Dataset Reports 3:1 (2024)
ISSN 2965-1182

RESEARCH ARTICLE

DATASET REPORTS

Identificação e mapeamento de hotspots de acidentes de trabalho no Brasil utilizando técnicas de machine learning e análise espacial

Identification and mapping of workplace accident hotspots in Brazil using machine learning and spatial analysis techniques

Raphael Hendrigo de Souza Gonçalves [a*], Wendel Marcos dos Santos [a]

[a] Especialização em Engenharia e Segurança do Trabalho, Faculdade Unyleya, 70711-040, Brasilia, DF, Brasil.
[b] Departamento de Informática e Turismo, Instituto Federal de Educação Ciência e Tecnologia de São Paulo, 01109-010, São Paulo, SP, Brasil.

Resumo
Este estudo tem como objetivo principal identificar e analisar os hotspots de acidentes de trabalho no Brasil, com foco em setores críticos como construção civil, transporte rodoviário, mineração, e energia elétrica. A metodologia adotada envolve a aplicação de algoritmos de machine learning, especificamente K-means, DBSCAN, HDBSCAN e Agglomerative Clustering, para a clusterização de dados de acidentes fornecidos pelo INSS. Além disso, foram utilizadas técnicas de análise espacial com o auxílio de ferramentas GIS para mapear e visualizar as áreas de maior incidência de acidentes. Os resultados revelaram que a maioria dos acidentes está concentrada em regiões metropolitanas, especialmente nas regiões Sudeste e Sul do Brasil. Os algoritmos de clusterização permitiram identificar padrões de risco em diferentes setores, destacando a falta de treinamento adequado e a não utilização de equipamentos de proteção individual (EPI) como fatores críticos. A análise espacial possibilitou a visualização clara dos hotspots, oferecendo subsídios para a formulação de políticas de segurança mais eficazes e direcionadas. Conclui-se que a combinação de técnicas de machine learning com análise espacial é uma abordagem poderosa para a identificação de hotspots de acidentes de trabalho, contribuindo significativamente para a redução de riscos e a promoção de ambientes de trabalho mais seguros. O estudo abre possibilidades para pesquisas futuras que integrem variáveis socioeconômicas e culturais na análise de acidentes de trabalho.

Palavras-chave: Pontos críticos de acidentes. Aprendizado de máquina. Análise espacial. Segurança ocupacional. Agrupamento.

Abstract
This study aims to identify and analyze workplace accident hotspots in Brazil, focusing on critical sectors such as construction, road transportation, mining, and electric energy. The adopted methodology involves applying machine learning algorithms, specifically K-means, DBSCAN, HDBSCAN, and Agglomerative Clustering, for clustering accident data provided by INSS. Additionally, spatial analysis techniques were used with GIS tools to map and visualize areas with the highest incidence of accidents. The results revealed that most accidents are concentrated in metropolitan regions, particularly in the Southeast and South of Brazil. The clustering algorithms identified risk patterns across different sectors, highlighting inadequate training and the non-use of personal protective equipment (PPE) as critical factors. Spatial analysis provided a clear visualization of hotspots, offering insights into the formulation of more effective and targeted safety policies. It is concluded that combining machine learning techniques with spatial analysis is a powerful approach for identifying workplace accident hotspots, significantly contributing to risk reduction and the promotion of safer work environments. The study opens possibilities for future research that integrates socioeconomic and cultural variables into workplace accident analysis.

Keywords: Accident hotspots. Machine learning. Spatial analysis. Occupational safety. Clustering.

Graphical Abstract



*Corresponding author: Wendel M. dos Santos. E-mail address: wendel.santos@ifsp.edu.br
Submitted: 10 September 2024; Accepted: 17 October 2024; Published: 17 October 2024.
© The Author(s) 2024. Open Access (CC BY 4.0).

1. Introdução

A segurança no ambiente de trabalho é um dos pilares fundamentais para o bem-estar dos trabalhadores e para o funcionamento eficiente das empresas. Em setores de alta complexidade e risco, como a construção civil, mineração, transporte rodoviário e energia elétrica, a necessidade de garantir um ambiente seguro é ainda mais crucial (Jaafar et al., 2018; Oliveira et al., 2010; Peinado, 2019). Esses setores são caracterizados por atividades que envolvem altos riscos operacionais, uso intensivo de maquinário pesado, exposição a substâncias perigosas e condições de trabalho muitas vezes adversas (Filgueiras, 2017). No Brasil, a questão da segurança no trabalho adquire uma dimensão ainda mais preocupante quando se observa a alta incidência de acidentes de trabalho, muitos dos quais resultam em lesões graves, incapacidades permanentes e até mortes (Menezes & Dal Magro, 2023; Ministério da Previdência Social, 2024; Soares, 2019).

