenvolvimento Rural Sustentável**. Emater. Porto Alegre (RS) 2004.

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Rev Adm Pública**, Rio de Janeiro, 42(3):495-528, 2008.

CARNEIRO, F. F.; AUGUSTO, L. G. S.; RIGOTTO, R. M.; et al. (Orgs.). **Dossiê ABRASCO: um alerta sobre os impactos dos agrotóxicos na saúde**. Rio de Janeiro: EPSJV; São Paulo: Expressão Popular, 2015.

CARVALHO, M. M. X.; NODARI, E. S.; NODARI, R. O. “Defensives” or “pesticides”? A history of the use and perception of pesticides in the state of Santa Catarina, Brazil, 1950-2002. **História, Ciências, Saúde**, Manguinhos, v. 24, n. 1, p. 75-91, 2017.

CARVALHO, D. B. **Árvore De Decisão: Algoritmo genético para tratar o problema de pequenos disjuntos em classificação de dados**. Tese. Rio de Janeiro, RJ – Brasil. Dezembro, 2005. Disponível em: http://www.ipardes.gov.br/biblioteca/docs/tese_deborah_carvalho.pdf. Acesso em 04/01/2023.

CASIDA, J. E.; QUISTAD, G. B. Serine hydrolase targets of organophosphorus toxicants. **Chemico-Biological Interactions**, Germany, 157- 158, 277, 2005.

CARVALHAIS, L. C.; DENNIS, P. G.; TYSON, G. W.; SCHENK, P. M. Application of metatranscriptomics to soil environments. **Journal of Microbiological Methods**, [s.l.], 91, n. 2, p. 246-251, 2012.

- CAVALCANTI, L. P. A. N.; AGUIAR, A. P.; LIMA, J. A.; LIMA, A. L. Intoxicação por Organofosforados: Tratamento e Metodologias Analíticas Empregadas na Avaliação da Reativação e Inibição da Acetilcolinesterase. **Rev. Virtual Quim**, Niterói – RJ, 8 (3), 739-766, 2016
- CAVICCHIOLI, R.; RIPPLE, W. J.; TIMMIS, K. N.; AZAM, F. et al. Scientists' warning to humanity: microorganisms and climate change. **Nature Reviews Microbiology**, Island, 17, n. 9, p. 569-586, 2019.
- CERVANTES, J.; LAMONT, F. G.; LÓPEZ-CHAU, A.; et al. Data selection based on decision tree for svm classification on large data sets. **Applied Soft Computing**, Japan, 37: 787–798, 2015.
- CHATFIELD, C. V. COLLINS, A. **Introduction to Multivariate Analysis**. New York: Chapman e Hall, 1992.
- CHAUHAN, H.; CHAUHAN, A. Implementation of decision tree algorithm C4. 5, **International Journal of Scientific and Research Publications**. [s.l.] 2013.
- CHIELLE, E. O., BERGMANN, W. C. K., & PEREIRA, D. L. Avaliação da expressão sérica da enzima colinesterase e de marcadores bioquímicos hepáticos e renais em plantadores de tabaco. **Unoesc & Ciência - ACBS**, Joçaba, SC, 6(2), 153–162, 2015.
- CIOS, K. J; PEDRYCZ, W; SWINIARSKI, R. W; KURGAN, L. A. **Data Mining - A Knowledge Discovery Approach**. Springer, 2007.
- CONEN, F.; LENG, P.; AHMED, S. Data Structure for association rule mining: t-tress and p-trees. **IEEE Transactions on Knowledge and Data Engeneering**. Vol 16, nº 6. Junho, 2004.
- CONSEA - CONSELHO NACIONAL DE SEGURANÇA ALIMENTAR E NUTRICIONAL. **Os impactos dos agrotóxicos na segurança alimentar e nutricional: contribuições do Consea**. Brasília, 2014.
- COSTA, L. F. Bioinformatics: perspectives for the future. **Genet Mol Res**, Ribeirão Preto, SP, 3 (4):564-74, 2004.
- CREMONESE, C.; FREIRE, C.; MEYER, A. et al. Exposição a agrotóxicos e eventos adversos na gravidez no Sul do Brasil. 1996-2000. **Cad. Saúde Pública**. 2012; 28(7):1263-1272.
- CROSBY, E. B.; BAILEY, J. M.; OLIVERI, U. N.; LEVIN, E. D. Neurobehavioral impairments caused by developmental imidacloprid exposure in zebrafish. **Neurotoxicology And Teratology**, [s.l.], v. 49, p. 81-90, maio 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.ntt.2015.04.006>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0892036215000422>. Acesso em: 01 fev. 2023.

CUHRA, M.; TRAAVIK, T.; BOHN, T. Clone and age dependent toxicity of a glyphosate commercial formulation and its active ingredient in *Daphnia magna*. **Ecotoxicology**, [s.l.], v. 22, n. 2, p. 251-262, 6 dez. 2012. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10646-012-1021-1>. Disponível em: <https://link.springer.com/article/10.1007/s10646-012-1021-1>. Acesso em: 01 mar. 2023.

CUNHA, A. L. B. N.; MODOTTI, M. M.; SETTI, J. R. Classificação de Caminhões Através de Agrupamento por Análise de Cluster. **ANPET**. ISBN: 978-8587893-14-7. Fortaleza CE. 2008.

DANTAS R. A. S.; SAWADA, N. O.; MALERBO M. B. Pesquisas sobre qualidade de vida: Revisão da produção científicas das universidades públicas do Estado de São Paulo. **Revista Latino-Americana de Enfermagem**, Ribeirão Preto, v. 11, n. 4, p. 532-538, 2003.

DELLAVECHIA, J. F.; CRUZ, C.; SILVA, A. F.; CERVEIRA JR, W. R.; GARLICH, N. Macrophyte bioassay applications for monitoring pesticides in the aquatic environment. **Planta Daninha**, Minas Gerais, v. 34, n. 3, p. 587-603, 2016.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society B**, Oxford, v. 39, n. 1, p. 1-38, 1977.

DOU. Diário Oficial da União. Brasil. **Lei nº 7802, de 11 de julho de 1989**. 11 julho 1989. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/agrotoxicos/legislacao/arquivos-de-legislacao/lei-7802-1989-lei-dos-agrotoxicos/view>. Acesso em: 03 maio 2022.

DUARTE, R. M. B. O.; MATOS, J. T. V.; SENESE, N. **Organic Pollutants in Soils**. In: *Soil Pollution: From Monitoring to Remediation*, 103-126p. 2018.

DUTRA, R. M. S.; SOUZA, M. M. O. Impactos negativos uso de agrotóxicos à saúde humana. **Revista Brasileira de Geografia Médica e da Saúde**. Hygeia, ISSN: 1980-1726. 13 (24): 127 - 140, 2017.

DOMINGUES, M. R.; BERNARDI, M. R.; ONO, E. Y. S. ONO, M. A. Agrotóxicos: Risco à Saúde do Trabalhador Rural. Semina: **Ciências Biológicas e da Saúde**, Londrina, v. 25, p. 45-54, 2004.

EDDLESTON, M.; BUCKLEY, N. A.; EYER, P. DAWSON, A. H. **Management of acute organophosphorus pesticide poisoning**. *Lancet*. 2008, 371, 597.

EDDLESTON, M.; HILLIPS, M. R. Self Poisoning with Pesticides. **British Medical Journal**. 328, 42. CrossRef. PubMed. 2004.

ELLIOTT, J. E.; WILSON, L. K.; LANGELIER, K. M.; MINEAU, P. et al. Secondary poisoning of birds of prey by the organophosphorus insecticide, phorate. **Ecotoxicology**, [s.l.], 6, n. 4, p. 219-231, 1997.

