



UNIVERSIDADE FEDERAL DO TOCANTINS
CAMPUS UNIVERSITÁRIO DE PALMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

MÔNICA COSTA BARROS

MINERAÇÃO DE DADOS EM SAÚDE:
PERSPECTIVAS PARA A VIGILÂNCIA EM SAÚDE DA POPULAÇÃO EXPOSTA AO
AGROTÓXICO.

PALMAS- TO
2021

MÔNICA COSTA BARROS

**MINERAÇÃO DE DADOS EM SAÚDE:
PERSPECTIVAS PARA A VIGILÂNCIA EM SAÚDE DA POPULAÇÃO EXPOSTA AO
AGROTÓXICO.**

Dissertação apresentada ao Mestrado Profissional em Ciências da Saúde da Universidade Federal do Tocantins para a obtenção do título de Mestre.

Profa. Orientadora: Dr^a Glêndara Aparecida de Souza
Martins

Prof. Co-orientador: Dr. Warley Gramacho da Silva

PALMAS - TO
2021

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

B277m Barros, Mônica Costa Barros.

Mineração de dados em saúde:: perspectivas para a vigilância em saúde da população exposta ao agrotóxico. . / Mônica Costa Barros Barros. – Palmas, TO, 2021.

77 f.

Dissertação (Mestrado Acadêmico) - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Pós-Graduação (Mestrado) em Ciências da Saúde, 2021.

Orientadora : Glêndara Aparecida de Souza Martins Martins

Coorientador: Warley Gramacho da Silva Silva

1. Mineração de dados. 2. Vigilância epidemiológica. 3. Praguicidas. 4. Envenenamento. I. Título

CDD 610

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).

FOLHA DE APROVAÇÃO


MÔNICA COSTA BARROS

MINERAÇÃO DE DADOS EM SAÚDE: PERSPECTIVAS PARA A VIGILÂNCIA EM SAÚDE DA POPULAÇÃO EXPOSTA AO AGROTÓXICO.

Dissertação apresentada ao Mestrado Profissional em Ciências da Saúde da Universidade Federal do Tocantins para a obtenção do título de Mestre.

Data de aprovação: 19 /10 /2021.

Banca examinadora:

Documento assinado digitalmente
 Glendara Aparecida de Souza Martins
Data: 21/10/2021 17:07:37-0300
Verifique em <https://verificador.itl.br>

Profa. Dra. Glêndara Aparecida de Souza Martins

Orientadora

Instituição: UNIVERSIDADE FEDERAL DO TOCANTINS – UFT

ARY HENRIQUE MORAIS DE OLIVEIRA:87889676115

Assinado de forma digital por ARY HENRIQUE MORAIS DE OLIVEIRA:87889676115
Dados: 2021.10.21 16:35:44 -03'00'

Prof. Dr. Ary Henrique Moraes de Oliveira

Examinador Externo

Instituição: UNIVERSIDADE FEDERAL DO TOCANTINS – UFT

Sandra M^a B. Mariano

Profa. Dra. Sandra Maria Botelho Mariano

Examinadora Interna

Instituição: UNIVERSIDADE FEDERAL DO TOCANTINS - UFT

Dedico essa dissertação à minha família que me ensinou os valores essenciais à vida, e que me movem no sentido de cumprir a minha missão designada por Deus, o de ser sal e luz ao próximo. Dedico, também, aos trabalhadores e trabalhadoras do campo, vítimas das consequências nefastas da exposição por agrotóxicos pelo modelo hegemônico produtivo químico-dependente.

AGRADECIMENTOS

Esta produção representa a capacidade de resiliência e perseverança no processo de construção de conhecimento e aprendizado, mesmo se apresentando como um desafio transloucado e utópico. Primeiro pela minha formação parecer tão divergente ao tema- uma fisioterapeuta pesquisando intoxicação por agrotóxico? – e o segundo, pela dificuldade de integrar e manipular a grande quantidade de dados em diferentes fontes.

Portanto, primeiramente, agradeço à DEUS pelos dons dados neste caminho. À minha família que são minha base de fé e pela compreensão nos momentos de surto. Aos meus filhos, Davi e Mariana que me motivam a buscar o meu melhor a fim de construir um serviço público de qualidade.

Homenagem e gratidão aos grandes mestres da Fiocruz, que desde a especialização me incentivaram a ‘sair da caixinha’ na vigilância e promoção da saúde em ambiente e trabalho. Quebrando tabus acerca da autonomia do meu processo de formação, enquanto trabalhadora do SUS e cidadã, possibilitando o enfrentamento do desafio posto nesta jornada acadêmica.

Agradeço a todos membros do Fórum Tocantinense de Combate aos impactos dos agrotóxicos pela confiança no meu trabalho, pelas colaborações e discussões que alimentaram a curiosidade e hipóteses enquanto pesquisadora, e assim, me impulsionou à pesquisa. Da mesma forma, agradeço aos servidores da Secretaria de Estado da Saúde do Tocantins que diretamente e indiretamente colaboraram na pesquisa, em especial à Magna Dias Leite.

À minha orientadora Glêndara A. Souza Martins e ao coorientador Warley Gramacho, gratidão e admiração pelos conhecimentos, apoio e exemplo dados. E ao Pedro que aceitou o desafio de compor esta equipe de pesquisa.

A todos colegas e professores do programa de Mestrado em Ciências da Saúde da Universidade Federal do Tocantins pelos conhecimentos e experiências compartilhados, além do trabalho em equipe em prol de ampliar a produção científica voltada para o SUS local.

Por fim, agradeço ao conjunto de organizações, movimentos sociais e tantos outros espaços e sujeitos (individuais e coletivos) que, apesar de não nomeados aqui, certamente foram significativos, tanto para a elaboração deste trabalho, quanto para a certeza de seguir na luta permanente contra os agrotóxicos e pela vida.

"Há um tempo em que é preciso abandonar as roupas usadas, que já tem a forma do nosso corpo, e esquecer os nossos caminhos que nos levam sempre aos mesmos lugares. É o tempo da travessia: se não ousarmos fazê-la, teremos ficado, para sempre, à margem de nós mesmos." (Fernando Pessoa)

RESUMO

O extensivo uso de agrotóxicos na agropecuária brasileira culmina em uma maior exposição da população a essas substâncias em maior ou menor grau, por diversas rotas e vias de exposição. Por isso, o Ministério da Saúde, no ano de 2012, com a Portaria GM/MS nº 2.938, implantou a Vigilância das populações expostas aos agrotóxicos (VSPEA). Contudo, a detecção da intoxicação crônica e o nexos com outras doenças a partir dos sistemas de informação em saúde é um desafio para o sistema de Vigilância em Saúde em seus princípios básicos de prevenção, controle e promoção da saúde. O objetivo geral da pesquisa é discutir os fatores associados à intoxicação por agrotóxico com base na mineração de dados aplicada aos bancos de dados dos sistemas de informação em saúde. A partir da aplicação de tratamentos estatísticos e de mineração de dados aos bancos do SINAN, SIM e SINASC para extrair associações importantes que explicassem os impactos da intoxicação por agrotóxico. Os achados da pesquisa consistem no levantamento de dados para a compreensão do contexto resultando na construção de um data-warehouse voltado para tomada de decisão e planejamento da VSPEA, na modelagem de *datasets* a partir dos sistemas de informação em saúde, possibilitando a modelagem preditiva de três matrizes de predição capazes de reconhecer padrões dos fatores associados para o tipo da exposição, classificação da intoxicação e evolução das intoxicações por agrotóxicos. Revela as potencialidades da interdisciplinaridade com a Ciência de Dados para a vigilância em saúde, como também levanta questões para pesquisas futuras e reflexões do processo de vigilância epidemiológica no tema.

Descritores: Mineração de dados, Vigilância epidemiológica, Praguicidas, Envenenamento.

ABSTRACT

The extensive use of pesticides in Brazilian agriculture culminates in a greater exposure of the population to these substances to a greater or lesser extent, through different routes and routes of exposure. Therefore, the Ministry of Health, in 2012, with Ordinance GM/MS No. 2,938, implemented the Surveillance of populations exposed to pesticides (VSPEA). However, the detection of chronic intoxication and the link with other diseases from health information systems is a challenge for the Health Surveillance system in its basic principles of prevention, control and health promotion. The general objective of the research is to discuss the factors associated with pesticide poisoning based on data mining applied to databases of health information systems. From the application of statistical treatments and data mining to SINAN, SIM and SINASC databases to extract important associations that could explain the impacts of pesticide poisoning. The research findings consist of data collection to understand the context resulting in the construction of a data-warehouse focused on decision-making and planning of the VSPEA, in the modeling of datasets from health information systems, enabling the predictive modeling of three prediction matrices capable of recognizing patterns of associated factors for the type of exposure, classification of poisoning and evolution of pesticide poisoning. It reveals the potential of interdisciplinarity with Data Science for health surveillance, as well as raising questions for future research and reflections on the process of epidemiological surveillance on the subject.

Keywords: Pesticides, Poisoning, Data Mining, Epidemiological Surveillance.

LISTA DE ILUSTRAÇÕES

Figura 1. Modelo de processos SEMMA (2016)	26
Figura 2. Modelo de processos KDD-PH.....	26
Figura 3. Escolha de variáveis para análise de mineração de dados entre os sistemas de informação, construção do <i>dataset Data-VSPEA</i>	37
Gráfico 1. Distribuição dos casos por ano e sexo de intoxicação por agrotóxicos, Tocantins, 2008 a 2019.....	46
Gráfico 2. Classificação final do diagnóstico por agente tóxico, residentes do Tocantins, 2008 a 2019.....	52
Figura 4. Apontamentos de uso da mineração de dados na saúde.....	57

LISTA DE QUADROS E TABELAS

Quadro 1. Informações relevantes para formulação, monitoramento e avaliação da vigilância VSPEA. Construção do data warehouse-VSPEA.	34
Tabela 1. Análise de completude dos atributos nos <i>datasets</i> coletados dos sistemas de informação SINAN, SIM e SINASC.....	38
Tabela 2. Representação do resultado da análise exploratória dos atributos no banco de dados de intoxicação exógena por agrotóxicos. Tocantins, de 2008 a 2019.....	40
Tabela 3. Análise de completude e confiabilidade da identificação de eventos repetidos por usuário associados à exposição aos agrotóxicos. SINAN, de 2008 a 2019.....	41
Tabela 4 Resultados estatísticos da associação dos fatores individuais do óbito por intoxicação por agrotóxico, Tocantins, 2008 a 2020	43
Tabela 5. Tendências positivas negativas para mulheres que apresentaram abortos e filhos com anomalia congênita. Tocantins, de 2008 a 2020.....	44
Tabela 6. Tendência negativas para mulheres que apresentam aborto e filhos com anomalia. Tocantins, de 2008 a 2020.....	44
Quadro 2. Configuração dos <i>datasets</i> para o treino do modelo preditivo.....	47
Tabela 7. Fatores com maior prevalência no banco de dados de intoxicação por agrotóxico, Tocantins, 2008 a 2019.....	48
Tabela 8. Ranking da ocupação por zona de exposição (10 ^o posições)	51
Tabela 9. Distribuição dos casos por agente tóxico e classificação final, Tocantins, 2008 a 2019.....	53
Tabela 10. Matriz de confusão do algoritmo k-NN no <i>dataset</i> <i>classi_fin</i>	55
Tabela 11. Matriz de confusão do algoritmo k-NN no <i>dataset</i> <i>evolução</i>	56
Tabela 12. Matriz de confusão do algoritmo k-NN no <i>dataset</i> <i>exposição</i>	56

LISTA DE SIGLAS

AAE	Avaliação Ambiental Estratégica
AIA	Avaliação de Impacto Ambiental
AIS	Avaliação de impacto à saúde
ANVISA	Agência Nacional de Vigilância Sanitária
BI	Business Intelligence
CAR	Cadastro Ambiental Rural
CAT-WEB	Sistema de Comunicação de Acidente de Trabalho Previdenciário
CEP	Comitê de Ética em Pesquisa
CID-10	10ª Revisão da Classificação Internacional de Doenças
CRISP-DM	Cross-Industry Standard Process for Data Mining
DAP	Dialquil-fosfatos
DN	Declaração de Nascido Vivo
DNA	Ácido Desoxirribonucleico
DO	Declaração de Óbito
DTA	Doenças Transmissíveis por Alimentos
DATASUS	Departamento de Informática do SUS
DRT	Doença relacionada ao trabalho
EDI	Ingestão Diária Estimada
e- SUS	Sistema da Atenção primária
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
GAL	Sistema Gerenciador de Ambiente Laboratorial
HI	Índice de Perigo
HQ	Quocientes de Perigo
IBAMA	Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IES	Instituições de Ensino Superior
KDD	Knowledge Discovery in Databases
KDD-PH	Knowledge Discovery in Databases for Public Health
kNN	K Nearest Neighbor (algoritmo)
MPE	Ministério Público Estadual
MPF	Ministério Público Federal
MPT	Ministério Público do Trabalho
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
OMS	Organização Mundial da Saúde
PAIR	Perda Auditiva Induzida por Ruído
PCDaS	Plataforma de Ciência de Dados aplicada à Saúde
PMA	Produção Municipal Agrícola
RNA	Ácido ribonucleico
RCBP	Registro de Câncer de base populacional
RHC	Registro de Câncer hospitalar
SEFAZ	Secretaria de Estado da Fazenda
SEMMA	Sample, Explore, Modify, Model, Assess
SESMT	Serviços Especializados em Engenharia de Segurança e em Medicina do Trabalho
SIA	Sistema de Informação Ambulatorial
SIH	Sistema de Informação Hospitalar

SIM	Sistema de Informação de Mortalidade
SINAN	Sistema Nacional de Agravos de Notificação Compulsória
SINASC	Sistema de Informação Nascidos Vivos
SISAGUA	Sistema do Programa Nacional de Vigilância da Qualidade da Água para Consumo Humano
SISSOLO	Sistema de Informação de Vigilância em Saúde de Populações Expostas a Solo Contaminado
SUS	Serviço Único de Saúde
USEPA	Agência de Proteção Ambiental dos Estados Unidos
VMP/ LMR	Valor máximo permitido
VSPEA	Vigilância em saúde da população exposta ao agrotóxico

SUMÁRIO

RESUMO	08
1. INTRODUÇÃO	145
1.1 Objetivos	17
2 REFERENCIAL TEÓRICO	18
2.1 A compreensão do problema sob o olhar interdisciplinar da Vigilância em Saúde	18
2.2 A Ciência de dados e suas possíveis contribuições para o problema	23
3. METODOLOGIA	29
3.1 Tipo de estudo	29
3.2 Fonte de dados	29
3.3 Análise de dados	30
3.4 Riscos e benefícios e os aspectos éticos legais	31
4. RESULTADOS E DISCUSSÃO	32
4.1 Revisão teórico- metodológica norteadora para o levantamento de dados	32
4.2 Análise dos sistemas de informação em saúde	38
4.3 Preparação dos dados para os experimentos de modelagem	46
4.4 Resultados da Mineração	48
5. CONSIDERAÇÕES FINAIS	58
REFERÊNCIAS	59
APENDICE A	66

1 INTRODUÇÃO

A pesquisa dos impactos ambientais e na saúde humana do uso de agrotóxicos no desenvolvimento da produção agrícola tem uma importância mundial desde as descobertas de Rachel Carson, publicadas no livro “Silent Spring” (1962), que os denominava como biocidas, uma vez que, ao serem utilizados, extingue toda forma de vida. Para além das culturas agrícolas, também foram adotados no controle dos vetores de doenças como dengue e leishmaniose, denominados como agrotóxicos de uso na saúde pública.

No Brasil, o coeficiente de mortalidade por agrotóxico é 1,52 por 100.000 habitantes, representando 136.665 Anos Potenciais de Vida Perdido (BOCHNER, FREIRE, 2020). Por estatística houve associação positiva e significativa entre o respectivo consumo (pulverização) de agrotóxicos com as incidências médias de intoxicação por agrotóxicos agrícolas entre todas as unidades federativas e respectivas regiões mais produtoras agrícolas dentro de cada estado brasileiro (PIGNATI, 2018; DA CRUZ FERNANDES *et al*, 2019).

Segundo estimativas da Organização Mundial da Saúde (OMS), anualmente, de 1,5% a 3,0% da população mundial é acometida por intoxicação exógena. No Brasil, segundo o último boletim epidemiológico divulgado, obteve um total de 42.309 notificações para intoxicação exógena, destas 17.299 (40,9%) se referiam a agrotóxicos diversos, dentre inseticidas (43,8%) e herbicidas (36,2%) relacionados ao trabalho. Destacando os trabalhadores agropecuários, das florestas e da pesca (9.697; 37,2%) como os mais frequentes (BRASIL, 2018). Outro ponto evidenciado pelo Ministério da Saúde no seu Boletim Epidemiológico nº 15/2019 destaca os suicídios por intoxicação exógena, apontando o percentual de homens que usaram raticida (17,5%) e agrotóxicos (14,1%) e dentre as mulheres registraram 9,4% e 5,1%, respectivamente. Em estudo sob a análise das intoxicações por agrotóxicos agrícolas do período de 2012 a 2015, aponta o Tocantins ocupando o segundo lugar do ranking com a incidência de 77 casos por 1 milhão de habitantes por ano, assim como ocupa a quarta posição nos indicadores ambientais de exposição ambiental com 11,8 L/hab. (LARA *et al*, 2019).

No modelo de desenvolvimento de monoculturas possui uma relação ‘químico-dependente’ que gera impactos como poluição de matrizes ambientais como ar, água e solo, transformando-as em vias de contaminação e aumentando a extensão de possíveis riscos sobre a saúde ambiental, em local distante dos centros de utilização dessas substâncias (KIM, KABIR, JAHAN, 2017; SIDHU *et al*, 2019). Por isso, a exposição humana a agrotóxicos representa um importante problema de saúde pública, para o qual o setor de saúde tem buscado definir e

implementar ações voltadas à vigilância em saúde. Contudo, as intervenções devidas seu caráter interinstitucional é complexo e desafiador na implantação por transcender o setor saúde (PORTO, SOARES, 2012; MINISTÉRIO SAÚDE, 2018).

A avaliação do potencial carcinogênico do agrotóxico é bastante complexa devido à diversidade de produtos, mudanças no modo de utilização do produto, longo período de exposição para a detecção da cronicidade, além das dificuldades de transparência de dados sobre esta exposição e a falta do programa de biomonitoramento em nosso país, como política de Estado. Apesar disso, a exposição a agrotóxicos já foi associada ao surgimento de tumores na infância, especialmente a leucemia (CURVO, PIGNATTI, PIGNATTI, 2013; PIGNATTI *et al*, 2014; PIGNATTI *et al*, 2017).

Nesse contexto, a insuficiência de dados sobre o consumo de agrotóxicos, seus tipos e volumes, utilizado nos municípios brasileiros, o desconhecimento do seu potencial tóxico, a carência de diagnósticos laboratoriais e a pressão/assédio de fazendeiros do agronegócio que ocupam cargos públicos, favorecem o ocultamento e a invisibilidade desse importante problema de saúde pública (PIGNATI *et al.*, 2017).

No Brasil, estudos epidemiológicos são fortemente dependentes de dados sobre as populações analisadas. Contudo, estes dados percorrem um longo caminho desde a ocorrência real de uma intoxicação, passando pelo sistema de saúde pública, preenchimento de fichas, entrada nos sistemas e posterior recuperação por parte dos pesquisadores em sistemas fragmentados e desatualizados, o que dificulta onexo causal e o monitoramento das intoxicações crônicas, já que o sistema SINAN permite a visualização somente da intoxicação aguda, ou seja, casos novos.

Desta maneira, pesquisas que se aprofundam nos modelos teóricos-metodológicos da Epidemiologia crítica e da Ciência de Dados - dada a última ser um campo disciplinar que utiliza métodos científicos, processos, algoritmos e sistemas para extrair conhecimentos de dados, tanto estruturados como não-estruturados (DHAR, 2013) - no intuito de coletar e experimentar as potencialidades interdisciplinares que a mesmas possam contribuir com o processo de vigilância em saúde no tratamento de dados e levantamento de métricas norteadoras para avaliação do impacto ao ambiente e saúde humana à exposição dos agrotóxicos, dada complexidade da atuação da Vigilância em Saúde, frente a dificuldade de processos e investimentos para implementar tecnologias para o biomonitoramento.