As estatísticas brasileiras de acidentes de trabalho são alarmantes e refletem a urgência de intervenções mais eficazes. De acordo com dados do Ministério da Previdência Social e do Instituto Nacional do Seguro Social (INSS), milhares de acidentes são registrados anualmente, resultando em perdas humanas e econômicas significativas (Ministério da Previdência Social, 2024). Esses acidentes não apenas afetam diretamente a vida dos trabalhadores e suas famílias, mas também impõem um ônus financeiro considerável às empresas e ao sistema público de saúde. Além disso, a frequência e a gravidade desses acidentes podem comprometer a reputação das empresas, afetar a moral dos trabalhadores e reduzir a produtividade (Menezes & Dal Magro, 2023; Simonelli et al., 2016).

Diante desse cenário, a adoção de medidas preventivas eficazes torna-se imperativa. No entanto, as abordagens tradicionais de segurança no trabalho, baseadas principalmente em inspeções regulares e treinamentos, têm se mostrado insuficientes para lidar com a complexidade e a dinâmica dos ambientes de alto risco. É nesse contexto que a aplicação de tecnologias avançadas, como técnicas de machine learning e análise espacial, emerge como uma solução promissora. Essas tecnologias permitem uma análise mais profunda e detalhada dos dados de acidentes, possibilitando a identificação de padrões e a previsão de áreas de risco com maior precisão (Cunha et al., 2024; Abdullah & Sofyan, 2023; Babalola et al., 2023).

Este estudo propõe uma abordagem inovadora para a prevenção de acidentes de trabalho, centrada na identificação de hotspots - áreas geográficas ou setores específicos com alta concentração de acidentes. Utilizando algoritmos de machine learning, como K-means, DBSCAN, HDBSCAN, e Agglomerative Clustering, em conjunto com técnicas de análise espacial, busca-se não apenas identificar essas áreas de risco, mas também fornecer insights que possam ser utilizados na formulação de políticas de segurança mais direcionadas e eficazes . Estudos recentes mostram que a aplicação de machine learning e análise espacial na segurança do trabalho é uma tendência crescente, principalmente em setores de alta complexidade, como a mineração e a construção civil (Akalonu et al., 2017; Akay et al., 2021; Deng et al., 2020; Lima et al., 2023).

A importância deste estudo reside na sua capacidade de fornecer um suporte técnico-científico para a tomada de decisões em segurança do trabalho. Ao mapear os hotspots de acidentes, este trabalho contribui para a compreensão dos fatores que levam à ocorrência de acidentes em setores de alta complexidade, permitindo que empresas e órgãos reguladores adotem medidas preventivas mais adequadas. As tecnologias emergentes, como a análise de big data e machine learning, têm sido aplicadas em várias indústrias para a identificação de riscos ocupacionais, possibilitando uma abordagem mais preditiva e eficiente (Babalola et al., 2023).

Além disso, ao integrar machine learning e análise espacial, esta pesquisa demonstra o potencial dessas tecnologias em transformar a maneira como a segurança do trabalho é gerida no Brasil, oferecendo uma solução baseada em dados que pode ser replicada em diferentes contextos e setores. Pesquisas semelhantes mostram que a combinação dessas ferramentas resulta em maior precisão na previsão de riscos, facilitando a criação de ambientes de trabalho mais seguros (Akay et al., 2021). Em particular, setores como energia elétrica e mineração têm se beneficiado significativamente dessas inovações (Kemajl et al., 2024; Queiroz et al., 2023; Souza et al., 2023).

Assim, este estudo não apenas reforça a relevância da segurança no trabalho como um tema central na engenharia e gestão de empresas, mas também explora novas fronteiras no uso de tecnologias emergentes para enfrentar desafios antigos e persistentes no ambiente de trabalho. Acredita-se que a implementação das soluções propostas neste estudo contribuirá para a redução do número de acidentes de trabalho, promovendo ambientes de trabalho mais seguros e saudáveis, e, em última análise, preservando vidas e recursos.