ELMAGARMID, A. K.; IPEIROTIS, P. G.; VERYKIOS, V. S. Duplicate record detection: a survey. **IEEE Trans Knowl Data Eng** [s.l.], 19(1):1-16, 2007.

GOMES, M. A. F.; BARIZON, R. R. M. **Panorama da Contaminação Ambiental por Agrotóxicos e Nitrato de origem Agrícola no Brasil: cenário 1992/2011**. Embrapa Meio Ambiente Jaguariúna, SP, 2014.

EVANGELINE, S. B.; SUDHASINI, P. Avaliação da performance do algoritmo J48 para construção de modelos baseados em árvores de decisão. **Revista Brasileira de Computação Aplicada**. Passo Fundo - RS, v.10, n.2, pp.80–90, 10(2):80-90. 2018.

FAGERIA, N. K.; STONE, L. F. **Qualidade do Solo e Meio Ambiente**. EMBRAPA. Santo Antônio de Goiás, GO. 2006.

FAO. **Ranking da FAO mostra que uso de defensivos no Brasil é menor que em diversos países da Europa**. Publicado em 26/06/2019. Disponível em: [https://www.gov.br/agricultura/pt-br/assuntos/noticias/ranking-da-fao-mostra-que-uso-de-defensivos-no-brasil-emenor-que-em-diversos-paises-da-europa#:~:text=Nesse%20ranking%2C%20est%C3%A3o%20na%20frente,Gr%C3%A9cia%20\(0%2C30\)](https://www.gov.br/agricultura/pt-br/assuntos/noticias/ranking-da-fao-mostra-que-uso-de-defensivos-no-brasil-emenor-que-em-diversos-paises-da-europa#:~:text=Nesse%20ranking%2C%20est%C3%A3o%20na%20frente,Gr%C3%A9cia%20(0%2C30).). Acesso em: 01 mar. 2023.

FARIA, N.; FASSA, A. U.; FACHINI, L. A. Intoxicação por Agrotóxicos no Brasil: Os sistemas oficiais de Informação e desafios para realização de estudos epidemiológicos. **Ciência e Saúde Coletiva**, São Paulo, 12 (001): 25-38, 2007.

FARIA, N. M.; FACCHINI L. A.; FASSA A. G.; TOMASI, E. Estudo transversal sobre a saúde mental de agricultores da Serra Gaúcha (Brasil). **Rev Saúde Pública**, São Paulo, 33(4):391-400,1999.

FARIA, N. M.; FACCHINI, L. A.; FASSA, A. G.; TOMASI, E. Pesticides and respiratory symptoms among farmers. **Rev Saúde Pública**, São Paulo, 39(6):973-981, 2005.

FARIA, N. M. X.; FASSA, A. C. G.; MEUCCI, R. D. Association between pesticide exposure and suicide rates in Brazil. **Neurotoxicology**. 45:355-362, 2014.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **American Association for Artificial Intelligence**, 1996.

FENZKE, M. N.; MELLO, M. C. V. A.; SANTOS, K. N. S. C.; CEZAR-VAZ, M. R. Adoecimentos e fatores relacionados à saúde do trabalhador rural. **Journal of Nursing**, Recife, v. 12, n. 7, p. 2214-2226, 2018. DOI: <https://dx.doi.org/10.5205/1981-8963-v12i8a231532p2214-2226-2018>. Acesso em: 03 abr. 2023.

FERMENT, G.; MALGAREJO, L.; FERNANDES, G. B.; FERRAZ, J. M. (Org.). **Lavouras transgênicas – riscos e incertezas: mais de 750 estudos desprezados**

pelos órgãos reguladores de OGMs. Brasília: Ministério do Desenvolvimento Agrário, 2015.

FERNANDES, F. T.; FILHO, A. D. P. C. Perspectivas de mineração de dados e aprendizado de máquina para segurança e saúde ocupacional. **Revista Brasileira de Saúde Ocupacional**, São Paulo, v.44, 2019.

FERREIRA, A.; MAROICO, E.; YONAMINE, M.; OLIVEIRA, M. L. F. **Organophosphate and Carbate poisonings in some Municipalities of Paraná State, Brazil from 1994 to 2002.** São Paulo. 2007.

FIGUEIREDO, G. M. **Efeitos na saúde de trabalhadores expostos a longo prazo a agrotóxicos atendidos no ambulatório de toxicologia do Hospital das Clínicas da UNICAMP nos anos de 2006 e 2007.** Dissertação (Programa de Enfermagem) - Universidade de Campinas UNICAMP. São Paulo, 2009. Disponível em: <https://repositorio.unicamp.br/acervo/detalhe/439470>. Acesso em 10/01/2023.

FIOCRUZ/CICT - Fundação Oswaldo Cruz/Instituto de Comunicação e Informação científica e Tecnológica em Saúde (Fiocruz/CICT). **Estatística anual de casos de intoxicação e envenenamento: Brasil: 1999.** Rio de Janeiro: Fiocruz/CICT; 2000.

FIOCRUZ - Revista Galileu. **Entenda por que o Brasil é o maior consumidor de agrotóxicos do mundo.** Disponível em: <https://renastonline.ensp.fiocruz.br/recursos/entenda-brasil-maior-consumidor-agrotoxicos-mundo>. 2018. Acesso em: 27 jun. 2022.

FIORILLO, C. A. P.; RODRIGUES, M. A. **Manual de direito ambiental e legislação aplicável.** São Paulo: Max Limonad, 1997.

FRARES, R. M. O. **Contribuição ao estudo das intoxicações por agrotóxicos registradas pelo Centro de Controle de Intoxicações de Maringá, no período de 1991-1995.** Dissertação (Programa de Pós-Graduação em Sustentabilidade) - Universidade Estadual de Maringá -UEM. Maringá, 1997.

FREIRE, C.; KOIFMAN, S. PESTICIDES, depression and suicide: a systematic review of the epidemiological evidence. **Int J Hyg Environ Health** 2013; 216(4):445-460. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1438463912001393>. Acesso em: 07/01/2023.

GABOARDI, S. C.; CANDIOTTO, L. Z. P.; RAMOS, L. M. Perfil do Uso de Agrotóxicos no Sudoeste do Paraná (2011-2016). **Rev. NERA.** Presidente Prudente v. 22, n. 46, pp. 13-40, 2019.

GALLUP, J. L.; SACHS, J. D. Agriculture, Climate, and Technology: Why Are the Tropics Falling behind? **American Journal of Agricultural Economics**, [s.l.], 82, n. 3, p. 731-737, 2000.