Frente aos problemas e hipóteses acima levantadas para análise de situação de saúde para o enfrentamento e promoção de ações que protejam a população dos riscos à exposição de produtos mutagênicos e carcinogênicos foi traçado a este estudo os objetivos abaixo.

1.1 Objetivos

1.1.1 Objetivo geral

Discutir os fatores associados à intoxicação por agrotóxico com base na mineração de dados aplicada aos bancos de dados dos sistemas de informação em saúde.

1.1.2 Objetivos específicos

1. Identificar os sistemas de informação em saúde de importância para a avaliação dos fatores associados à da intoxicação por agrotóxicos a partir da revisão teórico metodológica;
2. Analisar as variáveis e atributos dos bancos de dados dos sistemas de informação de relevância e descrever suas possíveis relações para compreender o problema;
3. Analisar o banco de dados de Intoxicação exógena, através da mineração de dados a fim de aprofundar a descoberta de conhecimentos de seus atributos;
4. Analisar a duplicidade de agravos e doenças dos casos notificados por intoxicação exógena por meio do relacionamento probabilístico entre os datasets do SINAN;
5. Analisar os *datasets* dos dados levantados dos diversos sistemas de informação em saúde para conhecer os padrões e *inputs* da modelagem preditiva para a vigilância da população exposta ao agrotóxico.

Neste propósito, o trabalho consiste na apresentação do referencial teórico da vigilância da população exposta ao agrotóxico com objetivo de compreender o contexto do problema, assim como também da Ciência de dados para discussão da prospecção tecnológica no processo de trabalho da vigilância epidemiológica. Da metodologia empregada na pesquisa e na análise de dados em saúde. Em seguida, os resultados da busca teórico-metodológica para a construção da base de dados (*data-VSPEA*), da análise dos sistemas de informação e fontes de dados, e da experimentação de métodos de mineração de dados na mesma. Por fim, nossas considerações finais acerca da pesquisa e perspectivas futuras.

2 REFERENCIAL TEÓRICO

Neste capítulo apresentamos todo arcabouço teórico utilizado para nortear a pesquisa em sua complexidade disciplinar para compreender o problema acerca da vigilância da população exposta ao agrotóxico e objeto do estudo estabelecido na análise de situação de saúde da mesma.

2.1 A compreensão do problema sob o olhar interdisciplinar da Vigilância em Saúde.

Intoxicação exógena pode ser definida como um conjunto de efeitos nocivos ao organismo produzidos pela interação de um ou mais agentes tóxicos com o sistema biológico, representados por manifestações clínicas ou laboratoriais que revelam desequilíbrio orgânico. Os agrotóxicos, gases tóxicos e metais pesados são exemplos de agentes químicos que podem causar esse tipo de intoxicação.

A vigilância epidemiológica desse agravo é monitorada pelo Sistema Nacional de Agravos de Notificação (SINAN) compulsória, estabelecida pela Portaria de Consolidação nº 4/ 2017, na ficha de Intoxicação exógena. Desta forma, sistemas de vigilância nas secretarias de saúde, nos três níveis federativos, monitoram e controlam casos agudos de intoxicação por agrotóxico, ou seja, novos casos, e seu prazo de encerramento é de 6 meses (BRASIL, 2019). Portanto, não permitem acompanhamento prolongado de exposição crônica.

A Vigilância em Saúde é um serviço de saúde constituído no Serviço Único de Saúde (SUS), a partir da Lei 8080/1988, que compreende um conjunto de ações integradas de proteção e promoção da saúde, vigilância, prevenção e controle das doenças e agravos à saúde, abrangendo a vigilância epidemiológica, vigilância sanitária, vigilância em saúde ambiental, vigilância da saúde do trabalhador, vigilância laboratorial, promoção da saúde e vigilância da situação de saúde (MINISTÉRIO DA SAÚDE, 2019). Com o objetivo de prevenir, controlar e mitigar riscos à saúde da população, em que sua “máxima” para atuação é: “Informação para ação, fundamenta-se no campo da Saúde Coletiva e da Epidemiologia, como ciências norteadoras de suas práticas.

O uso de agrotóxicos e as práticas de aplicação não são uniformes, dependendo do tipo de cultura, condições climáticas, pragas e outros fatores sociais. O armazenamento de pesticidas e o descarte de embalagens vazias de pesticidas também promove a exposição humana e ambiental. Devido ao manuseio incorreto e práticas inadequadas, podem ocorrer consequências perigosas para a saúde dos agricultores, contaminação da cadeia alimentar e acúmulo de

resíduos de agrotóxicos no corpo do consumidor (SILVA, XAVIER, CEZAR-VAZ, 2019; LOPES, ALBUQUERQUE, 2018). Compreender e identificar o modo crítico e determinante da exposição a agrotóxicos é de suma importância no desenvolvimento de estratégia de longo prazo e uso sustentável de pesticidas entre os agricultores nos esforços para proteger sua saúde, a saúde dos consumidores e o meio ambiente em geral (BORTOLOTTI, et al., 2020; MINISTÉRIO SAÚDE, 2018; VASCONCELLOS, et al., 2019).

O uso de agrotóxicos e suas implicações para a saúde pública em termos de segurança química e alimentar, pode ser considerado um problema complexo que demanda múltiplos enfoques e disciplinas. Desta forma, os agrotóxicos podem ser analisados de forma complexa e sistêmica a partir de uma avaliação integrada (articulação, interpretação e comunicação), de tal modo que sua cadeia global de causa-efeito possa ser avaliada a partir de uma perspectiva holística ou sinóptica. Tal perspectiva permite prover informação útil aos que tomam decisões – incluindo cidadãos e movimentos sociais – facilitando, desta forma, a definição e a implementação de políticas e estratégias para os desafios de se enfrentar os impactos dos agrotóxicos.

Torna-se necessária, além do aprofundamento de estudos epidemiológicos, toxicológicos e nutricionais, a integração de diferentes disciplinas e perspectivas articuladas à saúde coletiva, como a economia ecológica, as ciências sociais e ambientais para transformar o modelo de desenvolvimento. É preciso tornar visíveis os efeitos e os custos socioambientais e à saúde pública associados a esta lógica de produção e comércio internacional de alimentos. A visibilidade de tais impactos é estratégica para a viabilização de políticas públicas que fomentem o comércio justo (*fair trade*), a reforma agrária, a agroecologia e o consumo consciente e saudável de alimentos dentro dos princípios da sustentabilidade ambiental e da justiça ambiental, como para a regulação dos mesmos.

Visto que no Decreto nº 4.074/2002, não há uma previsão legal de periodicidade mínima para a revisão dos registros desses produtos e ocorrem, nos termos do inciso VI, art. 2, orientada por alertas internacionais, novos estudos científicos ou denúncias realizadas por instituições de referência (FRIEDRICH et al, 2021). Torna evidente a responsabilidade da Vigilância em Saúde no seu papel de monitorar, analisar e produzir informes e alertas quanto aos efeitos dos agrotóxicos a fim de fomentar e fundamentar o processo regulatório.

De acordo com Santos (2007) refere-se à ecologia dos saberes como a infinita pluralidade dos saberes e da necessidade de valorização dos mesmos, ou seja, reconhece que todo saber possui uma dosagem de ignorância e toda ignorância possui uma dosagem de

saber, necessitando que haja diálogo entre os diversos saberes e ciências para realização de ações que sejam de fato emancipatórias. “A partir do momento em que a comunidade científica perceber a importância da reflexividade sobre suas práticas, o conhecimento produzido não será mais insensível” (SANTOS, 1989, p. 118). Será um conhecimento esclarecedor, mais formativo do que informativo, criador e não destruidor da competência social dos não cientistas, interessado na transformação do que existe e respeitando a dialética histórica presente nas relações entre sujeito e objeto, ou melhor, “sujeito” (BREILH, 2008; CARNEIRO, 2015).

Pautado, também, em seu estudo das Sociologias com os aportes da Teoria da Complexidade, da Epidemiologia Crítica, da Hermenêutica de Profundidade e dos conceitos de território e territorialidades para compreender o macro fenômeno da modernização agrícola em suas inter-relações com a saúde, o trabalho, o ambiente e a vulnerabilidade das populações atingidas pelos processos de “desenvolvimento” no campo (BREILH, 2008).

Revisões de literatura constataam a presença de compostos de agrotóxicos originais no sangue humano, urina, leite materno, sêmen, tecido adiposo, líquido amniótico, mecônio infantil e sangue do cordão umbilical. E a bioacumulação dessas substâncias químicas no corpo ao longo do tempo foi associada a defeitos congênitos, tumores, alterações genéticas, distúrbios do sangue e dos nervos, dificuldade de concepção e até coma ou morte. Portanto, denotando o alto potencial genotóxico dessas substâncias, que possuem um efeito destrutivo no material genético de uma célula (cromossomo, DNA ou RNA) afetando sua integridade (NEVES, 2017; ALI, ISMAIL, *et al.*, 2018; DUJARDIN, BOCCA, 2019; SILVERIO, PINHEIRO, 2019). Representando alto perigo à saúde e biodiversidade e deve, de acordo com a nossa Constituição, ser monitorado e fiscalizado a fim de garantir os direitos fundamentais, ou seja, visa garantir ao ser humano, o respeito à vida, à liberdade, à igualdade e a dignidade, para o pleno desenvolvimento de sua personalidade por meio das Políticas Públicas.

Assim, sabendo que a associação dos fatores genéticos e ambientais pode desencadear doenças graves e crônicas, como Doença de Parkinson, efeitos adversos na gravidez, autismo, entre outras. Em que o aparecimento dos sintomas ocorre tardiamente sem uma origem específica e determinada, é de extrema importância realizar estudos para avaliar os fatores ambientais, em especial na população rural (DA SILVA, XAVIER, *et al.*, 2019; VASCONCELLOS, *et al.*, 2019; MUSTIELES, *et al.*, 2020).

Diversos estudos vêm se propondo no desenvolvimento de metodologias analíticas aprimoradas que aumentam os desempenhos de sensibilidade e robustez e reduzem o tempo de

análise (HAKME, *et al.*, 2018; NARENDERAN, MEYYANATHAN, BABU, 2020). Diante dessa preocupação, recentemente tem se iniciado na tentativa de detectar os múltiplos resíduos da diversidade de pesticidas utilizados e seus compostos hormonais que impactam na saúde humana: por métodos de cromatografia líquida de alta eficiência, visto sua capacidade de separar uma grande quantidade de compostos em diversos tipos de amostra no espaço de tempo de alguns minutos, exibindo alta resolução e detectabilidade a custo baixo (RELANGI, 2019); utilização de nanotecnologias e nanomateriais proporcionando ferramentas poderosas de detecção miniaturizadas para aplicação de campo proporcionando rapidez no rastreamento de resíduos (HU, *et al.*, 2020; RODRIGUEZ-SAONA, AYKAS, *et al.*, 2020). Porém, a escolha do sistema de extração e detecção permanece a mesma, que consiste no sistema tradicional de extração e detecção acoplado à massa para quantificação, sendo demorados e caros (NARENDERAN, MEYYANATHAN, BABU, 2020).

Para além da comprovação de resíduos nas diversas matrizes da rota de exposição, já mencionadas, outra preocupação é o risco cumulativo da exposição a pesticidas por dieta. Apesar de estudos disponíveis na literatura apontarem baixo risco para o limite de ingestão diária, a preocupação pública permanece. E relativamente, há incertezas nos estudos de modelagem e métodos de avaliação de risco cumulativo, além de limitações de acesso a dados de consumo e níveis de resíduos nos diferentes níveis de processamento dos alimentos (JARDIM *et al.*, 2018; LARSSON, *et al.*, 2018; LI *et al.*, 2018; MAGGIONI, *et al.*, 2018; SIEKE, 2018; NOUGADÈRE, *et al.*, 2020; BIRGHILĂ, *et al.*, 2020).

Outro ponto observado pelos estudos sobre o risco da dieta, está na incerteza do acúmulo e no nível de ingestão diária, já que o consumo é pessoal e individual, no entanto as mesmas comparam os níveis encontrados nos diversos tipos de alimentos e compara por meio de métodos estatísticos com os riscos relevantes por meio de Ingestão Diária Estimada (EDI), Quocientes de Perigo (HQ) e Índice de Perigo (HI) para a população, reguladas pela Organização Mundial da Saúde (OMS) e Agência de Proteção Ambiental dos Estados Unidos (USEPA) (ELGUETA, *et al.*, 2020; LU, *et al.*, 2020; WUMBEI, 2019).

Contudo, todo esse sistema enfrenta desafios importantes para sua efetividade e eficiência, desde a falta de recursos humanos qualificados na assistência, quanto a recursos de diagnóstico, acesso a dados e sistemas de informação integrados, a falta de abordagem estratégica e multidisciplinar (BRASIL, 2018; MINISTÉRIO SAÚDE, 2018; BRASIL, 2019; JARDIM *et al.*, 2018; PIGNATI, MACHADO, CABRAL, 2007), além da permissividade do valor máximo permitido (VMP/ LMR), 5000 vezes maiores que o estabelecido pelas portarias

de regulação do controle da água e dos alimentos, comparados aos valores praticados na União Europeia. Portanto, outras medidas para o consumo de pesticidas precisam ser avaliadas para garantir a saúde humana, a segurança alimentar e a proteção ambiental (BOMBARDI, 2017; O. GOMES, *et al.*, 2020).

Agrotóxicos e seus metabólitos podem ser medidos avaliando biomarcadores específicos. Diferentes biomarcadores podem ser usados isoladamente e / ou em combinação para estabelecer os níveis críticos de exposição. A depressão da colinesterase sanguínea é usada como biomarcador de exposição à toxicidade do organofosforado e de carbamato. No entanto, um método mais eficaz e quantitativo para avaliar a exposição a pesticidas é quantificar os níveis de dialquil-fosfatos urinários (DAP). Os níveis urinários e sanguíneos de DAPs são frequentemente considerados biomarcadores de exposição a pesticidas organofosforados parentais. Sua medição fornece informações sobre a exposição cumulativa a uma grande quantidade de pesticidas organofosforados (BENFORD, *et al.*, 2018; KAPELEKA, SAULI, NDAKIDEMI, 2019; MORE, *et al.*, 2019; RELANGI, 2019; SIDHU, *et al.*, 2019; SILVERIO, PINHEIRO, 2019; VASCONCELLOS, *et al.*, 2019).

A bioacumulação de agrotóxicos envolve o acúmulo desses produtos químicos tóxicos como resultado da absorção contínua de produtos químicos no corpo a uma taxa na qual o corpo não consegue metabolizar e excretar. Isso é alimentado pelo uso persistente de pesticidas que não se dividem em partes constituintes mais seguras, mas permanecem intactos por períodos prolongados de tempo, tornando-se facilmente acessíveis ao corpo humano. Muitas abordagens e técnicas de biomonitoramento humano foram desenvolvidas e aplicadas na detecção precoce de efeitos genotóxicos da exposição de longo prazo a pesticidas perigosos. No entanto, essas técnicas, embora relatadas como simples e rápidas, são complicadas e dificilmente aplicadas na detecção precoce de efeitos genotóxicos de agrotóxicos, principalmente entre agricultores que trabalham em ambiente descontrolado. Isso pode ser responsável pela falta geral de informações de seus efeitos prejudiciais, daí a necessidade de esforços conjuntos para desenvolver testes rápidos, técnicas simples e economicamente viáveis e abordadas na detecção dos danos ao DNA em nível de célula (LOPES, ALBUQUERQUE, 2018; SILVA, XAVIER, CEZAR-VAZ, 2019; SILVERIO, PINHEIRO, 2019).

2.2 A Ciência de dados e suas possíveis contribuições para o problema

Tradicionalmente, na Epidemiologia, a estatística é a ciência base referencial quanto ao método de coleta, organização e análise dos dados. Contudo termos como ciência de dados e big data vem se destacando nas mídias e meio acadêmico como possibilidade de vencer os desafios encontrados na análise de grande volume de dados. AALST (2016) afirma que o campo de ciência de dados se origina da estatística, que também deu origem ou impulsionou vários outros campos.

A partir da definição da ciência de dados como um campo interdisciplinar para o estudo de dados estruturados ou não estruturados (pequenos ou grande volume), e terem outras propriedades, como grande velocidade de produção e variedade de conteúdo, podendo ainda serem estáticos ou alimentados em tempo real. E que suas atividades incluem: a extração de dados, sua preparação, exploração, transformação, armazenamento, recuperação, manutenção das infraestruturas computacionais necessárias, aplicação de diversas estratégias de mineração de informação e aplicação de métodos de aprendizado de máquina, tal como o emprego de recursos de apresentação dos resultados e previsões (HAYASHI, 1998; PRESS, 2013; KHOURY; IOANNIDIS, 2014; AALST, 2016; HOUSEH; KUSHNIRUK; BORYCKI, 2019 apud SALDANHA, 2021). Tão certo, as duas ciências se aproximam e parecem ser sobrepostas. Contudo vários métodos de ciência de dados são inferenciais e dedutivos, partindo da amostra para a população, mas a ciência de dados também utiliza métodos indutivos, enquanto que a estatística, do ponto de vista mais tradicional, define-se pela inferência da parte para o todo (SALDANHA, 2021).

Na Vigilância epidemiológica, a partir de suas características e bases teóricas, utiliza-se a estatística por sua abordagem de inferência causal das doenças ou eventos em saúde pública em que se busca testar a hipótese de causalidade. A análise de dados se faz a partir de base dados limitados, em geral estáticos, limpos e de natureza pouco relacional, amostrados cientificamente e seguindo suposições claras, como independência, estacionariedade e normalidade (MILLER, 2010; SHMUELI, 2010).

Enquanto, “Ciência de Dados é um campo de estudo que se destaca pela capacidade de auxiliar a descoberta de informação útil a partir de grandes ou complexas bases de dados, bem como a tomada de decisão orientada por dados” (PCDaS, 2021). Aqui o desafio de análise é lidar com a abundância, exaustividade e variedade, sua dinâmica, desordem e incerteza, e a necessidade de lidar com dados que não foram gerados para responder uma questão específica (KITCHIN, 2014).

Estudos apontam para as potencialidades da incorporação da ciência de dados como ganho para a Saúde Pública no enfrentamento do aumento de sua complexidade, sendo capaz de aproveitar a oferta de dados e a diversidade de métodos disponíveis para a obtenção de novas fontes de dados, que requererem novos métodos de estudo e abordagens diferenciadas, enfrentadas por equipes transdisciplinares (DICLEMENTE et al., 2019; SALDANHA, 2021).

Salathé et al. (2012), aponta que a integração de diversas fontes de dados, em conjunto com dados já sistematizados, apresenta-se como uma oportunidade única para inovação da vigilância epidemiológica, preenchendo hiatos da infraestrutura de Saúde Pública e operando sem fronteiras e tempo real.

Saldanha (2021) apresenta em seu estudo as diferentes fases da epidemiologia, pontuando as inovações e transformações nos procedimentos analítico e preventivo e da apresentação dos dados. Pontuando, o momento atual com o paradigma do *big data*, que se apresenta com universo de dados estruturados em sistemas e não estruturados contidos em prontuários ou mesmo da internet, do qual o processo analítico deve avançar para a ciência de dados para atingir uma medicina de precisão e predição de cenários.

No contexto de aplicação dos modelos e processos de mineração de dados na saúde pública, a etapa de aquisição compreende a coleta de informações secundárias junto a repositórios de dados públicos ou privados, ou mesmo primários. Da etapa de armazenamento e recuperação de dados envolvem tecnologias específicas que necessitam considerar grandes massas de dados sem comprometer a velocidade das operações de leitura e apresentação dos arquivos. A etapa de mineração de dados envolve processos para a detecção de padrões, associações e tendências temporais e espaciais, envolvendo em geral algoritmos autônomos e autorregulados para a realização destas tarefas. A última etapa, referente à análise e visualização de dados, visa a extração e apresentação de informações permitindo ao utilizador a descoberta de novas relações (SALDANHA, 2021).

Neste passo, a utilização de tecnologias de ciência de dados ganha significado prático ao usuário, permitindo informar aos gestores sobre diferentes possibilidades de políticas públicas e na apresentação das informações para o público em geral (FUNG; TSE; FU, 2015), através de um grande e amorfo conjunto de dados que se transforma de maneira interativa, atribuindo significado (HUANG et al., 2015).