O tema deste trabalho se fundamenta na profunda relevância da segurança no trabalho, especialmente no contexto brasileiro, onde a alta incidência de acidentes de trabalho continua a ser um desafio crítico. A segurança no trabalho não apenas afeta diretamente a saúde e o bem-estar dos trabalhadores, mas também tem implicações econômicas significativas, impactando a produtividade (Schwambach et al., 2024), gerando custos elevados com tratamentos médicos, indenizações e afastamentos, e, em última análise, afetando a competitividade das empresas no mercado (Jozan et al., 2023; Liang & Cheng, 2023). Em setores de alta complexidade e risco, como a construção civil, mineração, transporte rodoviário e energia elétrica, a segurança no trabalho assume uma importância ainda maior devido às condições adversas e aos riscos inerentes às atividades desempenhadas (Carvalho et al., 2020; Sousa & Rodolpho, 2020).

As estatísticas recentes indicam que o Brasil enfrenta uma realidade alarmante em relação aos acidentes de trabalho, com milhares de casos registrados anualmente, muitos dos quais resultam em lesões graves, incapacidades permanentes ou mortes (Menezes & Dal Magro, 2023; Ministério da Previdência Social, 2024). Esta situação reflete a necessidade urgente de desenvolver e implementar estratégias de prevenção mais eficazes, capazes de mitigar os riscos e proteger a vida dos trabalhadores.

Nesse contexto, a utilização de técnicas avançadas de machine learning e análise espacial apresenta-se como uma abordagem promissora para enfrentar este desafio. Diferentemente dos métodos tradicionais de análise de segurança, que muitas vezes se limitam a uma visão retrospectiva e descritiva dos acidentes, essas técnicas permitem uma análise preditiva e mais detalhada dos dados, facilitando a identificação de padrões de risco que podem passar despercebidos em abordagens convencionais. Os algoritmos de machine learning, como K-means, DBSCAN, HDBSCAN, e Agglomerative Clustering, são capazes de processar grandes volumes de dados e identificar clusters ou agrupamentos que indicam áreas ou atividades com maior propensão a acidentes (Miraftabzadeh et al., 2023; Stewart & Al-Khassaweneh, 2022).

A análise espacial, por sua vez, permite a visualização geográfica dos dados, identificando hotspots de acidentes e facilitando a compreensão espacial dos riscos. Ao integrar essas duas abordagens, este trabalho propõe uma forma inovadora de mapear e entender a distribuição dos acidentes de trabalho no

Brasil, fornecendo insights valiosos para a formulação de políticas públicas e estratégias empresariais de prevenção. A possibilidade de prever onde e como os acidentes podem ocorrer abre novas frentes para a implementação de medidas preventivas mais direcionadas e eficazes, que podem salvar vidas e reduzir os custos associados aos acidentes (Mendonça et al., 2017; Rosa, 2011).

Além do impacto prático e imediato que este estudo pode ter na melhoria da segurança do trabalho, ele também contribui para a literatura acadêmica ao explorar a aplicação de técnicas de machine learning e análise espacial em um campo onde essas tecnologias ainda são pouco utilizadas (Abdullah & Sofyan, 2023). Este estudo, portanto, não só preenche uma lacuna existente na pesquisa acadêmica, mas também oferece um modelo que pode ser replicado em diferentes setores e regiões, ampliando seu alcance e impacto.

Dessa forma, o objetivo desta pesquisa é identificar e analisar hotspots de acidentes de trabalho no Brasil utilizando técnicas de machine learning e análise espacial, com o intuito de fornecer subsídios para a criação de políticas de segurança mais eficazes. Para isso, foram aplicados algoritmos de clusterização (K-means, DBSCAN, HDBSCAN e Agglomerative Clustering) para agrupar dados de acidentes de trabalho fornecidos pelo INSS; utilizaram-se ferramentas de análise espacial para mapear e visualizar áreas com maior incidência de acidentes de trabalho; identificaram-se padrões de risco em setores críticos, como construção civil, mineração, transporte rodoviário e energia elétrica; e propuseram-se medidas preventivas baseadas nos resultados obtidos, visando à redução do número de acidentes de trabalho.

2. Metodologia

Este trabalho caracteriza-se por uma abordagem quantitativa, uma vez que se baseia na análise de dados numéricos e na aplicação de métodos estatísticos e computacionais para identificar padrões e correlações. A pesquisa quantitativa é adequada neste contexto, pois permite a análise objetiva e sistemática de grandes volumes de dados, facilitando a generalização dos resultados e a identificação de tendências em ampla escala (Silva et al., 2014).

A pesquisa quantitativa aqui utilizada inclui a análise exploratória de dados (EDA), que foi realizada para compreender a distribuição e as características dos dados de acidentes de trabalho. Essa etapa preliminar foi essencial para a preparação dos dados e para a escolha das técnicas de análise que seriam aplicadas posteriormente. A escolha por uma abordagem quantitativa foi motivada pela natureza dos dados disponíveis, que incluem informações detalhadas sobre milhares de acidentes de trabalho registrados no Brasil, abrangendo diferentes setores, tipos de acidentes e dados geográficos.