- GONÇALVES, G. M. S.; GURGEL, I. G. D.; COSTA, A. M. et al. Uso de agrotóxicos e a relação com a saúde na etnia Xukuru do Ororubá, Pernambuco, Brasil. **Saúde Soc**, São Paulo, 21(4):1001-1002, 2012.
- GONZAGA, C. W. P.; BALDO, M. P.; CALDEIRA, A. P. Exposição a agrotóxicos ou práticas agroecológicas: ideação suicida entre camponeses do semiárido no Brasil. **Ciência & Saúde Coletiva**, Rio de Janeiro, 26(9):4243-4252, 2021.
- HAIR, J.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise Multivariada de Dados**. 5ª ed. Porto Alegre, RS: Bookman Companhia Editora, 593p. 2005.
- HALLMANN, C. A.; FOPPEN, R. P. B.; VAN TURNHOUT, C. A. M.; DE KROON, H. et al. Declines in insectivorous birds are associated with high neonicotinoid concentrations. **Nature**, [s.l.] 511, n. 7509, p. 341-343, 2014.
- HAN, J.; KAMBER, M. **Data Mining – Concepts and Techniques**. Morgan Kaufmann Publishers, New York. 1st Edition Edition. 2000.
- HAND, D.; Mannila, H; SMYTH, P. **Principles of Data Mining**, The MIT Press, Massachusetts. 2001.
- HAMZA, M. A.; ANDERSON, W. K. Soil compaction in cropping systems: A review of the nature, causes and possible solutions. **Soil and Tillage Research**, Lexington, 82, n. 2, p. 121-145, 2005.
- HARDING, T. W.; ARANGO, M. V.; BALTAZAR, J. et al. Mental disorders in primary health care: a study of their frequency and diagnosis in four development countries. **Psychological Medicine**, Cambridge, v. 10, n. 2, p. 231-241, May 1980.
- HEIL, D.; G. SPOSITO. Chemical attributes of soil quality. In E.G. Gregorich and M.R. Carter (eds.), **Soil Quality for Crop Production**, Chap. 3. Elsevier Science Publishers, Amsterdam, The Netherlands. 1997.
- HERCULANO, S.; FREITAS, C. M.; PORTO, M. F. S. Introdução: qualidade de vida e riscos ambientais como um campo interdisciplinar em construção. In: Herculano S, Freitas CM, Porto MFS, organizadores. **Qualidade de vida & riscos ambientais**. Niterói: EDUFF; 2000. p. 17-26.
- HIPP, J.; GUNTZER, U. e NAKHAEIZADEH, G. Algorithms for Association Rule Mining – A General Survey and Comparison. **SIGKDD**. 2000.
- HAN, J.; PEI, J.; YIN, Y. Mining Frequent Patterns Without Candidate Generation. **SIGMOD** 2000.
- HAN, J.; KAMBER, M. **“Data Mining – Concepts and Techniques”**; Morgan Kaufmann Publishers, Inc, 2001.
- HU, X.; ILLHOY, Y. (2004). **Cluster ensemble and its applications in gene expression analysis**. ACM International Conference Proceeding Series; Vol. 55

Proceedings of the second conference on Asia Pacific bioinformatics – v. 29, pp. 297-302, Nova Zelândia. 2004.

IARC. International agency for research on cancer. International agency for research on cancer. Monographs on the identification of carcinogenic hazards to humans. **Agents Classified by the IARC Monographs**, Volumes 1–129. Última atualização em 22 julho 2021. Disponível em: <https://monographs.iarc.who.int/agents-classified-by-the-iarc/>. 2021. Acesso em: 11 ago. 2022.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Uso de Agrotóxicos no Estado do Paraná**. Safra 1998/1999. ISBN 85 2400847-4. Rio de Janeiro, 2001.

IBGE - instituto brasileiro de geografia e estatística. **Censo Agro 2017**. Disponível em: <https://censos.ibge.gov.br/agro/2017/>. Acesso em: 25 set. 2022.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Indicadores de Desenvolvimento Sustentável: Brasil 2012**. Estudos e Pesquisas. Informação Geográfica, n. 9. Rio de Janeiro: IBGE; 2012.

INCA - Instituto Nacional de Câncer. **Exposição a Agrotóxicos no Ambiente e no Trabalho**. Publicado em 24 Agosto 2021. Disponível em: <https://www.inca.gov.br/exposicao-no-trabalho-e-no-ambiente/agrotoxicos>. Acesso em: 05 abr. 2022.

IPARDES - Instituto Paranaense de Desenvolvimento Econômico e Social. **Relação dos municípios do Estado ordenados segundo as Mesorregiões e as microrregiões geográficas do IBGE – PARANÁ – 2012**. Disponível em: http://www.ipardes.gov.br/pdf/mapas/base_fisica/relacao_mun_micros_mesos_parana.pdf. Acesso em: 18 ago. 2022.

JACOBSON, M. H.; WU, Y.; LIU, M.; KANNAN, K.; Li, A. J.; ROBINSON, M. et al. Organophosphate pesticides and progression of chronic kidney disease among children: a prospective cohort study. **Environ Int**. São Paulo -SP. 155:106597. 2021. doi: <http://dx.doi.org/10.1016/j.envint.2021.106597>. PMID:33951537.

JEYARATNAM, J. Acute pesticide poisoning: a major global health problem. **World Health Stat Q**, Geneva, 43:139-44, 1990.

JOBIM, P. F. C.; NUNES L. N.; GIUGLIANI R., et al. Existe uma associação entre mortalidade por câncer e uso de agrotóxicos? Uma contribuição ao debate. **Ciênc. Saúde Colet**, Rio de Janeiro, 15(1):277-288, 2010.

JOHNSON, R. A.; Wichern, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall. 1998.

JUNG, Y. G.; KANG, M. S.; HEO, J. Clustering performance comparison using K-means and expectation maximization algorithms. **Biotechnology, biotechnological equipment**, London, 28(sup1), S44–S48, 2014. <https://doi.org/10.1080/13102818.2014.949045>.

KARLEN, D. L.; DITZLER, C. A.; ANDREWS, S. S. Soil quality: why and how? **Geoderma**, [s.l.] 114, n. 3, p. 145-156, 2003.

KAUFMAN, L.; P. J. ROUSSEEUW. **Find Groups in Data: An Introduction to Cluster Analysis**. New York: Wiley, 368 p. 1990.

KAUR, S.; KUMAR, V.; CHAWLA, M.; CAVALLO, L. et al. Pesticides Curbing Soil Fertility: Effect of Complexation of Free Metal Ions. **Frontiers in Chemistry**, 5, n. 43, 2017- July-04 2017.

KIM, J.; KO Y.; LEE W. J. Depressive symptoms and severity of acute occupational pesticide poisoning among male farmers. **Occup Environ Med**, [s.l.], 70(5):303- 309, 2013.

KIM, Y. A.; YOON, Y. S.; KIM, H. S.; JEON, S. J.; COLE, E.; LEE, J.; CHO, Y. H. Distribution of fipronil in humans, and adverse health outcomes of in utero fipronil sulfone exposure in newborns. **International Journal of Hygiene and Environmental Health**, [s.l.], 222(3), 524–532. 2019.

KOUREAS, M. et al. Systematic review of biomonitoring studies to determine the association between exposure to organophosphorus and pyrethroid insecticides and human health outcomes. **Toxicology Letters**, Amsterdam, v. 210, n.2, p.155-168, abr. 2012.

LAL, R. Soil carbon dynamics in cropland and rangeland. **Environmental Pollution**, Germany, 116, n. 3, p. 353-362, 2002.

LAROSE, D. T. **Discovering Knowledge in Data: An Introduction to Data Mining**. John Wiley and Sons, Inc, New Jersey, Canadá. 2005.

LAUBER, C. L.; HAMADY, M.; KNIGHT, R.; FIERER, N. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. **Applied and Environmental Microbiology**, Washington, 75, n. 15, p. 5111, 2009.

LEE, Y. C.; HUANG, S. C. HUANG, C. H.; WU, H. H. A new approach to identify high burnout medical staffs by kernel K-means cluster analysis in a regional teaching hospital in Taiwan. **PubMed Inquiry**. Rockville US, 53(2), 2016.

BRASÍLIA, **Lei Federal 6.938/1981, de 31 de Agosto de 1981**: Dispõe sobre a Política Nacional do Meio Ambiente, seus fins e mecanismos de formulação e aplicação, e dá outras providências. Ministério de Meio Ambiente - Planalto Central. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l6938compilada.htm. Acesso em: 07 mar. 2022.