Contudo, a ciência de dados enfrenta paradigmas quanto a epistemologia em duas correntes (nova era do empirismo e a de *data driven science*), apontando uma abordagem nova

da compreensão do mundo, em que se apresenta novas ideias e intuições a partir da análise dos dados, ao invés de testar hipóteses (KITCHIN, 2014).

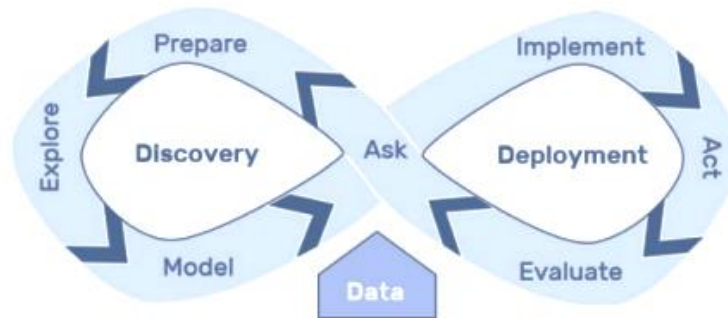
“Nesta nova configuração, a saúde pública se beneficia de um método misto, buscando na abordagem indutiva da ciência de dados a construção de novas hipóteses sem necessitar de apoio ou confirmações clínicas a priori, e na abordagem dedutiva para a verificação destas hipóteses com o rigor científico necessário para a construção de políticas públicas” (SALDANHA, 2021).

Apesar dos sistemas de informações em saúde apresentarem dados sobre o ciclo de vida por completo, não é possível atualmente percorrer o itinerário terapêutico de um paciente, através da disponibilização do DATASUS, por exemplo, no caso de um nascimento pré-maturo, procedimentos ambulatoriais, internação e óbito, devido a anonimização para preservar a privacidade e segurança. A vantagem de um maior volume de informações sobre os eventos apresenta consequências: os dados brutos são mais complexos, apresentam uma linguagem mais técnica e específica, as documentações podem ser deficientes, o volume de dados pode demandar soluções computacionais específicas e, em geral, não permitem o cruzamento entre as bases de dados. Por isso, o Ministério da Saúde tem se esforçado em estratégias para desenvolver, reestruturar e garantir a integração desses sistemas, de modo a permitir um registro da situação de saúde individualizado por meio do Cartão Nacional de Saúde (MINISTÉRIO DA SAÚDE, 2021).

Na ciência de dados existe três grandes modelos de processamento estabelecidos e reconhecidos: Knowledge Discovery in Databases (KDD), Sample, Explore, Modify, Model, Assess (SEMMA) e o Cross-Industry Standard Process for Data Mining (CRISP-DM). O processo KDD é definido como a “extração não trivial de conhecimento implícito, previamente desconhecido e potencialmente útil dos dados” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996c; FRAWLEY; PIATETSKYSHAPIRO; MATHEUS, 1992 apud SALDANHA, 2021).

O modelo SEMMA, introduzido em 1997, sendo aplicado na condução de projetos de mineração de dados, pode ser considerado como a implementação do modelo KDD, ou seja, direcionando para a compreensão de um problema (SALDANHA, 2021). E possui um modelo definido de passos definido e revisado pelo Instituto SAS em 2016:

Figura 1. Modelo de processos SEMMA (2016)

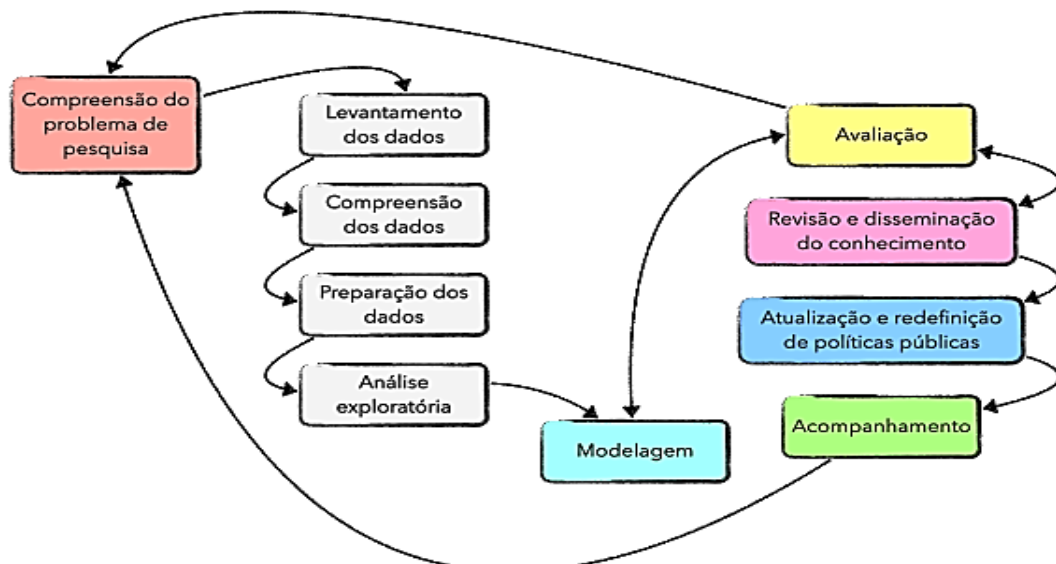


Fonte: Rotondo e Quilligan (2020)

Por fim, o modelo “CRISP-DM utiliza os elementos mais importantes do modelo KDD e os insere em um processo cíclico que prevê explicitamente a compreensão do negócio e a implementação de alterações no negócio segundo os resultados obtidos” (SALDANHA, 2021). Amplamente divulgado devido a sua amplitude na documentação de seus processos do modelo (ROTONDO; QUILLIGAN, 2020).

Mais recentemente, o modelo *Knowledge Discovery in Databases for Public Health* (KDD-PH), proposto por Saldanha (2021), no qual aproxima a linguagem e interesses da ciência de dados aos da Saúde Pública, formando um modelo aplicado à mineração de dados em saúde.

Figura 2. Modelo de processos KDD-PH



Fonte: SALDANHA (2021).

Tão logo visualizamos o diagrama acima, é possível resgatar o conceito de análise de situação de saúde do qual define-se por “*identificação, descrição, priorização e explicação dos problemas de saúde com o objetivo de identificar as necessidades sociais e determinar as*

prioridades de ação” (TEIXEIRA et al, 2010, p. 57). Um processo essencial na vigilância em saúde para o planejamento de prioridades de ações e formulação das políticas públicas, além de possibilitar o monitoramento e avaliação das políticas públicas sobre a situação de saúde da população. Que se correlaciona com outro conceito amplamente utilizado na ciência de dados, o *Business Intelligence* (BI).

O BI combina análise, mineração de dados, visualização de dados, ferramentas/infraestrutura de dados e práticas recomendadas para ajudar as organizações a tomar decisões impulsionadas por dados para gerar mudanças positivas, eliminar a ineficiência e se adaptar rapidamente às mudanças, partir dos seguintes processos:

- **Preparação de dados:** o processo de compilar várias fontes de dados, identificar as dimensões/medidas e prepará-las para a análise de dados.
- **Análise descritiva:** o uso da análise de dados em série histórica.
- **Análise visual:** a exploração dos dados através de histórias visuais para comunicar informações conforme necessário e manter-se no fluxo da análise.
- **Consultas:** fazendo perguntas específicas aos dados, o BI extrai as respostas dos conjuntos de dados.
- **Análise estatística:** a aplicação dos resultados da análise descritiva para explorar os dados em mais profundidade usando conceitos estatísticos (por exemplo, como e por que determinada tendência ocorreu).
- **Visualização de dados:** o processo de transformar a análise de dados em representações visuais, como gráficos, diagramas e histogramas, para facilitar o consumo dos dados.
- **Mineração de dados:** o uso de bancos de dados, estatísticas e aprendizado de máquina para revelar tendências em conjuntos de dados grandes.
- **Geração de relatórios:** o compartilhamento de análises de dados com as partes interessadas para que elas possam tirar conclusões e tomar decisões.
- **Benchmarking e métricas de desempenho:** a comparação de dados de desempenho atuais e históricos para acompanhar o desempenho em relação às metas, geralmente com o uso de painéis personalizados.

A mineração de dados constitui-se de técnicas que consistem na especificação de métodos que nos garantam como descobrir os padrões que nos interessam. Dentre as principais utilizadas, temos: técnicas estatísticas, técnicas de aprendizado de máquina e técnicas baseadas em crescimento-podavalidação. No processo da descoberta do conhecimento exige passos a

seguir, ou seja, tarefas na busca da compreensão do dado, que tipo de regularidades ou padrões pode-se encontrar. São elas:

1. Análise de Regras de Associação: é um padrão da forma $X \rightarrow Y$, onde X e Y são conjuntos de valores (*Support*).
2. Análise de Padrões Sequenciais: é uma expressão da forma $\langle I_1, \dots, I_n \rangle$, onde cada I_i é um conjunto de itens. A ordem em que estão alinhados estes conjuntos reflete a ordem cronológica em que aconteceram os fatos representados por estes conjuntos.
3. Classificação e Predição: é o processo de encontrar um conjunto de modelos (funções) que descrevem e distinguem classes ou conceitos, com o propósito de utilizar o modelo para prever a classe de objetos que ainda não foram classificados.

Alguns conceitos quanto à estrutura de sistemas de informação, linguagem, processamento e limpeza dos dados, também, se demonstram necessários permear nos espaços de trabalho dos profissionais de saúde que lidam com a busca e tratamento dos dados para a análises epidemiológicas e de situação de saúde. Pois no campo da Ciência de Dados, já existem soluções e recursos que ajudam na recuperação de dados, transformação de dados normalização, discretização, integração (*linkage*) dos bancos de dados de forma simples, prática e automatizada. O que melhora na eficiência e efetividade da análise de dados diminuindo tempo e recursos humanos para tais tarefas complexas principalmente com o enorme volume de registros (*big data*) em que é preciso processar em tempo real.

Em suma, não é o objeto desse estudo aprofundar nas técnicas e algoritmos em mineração de dados, mas de posse da aproximação e experimentação dos conhecimentos desta ciência, propiciar a análise crítica dos processos e ferramentas utilizados na vigilância epidemiológica atualmente, a fim de prospectar mudanças tecnológicas na *práxis* possibilitando superar os problemas, já há muito vivenciados e relatados por pesquisadores, na análise de dados em saúde.

3 METODOLOGIA

3.1 Tipo de estudo

Trata-se de estudo exploratório descritivo de revisão de literatura e experimental dos métodos de mineração de dados, baseado no modelo KDD-PH descrito por Saldanha (2021).

A pesquisa exploratória está apoiada na análise documental dos relatórios do serviço de Vigilância em Saúde e das pesquisas científicas encontradas através das estratégias de busca definidas para: a) para compreensão do problema e fundamentação da construção da base de dados *data-VSPEA* e b) busca de ferramentas *open source* (acesso aberto) e tecnologias da ciência de dados; para levantamento de requisitos, definição do estudo de caso e compreensão do domínio da aplicação das informações. E experimental das técnicas de processo de descoberta de conhecimento em bases de dados (KDD): higienização dos dados e pré-processamento dos dados, análise exploratória, em seguida, modelagem para a descrição das variáveis e das bases de dados. Portanto, as etapas abordadas e discutidas nesse estudo, conforme o modelo adotado para a metodologia do trabalho, compreendem as tarefas da compreensão do problema da pesquisa, ou seja, realizar a análise de situação de saúde para a vigilância da população exposta ao agrotóxico.

3.2 Coleta e fonte de dados

A pesquisa foi desenvolvida utilizando-se a coleta de dados secundários do Sistema de Informação de Agravos de Notificação Compulsória (SINAN) e demais sistemas de informação da Secretaria de Estado de Saúde do Tocantins, a fim de testar a integração dos sistemas e identificar a cronicidade da exposição de forma individualizada, após a aprovação do projeto de pesquisa pelo Comitê de ética em Pesquisa- UFT sob o número 4.381.097.

Os dados foram tabulados pelo e analisados com o uso do *Tabwin* versão 3.2 e Microsoft Office 2013®.

SINAN: casos notificados no período (2008 a 2019) dos residentes do Tocantins, pelos códigos da 10ª Revisão da Classificação Internacional de Doenças (CID-10).

SIM: registro de óbitos no período de 2008 a 2020, de residentes nos municípios do Tocantins, por causa básica de óbito dos códigos da 10ª Revisão da Classificação Internacional de Doenças (CID-10) associados à intoxicação: X48, X68, X87, Y18 (BOCHNER, FREIRE, 2020).

SINASC: registros de nascimentos com anomalia congênita detectada (campo 34- Sim) de residentes nos municípios do Tocantins, no período de 2008 a 2020.

3.3 Análise de dados

3.3.1 Seleção secundária

Para uma seleção secundária de análise dos dados utilizou-se os seguintes critérios:

- a) SINAN- linkage: período (2008 a 2019) dos residentes do Tocantins, pelos códigos da 10ª Revisão da Classificação Internacional de Doenças (CID-10) (C06, C18, C20, C34, C44, C50, C61, C80, F99, H83.3, L25.9, T65.9, Z91).
- b) Intox_ agrotóxico: Intoxicação exógena com o filtro dos casos de agrotóxicos selecionando no campo 49 (*p_agente toxico*) as opções: (02) Agrotóxico/uso agrícola; (03) Agrotóxico/uso doméstico; (04) Agrotóxico/uso saúde pública; (05) Raticida; (06) Produto Veterinário (BRASIL, 2018).

3.3.2 Análise de Completitude e Confiabilidade

Com os bancos de dados coletados a partir das seleções descritas (itens 3.2 e 3.3.1) aplicou-se a análises exploratórias com o Excel e o software *Orange Canvas*.

A análise de completude e confiabilidade inicia-se com a identificação do caso repetido ou crônico, mediante a verificação de casos duplicados na base de dados dos SINAN, a partir do banco SINAN-linkage baseando-se em padronização, blocagem e pareamento de registros (CAMARGO e COELI, 2002).

Para encontrar os pares verdadeiros (identificação do mesmo usuário em eventos diferentes) dos pares falsos (identificação de pessoas homônimas), tal etapa foi desenvolvida em três passos. No primeiro deles, a blocagem foi realizada com base na verificação de registros duplicados das variáveis nome do paciente e nome da mãe; no segundo, utilizou a combinação do nome do paciente + nome da mãe; e no terceiro, a combinação do nome do paciente + nome da mãe + sexo por correspondência. As variáveis usadas no pareamento foram nome, sexo, data de nascimento e município de residência. Foram revisados manualmente todos os registros com intuito de melhorar a captação de pares verdadeiros.

Para a completitude dos dados, foram analisadas todas as variáveis (já mencionadas acima) não preenchidas, ou seja, dispostas em brancos e ignoradas nas fichas de notificação. A partir do critério proposto por Romero e Cunha (2006), com os seguintes graus de avaliação: excelente (menor do que 5%), bom (5 a 10%), regular (10 a 20%), ruim (20 a 50%) e muito ruim (50% ou mais).

3.3.3 Mineração de Dados

A ferramenta utilizada no estudo foi *Orange Canvas*, que consiste em uma plataforma para execução de aprendizado de máquina, mineração de dados e visualização de dados. Foram utilizados algoritmos para discriminar as relações entre os atributos dos diversos bancos de dados para buscar padrões entre a exposição aos agrotóxicos e ocorrência de casos de óbitos, abortos e anomalias genéticas por município de residência.

O algoritmo aplicado foi o kNN, utilizando a distância euclidiana, com número de vizinhos parametrizado em 5, a partir de atributos-alvo. A técnica utilizada mensurar a capacidade de generalização do modelo, foi a validação cruzada por k-fold, onde o número de k (folds) foi definido em 10, treinando e testando o modelo com todos os dados disponíveis para evitar variância e obter um resultado de maior robustez.

3.4 Riscos e benefícios e os aspectos éticos legais

Quanto aos riscos envolvidos na pesquisa consistem em riscos mínimos inerentes à possibilidade de identificação dos participantes, no entanto, foi adotado codinomes para assegurar o anonimato dos dados individuais, conforme preconizado pela Resolução nº 466, homologada pelo Conselho Nacional de Saúde do Ministério da Saúde em 2012. Quanto aos benefícios, a realização desta pesquisa propiciou o estudo exploratório das intoxicações por agrotóxico no estado do Tocantins na perspectiva da mineração de dados, o que corrobora para a melhoria das medidas de controle e vigilância da saúde, além de fornecer subsídios para a melhoria na política de atenção à saúde.

O estudo foi submetido e tem a aprovação do Comitê de Ética em Pesquisa (CEP) sob o número 4.381.097, como também a autorização da Secretaria de Estado de Saúde do Estado do Tocantins para acesso e coleta dos bancos de dados secundários (SINAN e SIM). Foram adotadas medidas de segurança para garantir o sigilo, privacidade e confidencialidade dos participantes, tais como: a) armazenamento dos dados em um computador que possui acesso restrito por meio de senha, localizado na Vigilância em Saúde. Este computador possui algumas tecnologias como: antivírus, uma conexão segura e *antispywares*, favorecendo o sigilo, segurança dos dados, a proteção da privacidade e a confidencialidade do paciente; b) termo de compromisso de sigilo profissional, privacidade e confidencialidade dos dados secundários com os integrantes da equipe; c) restrição e anonimização dos dados pessoais dos usuários.

4.RESULTADOS E DISCUSSÃO

Os resultados apresentam-se nos passos executados na pesquisa, desde a busca e análise exploratória da revisão de literatura para compreender o contexto e produzir o levantamento de dados e pré-requisitos para a construção do data-warehouse que fornecesse o *business intelligence (BI)* necessário para a análise de situação de saúde da população exposta ao agrotóxico, até a experimentação das ferramentas/ instrumentos coletados na mineração.

4.1 Revisão teórico- metodológica norteadora para o levantamento de dados

Na revisão da literatura observou-se que os estudos que utilizam modelagem inferencial para avaliação de risco da cronicidade da exposição aos agrotóxicos concentram-se na avaliação de resíduos em alimentos (ELGUETA, *et al.*, 2020; LU, *et al.*, 2020; WUMBEI, 2019). Portanto, consideram somente a exposição dietética o que traz incertezas para as medidas já que o ser humano está sob a exposição por diversas vias em que não se mensura as doses-respostas. A abordagem de avaliação do risco em relação à exposição a agrotóxicos na saúde humana deve ser crônica, visto que o consumo se dará pela vida inteira do indivíduo, tanto por meio ocupacional, quanto pelos meios ambientais e alimentares, exposição essas, que muitas resultam em intoxicações agudas e crônicas, que podem levar ao surgimento de diversas doenças (CARNEIRO, 2015; CESAR-VAZ *et al.*, 2016).

Na literatura verifica-se o condicionante socioambiental e as exposições a certos riscos como os fatores que mais causam doenças e óbitos na população em geral. Por isso, é fundamental o conceito de promoção da saúde e não da doença, ou seja, de forma preventiva e não atacando as consequências que muitas vezes são nefastas, terríveis e quando não irreversíveis, principalmente para as populações mais vulneráveis (PORTO, SOARES, 2012; STURZA, PEDRON, BARRIQUELLO, 2021).

A Avaliação de impacto à saúde (AIS), que compreende uma combinação de procedimentos, métodos e instrumentos, pelos quais uma política, programa ou projeto, do setor público ou privado, pode ser julgada em termos dos seus potenciais efeitos na saúde da população, bem como quanto à distribuição desses efeitos na mesma população, tendo assim uma forte preocupação relativa a equidade. É na sua base construído como um instrumento de apoio à tomada de decisão, tendo como principais *outputs* recomendações para os decisores assegurarem que as políticas implementadas minimizam efeitos na saúde e na equidade negativos e maximizar positivos (MINISTÉRIO DA SAÚDE, 2014; ABE, MIRAGLIA, 2017). Ao longo das suas diversas fases, observou-se que a AIS é geralmente e muito frequentemente

uma abordagem qualitativa, sendo rara a aplicação de metodologias estatísticas multivariadas. No entanto, a necessidade de tratamentos dados ganha cada vez mais importância e valor na busca de respostas para enfrentar doenças e agravos tanto para a ciência quanto para as autoridades sanitárias.