A abordagem quantitativa foi complementada por aspectos qualitativos na interpretação dos resultados, especialmente ao considerar os contextos setoriais e regionais específicos dos acidentes analisados. Essa combinação permitiu uma visão mais holística dos dados, onde os números foram contextualizados dentro das realidades operacionais dos setores analisados.

2.1 Técnicas de Coleta de Dados

A coleta de dados foi realizada a partir de fontes secundárias, utilizando como base principal o dataset de acidentes de trabalho fornecido pelo Instituto Nacional do Seguro Social (INSS). Este dataset é considerado uma das fontes mais abrangentes e detalhadas disponíveis no Brasil, cobrindo um longo período e diferentes setores econômicos. Os dados incluem informações sobre o tipo de acidente, a gravidade, a localização geográfica, as características dos trabalhadores envolvidos (como idade e gênero), e o setor de atividade.

A escolha deste dataset foi fundamentada pela sua relevância e abrangência, bem como pela confiabilidade dos dados, que são sistematicamente coletados e reportados pelas autoridades brasileiras. Além disso, o dataset permite uma análise longitudinal dos dados, possibilitando o estudo de tendências ao longo do tempo e a avaliação do impacto de intervenções e políticas de segurança.

Os dados coletados passaram por um processo rigoroso de limpeza e preparação, incluindo a remoção de duplicatas, correção de inconsistências e tratamento de valores ausentes. Em particular, a normalização dos dados foi necessária para garantir que as variáveis pudessem ser comparadas e analisadas de forma adequada. Por exemplo, dados categóricos como o tipo de acidente foram codificados usando técnicas de codificação de variáveis (como One-Hot Encoding), enquanto dados numéricos foram normalizados para eliminar discrepâncias de escala que poderiam afetar os resultados das análises subsequentes.

2.2. Técnicas de Análise de Dados

A análise dos dados foi realizada em várias etapas, cada uma utilizando técnicas específicas de machine learning e análise espacial para atingir os objetivos propostos. As técnicas escolhidas foram selecionadas com base na natureza dos dados e nos resultados esperados. A seguir, detalham-se as principais técnicas e métodos empregados.

2.2.1 Algoritmos de Clusterização

A clusterização é uma técnica de machine learning não supervisionada utilizada para agrupar dados em clusters ou grupos com base em suas características similares (Abdullah & Sofyan, 2023; Mutlu et al., 2023; Tokuda et al., 2022). Este trabalho utilizou vários algoritmos de clusterização para identificar padrões nos dados de acidentes de trabalho, conforme detalhado na sequência.

• **K-means**: Este algoritmo foi utilizado como ponto de partida para a análise. K-means particiona os dados em um número pré-definido de clusters, minimizando a variação dentro de cada cluster e maximizando a diferença entre os clusters. Este método foi escolhido por sua simplicidade e eficiência, especialmente em conjuntos de dados grandes como o utilizado neste estudo. No entanto, uma limitação do K-means é que ele assume que os clusters são esféricos e de tamanho semelhante, o que pode não ser o caso em dados reais de acidentes de trabalho, onde os padrões de risco podem variar significativamente entre diferentes setores e regiões (Akay et al., 2021; F. Deng et al., 2020; Miraftabzadeh et al., 2023).

• **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise): Este algoritmo foi aplicado para identificar clusters baseados na densidade dos dados, o que é particularmente útil em contextos em que os acidentes ocorrem com maior frequência em certas áreas geográficas. DBSCAN é menos sensível à forma dos clusters e é capaz de identificar outliers, que podem representar áreas ou condições de trabalho com alto risco de acidentes. A aplicação do DBSCAN permitiu a identificação de regiões de alta densidade de acidentes (hotspots), que foram posteriormente analisadas em detalhes (Deng, 2020).

• **HDBSCAN** (Hierarchical DBSCAN): HDBSCAN é uma extensão hierárquica do DBSCAN que foi utilizada para identificar clusters com densidade variável . Isso foi especialmente

útil na análise de setores industriais com diferentes níveis de risco, permitindo a identificação de subclusters dentro de grandes grupos (Stewart & Al-Khassaweneh, 2022).