LIMA, R. A. F. **Estratégias de Seleção de Atributos para detecção de anomalias em transações eletrônicas**. Dissertação (Ciência da Computação) - Universidade Federal de Minas Gerais - Belo Horizonte, 2016. Disponível em: <https://repositorio.ufmg.br/handle/1843/ESBF-ACXKTA>. Acesso em: 07/01/2023.

LIN, S.W.; CHEN, S. C. Parameter determination and feature selection for C4. 5 algorithm using scatter search approach, **Soft Computing**: Kitakyushu Japan, 63-75. 2012.

LINDEN, R. Técnicas de Agrupamento. **Revista de Sistemas de Informação da FSMA**, Rio de Janeiro, n. 4 pp. 18-36 19, 2009.

LIONETTO, G. M. et al. Acetylcholinesterase as a biomarker in environmental and occupational medicine: new insights and future perspectives. **Biomed Research International**. Salento, v. 2013, n-7, p. 1-8, jul.2013.

LOPES, C. V. A., ALBUQUERQUE, G. S. C. Desafios e avanços no controle de resíduos de agrotóxicos no Brasil: 15 anos do Programa de Análise de Resíduos de Agrotóxicos em Alimentos. **Cadernos de Saúde Pública**, Rio de Janeiro, 37(2), 1-14. 2021.

MACHADO, P. A. L **Direito ambiental brasileiro**. São Paulo: Malheiros, 2013.

MANJABOSCO, C. W.; MORATA, T. C.; MARQUES, J. M. Perfil audiométrico de trabalhadores agrícolas. **Arq Int Otorrinolaringol**: São Paulo, 8(4):285-295, 2004

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. MIT Press. 2003.

MARDIA, K. D.; Kent, J. T.; Bibby, J. N. **Multivariate analysis**. London: Academic Press. 1979.

MARI, J.; WILLIAMS, P.A. Validity study of a psychiatric screening questionnaire (SRQ-20) in primary care in the city of Sao Paulo. **British Journal of Psychiatry**, Cambridge. 148, 23-26.doi:10.1192/bjp.148.1.23, 1986.

MARTINS, J. G.; BARISON, A.; OLIVEIRA, G. A. R.; LIÃO, L. M.; A Database for Automatic Classification of Gender in Araucaria angustifolia Plants. **Soft Computing**: Kitakyushu Japan, 25, 5503–5517, 2021.

MARTINS, E. L. **Previsão da lixiviação de agrotóxicos utilizados na cultura de algodão em Mato Grosso**. 82 f. Dissertação (Mestrado em Agricultura Tropical) – Universidade Federal de Mato Grosso, Cuiabá, 2006.

MATHWORKS. MatLab v. 6.5 – **User's Guide**. Mathworks Inc. Massachusetts, U.S. 2002.

MATOS, A. K. V. Revolução Verde, biotecnologia e tecnologias alternativas. **Cadernos da FUCAMP**, Monte Carmelo, v. 10, n. 12, p. 1-17, 2010.

MCCUE, C. **Data Mining and Predictive Analysis - Intelligence Gathering and Crime Analysis**. Elsevier, 2007.

MEDEIROS, M. **Seleção de Principais Atributos para Redes Neurais Artificiais do Tipo Mlp: Um Estudo De Caso Sobre Mineração de Dados Para Diagnóstico de Dengue**. Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Norte Caicó - RN. 2018. Disponível em: <https://repositorio.ufrn.br/handle/123456789/42876>. Acesso em: 01/02/2023.

MEFTAUL, I. Md., VENKATESWARLU, K., DHARMARAJAN, R., et al. Controversies over human health and ecological impacts of glyphosate: Is it to be banned in modern agriculture? **Environmental Pollution**, [s.l.], 263, 1-72, 2020.

MESA, Y. P. L. The decision-making process of synthetic pesticide use in agricultural communities in Colombia: A grounded theory approach. **Revista Facultad Nacional de Salud Pública**, Medellín, 38(2), 1-7, 2020.

MESNAGE, R.; RENNEY, G.; SÉRALINI, et al. Multiomics reveal non-alcoholic fatty liver disease in rats following chronic exposure to an ultra-low dose of Roundup herbicide. **Scientific Reports**, London, 7(1), 1-15, 2017.

MILLER, G. T. **Ciência ambiental**. 11ª Edição. São Paulo: Thomson Learning; 2007.

MILLS, P. J.; CAUSSY, C.; & LOOMBA, R. Glyphosate excretion is associated with steatohepatitis and advanced liver fibrosis in patients with fatty liver disease. **Clinical Gastroenterology and Hepatology**, [s.l.], 18(3), 741-743, 2020.

MORALES-OLMEDO, M.; ORTIZ, M.; SELLÉS, G. Effects of transient soil waterlogging and its importance for rootstock selection. **Chilean journal of agricultural research**, Chile, 75, p. 45-56, 2015.

MORAES, P. V. D.; ROSSI, P. Comportamento ambiental do glifosato. **Scientia Agraria Paranaensis**, Marechal Cândido Rondon, v. 9, n. 3, p. 22-35, 2010.

MORATA, T. C.; LITTLE, B. Suggested guidelines for studying the combined effect of occupational exposure to noise and chemicals on hearing. **Noise Health**, [s.l.], 4(14):73-87, 2002.

MOREIRA, J. C.; JACOB, S. C.; PERES, F. et al. Avaliação integrada do impacto do uso de agrotóxicos sobre a saúde humana em uma comunidade agrícola de Nova Friburgo, RJ. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 7, n. 2, p. 299-311, 2002. DOI: 10.1590/S1413-81232002000200010.

MOSTAFALOU, S.; ABDOLLAHI, M. Pesticides: an update of human exposure and toxicity. **Arch Toxicol**. [s.l.], v. 91, n. 2, p. 549-599, 2017.

MOTTA, C. G. L. **Sistema Inteligente para Avaliação de Riscos em Vias de Transporte Terrestre**. Dissertação (Programa de Engenharia Civil) - Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2004.

MURAKAMI Y.; PINTO N.F.; ALBUQUERQUE G.S.C.; PERNA P.O.; LACERDA A. Intoxicação crônica por agrotóxicos em fumicultores. **Saúde Debate**. 2017;41(113):563-76. doi: <http://dx.doi.org/10.1590/0103-1104201711317>.

NAGARAJU, R.; JOSHI, A. K. R.; VAMADEVA, S. G.; & RAJINI, P. S. Deregulation of hepatic lipid metabolism associated with insulin resistance in rats subjected to chronic monocrotophos exposure. **Journal of Biochemical and Molecular Toxicology**, [s.l.], 34(8), 1–9, 2020.

NASCIMENTO, D. Z.; MARQUES, G. M.; TREVISOL, D. J. O desafio das agências reguladoras ao redor do mundo no uso do glifosato. **Saúde e Sociedade**, São Paulo, 28(4), 297–298, 2019.

NAWAZ, M. F.; BOURRIÉ, G.; TROLARD, F. Soil compaction impact and modelling. A review. **Agronomy for Sustainable Development**, [s.l.], 33, n. 2, p. 291-309, 2013/04/01, 2013.

NETTO, O. P. **Um Filtro Iterativo Utilizando Árvores de Decisão**. Dissertação (Interunidades em Bioinformática) – Universidade de São Paulo - USP. Ribeirão Preto, 2013. Disponível em: <https://teses.usp.br/teses/disponiveis/95/95131/tde-10122013-185844/pt-br.php>. Acesso em 01/02/2023.

NEVES, P. D. M.; BELLINI, M. Intoxicações por agrotóxicos na mesorregião norte central paranaense, Brasil – 2002 a 2011. **Ciênc. Saúde Colet.** Rio de Janeiro, 18(11):3147- 3156, 2013.