Segundo Silveira e Fenner (2017), a AIS é uma metodologia preconizada pela Organização Mundial de Saúde (OMS) já bastante difundida em diversos países, porém com poucas iniciativas no Brasil e sendo, também, necessário investir em pesquisas do método aplicados aos diversos tipos de empreendimentos. E sugere a inserção do instrumento de Avaliação Ambiental Estratégica (AAE), além da Avaliação de Impacto Ambiental (AIA) no modelo adaptado para o Brasil (MINISTÉRIO DA SAÚDE, 2014). Pois o AAE possibilita análises preditivas, fundamentais para subsidiar a formulação de políticas públicas sob “enfoque analítico e participativo que visam integrar as considerações ambientais nos planos, políticas e programas, avaliando as interligações com as considerações econômicas e sociais” (OCDE, 2006).

Portanto, da revisão de literatura científica levantou-se: um quadro de doenças previstas e observadas relacionadas à exposição crônica a agrotóxicos, fatores epidemiológicos e vias de exposição; índices de contaminação do meio ambiente previstas, ou seja, valores de resíduos em matrizes como água e alimentos (LMR); dados da produção do agronegócio e comercialização de agrotóxicos; dados relativos à territorialização (populacionais e IDH); metodologias e instrumentos de apoio na avaliação.

Do levantamento de dados secundários e sistemas de informação de interesse para a pesquisa, encontrou-se no ambiente do DATASUS: Assistência à Saúde (internação hospitalar, produção ambulatorial, vigilância alimentar e nutricional); Epidemiológica e Morbidade (SINAN: intoxicação exógena, doenças de veiculação hídrica-DTA, DRT acidente de trabalho grave, DRT câncer relacionado ao trabalho, DRT dermatose ocupacional, DRT PAIR e surto; SIM); Estatísticas Vitais (natalidade, mortalidade, câncer); Relatórios do Programa de Análise de Resíduos em Alimentos. Demográficas e Socioeconômicas (população, condições socioeconômicas, saúde, trabalho, moradia) do Instituto Brasileiro de Geografia e Estatística (IBGE). Dados de ponto de monitoramento e índice da qualidade da água (SISAGUA), ciclo de vida de petições de agrotóxicos na ANVISA e quantidade de agrotóxico comercializado, pelo portal de dados abertos.

Resultando na construção de uma matriz de dados (*data warehouse*) relevantes para a compreensão do problema e do contexto para fomentar as análises de dados pertinentes para a

formulação, monitoramento e avaliação da vigilância da população exposta ao agrotóxico, a ser denominado *data-VSPEA* (quadro 1).

Quadro 1. Informações relevantes para formulação, monitoramento e avaliação da vigilância VSPEA. Construção do *data warehouse-VSPEA*.

Contexto	Dados	Possíveis Fontes	Banco de dados e fontes encontradas com variáveis sobre intoxicação/ envenenamento (dados abertos)
Saúde	<ul style="list-style-type: none"> ● Doenças agudas ● Doenças crônicas ● Exames laboratoriais (biomonitoramento) ● Assistenciais ● Morbidade ● Mortalidade 	<ul style="list-style-type: none"> ● SINAN ● GAL e sistemas da rede privada e conveniada laboratorial ● Prontuários ● SIA ● SIH ● SINASC ● RHC/ RCBP ● SIM ● e-SUS 	<ul style="list-style-type: none"> ● Intoxicação exógena ● DTA ● DRT- AT ● DRT- Câncer ● DRT- dermatose ● DRT- PAIR ● SIA ● SIH ● SIM
Ambiente /Território	<ul style="list-style-type: none"> ● Dados demográficos ● Licenças de uso e perfil de produção ● Produção agrícola ● Monitoramento de resíduos matrizes (água, solo, ar, alimentos, ecotoxicologia) 	<ul style="list-style-type: none"> ● IBGE ● CAR/ Meio Ambiente ● SISAGUA ● SISSOLO ● Laboratórios de pesquisa - IES ● Embrapa ● Cartografia do território (Vigilância popular) ● MPE ● MPF 	<ul style="list-style-type: none"> ● Produção Agrícola municipal (PMA) ● Censo agropecuário ● SISAGUA ● SISSOLO
Trabalho	<ul style="list-style-type: none"> ● Mapeamento dos estabelecimentos ● Pessoal ocupado ● Previdência ● Saúde ocupacional ● Vigilância em Saúde do Trabalhador 	<ul style="list-style-type: none"> ● IBGE ● CAT-WEB ● Sistemas de SESMT ● Relatórios de Vigilância em Saúde do trabalhador ● Sistemas de fiscalização do trabalho (MPT) 	<ul style="list-style-type: none"> ● Censo ● Censo agropecuário ● Cat-Web

Regulação	<ul style="list-style-type: none"> ● Ciclo de vida de petições de registro (autorização) ● Produção ● Comercialização ● Utilização 	<ul style="list-style-type: none"> ● ANVISA ● IBAMA ● Sistemas de órgãos estaduais de fiscalização agropecuária ● SEFAZ 	<ul style="list-style-type: none"> ● ANVISA ● IBAMA
------------------	--	---	---

Fonte: Elaboração da autora.

Outra fonte importante de dados e informação se faz a partir da própria população e comunidade, a fim de efetivar a participação popular na vigilância ativa e participativa da saúde além de fomentar e/ou fortalecer a vigilância popular que se constrói da proatividade, da luta e resistência dos povos que sofrem com os impactos em sua carne às vezes negligenciados e/ou invisibilizados nos territórios. Contudo ainda enfrentamos muita dificuldade de realizar o resgate e análise desses dados apesar de algumas estratégias terem sido adotadas como a busca por notícias (clipping de notícias) nos serviços de vigilância.

Portanto, o quadro 1 representa o desafio implicado aos trabalhadores da VSPEA para a organização da análise situacional de saúde, desde à insuficiência de ferramentas à falta de atualização e acesso a dados importantes como a comercialização e consumo de agrotóxicos nos territórios, cujos não estão disponíveis no portal de dados aberto (acessado e verificado pela autora de maio de 2020 a julho de 2021), ferindo o direito à informação. Além de muitas das informações relevantes não estarem disponíveis de forma estruturada em banco de dados, demonstra o quão é complexo a produção de informação e análise dos dados para a vigilância da população exposta ao agrotóxico.

Teixeira (2017) discorre em seu estudo a construção da argumentação de que o direito à informação sobre os dados de agrotóxicos, é um direito fundamental, e a disponibilização e relação entre os envolvidos está desequilibrada. Reforça o protagonismo da população (consumidor) na vigilância e atuação por meio de ação civil pública, além da ação intersetorial e interministerial nas três esferas de governo.

Apesar das dificuldades e desafios apresentados, a integração à ciência de dados e aproximação do trabalho multidisciplinar à ciência da computação possibilita o avanço de soluções práticas como atualmente vivenciados nos espaços do controle externo e educação como também na saúde (FIOCRUZ, 2017; UFT, 2021).

O Sistema de Informação de Agravos de Notificação Compulsória (SINAN) é alimentado, principalmente, pela notificação e investigação de casos de doenças e agravos que constam da lista nacional de doenças de notificação compulsória (Portaria de Consolidação nº

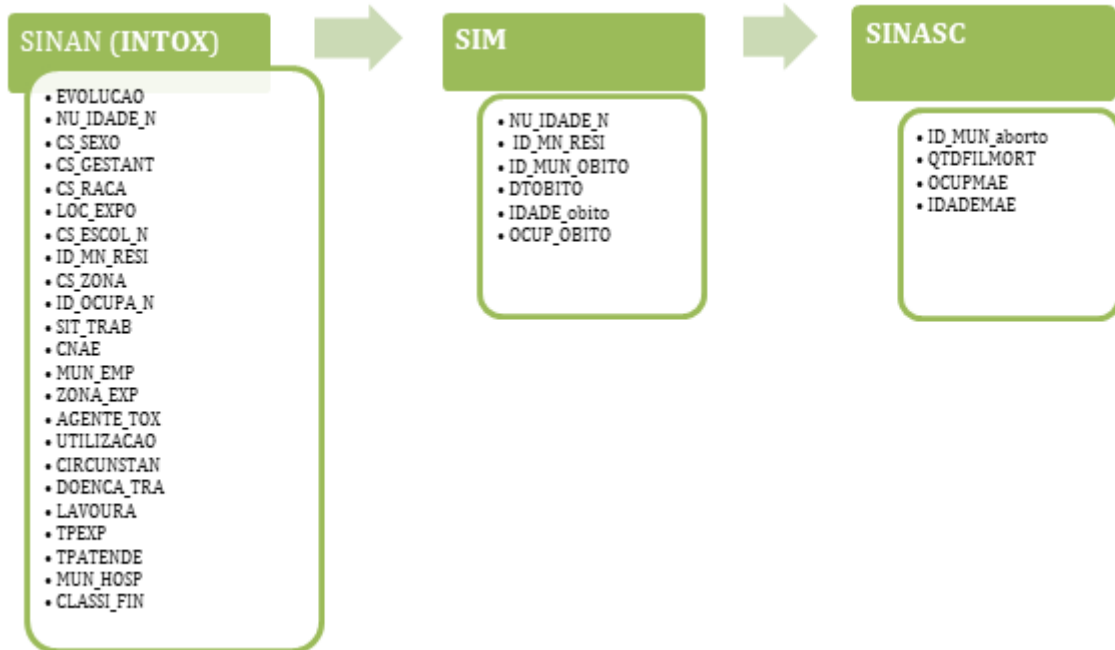
4, de 28 de setembro de 2017, anexo V - Capítulo I). Sua utilização efetiva e sistemática permite a realização do diagnóstico dinâmico da ocorrência de um evento na população, podendo fornecer subsídios para explicações causais dos agravos de notificação compulsória, além de vir a indicar riscos aos quais as pessoas estão sujeitas, contribui para a democratização da informação, permitindo que todos os profissionais de saúde tenham acesso à informação e as tornem disponíveis para a comunidade.

Desta maneira, o SINAN é, portanto, um instrumento relevante para auxiliar o planejamento da saúde, definir prioridades de intervenção, além de permitir que seja avaliado o impacto das intervenções. Dentro do escopo deste sistema de informação, nato da Vigilância em Saúde, podemos destacar alguns agravos de interesse que podem se relacionar com o objetivo do estudo, a intoxicação por agrotóxicos, são eles: DRT Câncer, Surto por Doenças diarreicas por alimentos- DTA, DRT PAIR, DRT Dermatoses e Acidente de trabalho; devido os sintomas agudos da intoxicação assemelhar com outras possíveis causas de doenças e agravos, podem, assim, ser confundidos e não relacionados com a exposição ao agrotóxico na avaliação realizada pelo serviço de saúde.

A partir da análise de dados dos sistemas identificados (Quadro 1), é de suma importância descobrir padrões com o objetivo de trazer *inputs*, ou seja, informações importantes que despertem na equipe de vigilância epidemiológica a investigação e ações de prevenção e controle dos riscos no território. Desta forma, segue um diagrama (Figura 3) das variáveis elencadas e a relação entre os diversos Sistemas de informação para a criação do *dataset* da pesquisa para os testes de modelagem da mineração de dados, a ser denominado *Data-VSPEA*. A seleção dessas variáveis teve como base a literatura científica e os protocolos de atendimento, observando os sintomas agudos e crônicos da intoxicação por agrotóxicos.

A categorização das variáveis de interesse se fez através do estudo do dicionário de dados dos diversos sistemas DATASUS, no intuito de minimizar possíveis casos de sub-registro pelo diagnóstico diferenciado. Haja visto a literatura já descrever casos de ‘chuvas de agrotóxico’ que podem causar intoxicações (PIGNATTI, MACHADO, CABRAL, 2007) e estas passem despercebidas pelo serviço médico devido a falta da informação de tal possível feito, em que os efeitos podem ser provocados à distância do local de pulverização pela deriva do ambiente rural podendo atingir a população urbana ou periurbana.

Figura 3. Escolha de variáveis para análise de mineração de dados entre os sistemas de informação, construção do *dataset Data-VSPEA*.



Fonte: Autora, adaptado do Datasus. Disponível em: <https://datasus.saude.gov.br/transferencia-de-arquivos/>.

Legenda: Dicionário de dados (apêndice A).

Nesse contexto, observa-se uma grande dispersão de dados em variados sistemas, dificultando o cruzamento de informações que norteiam tomadas de decisões e políticas públicas. Estudos recomendam associar os diversos dados de diferentes sistemas de informação para identificar padrões de consumo e áreas prioritárias com maior exposição a agrotóxicos, gerar análise exploratória, correlações estatística espacial e visual para o fortalecimento das ações de vigilância bem como ações integradas de órgãos de fiscalização da agricultura, meio ambiente, trabalho e saúde, além de ações de promoção voltadas à práticas agroecológicas (PIGNATI, OLIVEIRA, SILVA, 2014; PIGNATI, 2018; DA CRUZ FERNANDES, et al., 2019).

Dessa forma, nota-se diversos desafios a serem superados desde: a fragmentação dos sistemas de informação; a subnotificação (sub registro) e falta de completude das notificações; falta de acesso e transparência dos dados e a descontinuidade da atualização; a insuficiência na inserção de tecnologias para o efetivo controle de resíduos dos produtos químicos nas matrizes e vias de exposição; a permissividade da legislação/ regulamentação ambiental; a subutilização das evidências científicas no processo de trabalho e a falta de abordagem estratégica multidisciplinar (PIGNATI, MACHADO, CABRAL, 2007; BOMBARDI, 2017; BRASIL, 2018; JARDIM et al., 2018; MINISTÉRIO SAÚDE, 2018; BRASIL, 2019; O. GOMES, et al., 2020).

Nesse sentido, a ciência de dados apresenta soluções de integração de banco de dados com alto potencial de inovação no processamento e análise dos dados com ferramentas que automatizam os processos de limpeza, transformação e análise de dados, apresentando maior performance de visualização de dados com maior capacidade de atualização em tempo real, minimizando tempo e esforço de recursos humanos (HUANG, 2015; SALDANHA, 2021) .

4.2 Análise dos sistemas de informação em saúde

Da matriz de dados relevantes, o *data-VSPEA* (quadro 1), foram selecionados os sistemas de informação em saúde (DATASUS) para a compreensão do problema pelas informações geradas da rede de atenção pelo Sistema Único de Saúde (SUS), a partir da análise do SINAN, SIM e SINASC.

O *dataset* base/chave da pesquisa foi o SINAN pelo agravo de Intoxicação exógena (T65.9), no qual no período de 2008 a 2019, apresentou 14944 registros e 79 atributos. Então foi iniciada a fase de higienização e tratamento dos dados na composição da etapa de pré-processamento, onde os dados foram selecionados e preparados, em que foi identificado diversos atributos muitos registros incompletos, duplicados e ignorados. Então, o banco Intox_agrotoxico apresentou 3556 registros utilizados para análise das relações entre os demais *datasets* dos sistemas de informação.

Na tabela 1, apresenta-se o percentual de completude das variáveis/atributos dispostos nos bancos de dados dos sistemas de informação SINAN, SIM e SINASC. A completude indica o percentual de registros encontrados no atributo que possuem um valor/ dado, ou seja, não se considera dados faltantes ou ignorados.

Tabela 1. Análise de completude dos atributos nos *datasets* coletados dos sistemas de informação SINAN, SIM e SINASC.

Sistema / N° de instâncias	Atributos elencados	% de Completude	Atributos elencados	% de Completude
SINAN 3556 instâncias	LOC_EXPO	6,5	NU_IDADE_N	0
	ZONA_EXP	8,55	CS_SEXO	0,03
	AGENTE_TOX	0	CS_GESTANT	1,8
	UTILIZACAO	97,14	CS_RACA	1

	CIRCUNSTAN	1,75	CS_ESCOL_N	19,32
	TPATENDE	0,9	CS_ZONA	98,87
	MUN_HOSP	56,18	ID_MN_RESI	0
	CRITÉRIO	11,59	ID_OCUPA_N,	42
	TPEXP	6,36	SIT_TRAB	23,85
	DIAG_CONF	59,3	CNAE	7,19
	EVOLUCAO	9,5	MUN_EMP	7,65
	LAVOURA	24,32	DOENCA_TRA	2,27
	CLASSI_FIN	5,65	CAT	2,98
SIM 115 instâncias	ID_MUN_OBIT	0	Legenda: excelente (menor do que 5%) bom (5 a 10%) regular (10 a 20%) ruim (20 a 50%) muito ruim (50% ou mais) Romero e Cunha (2006)	
	DTOBITO	0		
	IDADE_obito	0		
	OCUP_OBITO	22,6		
SINASC 11.391	ID_MUN_aborto	0		
	IDADEMAE	0		
	OCUPMAE	6,5		
	QTDFILMORT	0		

Fonte: SINAN, acesso em abril/2020. SIM, SINASC/ SES-TO. Acesso em julho/2021.

Legenda: Dicionário de dados (apêndice A).

De acordo com a tabela 1 verifica-se a baixa completude de atributos essenciais para análise do entendimento dos fatores influenciadores do processo da intoxicação. A falta de completude e qualidade dos dados interferiram diretamente nas associações entre as variáveis, reduzindo o nível de acurácia na modelagem da predição.

Conforme o exemplo na tabela 2, os resultados estatísticos inferenciais dos atributos mais relevantes são observados quando utilizados os dados ignorados/ faltantes comparados

aos que possuem um valor categórico para a avaliação da política pública, demonstrando a importância que a investigação dos processos que melhoram a qualidade das informações são relevantes para posteriores análises para a tomada de decisão.

Tabela 2. Representação do resultado da análise exploratória dos atributos no banco de dados de intoxicação exógena por agrotóxicos. Tocantins, de 2008 a 2019.

Support	Confidence	Coverage	Strength	Lift	Leverage	Antecedent	Consequent
0,712	1	0,712	1,303	1,077	0,051	LOC_EXPO=Residência	CNAE=Ignorado
0,661	1	0,661	1,405	1,077	0,047	LOC_EXPO=Residência, DOENCA_TRA=não	CNAE=Ignorado
0,631	1	0,631	1,472	1,077	0,045	LOC_EXPO=Residência, TPEXP=Aguda-unica	CNAE=Ignorado
0,616	1	0,616	1,507	1,077	0,044	LOC_EXPO=Residência, ZONA_EXP=urbana	CNAE=Ignorado
0,615	1	0,615	1,51	1,077	0,044	LOC_EXPO=Residência, CS_ZONA=urbana	CNAE=Ignorado
0,015	1	0,015	16,5	4,145	0,011	AGENTE_TOX=agrotóxico saúde, CIRCUNSTAN=uso Habitual	TPATENDE=Ambul atorial
0,013	1	0,013	17,875	4,145	0,01	CIRCUNSTAN=uso Habitual, CRITERIO=laboratorial	TPATENDE=Ambul atorial
0,013	1	0,013	18,652	4,145	0,01	ZONA_EXP=urbana, AGENTE_TOX=agrotóxico saúde, CIRCUNSTAN=uso Habitual	TPATENDE=Ambul atorial

Fonte: Dados trabalhados. SINAN-INTOX/ SES-TO. Acesso: abril de 2021.

Legenda: Dicionário de dados (apêndice A).

O sub registro e a falta de completude de campos que prejudicam a qualidade dos estudos e a vigilância da população exposta aos agrotóxicos, necessitam de melhorias nos sistemas e processos. Este resultado revela a falta de significado aos trabalhadores do SUS, ou a insuficiência da organização do processo de notificação e o zelo pelos dados coletados, entre outras diversas causas que devem ser mais bem exploradas. Desvela a inoperância da investigação dos casos, em que se conforma ser uma atividade proforma dos serviços de saúde. Impactando sensivelmente nas análises epidemiológicas, tão necessárias para a análise de situação de saúde que recaem sobre o planejamento e organização dos serviços para o atendimento da população. Retoma a urgência de estratégias que efetivamente ataquem o problema, retoma a ação de gestores frente a essa questão tão bem destacada em diversos estudos epidemiológicos (BRASIL, 2018; DA SAÚDE, 2018; BRASIL, 2019; FERNANDES, MOURA, DE OLIVEIRA, 2021).

4.2.1 Da análise do *linkage* de agravos notificados no SINAN

Apesar da notificação de intoxicação exógena possibilitar esta informação através do campo *tp_exposicao*, a análise do banco de dados de intoxicação por agrotóxico de 2008 a 2019 (N=14.872 registros) apresenta: 1473 (9,77%) registros em branco ou ignorado, 12.174 (81,18%) como intoxicação aguda. 1058 (7,11%) de intoxicações repetidas e 120 (0,8%) crônicas. O que leva a considerar que este campo da notificação não é efetivo para esta análise de exposição crônica, reforçando que o serviço de saúde notifica casos novos, sendo necessário desenvolver métodos e modelos de identificação dos casos crônicos.