• **Agglomerative Clustering**: Este algoritmo de clusterização hierárquica foi utilizado para criar uma hierarquia de clusters, permitindo a análise em diferentes níveis de granularidade. Agglomerative Clustering agrupa os dados de maneira iterativa, começando com cada ponto de dados como seu próprio cluster e fundindo-os com os clusters mais próximos até que todos os pontos estejam em um único cluster. Este método é útil para entender como os clusters se relacionam entre si e para identificar padrões complexos de risco que podem ser observados em diferentes níveis hierárquicos. Este método é útil para entender como os clusters se relacionam entre si e para identificar padrões complexos de risco que podem ser observados em diferentes níveis hierárquicos (Ackermann et al., 2014; Tokuda et al., 2022).

2.2.2. Análise Espacial

A análise espacial foi realizada utilizando Sistemas de Informação Geográfica (GIS), que permitiram a visualização e análise geográfica dos dados de acidentes de trabalho. A integração dos resultados da clusterização com os dados geoespaciais facilitou a identificação de hotspots e a compreensão da distribuição espacial dos riscos.

• **Mapas de Calor** (Heatmaps): Os mapas de calor foram gerados para visualizar a concentração de acidentes de trabalho em diferentes regiões do Brasil. Esses mapas destacam as áreas com maior densidade de acidentes, permitindo a identificação de hotspots e a priorização de intervenções de segurança.

• **Análise de Interpolação Espacial**: Foi realizada a interpolação espacial para estimar os níveis de risco em áreas sem dados disponíveis. Essa técnica permitiu preencher lacunas nos dados e fornecer uma visão mais completa da distribuição geográfica dos riscos de acidentes de trabalho. Métodos de interpolação, como o Kriging, foram aplicados para gerar superfícies contínuas de risco, que foram analisadas em conjunto com os resultados da clusterização.

• **Estatísticas Espaciais**: Técnicas de estatísticas espaciais, como o Índice de Moran, foram utilizadas para avaliar a autocorrelação espacial dos acidentes de trabalho, ajudando a identificar padrões espaciais significativos. Essas técnicas complementaram os resultados da clusterização e forneceram insights adicionais sobre a distribuição geográfica dos riscos. O uso do Índice de Moran, em particular, permitiu identificar áreas com maior concentração de acidentes, fornecendo uma base robusta para a análise espacial e sugerindo regiões prioritárias para intervenções de segurança.

2.2.3. Validação dos Resultados

A validação dos resultados obtidos foi realizada utilizando várias métricas e técnicas para garantir a robustez e a confiabilidade das conclusões.

• **Índice de Silhueta**: O Índice de Silhueta foi utilizado para avaliar a coesão e a separação dos clusters identificados. Tal métrica fornece uma medida de quão bem os dados estão agrupados, com valores próximos de 1 indicando boa coesão dentro dos clusters e boa separação entre eles. O Índice de Silhueta foi calculado para cada uma das técnicas de clusterização aplicadas, permitindo a comparação de seu desempenho.

• **Validação Cruzada**: A validação cruzada foi aplicada para avaliar a robustez dos modelos de clusterização. Este método divide os dados em subconjuntos, treinando os modelos em uma parte dos dados e testando-os na outra. A validação cruzada é importante para garantir que os resultados não sejam específicos para o conjunto de dados original e possam ser generalizados para outros contextos.

• **Comparação com Estudos Anteriores**: Os resultados obtidos foram comparados com os achados de estudos anteriores na área de segurança do trabalho. Essa comparação ajudou a verificar a consistência dos resultados e a contextualizar as conclusões dentro do corpo existente de conhecimento. Além disso, a comparação com estudos anteriores permitiu identificar novas contribuições do presente estudo e áreas onde os resultados divergem das pesquisas existentes.

3. Resultados e Discussão

3.1 Análise de Dados

Esta seção se concentra na apresentação detalhada dos dados coletados, na aplicação das técnicas de análise escolhidas e na interpretação dos resultados obtidos. A análise dos dados é fundamental para compreender os padrões de acidentes de trabalho e identificar hotspots em setores críticos, fornecendo insights valiosos para a formulação de políticas de segurança.

3.2 Apresentação dos Dados Coletados

A análise de dados foi conduzida utilizando um conjunto de dados extenso, obtido do Instituto Nacional do Seguro Social (INSS), que documenta os acidentes de trabalho ocorridos no Brasil ao longo dos anos. Este dataset inclui informações detalhadas sobre cada acidente, como o tipo de acidente, gravidade, localização geográfica, setor de atividade (CNAE), e características demográficas dos trabalhadores envolvidos. A preparação dos dados envolveu uma série de etapas, desde a limpeza e normalização dos dados até a aplicação de técnicas de codificação para transformar variáveis categóricas em formatos adequados para análise.