NEVES, P. D. M.; MENDONÇA, M. R.; BELLINI, M., & PÔSSAS, I. B. Intoxicação por agrotóxicos agrícolas no estado de Goiás, Brasil, de 2005-2015: Análise dos registros nos sistemas oficiais de informação. **Ciência & Saúde Coletiva**, Rio de Janeiro, 25(7), 2743–2754, 2020.

NISHIYAMA, P. **Utilização de agrotóxicos em áreas de reforma agrária no estado do Paraná**. Tese (Saúde Coletiva) - Campinas: Universidade Estadual de Campinas, São Paulo, 2003. Disponível em: <https://repositorio.unicamp.br/Busca/Download?codigoArquivo=455390>. Acesso em: 01/02/2023.

OLIVEIRA, C. J. S. **Classificação de Imagens Coletadas na Web**. Dissertação (Ciência da Computação) - Universidade Federal de Minas Gerais, MG, 2001. Disponível em: https://www.cin.ufpe.br/~fnj/CG/proposta_pms.tex. Acesso em: 01/02/2023.

OLIVEIRA, M. L. F.; ZAMBRONE, F. A. D. Vulnerabilidade e Cuidado na Utilização de Agrotóxicos por Agricultores Familiares. **Ciência, Cuidado e Saúde**, Maringá; 5:99-196, 2006. Disponível em: <https://repositorio.unicamp.br/Busca/Download?codigoArquivo=452802>. Acesso em: 01/03/2023.

OLIVEIRA, R. R.; CARVALHO, C. L. **Algoritmos de agrupamento e suas aplicações**. Technical report, Universidade Federal de Goiás, 2008. Disponível em: <https://datascience.eu/pt/aprendizado-de-maquina/algoritmos-de-agrupamento-e-o-seu-significado-na-aprendizagem-de->

maquinas/#:~:text=Os%20algoritmos%20de%20agrupamento%20s%C3%A3o,para%20destacar%20um%20bom%20agrupamento. Acesso em: 01/02/2023.

OLIVEIRA FILHO, L. C. I.; BARETTA, D. Por que devemos nos importar com os colêmbolos edáficos? **Scientia Agraria**, [s.l.], (2): 21-40, 2017. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/6113588.pdf>. Acesso em: 04/02/2023.

OLIVEIRA, L. B. **Avaliação da prevalência de sintomas gastrointestinais e cardiovasculares em trabalhadores rurais expostos e não expostos à agrotóxicos no Vale de Cuncas, Barro-CE**. Dissertação (Mestrado em Saúde Coletiva) – Universidade Católica dos Santos. SANTOS – SP, 2017. Disponível em: <https://tede.unisantos.br/handle/tede/3801>. Acesso em: 04 fev. 2023.

OLSON, R.; THOMPSON, S. V.; WIPFLI, B.; et al. Sleep, dietary, and exercise behavioral clusters among truck drivers with obesity: implications for interventions. **J Occup Environ Med**, Philadelphia, 58(3):314-21, 2016.

OMS – Organização Mundial da Saúde. **Manual de vigilância da saúde de populações expostas a agrotóxicos**. Brasília. Repartição sanitária pan-americana, escritório regional da organização mundial da saúde representação do brasil setor de embaixadas norte. Lote19-brasília-df-70800-400. Opas/brahep/002/97. 1996.

OULAS, A.; PAVLOUDI, C.; POLYMENAKOU, P.; et al. Metagenomics: Tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. **Bioinformatics and Biology Insights**, [s.l.], 9, p. 75-88, 2015.

GONZAGA, P. W. C.; BALDO, P. M.; CALDEIRA, P. A. Exposição a agrotóxicos ou práticas agroecológicas: ideação suicida entre camponeses do semiárido no Brasil. DOI: 10.1590/1413-81232021269.09052020. **Ciência & Saúde Coletiva**, 26(9):4243-4252, 2021.

PALMA, D. C. A. **Agrotóxicos em leite humano de mães residentes em Lucas do Rio Verde – MT**. Dissertação (Saúde Coletiva) - Universidade Federal de Mato Grosso. Cuiabá, 2011. Disponível em: <https://pt.scribd.com/document/94723797/41-Agrotoxicos-Em-Leite-Humano-de-Maes-Residentes-Em-Lucas-Do-Rio-Verde-Mt>. Acesso em: 02/02/2023.

PANDEY, A.; DABHADE, P.; & KUMARASAMY, A. Inflammatory effects of subacute exposure of roundup in rat liver and adipose tissue. **Dose-Response**, Florida, 17(2), 1–11. 2019.

PARANÁ. Secretaria de Estado da Saúde do Paraná. Superintendência de Atenção à Saúde. **Linha Guia da Atenção às Populações Expostas aos Agrotóxicos**. – 1 ed. – Curitiba: SESA, 2018. 72 p. Disponível em: https://www.saude.pr.gov.br/sites/default/arquivos_restritos/files/documento/2020-04/linhaguiaagrotoxicos.pdf. Acesso em: 01 mar. 2023.

PAVAO-ZUCKERMAN, M. A. Soil Ecology. In: JORGENSEN, S. E. e FATH, B. D. (Ed.). **Encyclopedia of Ecology**. Oxford: Academic Press, 2008. p. 3277-3283.

PEREIRA, P.; BOGUNOVIC, I.; MUÑOZ-ROJAS, M.; BREVIK, E. C. Soil ecosystem services, sustainability, valuation and management. **Current Opinion in Environmental Science & Health**, London, 5, p. 7-13, 2018.

PARK, M. G.; BLITZER, E. J.; GIBBS, J.; LOSEY, J. E. et al. Negative effects of pesticides on wild bee communities can be buffered by landscape context. *Proceedings of the Royal Society B: Biological Sciences*, [s.l.], 282, n. 1809, p. 20150299, 2015.

PASCHOAL, A. D. **Pragas, praguicidas e a crise ambiental: problemas e soluções**. Rio de Janeiro: Fundação Getúlio Vargas; 1979.

PEREIRA, E. F.; TEIXEIRA, C. S.; SANTOS, A. Qualidade de vida: abordagens, conceitos e avaliação. **Revista Brasileira de Educação Física e Esporte**, São Paulo, v. 26, n. 2, p. 241-250, 2012.

PERES, F.; MOREIRA, J. C.; DUBOIS, G. S. Agrotóxicos, saúde e ambiente: uma introdução ao tema. *In*: PERES, F.; MOREIRA, J. C. (Org.). **É veneno ou é remédio? agrotóxicos, saúde e ambiente**. Rio de Janeiro: Fiocruz, 2003. p. 21-41. DOI: 10.7476/9788575413173.

PERES, F.; OLIVEIRA-SILVA, J. J.; DELLA-ROSA, H. V.; LUCCA, S. R. de. Desafios ao estudo da contaminação humana e ambiental por agrotóxicos. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 10, p. 27-37, set./out. 2005. DOI: 10.1590/S1413-81232005000500006.

PIGNATI, W. A. **Os riscos, agravos e vigilância em saúde no espaço de desenvolvimento do agronegócio no Mato Grosso**. Tese. Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro. 2007. Disponível em: <https://www.arca.fiocruz.br/handle/icict/4567>. Acesso em: 01/02/2023.

PIGNATI, W. A.; LIMA, F. A. N. S.; LARA, S. S.; et al. Distribuição espacial do uso de agrotóxicos no Brasil: uma ferramenta para vigilância em saúde. **Ciência e Saúde Coletiva**, São Paulo, n. 22, v. 10, 2017. Disponível em: <https://www.scielo.br/j/csc/a/grnBRDjmtcBhm6CLprQvN/?lang=pt>. Acesso em: 02/02/2023.