Assim, analisando o **filtro_link** segue os resultados encontrados nos processos de padronização, blocagem, pareamento dos registros encontrados no SINAN, em que pares verdadeiros foram identificados automaticamente, dentre eles estão as duplicidades e os eventos repetidos/crônicos; pares falsos (homônimos) não há correspondência da variável nome da mãe, data de nascimento; e na revisão manual visualizamos eventos de pares verdadeiros que não foram identificados devido a erros de digitação e dados faltantes.

Dentro da rotina do sistema SINAN, conforme seu manual, verifica-se as duplicidades e as exclui do sistema, contudo deve se ater ao conceito de duplicidade: o mesmo evento é notificado mais de uma vez, o qual é considerado uma sujeira no banco de dados. Contudo, a pesquisa vem discutir essa rotina no sentido de melhorar a sensibilidade para as “falsas duplicidades” que caracterizam repetições de eventos na linha de vida do usuário e se as mesmas possuem relação com a intoxicação.

Tabela 3. Análise de completude e confiabilidade da identificação de eventos repetidos por usuário associados à exposição aos agrotóxicos. SINAN, de 2008 a 2019.

Atributos	Completeness	Identificação dos eventos	Distribuição de eventos	Nº de eventos combinados (repetidos/associados à T65.9)
Nome	0%	Aleatórios	31976	—
Nome da mãe	0,0031%	Pares verdadeiros	2115	451
Data de nascimento	0,0026%	Pares falsos (homônimos)	817	—
Sexo	5,51%	Pares verdadeiro-	1353	52

		falsos		
Município_res	2,75%	Duplicidade	-	67
N	36.261	Total	36.261	503 casos
Legenda: excelente (menor do que 5%); bom (5 a 10%); regular (10 a 20%); ruim (20 a 50%); muito ruim (50% ou mais). Romero e Cunha (2006).				

Fonte: Dados trabalhados. SINAN/ SES-TO. Acesso: abril de 2021.

Da análise mais aprofundada da duplicidade observamos 503 (1,38%) de registros de intoxicações repetidas para 224 usuários, ou seja, uma média de 2,24 eventos por usuário; destes 52 (10,3%) registros foram recuperados pela análise manual que por erros de digitação não conseguem ser reconhecidas pelo algoritmo e 67 (0,18%) dos registros realmente são as duplicidades, em que o manual SINAN orienta que sejam retiradas do sistema. O resultado nos alerta da fragilidade do sistema para realização da limpeza do banco de dados, necessitando de uma análise manual mais aprofundada para que não haja ruídos no banco de dados, mas principalmente para que não haja a deleção de casos que possam ser um indicador da intoxicação crônica.

Dentre os 503 casos registrados associados à intoxicação por agrotóxico (T65.9) observou-se a notificação de 1,19% (6 casos) de dermatose ocupacional (L25.9), 4,37% (22 casos) notificados como acidente de trabalho (Y96) e 18,48% (93 casos) fatores pessoais de risco não classificados (Z91). Os resultados apontam para a possibilidade de sub-registro de notificações de intoxicação por agrotóxico por falta do diagnóstico correto, como também para sinais e possíveis sequelas, confirmando a necessidade do acompanhamento mais detalhado individualizado do usuário.

Desta análise, também, aponta-se possibilidades de pesquisa na ciência de dados a fim de apresentar soluções para resolver a busca e o retorno de relatórios aos profissionais de saúde de intoxicações crônicas como o acompanhamento do usuário desde a primeira notificação, ainda inexistente no sistema e tabulador atual.

4.2.2 SIM

O Sistema de Informações de Mortalidade (SIM) é um sistema de vigilância epidemiológica nacional, cujo objetivo é captar dados da Declaração de Óbito (DO) a fim de fornecer informações sobre mortalidade para todas as instâncias do sistema de saúde. Deste, observou-se a possibilidade de selecionar os possíveis óbitos relacionados à intoxicação por

agrotóxicos selecionando o campo da causa básica dentre os possíveis achados na literatura.

Do recorte do estudo o banco de dados do SIM (descrição no item 3.3.1) apresentou 115 óbitos com causa básica definida com os códigos do CID-10 relacionados aos agrotóxicos ocorridos nos municípios do Tocantins de (2008 a 2020). Da análise de correlação entre as variáveis sexo, faixa etária e cor observa-se o padrão de ocorrência do óbito por intoxicação por agrotóxico dos fatores individuais associados na tabela 4.

Tabela 4. Resultados estatísticos da associação dos fatores individuais do óbito por intoxicação por agrotóxico, Tocantins, 2008 a 2020.

Correlation	Feature 1	Feature 2
0,951	Mas	Parda
0,945	15-49a	Mas
0,933	50 e+	Mas
0,92	15-49a	Parda
0,916	50 e+	Parda
0,848	Fem	Parda
0,845	Branca	Mas
0,843	15-49a	50 e+
0,829	15-49a	Branca
0,806	15-49a	Fem

Fonte: Dados trabalhados. SIM/ SES. Acesso em julho de 2021.

De acordo com os resultados (tabela 4) podemos verificar que o padrão de mortalidade por intoxicação de agrotóxicos no Tocantins é maior em homens pardos e brancos na faixa etária de 15 a 49 anos e mulheres pardas na faixa de 15 a 49 anos. Os resultados concordam com as características populacionais (IBGE, Censo Demográfico 2010) e os estudos epidemiológicos de mortalidade (BOCHNER, FREIRE, 2020).

4.2.3 SINASC

O Sistema de Informações sobre Nascidos Vivos (SINASC) tem por objetivo fornecer dados sobre natalidade para todas as instâncias do sistema de saúde a partir da Declaração de Nascido Vivo (DN), padronizada em todo o país. Neste sistema, foi observado dados quanto ao aborto e anomalias congênitas informadas pelo território que podem ser relevantes para o monitoramento associado ao crescimento da utilização de agrotóxico e/ou da área plantada de culturas permanentes e temporárias.

O banco de dados coletado do SINASC a partir do filtro do campo 34 com a opção SIM, seleciona os nascimentos com anomalia congênita ou malformações, em que no período de 2008 a 2020, retornou 11.391 registros com maior frequência nos municípios de Palmas (23%) e Araguaína (13,21%). Da análise de frequência das variáveis faixa etária, quantidade de filhos mortos e ocupação da mãe observa-se que a média de filhos mortos é de 2,37 (σ 0,858) em mulheres com 29,53 anos (σ 6,33) com a ocupação mais frequente sendo dona de casa em 42,78% (4874 casos) e 15,7% de trabalhadoras da agropecuária.

Das correlações entre as variáveis encontradas observa-se na tabela 5 tendências positivas altas da faixa etária de 15- 19 anos que apresentam filhos com anomalia há presença de aborto de 2 a 4 filhos com significância de 0,96, na faixa etária 20-34 anos de 4 a 6 filhos mortos (0,97), de 35-39 anos apresentam de 4 a 6 filhos mortos (0,95), da faixa etária de 40-44 anos possui 5 ou 6 filhos mortos (0,92) e por último, na faixa etária de 45- 49 anos (0,76).

Na tabela 6, apresenta-se as relações sobre as ocupações e a faixa etária com nascimentos de anomalias congênitas no qual a faixa mais adulta de mulheres (45-49 anos) trabalhadoras da produção de bens e serviços industriais e serviços administrativos apresentam filhos com anomalia, enquanto a faixa etária mais jovem que apresentam filhos com anomalia são trabalhadoras da agropecuária e pesca ou trabalham em serviços técnicos de nível médio.

Tabela 5. Tendências positivas negativas para mulheres que apresentaram abortos e filhos com anomalia congênita. Tocantins, de 2008 a 2020.			Tabela 6. Tendência negativas para mulheres que apresentam aborto e filhos com anomalia. Tocantins, de 2008 a 2020.		
Correlation	Feature 1	Feature 2	Correlation	Feature 1	Feature 2
0,974	15-19a	2 filhos morto	-0,045	< 15a	Técnicos de nível médio
0,949	15-19a	6 filhos morto	-0,039	< 15a	Trabalhadores serviços administrativos
0,949	15-19a	4 filhos morto	-0,031	< 15a	Trab agropec, florestais e pesca
0,919	15-19a	5 filhos morto	-0,041	15-19a	Trab da produção de bens e serv industriais
0,901	15-19a	7 filhos morto	-0,044	35-39a	Trab da produção de bens e serv industriais

0,982	20-34a	4 filhos morto	-0,03	35-39a	Membros Superiores do Poder Público
0,969	20-34a	6 filhos morto	-0,049	45-49a	Trab da produção de bens e serv industriais
0,968	20-34a	5 filhos morto	-0,042	45-49a	Trabalhadores servicos administrativos
0,934	20-34a	7 filhos morto	-0,028	45-49a	Técnicos de nível médio
0,959	35-39a	4 filhos morto	-0,027	45-49a	Trab serv, Vendedores Comercio e lojas
0,954	35-39a	5 filhos morto			
0,935	35-39a	6 filhos morto			
0,923	35-39a	7 filhos morto			
0,922	40-44a	6 filhos morto			
0,92	40-44a	5 filhos morto			
0,769	45-49a	6 filhos morto			
0,495	45-49a	9 filhos morto			

Fonte: Dados trabalhados. SINASC/ SES-TO. Acesso: julho de 2021.

Estes resultados desvelam o senso comum em que se diz: “quanto mais velha você engravidar mais chances você terá abortos ou filhos com defeito”. Observa-se neste cenário atual são mulheres cada vez mais jovens com a ocorrência maior de abortos e filhos com anomalias congênitas. Estes padrões corroboram com os estudos de hereditariedade das mutações genéticas provocadas pela contaminação dos agrotóxicos em que deve ser considerado as doses diárias e o tempo de latência para o reconhecimento das sequelas daquilo que produzimos hoje (COLBORN et al., 2002; AGUIAR, 2017; ALI, ISMAIL, *et al.*, 2018; KAPELEKA, SAULI, NDAKIDEMI, 2019; SILVERIO E PINHEIRO, 2019) e nos alerta para influência do condicionante socioambiental para as consequências das gerações futuras, desde a densidade demográfica como sociais (STURZA, PEDRON, BARRIQUELLO, 2021).

Assim, sabendo que a associação dos fatores genéticos e ambientais pode desencadear doenças graves e crônicas, como Doença de Parkinson, efeitos adversos na gravidez, autismo, entre outras. Em que o aparecimento dos sintomas ocorre tardiamente sem uma origem específica e determinada, é de extrema importância realizar estudos para avaliar os fatores

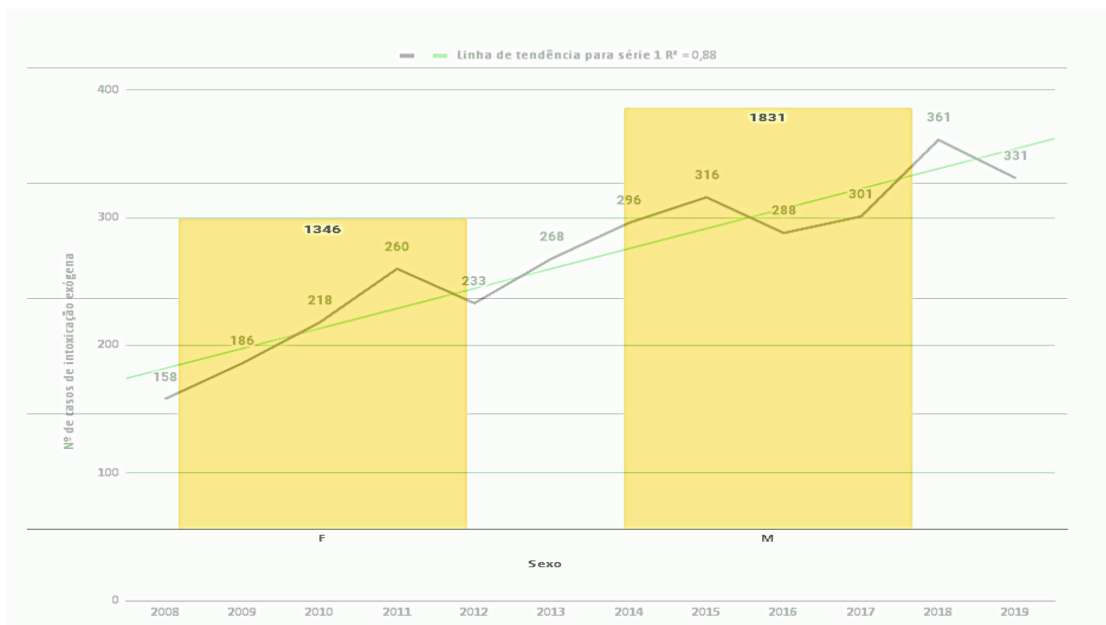
ambientais, em especial na população rural (SILVA, XAVIER, et al., 2019; VASCONCELLOS, et al., 2019; MUSTIELES, et al., 2020).

Os dados apresentados validam a matriz de dados para as associações apresentado no data-VSPEA (Quadro 1) pela a autora a fim de compreender o contexto em que este dado/indicador acontece, entender o ciclo de vida dos usuários. Estratégias de coleta de dados socioambientais se mostram necessárias, já que as mesmas não são observadas pelo sistema de informação em saúde. Indicam a necessidade de investigações aprofundadas como a observada pelo estudo (AGUIAR, 2017) junto aos territórios apontados pela análise para a compreensão do impacto mutagênico do agrotóxico nos genitores a fim de prevenir e controlar gerações futuras apresentando deformidades graves.

4.3 Preparação dos dados para os experimentos da modelagem

Do banco de dados Intox_exog, no período de 2008 a 2019, observa-se 14.872 registros, destes 3216 (21,62%) são por agrotóxico, apresentando tendência positiva no aumento dos casos de acordo com linha do tempo (Gráfico 1), distribuídos em 1346 (41,85%) mulheres e 1831 (56,93%) em homens.

Gráfico 1. Distribuição dos casos por ano e sexo de intoxicação por agrotóxicos, Tocantins, 2008 a 2019.



Fonte: Dados trabalhados. SINAN-INTOX/ SES-TO. Acesso: abril de 2021.

Os atributos em sua maioria são categóricos, sendo analisados frequentemente por contagem de frequência das subcategorias. Neste estudo a análise foi realizada pelo conjunto de dados relacionando os atributos a partir de um atributo- alvo para reconhecer os padrões e treinar o modelo de predição.

Dentre as probabilidades de combinação dos atributos que existe no banco de dados intox_agrotoxico e os demais atributos do SIM e SINASC, optou-se por criar os datasets de análise dos fatores associados (quadro 2) para: diagnóstico de intoxicação (classi_fin), evolução, exposição, da exposição nos municípios de residência (risco_local), da ocupação (risco_ocup), da mortalidade por envenenamento (intox_SIM) e dos abortos e anomalias congênicas (intox_Sinasc).

A escolha dos atributos se fez pela definição contidas nas fichas de notificação e manuais instrutivos de seus objetivos e descrição do campo para a análise do objetivo o qual o atributo-chave representa.

Quadro 2. Configuração dos datasets para o treino do modelo preditivo.

Dataset	Atributos
Classi_fin	ID_OCUPA_N ZONA_EXP AGENTE_TOX UTILIZACAO CIRCUNSTAN TPATENDE MUN_HOSP CRITERIO
Evolução	NU_IDADE_N, CS_SEXO, CS_GESTANT, CS_RACA, LOC_EXPO CS_ESCOL_N ID_MN_RESI CS_ZONA ID_OCUPA_N SIT_TRAB CNAE MUN_EMP ZONA_EXP AGENTE_TOX, UTILIZACAO, CIRCUNSTAN, DOENCA_TRA LAVOURA TPEXP TPATENDE MUN_HOSP CLASSI_FIN
Exposição	NU_IDADE_N CS_SEXO, CS_GESTANT CS_RACA LOC_EXPO, CS_ESCOL_N ID_MN_RESI CS_ZONA ID_OCUPA_N SIT_TRAB CNAE MUN_EMP ZONA_EXP AGENTE_TOX, UTILIZACAO CIRCUNSTAN DOENCA_TRA LAVOURA TPATENDE MUN_HOSP CLASSI_FIN DIAG_CONF CRITERIO EVOLUCAO
Risco_local	CS_ZONA ZONA_EXP AGENTE_TOX UTILIZACAO CIRCUNSTAN LAVOURA TPEXP TPATENDE MUN_HOSP, CLASSI_FIN DIAG_CONF CRITERIO EVOLUCAO NU_IDADE_N CS_SEXO CS_GESTANT CS_RACA LOC_EXPO CS_ESCOL_N
Risco_ocup	NU_IDADE_N CS_SEXO CS_GESTANT CS_RACA LOC_EXPO CS_ESCOL_N ID_OCUPA_N SIT_TRAB CNAE MUN_EMP ZONA_EXP AGENTE_TOX UTILIZACAO CIRCUNSTAN DOENCA_TRA LAVOURA TPEXP CLASSI_FIN EVOLUCAO CAT

Intox_SIM	ID_MN_INTOX ANO_INTOX ID_OCUPA_INTOX NU_IDADE_INTOX ID_MUN_OBIT DTOBITO IDADE_obito OCUP_OBITO
Intox_Sinasc	ID_MN_INTOX ID_OCUPA_INTOX NU_IDADE_INTOX ID_MUN_aborto QTDFILMORT OCUPMAE IDADEMAE

Fonte: Autora. Legenda: Dicionário de dados (apêndice A).

Os *datasets*, então, foram processados com aplicação de técnicas da mineração de dados a fim de se extrair os padrões de interesse, através das tarefas de análise por regras de associação, padrões sequenciais, classificação e predição. Normalmente, os dados considerados ruídos são excluídos para que não interfira na descoberta dos padrões (CAMILO, SILVA, 2009; GALVÃO, 2009), contudo neste trabalho eles foram mantidos para análise mais ampla da descoberta do processo de investigação e vigilância da intoxicação por agrotóxico.

4.4 Resultados da mineração dos dados

4.4.1. Os fatores associados à intoxicação por agrotóxicos

Das associações encontradas que analisaram os fatores associados para definição do diagnóstico, exposição, do local de exposição e da ocupação; os achados mais prevalentes foram descritos na tabela 7 a partir das métricas de associação *Support* e *Coverage*, representam a frequência em que os dados se apresentam no conjunto de dados e a regra é aplicada, sob nível de significância 100% (1).

Tabela 7. Fatores com maior prevalência no banco de dados de intoxicação por agrotóxico, Tocantins, 2008 a 2019.

Support	Confidence	Coverage	Fatores
0,712	1	0,712	LOC_EXPO=Residência
0,661	1	0,661	LOC_EXPO=Residência, DOENCA_TRA=não
0,631	1	0,631	LOC_EXPO=Residência, TPEXP=Aguda-unica
0,62	1	0,62	LOC_EXPO=Residência, EVOLUCAO=cura s/sequela
0,581	1	0,581	CS_SEXO=M
0,468	1	0,468	CS_SEXO=M, CS_RACA=parda
0,243	1	0,243	ZONA_EXP=urbana, AGENTE_TOX=raticida
0,201	1	0,201	ZONA_EXP=urbana, AGENTE_TOX=raticida, TPATENDE=Hospitalar
0,195	1	0,195	AGENTE_TOX=raticida, CIRCUNSTAN=tentativa de suicídio
0,104	1	0,104	AGENTE_TOX=produto veterinário, TPATENDE=Hospitalar

0,016	1	0,016	ZONA_EXP=rural, AGENTE_TOX=produto veterinário
0,015	1	0,015	ZONA_EXP=rural, AGENTE_TOX=raticida, CIRCUNSTAN=tentativa de suicídio, TPATENDE=Hospitalar
0,015	1	0,015	AGENTE_TOX=agrotóxico saúde, CIRCUNSTAN=uso Habitual
0,012	1	0,012	ZONA_EXP=rural, UTILIZACAO=Herbicida, CIRCUNSTAN=Acidental, CRITERIO=clinico
0,011	1	0,011	ZONA_EXP=urbana, AGENTE_TOX=agrotóxico saúde, UTILIZACAO=Inseticida, CIRCUNSTAN=uso Habitual

Fonte: SINAN- INTOX/ SES-TO. Acesso: abril de 2021. Legenda: Dicionário de dados (apêndice A).