3.3. Limpeza e Preparação dos Dados

Antes de realizar a análise, foi necessário preparar os dados, o que incluiu a remoção de duplicatas, tratamento de valores ausentes e normalização das variáveis. A limpeza dos dados é uma etapa crítica, pois garante a integridade e a qualidade dos dados que serão analisados. Os dados geográficos (latitude e longitude) foram verificados quanto à precisão e consistência, uma vez que a análise espacial depende da exatidão dessas coordenadas. Além disso, dados categóricos, como o tipo de acidente e o setor de atividade, foram codificados utilizando técnicas como One-Hot Encoding, que transformam essas categorias em variáveis binárias, permitindo sua inclusão em algoritmos de machine learning.

3.4. Distribuição dos Acidentes por Setor

Uma análise inicial dos dados revelou que os acidentes de trabalho estão distribuídos de forma desigual entre os diferentes setores econômicos. Setores como a construção civil, mineração, transporte rodoviário e energia elétrica apresentaram uma incidência significativamente maior de acidentes, confirmando as expectativas baseadas na revisão da literatura.

• **Construção Civil**: Este setor foi responsável por uma proporção significativa dos acidentes registrados, especialmente aqueles relacionados a quedas de altura, manuseio de equipamentos pesados e exposição a materiais perigosos. A

alta incidência de acidentes neste setor destaca a necessidade de intervenções específicas, como treinamento adequado e supervisão rigorosa.

• **Mineração**: Os dados confirmaram que a mineração é um dos setores mais perigosos, com uma alta taxa de acidentes fatais. Os acidentes neste setor estão frequentemente relacionados à instabilidade do solo, explosões e exposição a substâncias tóxicas. A análise dos dados de mineração sugere que a gestão de riscos integrada e o uso de tecnologias avançadas de monitoramento podem ser essenciais para reduzir os incidentes.

• **Transporte Rodoviário**: Este setor também apresentou uma alta incidência de acidentes, muitos dos quais resultaram de colisões e acidentes relacionados à fadiga dos motoristas. A análise dos dados deste setor sugere que a implementação de políticas de descanso obrigatório e monitoramento de rotas pode ajudar a mitigar esses riscos.

• **Energia Elétrica**: O setor de energia elétrica mostrou uma alta taxa de acidentes relacionados a choques elétricos e queimaduras graves. A análise dos dados indica que o uso inadequado de equipamentos de proteção e a falta de treinamento são fatores contribuindo para esses incidentes.

3.5. Análise dos Dados Coletados

A análise dos dados foi realizada em várias etapas, utilizando diferentes técnicas de machine learning e análise espacial para identificar padrões e hotspots de acidentes de trabalho. A seguir, detalha-se a aplicação de cada técnica e os resultados obtidos.

3.5.1. Clusterização com K-means

O algoritmo K-means foi o primeiro a ser aplicado aos dados, com o objetivo de identificar clusters de acidentes de trabalho com base em características comuns, como o tipo de acidente e a localização geográfica. O K-means foi configurado para criar um número pré-definido de clusters, e os resultados indicaram a formação de grandes clusters em áreas metropolitanas, onde a concentração de trabalhadores é maior e, consequentemente, o risco de acidentes é mais elevado.

Os clusters identificados pelo K-means destacaram regiões como São Paulo, Rio de Janeiro e Minas Gerais como áreas de alto risco, especialmente em setores como a construção civil e mineração. A análise dos clusters revelou que, nessas áreas, os acidentes tendem a ocorrer em ambientes de trabalho com alta densidade de trabalhadores, condições perigosas e falta de equipamentos de segurança adequados (Konzen et al., 2023).

Um dos principais desafios enfrentados com o uso do K-means foi a sua sensibilidade à escolha do número de clusters (k) e à distribuição dos dados. Embora o K-means tenha sido eficaz em identificar grandes grupos de acidentes, ele se mostrou menos eficiente na identificação de clusters menores ou de formas irregulares, que são comuns em setores como mineração e energia elétrica.

3.5.2. Análise com DBSCAN

Para superar as limitações do K-means, o algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) foi aplicado aos dados. O DBSCAN é particularmente adequado para identificar clusters de densidade variável e outliers, que podem representar áreas ou condições de trabalho com alto risco de acidentes.

Os resultados do DBSCAN foram reveladores, identificando vários clusters de alta densidade em regiões industriais, onde os acidentes são mais frequentes devido à natureza das atividades desenvolvidas. O DBSCAN também identificou outliers em áreas remotas, onde os acidentes são menos frequentes, mas tendem a ser mais graves. Esses outliers foram posteriormente analisados em detalhes, revelando que muitos desses incidentes estavam relacionados a condições extremas de trabalho e à falta de acesso rápido a serviços de emergência.