PIGNATI, W. A.; MACHADO, J. M. H. O agronegócio e seus impactos na saúde dos trabalhadores e da população do Estado de Mato Grosso. *In*: GOMEZ, MACHADO, PENA. (Orgs.) **Saúde do trabalhador na sociedade brasileira contemporânea**. Rio de Janeiro: Editora Fiocruz, 2011, p. 245-272.

PIGNATI, W. A.; MACHADO, J. M. H.; CABRAL, J. F. Acidente rural ampliado: o caso das “chuvas” de agrotóxicos sobre a cidade de Lucas do Rio Verde. **Ciência & Saúde & Coletiva**, Rio de Janeiro, v. 12, n. 1, p. 105-114, jan./mar. 2012. DOI: 10.1590/S1413-81232007000100014.

PORTO, M. F. **Injustiça ambiental no campo e nas cidades: do agronegócio químico dependente às zonas de sacrifício urbana**. *In*: PORTO, M. F.; PACHECO,

T.; LEROY, J. P. (Orgs). Injustiça Ambiental e Saúde no Brasil: o mapa de conflitos. São Paulo, Editora FIOCRUZ, 2013.

PORTO, M. F.; SOARES, W. L. Modelo de desenvolvimento, agrotóxicos e saúde: um panorama da realidade agrícola brasileira e propostas para uma agenda de pesquisa inovadora. **Revista Brasileira de Saúde Ocupacional**, São Paulo, 37(125), 31–39, 2012.

PORTUGAL, A. F.; COSTA, O. D. V.; COSTA, L. M. Propriedades Físicas e Químicas do Solo em Áreas com Sistemas Produtivos e Mata na Região da Zona da Mata Mineira. **Revista Brasileira de Ciência do Solo**. Viçosa – MG, Brasil, 34:575-585, 2010.

PORTIER, C. J. A comprehensive analysis of the animal carcinogenicity data for glyphosate from chronic exposure rodent carcinogenicity studies. **Environmental Health**, Colorado, 19(1), 1-18. 2020.

PRADO, J. A. F.; SIEGLOCH, A. E.; SILVA, B. F.; AGOSTINETTO, L. Exposição de trabalhadores rurais aos agrotóxicos. **Gaia Scien**. João Pessoa, 2021;15(1):141-57. doi: <https://doi.org/10.22478/ufpb.1981-1268.2021v15n1.56075>.

PULIDO MONCADA; GABRIELS, D.; LOBO, D.; REY, J. C. et al. **Visual field assessment of soil structural quality in tropical soils**. Soil and Tillage Research, 139, p. 8-18, 2014/06/01/ 2014.

QUINLAN, J. R. **C4.5**: Programs for machine learning. San Mateo, CA: Morgan Kaufmann Publishers, 1993.

QUONIAM, L.; TARAPANOFF, K.; ARAÚJO, J. R. H.; ALVARES, L. Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. **Ci Inf**. Brasília, 30(2):20-8, 2001.

RAIMUNDO, L. R.; MATTOS, M. C.; SIMÕES, P. W. T. A.; CECHINEL, C. **O Algoritmo de Classificação CART em uma Ferramenta de Data Mining**. Universidade do Extremo Sul Catarinense (UNESC) – Criciúma, SC, Brasil. 2008. Disponível em: <https://periodicos.unesc.net/ojs/index.php/sulcomp/article/view/1994/1888>. Acesso em: 01/02/2023.

RAMAI, D.; TAI, W.; RIVERA, M.; FACCIORUSSO, A.; TARTAGLIA, N.; PACILLI, M.; AMBROSI, A.; COTSOGLU, C.; & SACCO, R. Natural progression of nonalcoholic steatohepatitis to hepatocellular carcinoma. **Biomedicines**, China, 9(2), 1-16, 2021.

RAMYA, M.; LOKESH, V.; MANJUNATH, T; HEGADI, R. S. A predictive model construction for mulberry crop productivity. **Procedia Computer Science**, Manchester, 45: 156–165, 2015.

RAMPAZZO, T. G.; RAMPAZZO, N.; MENTLER, A.; BLUM, W. E. H. et al. Influence of soil tillage and erosion on the dispersion of glyphosate and

aminomethylphosphonic acid in agricultural soils. **International Agrophysics**, Austria, 1, p. 93- 100, 2014. Disponível em: <http://www.international-agrophysics.org/Influence-of-soil-tillage-and-erosion-on-the-dispersion-of-glyphosate-and-aminomethylphosphonic,104054,0,2.html>. Acesso em: 01 fev. 2023.

RIBEIRO, M. X. **Suporte a sistemas de auxílio ao diagnóstico e de recuperação de imagens por conteúdo usando mineração de regras de associação**. Tese. São Carlos, São Paulo, 2008. Disponível em: <https://teses.usp.br/teses/disponiveis/55/55134/tde-16022009-144432/fr.php>. Acesso em: 01/02/2023.

RICHARD O. D.; PETER, E. H.; DAVID G. **Stork, Pattern Classification**. 2ed. Wiley-Interscience Publication, John Wiley & Sons, Inc, 2001.

RICHART, A.; FILHO, J. T.; BRITO, O. R.; LLANILLO, R. F.; FERREIRA, R. Compactação do solo: causas e efeitos. **Semina: Ciências Agrárias**, v. 26, núm. 3, p. 321-343, Londrina, Paraná, 2005.

RICHENDRFER, H.; CRETON, R. Chlorpyrifos and malathion have opposite effects on behaviors and brain size that are not correlated to changes in AChE activity. **Neurotoxicology**, Bethesda, v. 14, p. 50-58, 2015. DOI: 10.1016/j.neuro.2015.05.002.

RIGOTTO, R. M.; VASCONSELOS, D. P.; ROCHA, M. M. 2014. Uso de agrotóxicos no Brasil e problemas para a saúde pública. **Cad. Saúde Pública**, Rio de Janeiro, 30(7):1-3, jul, 2014.

ROCHA, L. M.; CAPPABIANCO, F. A. M.; FALCÃO, A. X., Data clustering as an optimum-path forest problem with applications in image analysis. **International Journal of Imaging Systems and Technology**, Minnesota, vol. 19, no. 2, pp. 50-68, 2009. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/ima.20191/abstract>. Acesso: em 01 abr. 2023.

ROMÃO, J. E. JR. Doença renal crônica: definição, epidemiologia e classificação. *J Bras Nefrol*. São Paulo, 2004;26(3 suppl. 1):1-3, 2004. Disponível em: <https://www.bjnephrology.org/article/doenca-renal-cronica-definicao-epidemiologia-e-classificacao/>. Acesso: em 01 fev. 2023.

SANTANA E. L.; MACHINSKI J. M. O uso de praguicidas por trabalhadores de setor agrícola atendidos ambulatorialmente em Maringá no período de 2002 e 2003. **Acta sci Health sci**, Maringá, 26(2):325-329, 2004.

SANTANA, V. S.; MOURA, M. C. P.; NOGUEIRA, F. F. Mortalidade por intoxicação ocupacional relacionada a agrotóxicos, 2000-2009, Brasil. **Rev. Saúde Pública**, São Paulo, 47(3):598-606, 2013.

SANTOS, M. A. T.; AREAS, M. A.; REYES, F. G. R. Piretróides, uma visão geral. **Alim. Nutr**, Araraquara, v.18, n.3, p. 339-349, 2007.