A maior prevalência dos fatores associados às intoxicações por agrotóxico no Tocantins, ocorrem nas residências (*support* 0,712), zona urbana do tipo aguda (*support* 0,631) em pardos do sexo masculino, não relacionado ao trabalho e evoluem com atendimento hospitalar (*support* 0,521), denotando maior gravidade por estar associado com o uso de raticida sob a circunstância tentativa de suicídio e produto veterinário por acidente. Apesar da combinação destes fatores necessitar de um atendimento mais complexo, apresentam uma evolução maior de cura sem seqüela (*support* 0,379).

Um segundo grupo de vulnerabilidade encontrado é da zona rural, trabalhadores da agropecuária (44,34%) que utilizam de produto veterinário ou raticida sob a circunstância de tentativa de suicídio utilizando atendimento hospitalar, ou ainda, na utilização de herbicida por circunstância acidental, fazendo a busca de atendimento ambulatorial. O terceiro grupo classificado é o de zona urbana que utiliza agrotóxico de uso na saúde (inseticida) por uso habitual, caracterizando o perfil dos agentes de endemias.

Comparando os resultados à literatura, especialmente ao estudo De Oliveira e Costa (2018), que traz uma caracterização das intoxicações por agrotóxicos no Tocantins mais recente, observou-se concordâncias de fatores apresentados como sexo e faixa etária, tipo de exposição e atendimento, critério de diagnóstico e evolução; no entanto, traz divergências importantes quanto a local de exposição, da relação com o trabalho, quanto a prevalência do agente tóxico e circunstância, e inferências que não reconhecidas como o padrão constante no mesmo banco de dados estudado. Contudo, corroboram com o padrão descrito nas análises epidemiológicas de *coorte* retrospectivas sem recorte territorial (RAMOS et al., 2021). Deve-se ao monitoramento epidemiológico por sua fragmentação das notificações dos agravos, sub-registro e pouca qualidade nos dados informados que impossibilitam determinar um cenário real, com isso, dificulta assegurar a implementação de ações de cuidado com vistas à prevenção de outras tentativas de suicídio (e de óbitos) (MINISTÉRIO DA SAÚDE, 2019).

Esse achado na mineração traz uma reflexão crítica da forma e posicionamento a que se utiliza os dados brutos nos serviços e na pesquisa, sem compreensão mais aprofundada do contexto em que estão inseridos. Estudos observacionais descritivos, são recortes de eventos em um período e espaço, importantes para análise de situação de saúde principalmente para avaliações e monitoramento dos planos/ projetos em saúde pública. Contudo, não podem ser utilizados com viés de associação e generalização da compreensão do problema para a intervenção conforme verificado nas regras de associações.

Outras possíveis métricas de associação entre os conjuntos de dados ($X \rightarrow Y$) são *Strength* representa a razão Y/X , *Lift* que representa a frequência com que a regra é verdadeira e *Leverage* que representa a probabilidade da condicional dos eventos (AGRAWAL; SRIKANT, 1994; GALVÃO, 2009).

Portanto, resultados como os exemplos abaixo, sobre a exposição e sua gravidade podem contribuir de forma significativa para o planejamento de ações de prevenção e promoção mais assertivas e na identificação de grupos vulneráveis:

1. ZONA_EXP=urbana, AGENTE_TOX=agrotóxico saúde, CIRCUNSTAN=uso Habitual, UTILIZACAO=Inseticida \Rightarrow TPATENDE=Ambulatorial

Estas características descrevem a exposição dos agentes de combate a endemias em que com maior frequência as intoxicações são de menor gravidade exigindo o atendimento ambulatorial. Contudo, não é menos preocupante pois é um grupo de trabalhadores com exposição direta e frequente o que aumenta o risco de sequelas crônicas pela dose diária.

2. ZONA_EXP=rural, UTILIZACAO=Herbicida, CIRCUNSTAN=Acidental, CRITERIO=clinico, TPATENDE=Ambulatorial \Rightarrow AGENTE_TOX=agrotóxico agrícola

O item 2 apresenta padrões importantes quanto a intoxicação por agrotóxico de uso agrícola, que por legislação devem estar em ambientes controlados e seguros e seu uso deve ser realizado por trabalhadores treinados com o uso de proteção.

3. ZONA_EXP=urbana, AGENTE_TOX=raticida, CIRCUNSTAN=tentativa de suicidio
ZONA_EXP=urbana, AGENTE_TOX=produto veterinário, CIRCUNSTAN=Acidental

As associações apresentadas no item 3 permitem o questionamento sobre o acesso, comercialização e fiscalização destes produtos, que geram mortalidade quando comparados à análise de mortalidade contidas no SIM (Tabela 5).

A distribuição da ocupação por zona de exposição (Tabela 8) apresenta altos percentuais de dados ignorados, principalmente na zona urbana apresentou 31,97% dos registros e 6,02% na zona rural, prejudicando os índices de correlação entre as variáveis resultando no baixo nível de acurácia para o modelo preditivo. Portanto, o trabalho de análise

de dados da intoxicação por agrotóxicos anseia técnicas mais apuradas para a normalização e recuperação dos dados nestes atributos, a fim de possibilitar verificar as possíveis relações dos determinantes sociais à exposição e evolução dessas intoxicações.

Tabela 8. Ranking da ocupação por zona de exposição (10^o posições).

Ranking	Urbana	2490		70,02%	Rural	733		20,61%
			Grupo %	Total %			Grupo %	Total %
1 ^o	Ignorado	1137	45,66	31,97	Trabalhador agropecuário	325	44,34	9,14
2 ^o	Estudante	369	14,82	10,38	Ignorado	214	29,20	6,02
3 ^o	Fora da CBO ¹	347	13,94	9,76	Fora da CBO	55	7,50	1,55
4 ^o	Trabalhador agropecuário	173	6,95	4,87	Estudante	47	6,41	1,32
5 ^o	Trabalhador serviços e comércio	96	3,86	2,70	Trabalhador construção civil	19	2,59	0,53
6 ^o	Trabalhador construção civil	71	2,85	2,00	Trabalhador serviços e comércio	12	1,64	0,34
7 ^o	Trabalhador administrativo	63	2,53	1,77	Serviços de transporte	10	1,36	0,28
8 ^o	Trabalhadores da limpeza	41	1,65	1,15	Trabalhadores da limpeza	10	1,36	0,28
9 ^o	Professores de ciências biológicas e saúde	27	1,08	0,76	Professores de ciências exatas	10	1,36	0,28
10 ^o	Trabalhadores fabricação de alimentos, bebidas e fumo	27	1,08	0,76	Trabalhadores serviços industriais	10	1,36	0,28

Fonte: SINAN- INTOX/ SES-TO. Acesso: abril de 2021.

A tabela 8, traz em seus resultados alertas importantes para as intoxicações em estudantes e usuários que não estão ocupadas, provavelmente relacionadas com o padrão de evidência das tentativas de suicídio, apresentadas na tabela 7. Verifica-se, então, um fator importante a ser observado para o planejamento de ações de prevenção e promoção da saúde

¹ As ocupações denominadas fora da CBO consiste naquelas que não podem ser identificadas dentro Classificação Brasileira de Ocupação, como: dona de casa, desempregado, aposentado e presidiário.

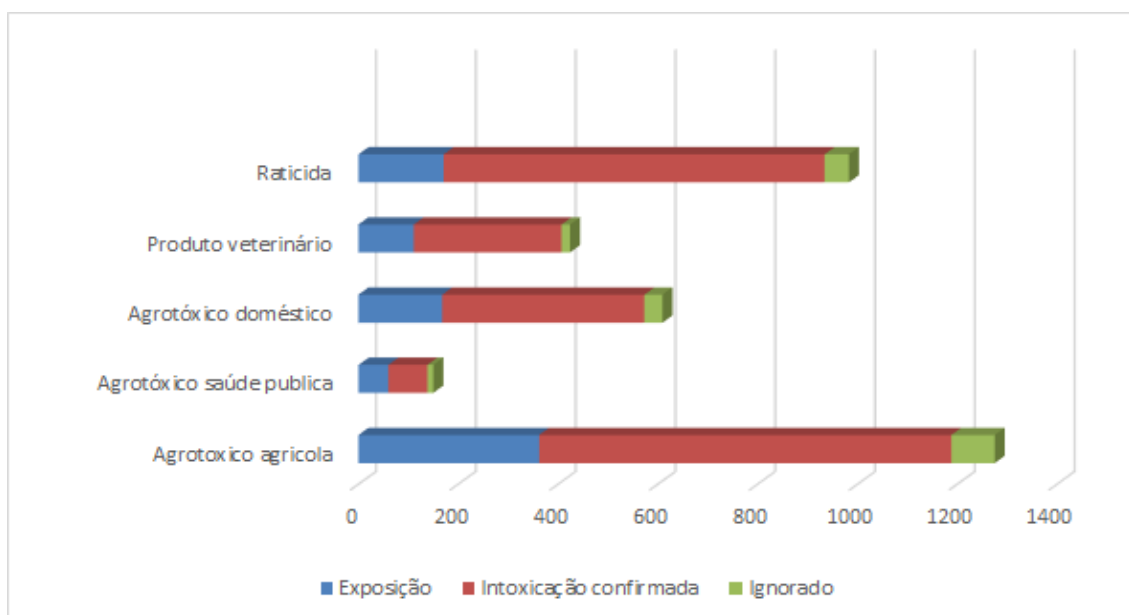
nas escolas e comunidades.

Outro ponto curioso é o percentual de trabalhador agropecuário intoxicado, também, na zona urbana, indicando ser a população mais vulnerável e de maior prioridade para a vigilância em saúde do trabalhador. Ademais, temos os trabalhadores de serviços e comércio e subsequente os da construção civil como as ocupações que mais apresentam intoxicações por agrotóxicos ocupando a quarta e quinta posições no ranking para a zona urbana e rural, denunciando a necessidade de investigação destes processos e ambientes de trabalho para tal exposição e uso destes produtos dos quais não estão evidentes para o risco.

A frequência resultante na nona posição, também denuncia um fator de preocupação não evidente na literatura, que é a intoxicação por agrotóxicos em professores das ciências biológicas e da saúde na zona urbana e das ciências exatas na zona rural possivelmente relacionados a transtornos mentais que levam à tentativa de suicídio, contudo todas as hipóteses aqui levantadas devem ser testadas a partir de estudos estatísticos de correlação posteriormente.

Da distribuição da classificação final do diagnóstico por tipos de agente tóxico (Gráfico 2), tem-se o grupo de agrotóxico agrícola com 37,23% (1324) dos casos, raticida com 28,18% (1002 casos), agrotóxico doméstico com 17,66% (628 casos), 12,51% (445 casos) de produto veterinário e por último agrotóxico de uso na saúde pública com 4,42% (157) registros de intoxicação exógena.

Gráfico 2. Classificação final do diagnóstico por agente tóxico, residentes do Tocantins, 2008 a 2019.



Fonte: SINAN- INTOX/ SES-TO. Acesso: abril de 2021

Revela-se no gráfico 2, o maior percentual de intoxicações confirmadas do que exposições para o grupo de agentes: raticida, produtos veterinários e agrotóxico doméstico; indicando a maior facilidade dos profissionais de saúde em reconhecer os sinais e sintomas da intoxicação nesses agentes. Outra hipótese a ser explorada seria a do sub registro quanto a classificação e confirmação do diagnóstico devido à pressão dos empregadores sobre os serviços de saúde para garantir a invisibilidade da situação de saúde dos trabalhadores (PORTO, SOARES, 2012; DANROCHO, 2019).

Tabela 9. Distribuição dos casos por agente tóxico e classificação final, Tocantins, 2008 a 2019.

Agrotóxico agrícola:	1324		37,23%	Agrotóxico saúde pública	157		4,42%
		Grupo %	Total %			Grupo %	Total %
Exposição	362	27,34	10,18	Exposição	60	38,22	1,69
Ignorado	87	6,57	2,41	Ignorado	12	7,64	0,34
Intoxicação confirmada	826	62,39	23,23	Intoxicação confirmada	78	49,68	2,19
Outro diagnóstico	6	0,45	0,17	Outro diagnóstico	1	0,64	0,03
Reação adversa	42	3,17	1,18	Reação adversa	6	3,82	0,17
Síndrome de abstinência	1	0,08	0,03	Síndrome de abstinência	0	0	0
Agrotóxico doméstico	628		17,66%	Produto veterinário	445		12,51%
Exposição	168	26,75	4,72	Exposição	111	24,94	3,12
Ignorado	36	5,73	1,01	Ignorado	17	3,82	0,48
Intoxicação confirmada	405	64,49	11,39	Intoxicação confirmada	296	66,52	8,32
Outro diagnóstico	2	0,32	0,06	Outro diagnóstico	7	1,57	0,20
Reação adversa	16	2,55	0,45	Reação adversa	14	3,15	0,39
Síndrome de	1	0,16	0,03	Síndrome	0	0	0

abstinência				de abstinência			
Raticida	1002		28,18%				
Exposição	171	17,07	4,81				
Ignorado	49	4,89	1,38				
Intoxicação confirmada	763	76,15	21,46				
Outro diagnóstico	7	0,70	0,20				
Reação adversa	12	1,20	0,34				
Síndrome de abstinência	0	0	0				

Fonte: SINAN- INTOX/ SES-TO. Acesso: abril de 2021.

Vale destacar ainda os registros como diagnóstico de reação adversa e síndrome de abstinência para este grupo de agentes dos quais não são de uso pessoal, configurando erros nos registros desses campos e que impactam nas análises epidemiológicas.

4.4.2 Modelagem de predição

No geral, os índices de acurácia apresentados nos datasets modelados foram baixos, as justificativas precisam ser melhor apuradas, mas é comum em bancos de dados reais em que há grande quantidade de dados ignorados e faltantes, ou seja, ruídos (CARVALHO, ESCOBAR, TSUNODA, 2014). Nos *datasets* de junção de diferentes sistemas de informação, como o Intox_sim, o valor obtido foi de 0,417 e o do Intox_sinasc com 0,152, não sendo possível a geração do modelo preditivo. Provavelmente as diferenças estruturais entre os bancos de dados e a divergência do número de instâncias/ registros entre eles impactaram a análise. Técnicas e estudos mais aprofundados a partir dos resultados da pesquisa são demandas de pesquisas futuras para a recuperação e normalização dos dados no pré-processamento. Segundo Carvalho, Escobar e Tsunoda (2014) apontaram o uso de diversos algoritmos possíveis na busca de resultados mais acurados.

Da modelagem de classificação para predição possíveis na primeira experimentação, com a aplicação do algoritmo kNN, utilizando a distância euclidiana, parametrizado em 5, validação cruzada por k-fold definido em 10, foi obtido para o atributo- alvo os seguintes resultados de acurácia: CLASSI_FIN (0,678), EVOLUCAO (0,848), TPEXP (0,845),

ID_MN_RESI (0,306), ID_OCUPA_N (0,518). Logo, seguem descritos nas tabelas 10, 11 e 12 as matrizes geradas pelo algoritmo k-NN:

A partir do treino do modelo preditivo através dos padrões encontrados no *dataset*, a matriz para a classificação final do diagnóstico (Tabela 10) detectou 523 registros classificados como exposição pelos serviços de saúde, mas que assumem padrão para intoxicação confirmada. Da mesma forma para intoxicação confirmada, foram detectados 376 registros que se confirmam com o padrão preditivo da exposição. Representando um aumento de 1,07 % dos casos classificados como intoxicação confirmada na base de dados.

Tabela 10. Matriz de confusão do algoritmo k-NN no dataset *classi_fin*

		Predicted						Σ
		Exposição	Ignorado	Intoxicação confirmada	Outro diag	Reação adversa	sind abstinência	
Actual	Exposição	338	11	523	0	0	0	872
	Ignorado	21	119	61	0	0	0	201
	Intoxicação confirmada	376	29	1953	1	9	0	2368
	Outro diag	4	0	19	0	0	0	23
	Reação adversa	22	0	68	0	0	0	90
	sind abstinência	1	0	1	0	0	0	2
	Σ	762	159	2625	1	9		3556

Fonte: Dados Trabalhados. SINAN- INTOX/ SES-TO. Acesso: abril de 2021.

Denota a dificuldade e erro no julgamento do diagnóstico do caso que gera sub registro de casos, não orienta as práticas corretas de tratamento e fomentam a falta de cuidado e erros com o manuseio e utilização destes agentes tóxicos.

Na tabela 11, o modelo prediz as categorias da evolução, foram registrados 316 casos como ignorado na qual assumem o padrão da categoria cura sem sequela, aumentando em 1,1% a frequência de casos de cura sem sequela. Os resultados não corroboram com os dados encontrados no SIM de mortalidade que registra 115 óbitos por envenenamento. Apesar do banco *Intox_sim* não obter acurácia relevante e métricas de associações fortes o mesmo pode ser usado para recuperar dados faltantes.

Tabela 11. Matriz de confusão do algoritmo k-NN no *dataset* evolução.

		Predicted						Σ
		Ignorado	cura c/sequela	cura s/sequela	perda de seguimento	óbito outra causa	óbito por intox	
Actual	Ignorado	15	5	316	0	0	1	337
	cura c/sequela	0	19	90	0	0	0	109
	cura s/sequela	22	14	2981	0	0	1	3018
	perda de seguimento	1	0	48	0	0	0	49
	óbito outra causa	0	0	6	0	0	0	6
	óbito por intox	0	0	37	0	0	0	37
	Σ	38	38	3478			2	3556

Fonte: Dados Trabalhados. SINAN- INTOX/ SES-TO. Acesso: abril de 2021.

Para o *dataset* de exposição (Tabela 12), 226 casos registrados como aguda repetida assumem no modelo predito a categoria aguda-única e 219 registros como ignorado poderiam ser qualificados como aguda-única. Neste modelo, o sistema possui fragilidades nos atributos para reconhecer a categoria repetida e crônica, os campos e seus dados não fornecem subsídios suficientes para predição, tornando-se dependente da anotação do profissional de saúde e do relato do paciente.

Tabela 12. Matriz de confusão do algoritmo k-NN no *dataset* exposição.

		Predicted					Σ
		Aguda-repetida	Aguda-única	Crônica	Crônica-aguda	Ignorado	
Actual	Aguda-repetida	27	226	5	0	2	260
	Aguda-única	34	2970	1	0	16	3021
	Crônica	10	14	6	0	0	30
	Crônica-aguda	4	13	2	0	0	19
	Ignorado	3	219	1	0	3	226
Σ	78	3442	15		21	3556	

Fonte: Dados Trabalhados. SINAN- INTOX/ SES-TO. Acesso: abril de 2021.

O modelo preditivo entregue pelo algoritmo k-NN demonstrou a possibilidade de qualificar os registros já existentes no banco de dados auxiliando o processo de investigação epidemiológica, como também podem ser formatados em interface de aplicativos a fim de contribuir com os profissionais da assistência no diagnóstico e nas condutas terapêuticas. Ademais, é uma ferramenta potente para análise de dados para a formulação de políticas públicas e planejamento de ações de intervenção para o controle e prevenção do agravo, desde que seja a partir de bancos de dados com qualidade que representem a realidade local. Que se faça complementações de dados através de outras fontes possíveis, como das redes sociais e dados de prontuários, para fomentar a busca do conhecimento do passado e construir predição

do futuro. Conforme Carvalho, Escobar e Tsunoda (2014) em sua revisão traz apontamentos (Figura 4) de uso da mineração de dados em saúde colaborando em diversos aspectos das políticas públicas, inclusive para a tomada de decisão.

Figura 4. Apontamentos de uso da mineração de dados na saúde.



Fonte: CARVALHO, ESCOBAR, TSUNODA (2014).

Em suma, os desafios do contexto para a produção de informação para o enfrentamento contra hegemônico do agronegócio não se resolvem com a mineração de dados, mas anuncia caminhos para desvelar conhecimento/ evidências contra as falácias produzidas cotidianamente a fim de fomentar políticas públicas baseadas em evidências (DANROCHO, 2019; VAINIO, 2020).