Uma das vantagens do DBSCAN é sua capacidade de lidar com dados ruidosos e clusters de forma arbitrária, o que foi particularmente útil na análise dos dados de acidentes em setores como mineração e transporte rodoviário, onde os riscos estão espalhados de maneira não uniforme.

3.5.3. Análise com HDBSCAN

O HDBSCAN (Hierarchical DBSCAN) foi então aplicado para fornecer uma visão mais detalhada dos dados, permitindo a identificação de subclusters dentro dos grandes grupos identificados pelo DBSCAN. O HDBSCAN é uma extensão hierárquica do DBSCAN que identifica clusters de densidade variável, oferecendo uma análise mais granular dos dados.

Os resultados do HDBSCAN revelaram subclusters em setores como a construção civil e energia elétrica, onde diferentes níveis de risco foram identificados dentro das mesmas áreas geográficas. Por exemplo, em canteiros de obras, os subclusters indicaram que os riscos variam significativamente dependendo da fase do projeto e do tipo de atividade sendo realizada. Da mesma forma, no setor de energia elétrica, os subclusters revelaram que os acidentes tendem a se concentrar em áreas específicas, como estações de transformação e linhas de transmissão de alta tensão.

Essa análise detalhada permitiu uma compreensão mais profunda dos padrões de risco, fornecendo insights que podem ser utilizados para direcionar as intervenções de segurança de maneira mais precisa e eficaz.

3.5.3. Clusterização Hierárquica com Agglomerative Clustering

Finalmente, o Agglomerative Clustering foi utilizado para criar uma hierarquia de clusters, permitindo a análise dos dados em diferentes níveis de granularidade. Este método de clusterização hierárquica foi particularmente útil para entender como os clusters se relacionam entre si e para identificar padrões complexos de risco que poderiam não ser detectados por métodos de clusterização mais simples.

Os resultados do Agglomerative Clustering indicaram que os acidentes de trabalho podem ser agrupados em níveis hierárquicos, com clusters menores dentro de grandes grupos que compartilham características semelhantes. Por exemplo, dentro do setor de mineração, os dados revelaram que os acidentes podem ser agrupados em diferentes níveis, dependendo da profundidade da mina, do tipo de mineral extraído e das condições ambientais.

A aplicação do Agglomerative Clustering também permitiu a identificação de padrões temporais, onde certos tipos de acidentes tendem a ocorrer em períodos específicos do ano, como durante a estação chuvosa em áreas de mineração ou durante picos de demanda de energia elétrica. Esses padrões temporais são cruciais para a formulação de estratégias de prevenção sazonais e específicas para cada setor.

3.5.4. Análise Espacial dos Hotspots

A análise espacial foi conduzida utilizando Sistemas de Informação Geográfica (GIS), que permitiram a visualização dos clusters identificados pelos algoritmos de machine learning em um contexto geográfico. A integração dos resultados da clusterização com os dados geoespaciais facilitou a identificação de hotspots de acidentes de trabalho, que foram visualizados em mapas de calor. Esses mapas destacaram regiões críticas, fornecendo uma ferramenta poderosa para gestores de segurança e formuladores de políticas.

Os mapas de calor gerados confirmaram as áreas de alta densidade de acidentes identificadas pelos algoritmos de clusterização, como hotspots de risco. Essas áreas incluíram grandes centros urbanos e regiões industriais, onde a concentração de trabalhadores é maior e as condições de trabalho são mais perigosas. São Paulo e Rio de Janeiro, por exemplo, mostraram-se como regiões com alta prevalência de acidentes nos setores de construção civil e manufatura.

A análise espacial também permitiu a identificação de padrões regionais, onde certos tipos de acidentes são mais prevalentes em determinadas áreas geográficas. No Norte do Brasil, os dados indicaram uma alta incidência de acidentes relacionados ao desmatamento e à exploração madeireira, enquanto no Sudeste, os acidentes estavam mais frequentemente associados à construção civil e à manufatura. Essas variações regionais reforçam a importância de adaptar as estratégias de prevenção de acordo com as características específicas de cada área.

Além disso, a análise de interpolação espacial foi utilizada para estimar os riscos em áreas sem dados disponíveis, fornecendo uma visão mais completa da distribuição geográfica dos acidentes de trabalho. A interpolação foi fundamental para preencher lacunas de dados em regiões remotas e rurais, particularmente em áreas com pouca infraestrutura de coleta de dados, como setores de mineração e energia elétrica. Essa técnica permitiu a criação de superfícies contínuas de risco, proporcionando insights mais detalhados para a formulação de políticas de segurança.