SANTOS, G. D.; FRANCISCO, C. N.; ALMEIDA, C. M. Mineração de dados aplicada à discriminação da cobertura da terra em imagem landsat 8 oli. **Bol. Ciênc. Geod.**, Curitiba, v. 21, no 4, p.706-720, 2015.

SANTOS, F. S. **Sistema para descobrimento e visualização de informações utilizando índices sociais de bairros e georreferenciamento no município de Porto Alegre**. Dissertação (Engenharia da Computação) - Universidade Federal do Rio Grande do Sul. Porto Alegre, 2018. Disponível em: <https://lume.ufrgs.br/handle/10183/190156>. Acesso em: 05/02/2023.

SANTOS, G. R. **Hierarquização geométrica dos preditores geoestatísticos**. Tese de Doutorado em Estatística e Experimentação Agropecuária (Setor de Ciências Exatas) - Universidade Federal de Lavras, Lavras, 2010.

SECRETARIA DA SAUDE – PR. **Intoxicação Aguda por Agrotóxicos**. Publicado em: 08 Maio 2018. Disponível em: <https://www.saude.pr.gov.br/Pagina/Intoxicacao-Aguda-por-Agrotoxicos>. Acesso em: 22 maio 2022.

SILVA, J. A. **Direito ambiental constitucional**. São Paulo: Malheiros, 2019.

SILVA, A. S.; WESSELING, C. Pesticide exposures and chronic kidney disease of unknown etiology: an epidemiologic review. **Environ Health**. [s.l.], 2017;16(1):49. doi: <http://dx.doi.org/10.1186/s12940-017-0254-0>. PMID: 28535811.

SILVÉRIO, A. C. P.; MARTINS, I.; NOGUEIRA, D. A. et al. Assessment of Primary Health Care for rural workers exposed to pesticides. **Revista de Saúde Pública**, São Paulo, 54, 9, 2020.

SIMAS, J. M. M. **Saúde e trabalho na bananicultura em uma região do Vale do Ribeira**. 2020. 160 f. Tese de Doutorado (Programa Interdisciplinar em Ciências da Saúde - Instituto de Saúde e Sociedade) - Universidade Federal de São Paulo, Santos, 2020.

SINITOX - Sistema Nacional de Informações Tóxico-Farmacológicas (SINITOX). **Dados de Intoxicação**. 2017. Disponível em: <https://sinitox.icict.fiocruz.br/dados-nacionais>. Acesso em: 13 jun. 2022.

SIQUEIRA, D. F.; MOURA, R. M.; CARNEIRO, G. E. et al. Análise da exposição de trabalhadores rurais a agrotóxicos. **Rev. Bras. Prom. Saúde**, Fortaleza, 26(2):182-191, 2013.

SIQUEIRA, S. L.; KRUSE, M. H. L. Agrotóxicos e saúde humana: contribuição dos profissionais do campo da saúde. **Revista da Escola de Enfermagem da USP**, São Paulo, v. 42, n. 3, p. 584-590, set. 2008. DOI: 10.1590/S0080-62342008000300024.

SOARES, W. L.; PORTO, A. O. Estimating the social cost of pesticide use: An assessment from acute poisoning in Brazil. **Ecological Economics**, Turkey (68):2721-2728, 2009.

SOARES, W. I.; PORTO, M. F. S. Uso de agrotóxicos e impactos econômicos sobre a saúde. **Rev. Saúde Pública**, São Paulo, 46(2):209-217, 2012.

SOARES, W. L. **Uso dos agrotóxicos e seus impactos à saúde e ao ambiente: uma avaliação integrada entre a economia, a saúde pública, a ecologia e a agricultura**. Tese. Escola Nacional de Saúde Pública, Sérgio Arouca, Fiocruz. Rio de Janeiro. 2010.

SOBRAL, A. M. A. R. G. **Identificação de segmentos de risco numa Instituição Financeira. Aplicação de técnicas de Data Mining a dados de Risco Operacional**. Dissertação (Gestão da Informação) - Universidade Nova de Lisboa – Portugal, 2015.

SOUDANI, N.; CHAÂBANE, M.; GHORBEL, I. et al. Glyphosate disrupts redox status and up-regulates metallothionein I and II genes expression in the liver of adult rats. Alleviation by quercetin. **General physiology and biophysics**, Bratislava, 2019, 38(02), 123–134.

SOUZA, R. C. T. **Uma metodologia para classificação de dados nominais baseada no processo kdd: ênfase aos algoritmos culturais, estimação de distribuição e análise de correspondência múltipla**. Tese (Métodos Numéricos em Engenharia) - Universidade Federal do Paraná. CURITIBA, 2013.

STORM, J. E.; ROZMAN, K. K.; DOULL, J. Occupational exposure limits for 30 organophosphate pesticides based on inhibition of red blood cell acetylcholinesterase. **Toxicology**, Hamburg, v.150, n.7, p.1-29, set, 2000.

STOROPOLI, J. E. **O Uso do Knowledge Discovery in Database (KDD) de Informações Patentárias sobre Ensino a Distância: Contribuições para Instituições de Ensino Superior**. Tese (Programa de Pós-Graduação em Administração) - Universidade Nove de Julho – UNINOVE. São Paulo, 2016.

TAKAMOTO, M.; TOMAZELA, M. G. J. M. Aplicação de técnicas de mineração de dados para planejamento agrícola no estado de São Paulo. **Reverte-Revista de Estudos e Reflexões Tecnológicas da Faculdade de Indaiatuba**. Indaiatuba, n. 9, 2011.

TAN, P. N.; KUMAR, V. **Interestigness Measures for Association Patterns: A Perspective**. KDD 2000 Workshop on Postprocessing in Machine Learning and Data Mining. 2000.

THOMAZ J. A. Trabalho e Saúde no Ambiente Destrutivo do Agrohidronegócio Canavieiro no Pontal do Paranapanema (SP). **Revista Pegada Eletrônica**, São Paulo, v. 15, p. 3-15, 2014.

UDAWATTA, R. P.; GANTZER, C. J.; JOSE, S. Chapter 14 - Agroforestry Practices and Soil Ecosystem Services. In: AL-KAISI, M. M. e LOWERY, B. (Ed.). Soil Health and Intensification of Agroecosystems: **Academic Press**, [s.l.], p. 305-333, 2017.

SIQUEIRA, L. S.; KRUSE, L.H.H M. Universidade de São Paulo. Agrotóxicos e saúde humana: contribuição dos profissionais do campo da saúde. **Rev Esc Enferm USP**, São Paulo-SP, 2008; 42(3):584-590.

VALLI, M. Análise de Cluster. **Augusto Guzzo Revista Acadêmica**, São Paulo, n. 4, p. 77-87, Ago. 2012. Disponível em: http://www.fics.edu.br/index.php/augusto_guzzo/article/view/107. Acesso em: 24 out. 2022.

VANDERHOEK, W.; KONRADSEN, F. Analysis of 800 Hospital admissions for acute poisoning in a rural area of Sri Lanka. **Clin Toxicol**. Bethesda, 2006. 44:225-231.

VASCONSELOS, L. M. R.; CARVALHO, C. L. Aplicação de Regras de Associação para Mineração de Dados na Web. **Relatório Técnico RT-INF_004-04**. Instituto de Informática, Universidade Federal de Goiás. 2004.

VASCONCELLOS, P. R. O.; RIZZOTTO, M. L. F.; MACHINESKI, G. G.; & COSTA, R. M. Condições da exposição a agrotóxicos de portadores da doença de Parkinson acompanhados no ambulatório de neurologia de um hospital universitário e a percepção da relação da exposição com o adoecimento. **Saúde em Debate**, São Paulo, 2019. 43(123), 1084–1094.