Os resultados observados e os produtos entregues pela experimentação de soluções em BI fornecem inovação nos processos e alento para a angústia de problemas do cotidiano dos trabalhadores que realizam a vigilância em saúde da população exposta ao agrotóxico. Produz uma reflexão e movimento para prospecção tecnológica no SUS e provoca o avanço das pesquisas em ciência de dados aplicadas à saúde.

5. CONSIDERAÇÕES FINAIS

No percurso da pesquisa, tornou-se evidente a relevância da informação em todo o processo da vigilância da população exposta ao agrotóxico. Desde priorizar os municípios para alocar recursos e ações de prevenção e promoção, quanto para subsidiar as discussões de ações, leis e projetos para a proteção da população dos riscos nefastos à saúde e ao ambiente que o uso indiscriminado e irracional dos agrotóxicos exerce sobre os territórios.

Destaca-se, ainda, a necessidade em aprofundar nos conhecimentos dos bancos de dados que produzidos pelos sistemas de saúde e assim utilizá-los, denotando a importância da interdisciplinaridade na Saúde Pública, na inserção de profissionais da Ciência de Dados no serviço de vigilância em saúde, fomentando a importância da educação permanente.

O projeto observou claramente as fragilidades de recursos tecnológicos entregues aos profissionais da vigilância epidemiológica, que tornam o trabalho moroso e impreciso. Nesse sentido, a pesquisa vem colaborar com questões e descobertas para que gestores do SUS e pesquisadores possam repensar o processo de trabalho da vigilância à população exposta ao agrotóxico considerando os seguintes fatores:

- A) A governança dos dados, que devido a fragmentação dos sistemas de informação, as áreas técnicas não possuem acesso a dados importantes para suas análises.
- B) Acesso e prospecção tecnológica: sensibilidade dos gestores para a inovação e disponibilidade de recursos e insumos tecnológicos para a execução do trabalho.
- C) Formação e inserção de trabalhadores nos serviços do SUS na perspectiva multidisciplinar para o avanço da vigilância epidemiológica e análise dos dados dentro das organizações.
- D) O olhar cartesiano e determinista da forma em que se manuseia o dado bruto, sem conhecê-lo e contextualizá-lo.
- E) Parceria intersetorial que fomenta a tomada de decisão de ações mais assertivas nos territórios.

De maneira geral, configura-se um estudo experimental de modelagem inicial para a introdução da mineração de dados na compreensão da exposição pelo agrotóxico, tão ansioso pela produção de informações que auxiliem na descoberta e enfrentamento dos impactos do uso indiscriminado de agrotóxicos pelo país.

REFERÊNCIAS

AALST, W. van der. *Process Mining*. 2. ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. v. 0. ISBN 978-3-662-49850-7.

ABE, K C; MIRAGLIA, S G El K. Avaliação de Impacto à Saúde (AIS) no Brasil e América Latina: uma ferramenta essencial a projetos, planos e políticas. *Interface - Comunicação, Saúde, Educação* [online]. 2017 [Acessado 12 Outubro 2020]. Disponível em: <<https://doi.org/10.1590/1807-57622016.0802>>. Epub 20 Jul 2017. ISSN 1807-5762.

AGUIAR, A. C. P. Más-formações congênitas, puberdade precoce e agrotóxicos: uma herança maldita do agronegócio para a Chapada do Apodi (CE). 2017. 199 f. Dissertação (Mestrado em Saúde Pública) - Faculdade de Medicina, Universidade Federal do Ceará, Fortaleza, 2017.

AGRAWAL, Rakesh; SRIKANT, Ramakrishnan. Fast algorithms for mining association rules in large databases. In: *International conference on very large data bases*, 20. 1994, San Francisco. Proceeding... San Francisco: Morgan Kaufmann, 1994. p. 478-499.

ALI, T. *et al.* Pesticide genotoxicity in cotton picking women in Pakistan evaluated using comet assay. **Drug and Chemical Toxicology**, v. 41, n. 2, p. 213–220, 3 abr. 2018.

BENFORD, D. *et al.* The principles and methods behind EFSA’s Guidance on Uncertainty Analysis in Scientific Assessment. **EFSA Journal**, v. 16, n. 1, 1 jan. 2018.

BIRGHILĂ, S. *et al.* Levels, Sources and Risk Characterization of Organochlorine Pesticides in Beer Samples from Romania. **REV.CHIM.(Bucharest)** ♦. [s.l: s.n.]. Disponível em: <<http://www.revistadechimie.ro>>. Acesso em: 20 maio. 2020.

BOCHNER, R.; FREIRE, M. M. Análise dos óbitos decorrentes de intoxicação ocorridos no Brasil de 2010 a 2015 com base no Sistema de Informação sobre Mortalidade (SIM). **Ciência & Saúde Coletiva**, v. 25, p. 761-772, 2020.

BOMBARDI, LM. **Geografia do uso de agrotóxicos no Brasil e conexões com a União Europeia**. FFLCH-USP, 2017.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Intoxicações exógenas relacionadas ao trabalho no Brasil, 2007-2016**. Boletim epidemiológico nº 58. v 49. dez. 2018. Disponível em: <https://portalarquivos2.saude.gov.br/images/pdf/2018/dezembro/26/2018-027.pdf>

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. Resolução MS/CNS nº 588, de 12 de julho de 2018. Fica instituída a Política Nacional de Vigilância em Saúde (PNVS), aprovada por meio desta resolução. Diário Oficial da República Federativa do Brasil, Brasília (DF), 2018 ago 13; Seção 1:87.

BREILH J. Una Perspectiva Emancipadora de la Investigación y Acción, Basada en la Determinación Social de la Salud. **Taller Latinoamericano de Determinantes Sociales de la Salud**, de 30 de setembro a 2 de outubro de 2008, na Universidade Autônoma do México, Cidade do México.

CAMARGO JRKR, COELI CM. Avaliação das diferentes estratégias de blocagem no relacionamento probabilístico de registros. **Rev Bras Epidemiol** 2002;5:185-96.

CAMILO, C. O.; SILVA, J. C. d. **Mineração de dados: Conceitos, tarefas, métodos e ferramentas**. Universidade Federal de Goiás (UFG), v. 1, n. 1, p. 1–29, 2009.

CARNEIRO, Fernando Ferreira (Org.) **Dossiê ABRASCO: um alerta sobre os impactos dos agrotóxicos na saúde** / Organização de Fernando Ferreira Carneiro, Lia Giraldo da Silva Augusto, Raquel Maria Rigotto, Karen Friedrich e André Campos Búrigo. Rio de Janeiro: EPSJV; São Paulo: Expressão Popular, 2015.

CARVALHO, D. R.; ESCOBAR, L. F. A; TSUNODA, D. Pontos de Atenção para o Uso da Mineração de Dados da Saúde. **Informação & Informação**, [S.l.], v. 19, n. 1, p. 249-272, fev. 2014. ISSN 1981-8920. Disponível em: <<https://www.uel.br/revistas/uel/index.php/informacao/article/view/16532/14216>>. Acesso em: 03 set. 2021. doi:<http://dx.doi.org/10.5433/1981-8920.2014v19n1p249>.

COLBORN ET AL, T., DUMANOSKI, D., MYERS, J. P. **O Futuro Roubado**. Porto Alegre: L&PM. 2002

CURVO, PIGNATTI e PIGNATTI. Morbimortalidade por câncer infanto-juvenil associada ao uso agrícola de agrotóxicos no estado de Mato Grosso, Brasil. **Cad. Saúde. Colet.** 2013; 21(1): 10-7.

DA CRUZ FERNANDES, Amarildo et al. Modelagem da utilização de agrotóxicos no Brasil: sua toxicologia e correlação quanto a intoxicações nos trabalhadores. **X Congresso Brasileiro de Gestão Ambiental**, IBEAS:2019.

DANROCHO, Leomar et al. Agrotóxicos: a ciência e o ataque aos mensageiros. **Cadernos Ibero-Americanos de Direito Sanitário**, v. 8, n. 1, p. 141-146, 2019.

DE OLIVEIRA SILVA, Sérgio Luís; COSTA, Ediná Alves. Intoxicações por agrotóxicos no estado do Tocantins: 2010–2014. **Vigilância Sanitária em Debate: Sociedade, Ciência & Tecnologia**, v. 6, n. 4, p. 13-22, 2018.

DHAR, Vasant. Ciência de dados e previsão. **Comunicações da ACM** , v. 56, n. 12, pág. 64-73, 2013.

DICLEMENTE, R. et al. Need for Innovation in Public Health Research. **American journal of public health**, v. 109, n. S2, p. S117–S120, 2019. ISSN 15410048.

DUJARDIN, B.; BOCCA, V. Cumulative dietary exposure assessment of pesticides that have chronic effects on the thyroid using SAS® software. **EFSA Journal**, v. 17, n. 9, 1 set. 2019.

ELGUETA, S. *et al.* Pesticide Residues and Health Risk Assessment in Tomatoes and Lettuces from Farms of Metropolitan Region Chile. **Molecules**, v. 25, n. 2, p. 355, 15 jan. 2020.

FERNANDES, Andréia Ribeiro Costa; MOURA, Gabriela Purificação; DE OLIVEIRA, Gisele Lopes. Intoxicação por agrotóxico: ausência de registros no Sistema de Saúde do Estado da Bahia. **Brazilian Journal of Development**, v. 7, n. 5, p. 44574-44586, 2021.

FRIEDRICH, Karen et al. Situação regulatória internacional de agrotóxicos com uso autorizado no Brasil: potencial de danos sobre a saúde e impactos ambientais. **Cadernos de Saúde Pública**, v. 37, 2021.

FUNG, I. C.-H.; TSE, Z. T. H.; FU, K.-W. Converting Big Data into public health. **Science**, v. 347, n. 6222, p. 620–620, feb 2015. ISSN 0036-8075.

FUNDAÇÃO OSWALDO CRUZ. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde. V Encontros Pré-ConfOA: dados abertos, ciência de dados aplicada à saúde, CIDACS [gravação de vídeo]. Direção: Icict. Rio de Janeiro: VideoSaúde, 2017. 1 arquivo MP4 (146 min) ntsc, son., color.

GALVÃO, N. D., MARIN, H. de F.. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem** [online]. 2009, v. 22, n. 5 Epub 26 Feb 2010. ISSN 1982-0194. <https://doi.org/10.1590/S0103-21002009000500014>.

HAKME, E. *et al.* Further improvements in pesticide residue analysis in food by applying gas chromatography triple quadrupole mass spectrometry (GC-QqQ-MS/MS) technologies. **Analytical and Bioanalytical Chemistry**, v. 410, n. 22, p. 5491–5506, 1 set. 2018.

HU, B. *et al.* Rapid nondestructive detection of mixed pesticides residues on fruit surface using SERS combined with self-modeling mixture analysis method. **Talanta**, v. 217, p. 120998, 1 set. 2020.

HUANG, T. et al. Promises and Challenges of Big Data Computing in Health Sciences. **Big Data Research**, Elsevier Inc., v. 2, n. 1, p. 2–11, 2015. ISSN 22145796. Disponível em: .

International Agency for Research on Cancer (IARC). Cancer Incidence and Mortality Worldwide: **IARC Cancer Base** N°11, v1, 2012. [Internet]. Available from: <http://globocan.iarc.fr>, accessed on 12/september/2020.

Instituto Brasileiro de Geografia e Estatística (BR). Brasil, Tocantins. <https://cidades.ibge.gov.br/brasil/mt/panorama>. Acesso em 26/09/2019.

Instituto Nacional de Câncer José Alencar Gomes da Silva (BR). Registros hospitalares de câncer: planejamento e gestão. 2a ed. **rev. Rio de Janeiro: INCA**; 2010.

JARDIM, A. N. O. *et al.* Dietary cumulative acute risk assessment of organophosphorus, carbamates and pyrethroids insecticides for the Brazilian population. **Food and Chemical Toxicology**, v. 112, p. 108–117, 1 fev. 2018.

KAPELEKA, J. A.; SAULI, E.; NDAKIDEMI, P. A. Pesticide exposure and genotoxic effects as measured by DNA damage and human monitoring biomarkersInternational. **Journal of Environmental Health Research**, Taylor and Francis Ltda., 2019. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/09603123.2019.1690132>>. Acesso em: 15 ago. 2020

KIM, K. H.; KABIR, E.; JAHAN, S. A. Exposure to pesticides and the associated human health effects. **Science of the Total Environment**, Elsevier B.V., , 1 jan. 2017.

LARA, S. S. DE; PIGNATI, W. A.; PIGNATTI, M. G.; LEÃO, L. H. DA C.; MACHADO, J.

M. H. A agricultura do agronegócio e sua relação com a intoxicação aguda por agrotóxicos no Brasil. **Hygeia - Revista Brasileira de Geografia Médica e da Saúde**, v. 15, n. 32, p. 1-19, 22 out. 2019.

LARSSON, M. O. *et al.* Refined assessment and perspectives on the cumulative risk resulting from the dietary exposure to pesticide residues in the Danish population. **Food and Chemical Toxicology**, v. 111, 2018.

LI, Z. *et al.* A monitoring survey and dietary risk assessment for pesticide residues on peaches in China. **Regulatory Toxicology and Pharmacology**, v. 97, p. 152–162, 1 ago. 2018.

LOPES, C. V. A.; ALBUQUERQUE, G. S. C. DE. Agrotóxicos e seus impactos na saúde humana e ambiental: uma revisão sistemática. **Saúde em Debate**, 2018.

LU, E. H. *et al.* Systematic probabilistic risk assessment of pesticide residues in tea leaves. **Chemosphere**, v. 247, p. 125-692, 1 maio 2020.

MAGGIONI, D. A. *et al.* National short-term dietary exposure assessment of a selected group of pesticides in Argentina. **Journal of Environmental Science and Health - Part B Pesticides, Food Contaminants, and Agricultural Wastes**, v. 53, n. 10, p. 639–651, 3 out. 2018.

MILLER, H. J. The data avalanche is here. Shouldn't we be digging? **Journal of Regional Science**, v. 50, n. 1, p. 181–201, 2010. ISSN 00224146.

Ministério da Saúde (BR). Secretaria de Vigilância Ambiental em Saúde. Departamento de Vigilância em Saúde Ambiental e Saúde do Trabalhador. **Avaliação de Impacto à Saúde – AIS: metodologia adaptada para aplicação no Brasil**. Brasília; 2014.

Ministério da Saúde (BR). **Relatório Nacional de Vigilância em Saúde de Populações Expostas a Agrotóxicos**. Brasília: [s.n.]. Disponível em: <<http://portalsaude.saude.gov.br/index>>. Acesso em: 24 ago. 2020.

Ministério da Saúde. e-SUS notifica. 2021. Disponível em: <https://datasus.saude.gov.br/notifica/>.

MORE, S. J. *et al.* Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. **EFSA Journal**, v. 17, n. 3, 1 mar. 2019.

MUSTIELES, V. *et al.* Maternal and paternal preconception exposure to phenols and preterm birth. **Environment International**, v. 137, 1 abr. 2020.

NARENDERAN, S. T.; MEYYANATHAN, S. N.; BABU, B. Review of pesticide residue analysis in fruits and vegetables. Pre-treatment, extraction and detection techniques. **Food Research International**, Elsevier Ltd, , 1 jul. 2020.

NEVES, F. R. M. (2017). Efeito dos agrotóxicos e seus metabólitos em células sanguíneas. Master's Dissertation, Faculdade de Medicina de Ribeirão Preto, University of São Paulo, Ribeirão Preto. doi:10.11606/D.17.2017.tde-07062017-152230. Retrieved 2019-11-25, from www.teses.usp.br.

NOUGADÈRE, A. *et al.* Dietary exposure to pesticide residues and associated health risks in infants and young children – Results of the French infant total diet study. **Environment International**, v. 137, p. 105529, 1 abr. 2020.

O. GOMES, H. DE *et al.* A socio-environmental perspective on pesticide use and food production. **Ecotoxicology and Environmental Safety**, v. 197, p. 110627, 1 jul. 2020.

Organisation de Coopération et de Développement Économiques (OCDE). **L'évaluation environnementale stratégique**. Guide de Bonnes Pratiques Dans le Domaine de la Coopération pour le Développement. Paris: OCDE; 2006.

Organização Mundial da Saúde (OMS). **Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde**. São Paulo: Edusp; 1995. Vol 1.

PCDaS. Plataforma de Ciência de Dados aplicada à Saúde. Laboratório de Informação em Saúde (Lis). Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict). Fundação Oswaldo Cruz (Fiocruz). url: <https://bigdata.icict.fiocruz.br>.

PIGNATI, W. A.; MACHADO, J. M. H.; CABRAL, J. F. Acidente rural ampliado: O caso das “chuvás” de agrotóxicos sobre a cidade de Lucas do Rio Verde - MT. **Ciência e Saúde Coletiva**, v. 12, n. 1, p. 105–114, jan. 2007.

PIGNATI, WA. Resumo executivo dos Impactos dos agrotóxicos na saúde e no ambiente nos municípios do “interior” de Mato Grosso, Brasil. 2014.

PIGNATI, WA; OLIVEIRA, NP; SILVA, AMC. Vigilância aos agrotóxicos: quantificação do uso e previsão de impactos na saúde-trabalho-ambiente para os municípios brasileiros. **Ciência & Saúde Coletiva**, v.19, n.12, 2014. p. 4670.

PIGNATI, W. A. *et al.* Residues, fate and risk assessment of spirotetramat and its four metabolites in pineapple under field conditions. **Human and Ecological Risk Assessment: An International Journal**, v. 12, n. 1, p. 1–11, 2015.

PIGNATI, WA, et al. Distribuição espacial do uso de agrotóxicos no Brasil: uma ferramenta para a vigilância em saúde. **Ciência Saúde Coletiva**. [Periódico na internet] (2017/Jul).

Portal Brasileiro de dados abertos: dados.org. Disponível: <<https://dados.gov.br/>>

PORTO, Marcelo Firpo; SOARES, Wagner Lopes. 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**, v. 37, p. 17-31, 2012.

RAMOS, Maria Lúcia Henrique et al. Perfil epidemiológico dos casos de intoxicação por agrotóxicos de 2013 a 2017 no Brasil. **Brazilian Journal of Development**, v. 6, n. 7, p. 43802-43813, 2020.

RELANGI, P. K. LC/LC-MS methods for detection of multiple pesticide residues and endocrine disrupting chemicals. effects of ozonization on degradation and removal. [s.l: s.n.]. 2019.

RODRIGUEZ-SAONA, L. *et al.* Miniaturization of Optical Sensors and their Potential for High-Throughput Screening of Foods. **Current Opinion in Food Science**, abr. 2020.

ROTONDO, A.; QUILLIGAN, F. Evolution Paths for Knowledge Discovery and Data Mining Process Models. **SN Computer Science**, Springer Singapore, v. 1, n. 2, p. 1–19, 2020. ISSN 2662-995X. Disponível em: <https://doi.org/10.1007/s42979-020-0117-6>.

SALDANHA, RF; PEDROSO, M.; BARCELLOS, C. Ciência de dados e big data: o que isto significa para estudos populacionais e da saúde. **Cadernos Saúde Coletiva**. 2019.

_____. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). **Cadernos de Saúde Pública [online]**. 2019, v. 35, n. 9, 2019. ISSN 1678-4464. Disponível em: <<https://doi.org/10.1590/0102-311X00032419>>.

SALDANHA, R. de F. Da aquisição à visualização de dados: aplicações da ciência de dados em saúde. 2021. 167f. Tese (Doutorado em Informação e Comunicação em Saúde) - Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2021.

SANTOS, Boaventura de Sousa. **Da ideia da universidade à universidade de ideias**. 1989.

_____. Boaventura de Sousa. **Para além do pensamento abissal: das linhas globais a uma ecologia de saberes**. Novos estudos-CEBRAP, n.79, p.71-94, 2007.

SHMUELI, G. To Explain or to Predict? **Statistical Science**, v. 25, n. 3, p. 289–310, 2010.

SIDHU, G. K. *et al.* Toxicity, monitoring and biodegradation of organophosphate pesticides: A review. **Critical Reviews in Environmental Science and Technology**, v. 49, n. 13, p. 1135–1187, 3 jul. 2019.

SIEKE, C. Probabilistic cumulative dietary risk assessment of pesticide residues in foods for the German population based on food monitoring data from 2009 to 2014. **Food and Chemical Toxicology**, v. 121, p. 396–403, 1 nov. 2018.