VELOSO, M. J. P. S. A. **Regras de associação aplicadas a um método de apoio ao planejamento de recursos humanos**. Tese (Análise de Dados e Sistemas de Apoio à Decisão) - Faculdade de Economia Universidade do Porto. 2003.

VERVAET, B. A.; NAST, C. C.; JAYASUMANA, C. et al. Chronic interstitial nephritis in agricultural communities is a toxin-induced proximal tubular nephropathy. **Kidney International**. Antuérpia -Bélgica, 97(2):350-369. 2020. Disponível em: <https://www.kidney-international.org/action/showPdf?pii=S0085-2538%2819%2931150-0>.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. *In*: NOVAIS, R. F.; ALVAREZ V. V. H; SCHAEFER, C. E. G. R. **Tópicos em ciência do solo**. Viçosa, Sociedade Brasileira de Ciência do Solo, 2000.

WAIKATO, U. O. **WEKA**. Arquivo ARFF. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/arff.html>. Acesso em: 23 nov. 2022.

WAISSMANN W. Agrotóxicos e doenças “não” transmissíveis. **Cien Saude Colet**, São Paulo, 12(1):20-21, 2007.

WALL, D. H.; KNOX, M. A. Soil Biodiversity. *In*: **Reference Module in Earth Systems and Environmental Sciences**, Encyclopedia of Soils in the Environment, p. 136-141: Elsevier, Colorado State University, Fort Collins, USA, 2014.

WAHLANG, B.; JIN, J.; BEIER, J. I. et al. Mechanisms of environmental contributions to fatty liver disease. **Current Environmental Health Reports**, Digital Library, 6(3), 80–94, 2019.

WANG, Y.; CHEN, C.; ZHAO, X. et al. Assessing joint toxicity of four organophosphate and carbamate insecticides in common carp (*Cyprinus carpio*) using acetylcholinesterase activity as an endpoint. **Pesticide Biochemistry and Physiology**, [s.l.], v. 122, p. 81-85, July 2015. DOI: 10.1016/j.pestbp.2014.12.017.

WHO, World Health Organization. **Public Health Impact of Pesticides Used in Agriculture**. Geneva/Switzerland: WHO, 1990.

WHO, World Health Organization. Division of Mental Health. A users guide to the self-reporting questionnaire (SRQ-20). Geneva/Switzerland: WHO, 1994.

WOŁĘJKO, E.; JABŁOŃSKA-TRYPUĆ, A.; WYDRO, U. et al. Soil biological activity as an indicator of soil pollution with pesticides – A review. **Applied Soil Ecology**, Oklahoma, 147, 2020.

WITTEN, I. H.; FRANK, E. **Data Mining - Practical Machine Learning Tools and Techniques**. Elsevier, 2005.

WIVES, L. K. **Um Estudo sobre Agrupamento de Documentos Textuais em Processamento de Informações “não” estruturadas usando técnicas de Clustering**. Dissertação (Programa de Pós-Graduação em Computação) - Universidade do Rio Grande do Sul. Porto Alegre, 1999.

WU, C.; STEINBAUER, J. R.; KUO, G. M. Clustering analysis of diabetes patients basic diagnosis index. AMIA. Annual Symposium proceedings. **AMIA Symposium**, 1158, 2005.

WU, X. **Knowledge Acquisition from Databases**. USA: Ablex Publishing Corporation. 1995.

ZHANG, W.; JIANG, F.; OU, J. **Global pesticide consumption and pollution: With China as a focus**. Proceedings of the International Academy of Ecology and Environmental Sciences, London, 1, p. 125-144, 08/01, 2011.

ZHAO, Z.; ZHANG, L.; DENG, J.; WU, J. The potential effects of phytoplankton on the occurrence of organochlorine pesticides (OCPs) and polycyclic aromatic hydrocarbons (PAHs) in water from Lake Taihu, China. **Environmental Science**, London, v. 17, n. 6, p. 1150-1156, 2015. DOI: 10.1039/c5em00025d.

ZILLI, J. É.; BOTELHO, G. R.; NEVES, M. P.; RUMJANEK, N. G. Efeito de glyphosate e imazaquin na comunidade bacteriana do rizoplano de soja (*Glycine max* (L.) Merrill) e em características microbiológicas do solo. **Revista Brasileira de Ciência do Solo**, Viçosa – MG, v. 32, n. 2, p. 633-642, mar./abr. 2008. DOI: 10.1590/S0100-06832008000200018.

ZHOU, Z.-H. Three perspectives of data mining. **Artificial Intelligence Journal**, Australia, p. 139–146, 2003.

APÊNDICE A - Dados de elementos e características do solo do município de Santa Tereza do Oeste – PR

Categoria	Tipo	Campo	Descrição
Fertilidade	Numérico	Al	Alumínio
	Numérico	Ba	Bário
	Numérico	Ca	Cálcio
	Numérico	Cr	Cromo
	Numérico	C	Carbono
	Numérico	Cu	Cobre
	Numérico	Fe	Ferro
	Numérico	P	Fósforo
	Numérico	K	Potássio
	Numérico	Mg	Magnésio
	Numérico	Mn	Manganês
	Numérico	Na	Sódio
	Numérico	Ni	Níquel
	Numérico	Pb	Chumbo
	Numérico	Zn	Zinco
	Numérico	In_Alumin	Íon alumínio
	Numérico	Nit	Nitrogênio
Agrotóxicos	Numérico	Ampa	Subproduto AMPA ácido aminometilfosfônico
	Numérico	Glyphosate	Glyphosato
	Numérico	Organoclorados	organoclorados
Física	Numérico	Ph	pH
	Numérico	Som	Matéria orgânica no solo
	Numérico	tefetiva	Tefetiva
	Numérico	SB	Soma base
Granulometria	Numérico	Sand	Areia
	Numérico	Silt	Silte
	Numérico	Clay	Argila

APÊNDICE B - Dados de saúde da população da zona rural e urbana do município de Santa Tereza do Oeste - PR

Tipo	Descrição do Campo
num	Data de Nascimento
texto	Sexo
texto	Escolaridade
texto	Zona rural
texto	Número de pessoas na família
texto	Você tem dores de cabeça frequentes?
texto	Você tem falta de apetite?
texto	Você dorme mal?
texto	Você se assusta com facilidade?
texto	Você tem tremores musculares?
texto	Você se sente nervoso, tenso ou preocupado?
texto	Você tem má digestão?
texto	Você sente que suas ideias ficam embaralhadas de vez em quando?
texto	Você tem se sentido triste ultimamente?
texto	Você tem chorado mais do que de costume?
texto	Você tem dificuldade para tomar decisões?
texto	Você tem dificuldade para tomar decisões?
texto	Você acha que seu trabalho é penoso e lhe causa sofrimento?
texto	Você se acha incapaz de desempenhar um papel útil na sua vida?
texto	Você tem perdido o interesse pelas coisas?
texto	Você se sente uma pessoa sem valor?
texto	Alguma vez você pensou em acabar com sua vida?
texto	Você se sente cansado o tempo todo?
texto	Você sente alguma coisa desagradável no estômago?
texto	Você se cansa com facilidade?
texto	Teve contato com agrotóxicos?
texto	Você tem problemas ortopédicos?
texto	Você tem dores musculares?
texto	Você tem fadiga?
texto	Você tem doença respiratória?
texto	Você tem doença cardiovascular?
texto	Você tem alterações endócrinas?
texto	Você tem doença hepática?
texto	Você tem doença renal urinária?
texto	Você tem problemas neurológicos?
texto	Você tem problemas de pele?
texto	Tabagista
texto	Faz consumo de Álcool