SILVA, I. R. DA; XAVIER, D. M.; CEZAR-VAZ, M. R. Os impactos relacionados ao uso de agrotóxicos na saúde dos trabalhadores rurais: uma revisão sistemática. **REMEA - Revista Eletrônica do Mestrado em Educação Ambiental**, v. 36, n. 3, p. 160–177, 17 dez. 2019.

SILVEIRA, M; FENNER, A L D. Avaliação de Impactos à Saúde (AIS): análises e desafios para a Vigilância em Saúde do Brasil. **Ciência & Saúde Coletiva [online]**. 2017, v. 22, n. 10 [Acessado 12 Outubro 2020] , pp. 3205-3214. Disponível em: <<https://doi.org/10.1590/1413-812320172210.18272017>>. ISSN 1678-4561. <https://doi.org/10.1590/1413-812320172210.18272017>.

SILVERIO, A. M.; PINHEIRO, P. B. A biociência dos agrotóxicos e seu impacto na saúde. **Revista Ouricuri**, v. 9, n. 2, p. 16–33, 4 jun. 2019.

SINAN. Secretaria de Estado da Saúde do Tocantins. Tabwin. Acesso em abril,2021.

SINASC. Secretaria de Estado da Saúde do Tocantins. Tabwin. Acesso em julho,2021.

SIM. Secretaria de Estado da Saúde do Tocantins. Tabwin. Acesso em julho, 2021.

STURZA, Janaína Machado; PEDRON, Aline Michele; BARRIQUELLO, Carolina Andrade.

A dinâmica de conjugação entre controle de riscos e promoção da saúde: o direito humano fundamental à saúde em tempos de sociedade globalizada. **RFD-Revista da Faculdade de Direito da UERJ**, n. 39, p. 70-89, 2021.

TEIXEIRA, Carmen Fontes. **Planejamento em saúde: conceitos, métodos e experiências**. 2010.

TEIXEIRA, Tarcisio Miguel. Saúde e direito à informação: o problema dos agrotóxicos nos alimentos. **Revista de Direito Sanitário**, v. 17, n. 3, p. 134-159, 2017.

UNIVERSIDADE FEDERAL DO TOCANTINS. Workshop em Ciência de Dados em Controle Externo. Publicado Canal UFT oficial. 4 vídeos (3h). Disponível em: <https://www.youtube.com/watch?v=5BW2pV6x-AQ>. Acesso em 19 novembro 2021.

VAINIO, H. Public health and evidence-informed policy-making: The case of a commonly used herbicide *Scandinavian Journal of Work, Environment and Health Nordic Association of Occupational Safety and Health*, 2020.

VASCONCELLOS, P. R. O. *et al.* 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**, v. 43, n. 123, p. 1084–1094, out. 2019.

World Health Organization (WHO). International Agency for Research on Cancer (IARC). **IARC Handbooks of Cancer Prevention: Breast Cancer Screening**. Lyon: IARC Press; 2002a.

_____. **Diet, nutrition and the prevention of chronic diseases: report of a Joint WHO/FAO Expert Consultation**. WHO Technical Report Series, No. 916. Geneva: World Health Organization; 2003.

WUMBEI, A. Consumption Risk Assessment of Pesticides Residues in Yam Agriculture represents an important economic sector in Ghana, accounting for about 40% of the country's gross domestic. **Tunisian Journal of Plant Protection**. [s.l.: s.n.].

APÊNDICE A – Dicionário de dados *dataset* VSPEA

Nº de notificação e campos que correspondem aos campos de 1 a 30 dos blocos “**Dados Gerais**”, “**Notificação Individual**” e “**Dados de residência**” correspondem aos mesmos campos da ficha de notificação (são comuns a todos os agravos no SINAN) e que podem ser usados para *linkage* dos demais bancos na identificação dos indivíduos e integração para seguimento do caso.

Nome do campo	Campo	Categorias	Descrição	Características	DBF
SINAN					
1. Ocupação (c.b.o)	co_cbo_ocupacao		Informar a atividade exercida pelo paciente no setor formal, informal ou autônomo ou sua última atividade exercida quando desempregado	Campo essencial	ID_OCUPA_N
2 Local de ocorrência da Exposição	tp_ocorrendia	1. Residência 2. Ambiente de Trabalho 3. Trajeto do trabalho 4. Serviços de saúde 5. Escola/creche 6. Ambiente Externo 7. Outro 9 Ignorado	Informar o local de ocorrência da exposição	Campo essencial Se campo = 1, Preencher automaticamente nos “Dados de Exposição” os campos 40, 41,42, 43, 44, 45, 46, 47 e 48 com os “Dados de Residência” da Notificação. Se campo = 7, preencher o item outro.	LOC_EXPO
3. Local de ocorrência da exposição (outro) especificar	ds_ocorrendia_outro		Especificar outro local de ocorrência da exposição	Habilitar se campo 34 – Local de ocorrência da exposição = 7 – outro	LOC_EXP_DE
4. Nome do local / estabelecimento de ocorrência	no_local_exposicao		Informar o nome do local/ estabelecimento de ocorrência do acidente/exposição	Campo essencial	NOEMPRESA

5. Código da atividade econômica (cnae)	co_cnae		Informar o nome do local/ estabelecimento de ocorrência do acidente/exposição	Campo essencial	CNAE
6. Zona de exposição	tp_zona_exposicao	1 Urbana 2 Rural 3 Periurbana 9 Ignorado	Informar a zona do local de ocorrência do acidente/ exposição	Campo essencial	ZONA_EXP
7. Grupo do agente tóxico / classificação geral	tp_agente_toxico	01 Medicamento 02 Agrotóxico/uso agrícola 03 Agrotóxico/uso doméstico 04 Agrotóxico/uso saúde pública 05 Raticida 06 Produto Veterinário 07 Produto de uso domiciliar 08 Cosmético / higiene pessoal 09 Produto químico de uso industrial 10 Metal 11 Drogas de abuso 12 Planta tóxica 13 Alimento e bebida 14 Outro 99 Ignorado	Informar o grupo do agente tóxico /classificação geral	Campo essencial Se preencher este campo = 2, 3 ou 4, habilitar os campos 51 - Se agrotóxico, qual a finalidade da utilização, 52 - Se agrotóxico, quais as atividades exercidas na exposição atual (1ª, 2ª e 3ª opção) Se campo = 14, preencher o item outro.	AGENTE_TOX

8. Grupo do agente tóxico / Classificação geral Outro (especificar)	ds_agente_toxico_outro co_ag_tox_marca ds_agente_toxico ds_principio_ativo	Especificar outro agente tóxico	Habilitar se campo 49 – Grupo do agente tóxico/ classificação geral = 14 - outro	OUT_AGENTE COAGTOXMA P_ATIVO_
9. Se agrotóxico, qual a finalidade da utilização	tp_utilizacao	1 Inseticida 2 Herbicida 3 Carrapaticida 4 Raticida 5 Fungicida 6 Preservante para madeira 7 Outro 8 Não se aplica 9 Ignorado Informar, caso o produto químico seja agrotóxico, qual a finalidade da utilização	Se no campo 49- (Grupo do agente tóxico / classificação Geral) a opção for diferente de 02 - Agrotóxico (uso Agrícola); 03 - Agrotóxico (uso doméstico) ou 04 - Agrotóxico (uso saúde pública), pular para o campo 54 (Via de exposição /contaminação) (1ª opção) Só habilitar esse campo se o campo Grupo do Agente = 02, 03 ou 04 Se campo = 7, preencher item outro.	UTILIZACAO
10. Se agrotóxico, qual a finalidade da utilização Outro (especificar)	ds_utilizacao_outro	Especificar outra finalidade	Habilitar se campo 51 - Se agrotóxico, qual a finalidade da utilização = 7 - outro	UTIL_DESC
11. Se agrotóxico, quais as atividades exercidas na exposição atual	tp_atividade_2	01 Diluição 02 Pulverização 03 Tratamento de sementes 04 Armazenagem Informar, caso o produto químico seja agrotóxico, quais as atividades exercidas	Só habilitar esse campo se o campo 49- Grupo do Agente = 02, 03 e 04 e 52.	ATIVIDA_1

		05 Colheita 06 Transporte 07 Desinsetização 08 Produção / formulação 09 Outros 10 Não se aplica 99 Ignorado	na exposição atual (2ª opção)		
12. Se agrotóxico de uso agrícola, qual a cultura / lavoura	ds_cultura_lavoura		Informar, caso o produto químico seja agrotóxico, qual a cultura/lavoura predominante,	Habilitar se campo 49 Grupo do agente tóxico / classificação geral = 2 - Agrotóxico/uso agrícola	LAVOURA
13. Via de exposição / contaminação	tp_via_contaminacao_1	1 Digestiva 2 Cutânea 3 Respiratória 4 Ocular 5 Parenteral 6 Vaginal 7 Transplacentária 8 Outra 9 Ignorado	Informar a via de exposição/ contaminação	Campo essencial	VIA_1 VIA_2
14. Circunstância da exposição / Contaminação	tp_contaminacao	01 Uso habitual 02 Acidental 03 Ambiental 04 Uso terapêutico 05 Prescrição médica inadequada 06 Erro de administração 07 Automedicação 08 Abuso	Informar em que circunstância ocorreu a exposição/ contaminação	Campo essencial se campo = 13, preencher item outra.	CIRCUNSTAN

		09 Ingestão de alimento ou bebida 10 Tentativa de suicídio 11 Tentativa de aborto 12 Violência / homicídio 13 Outra 99 Ignorado			
15. A exposição/contaminação pode ser considerado acidente de trabalho/ocupacional	st_acidente_trabalho	1 Sim 2 Não 9 Ignorado	Informe, em caso de Acidente de Trabalho, se foi feita comunicação do mesmo	Campo essencial	DOENCA_TRA
16. Tipo de exposição	tp_exposicao	1 Aguda – única 2 Aguda – repetida 3 Crônica 4 aguda sobre crônica 9 Ignorado	Informar o tipo de exposição	Campo essencial	TPEXP
17. Tempo decorrido entre a exposição e o atendimento (número / tempo)	tp_tempo	1 = (hora) 2 = (dia) 3 = (mês) 4 = (ano) 9 = Ignorado	Informar o tempo trabalho na ocupação.	Campo essencial	NUTEMPO TPTEMPO
18. Tipo de atendimento	tp_atendimento	1 Hospitalar 2 Ambulatorial 3 Domiciliar 4 Nenhum 9 Ignorado	Informar o tipo de atendimento	Campo essencial	TPATENDE
19. Classificação final	tp_classificacao_final	1 Intoxicação Confirmada 2 Exposição	Informar a classificação final do caso	CAMPO OBRIGATÓRIO	CLASSI_FIN

		3 Reação adversa 4 Diagnóstico diferencial 5 Síndrome de abstinência 9 Ignorado		quando o campo data de encerramento estiver preenchido Se campo diferente de 1, pular para o campo 67- Critério de confirmação	
20. Se intoxicação confirmada, qual o diagnóstico (cid – 10)	co_cid_diagnostico		Informar o código do cid 10, se o caso for de intoxicação confirmada	Campo essencial se campo Classificação final=1 (Intoxicação Confirmada)	DIAG_CONF
21. Critério de confirmação	tp_criterio_confirmacao	1 Clínico laboratorial 2 Clínico epidemiológico 3 Clínico	Informar o critério de confirmação do caso	Campo essencial	CRITERIO
22. Evolução do caso	tp_evolucao_caso	1 Cura sem sequela 2 Cura com sequela 3 Óbito por intoxicação exógena 4 Óbito por outra causa 5 Perda do seguimento 9 Ignorada	Informar a evolução do caso	Campo essencial Se evolução do caso for = 3 ou 4 habilitar 69-Data do óbito.	EVOLUCAO
23. Data óbito	dt_obito		Informar data do óbito	Campo essencial Se campo 68- Evolução do caso=3 (Óbito por intoxicação exógena) ou 4 (Óbito por outra causa). Data do óbito deve	DT_OBITO

	ser >= Data dos primeiros sintomas.				
24. Comunicação de acidente de trabalho (cat)	tp_comunicacao_acidente_trab	1 Sim 2 Não 3 Não se aplica 9 Ignorado	Informe, em caso de Acidente de Trabalho(CAT), se foi feita comunicação do mesmo	Campo essencial	CAT
25. Informações complementares e observações	ds_observacao				Dado disponível pela consulta do Sinan, variável não exportada.
26. Principal agente causador da Dermatose	tp_agente_causador	01 – Cimento 02 – Borracha 03 – Plástico 04 – Solventes Orgânicos 05 – Graxas 06 – Óleo de Corte 07 – Resinas 08 – Níquel 09 – Cosméticos 10 – Madeiras 11 – Cromo 12 – Outros 99 – Ignorado	Principal Agente causador da Dermatose	Campo obrigatório	AGENTE
27. Tempo de exposição ao agente de risco	nu_tempo_exposicao_risco tp_tempo_exposicao_risco	H (hora) D(dias) M(meses) A(anos)			NUTEMPORIS TPTEMPORIS
28. Localização da lesão (parte do corpo atingida)	tp_local_lesao	01 – Mão 02 – Membro Superior 03 – Cabeça	Localização da lesão	Campo Essencial	LESAO

		04 – Pescoço 05 – Tórax 06 – Abdome 07 – Membro inferior 08 – Pé 09 – Todo o corpo 10 – Outro 99 – Ignorado			
29. Diagnostico Específico CID 10	co_cid_diagnostico		Diagnóstico específico - CID 10	Campo Obrigatório	DIAG_ESP
30. Com afastamento do trabalho	tp_afastamento_trabalho	1- melhora 2- piora 9- ignorado	Evolução da situação de saúde do paciente com afastamento do trabalho	Campo Essencial	EVOL_AFAST
31. Há ou houve Outros trabalhadores com a mesma doença no local de trabalho	st_trabalhador_mesma_doenca	1- sim 2- não 9- ignorado	Outros trabalhadores sofreram ou não o mesmo agravo no local de trabalho	Campo Essencial	TRAB_DOE
32. Evolução derm.	tp_evolucao_caso	1- Cura 2- Cura não confirmada 3- Incapacidade temporária 4- Incapacidade permanente parcial 5- Incapacidade permanente total 6- Óbito por doença relacionada ao trabalho 7- Óbito por outra causa 8- Outro	Como evoluiu a situação do paciente	Campo Essencial Se campo <> 6 ou 7, pular para campo 60 (Foi feita comunicação de acidente do trabalho).	EVOLUCAO

	9- Ignorado			
33. Informações complementares e observações	ds_observacao		Informar observações a respeito do caso se necessário.	DS_OBS
34. Tempo de exposição ao agente de risco PAIR	nu_tempo_exposicao_risco tp_tempo_exposicao_risco	H (hora) D(dias) M(meses) A(anos)		Campo Essencial NUTEMPORIS TPTEMPORIS
35. Exposição Concomitante a Ruído e Solvente a Base de Tolueno	st_exposicao_solvente	1 – Sim 2 – Não 9 – Ignorado	Houve exposição concomitante a ruído e Solvente a Base de Tolueno	Campo Essencial SOLVENTE
36. Exposição Concomitante a Ruído e Metais pesados	st_exposicao_metal	1 – Sim 2 – Não 9 – Ignorado	Houve exposição concomitante a ruído e Meais Pesados	Campo Essencial METAL
37. Exposição Concomitante a Ruído e Outros Se Outros Especificar	ds_exposicao_outro		Houve exposição concomitante a ruído e outros produto/outra situação especificar	Campo Essencial Habilitar se campo Exposição Concomitante a Ruído e Outros=1(sim) RUI_OUTDES
38. . Sintomas Zumbido	st_sintoma_tontura	1 – Sim 2 – Não 9 – Ignorado	Sintomas Zumbido	Campo Essencial ZUMBIDO
39. Sintomas Tontura	st_sintoma_tontura	1 – Sim 2 – Não 9 – Ignorado	Sintomas Tontura	Campo Essencial TONTURA
40. Sintomas Dificuldade para compreensão da fala	st_sintoma_dificuldade_fal a	1 – Sim 2 – Não 9 – Ignorado	Sintomas Dificuldade para compreensão da fala	Campo Essencial FALA
41. Sintomas Cefaléia	st_sintoma_cafaleia	1 – Sim 2 – Não 9 – Ignorado	Sintomas Cefaléia	Campo Essencial CEFALEIA
42. Diagnostico Específico CID 10	co_cid_diagnostico		Diagnóstico específico - CID 10	Campo Obrigatório DIAG_ESP

43. Há ou houve Outros trabalhadores com a mesma doença no local de trabalho	st_trabalhador_mesma_doenca	1- sim 2- não 9- ignorado	Outros trabalhadores sofreram ou não o mesmo agravo no local de trabalho	Campo Essencial	TRAB_DOE
44. Evolução do caso PAIR	tp_evolucao_caso	1- Cura 2- Cura não confirmada 3- Incapacidade temporária 4- Incapacidade permanente parcial 5- Incapacidade permanente total 6- Óbito por doença relacionada ao trabalho 7- Óbito por outra causa 8- Outro 9- Ignorado	Como evoluiu a situação do paciente	Campo Essencial Se campo < 6 ou 7, pular para campo 60 (Foi feita comunicação de acidente do trabalho).	EVOLUCAO
45. Houve exposição nos locais de trabalho, durante toda a sua vida profissional, a algum dos itens abaixo relacionados? Aminas Aromáticas	st_expamina_aromatica	1 –Sim 2 –Não 9 –Ignorado	Ocorreu exposição nos locais de trabalho, durante toda a sua vida profissional Aminas Aromáticas.	Campo Essencial	AMINA
46. Houve exposição nos locais de trabalho, durante toda a sua vida profissional, a algum dos itens abaixo relacionados? Hidrocarbonetos alifáticos ou aromáticos (seus derivados halogenados)	st_exp_hidrocarboneto	1 –Sim 2 –Não 9 –Ignorado	Ocorreu exposição nos locais de trabalho, durante toda a sua vida profissional Hidrocarbonetos alifáticos ou aromáticos (seus derivados halogenados)	Campo Essencial	HIDROCARBO

47. Houve exposição nos locais de trabalho, durante toda a sua vida profissional, a algum dos itens abaixo relacionados? Óleos Minerais	st_exp_oleo_mineral	1 –Sim 2–Não 9 – Ignorado	Ocorreu exposição nos locais de trabalho, durante toda a sua vida profissional Óleos Minerais	Campo Essencial	OLEOS
48. Houve exposição nos locais de trabalho, durante toda a sua vida profissional, a algum dos itens abaixo relacionados? Outros (especificar)	ds_exp_outro		Ocorreu exposição nos locais de trabalho, durante toda a sua vida profissional Outros (Especificar)	Campo Essencial	OUT_EXP_DE
49. Diagnóstico específico Câncer RT	co_cid_diagnóstico	CID 10		Campo Essencial	DIAG_ESP
50. Evolução do caso Câncer	tp_evolucao_caso	1- Cura 2- Cura não confirmada 3- Incapacidade temporária 4- Incapacidade permanente parcial 5- Incapacidade permanente total 6- Óbito por doença relacionada ao trabalho 7- Óbito por outra causa 8- Outro 9- Ignorado	Como evoluiu a situação do paciente	Campo Essencial Se campo <> 6 ou 7, pular para campo 60 (Foi feita comunicação de acidente do trabalho).	EVOLUCAO
SINASC					
1.Município de residência da parturiente	ID_MUN_aborto	Categórico (IBGE)	Indica o município em que a mulher que possuem	Campo Obrigatório	

			filhos com anomalias ou aborto reside	
2. Quantidade de abortos	QTDFILMORT	Numérico	Indica a quantidade de abortos já existentes até este filho com anomalia presente.	Campo Essencial
3. Ocupação da parturiente	OCUPMAE	Categórico (CBO)	Indica a ocupação atual	Campo Essencial
4. Idade da parturiente	IDADEMAE	Numérico	Indica a idade da parturiente	Campo Obrigatório
SIM				
1. Município do óbito	ID_MUN_OBIT	Categórico (IBGE)	Indica o município em que o óbito por intoxicação por agrotóxico	Campo Obrigatório
2. Data do óbito	DTOBITO	Numérico	Indica a data do óbito	Campo Obrigatório
3. Idade em que ocorreu o óbito	IDADE_obito	Numérico	Indica a idade do individuo na data do óbito	Campo Essencial
4. Ocupação	OCUP_OBITO	Categórico (CBO)	Indica a ocupação atual	Campo Essencial