4. Observações Finais

Os principais achados deste estudo revelaram que os acidentes de trabalho no Brasil estão concentrados em determinadas áreas geográficas e setores econômicos, com uma distribuição desigual que reflete as características inerentes de cada setor e região. A construção civil emergiu como um dos setores mais perigosos, com uma alta incidência de acidentes relacionados a quedas de altura, manuseio de equipamentos pesados e exposição a materiais perigosos. A mineração, por sua vez, apresentou uma elevada taxa de acidentes fatais, frequentemente associados à instabilidade do solo, explosões e exposição a substâncias tóxicas. O transporte rodoviário destacou-se pelos acidentes decorrentes de colisões e fadiga dos motoristas, enquanto o setor de energia elétrica mostrou uma prevalência de acidentes relacionados a choques elétricos e queimaduras graves.

A aplicação do algoritmo K-means permitiu a identificação de grandes clusters de acidentes em áreas metropolitanas e industriais, onde a densidade de trabalhadores é maior. No entanto, a sensibilidade do K-means à escolha do número de clusters e sua limitação em identificar clusters de forma irregular indicou a necessidade de utilizar algoritmos mais avançados. O DBSCAN, por exemplo, foi eficaz em identificar clusters de alta densidade e outliers em áreas remotas, enquanto o HDBSCAN ofereceu uma visão mais detalhada dos subclusters, revelando diferentes níveis de risco dentro de um mesmo setor ou região.

A análise espacial, realizada através de GIS, foi crucial para a identificação dos hotspots de acidentes, permitindo a visualização geográfica dos riscos e facilitando a comunicação dos resultados para diferentes stakeholders. Os mapas de calor e as análises de interpolação espacial destacaram áreas de alta concentração de acidentes, proporcionando uma base sólida para a implementação de medidas de segurança direcionadas.

5. Implicações Práticas

Os achados deste estudo têm implicações práticas significativas para a gestão da segurança no trabalho no Brasil. A identificação de hotspots de acidentes fornece informações cruciais para os gestores de segurança, que podem direcionar seus esforços para as áreas e setores onde os riscos são mais elevados. A aplicação de algoritmos de clusterização e análise espacial oferece uma abordagem baseada em dados para a prevenção de acidentes, permitindo que as empresas e os órgãos reguladores adotem medidas mais precisas e eficazes (Ackermann et al., 2014).

Uma das principais implicações práticas deste estudo é a possibilidade de implementar sistemas de monitoramento contínuo em setores de alto risco, utilizando tecnologias de machine learning e GIS para identificar e reagir rapidamente a mudanças nos padrões de risco. Por exemplo, em canteiros de obras, onde os riscos variam significativamente durante as diferentes fases de um projeto, sistemas de monitoramento baseados em dados podem fornecer alertas em tempo real, ajudando a prevenir acidentes antes que eles ocorram.

Além disso, os resultados deste estudo sugerem a necessidade de políticas públicas mais direcionadas, que levem em consideração as especificidades regionais e setoriais dos acidentes de trabalho. Políticas que incentivem o uso de tecnologias avançadas de segurança, como equipamentos de proteção individual (EPI) inteligentes e sistemas de monitoramento de condições ambientais, podem ser particularmente eficazes em setores como mineração e energia elétrica, onde os riscos são elevados e os acidentes podem ter consequências graves (Konzen et al., 2023).

A análise detalhada dos subclusters, especialmente em setores como construção civil e energia elétrica, também aponta para a importância de uma gestão de segurança adaptada às condições específicas de cada local de trabalho. A implementação de programas de treinamento contínuo, que sejam atualizados regularmente com base nos padrões de risco identificados, pode ajudar a reduzir a incidência de acidentes e a melhorar a segurança dos trabalhadores (Konzen et al., 2023; Queiroz et al., 2023).

6. Sugestões para Pesquisas Futuras

Embora este estudo tenha fornecido uma análise abrangente dos acidentes de trabalho no Brasil, há várias áreas que merecem investigação adicional. Uma das principais sugestões para pesquisas futuras é a exploração de variáveis socioeconômicas e culturais como fatores de risco para acidentes de trabalho. Estudos que integrem essas variáveis podem oferecer uma compreensão mais profunda dos fatores subjacentes que contribuem para a ocorrência de acidentes em diferentes contextos regionais e setoriais.

Outra área promissora para futuras pesquisas é a aplicação de técnicas mais avançadas de machine learning, como redes neurais artificiais e aprendizado profundo, que podem ser

utilizadas para prever padrões de risco de forma ainda mais precisa. Essas técnicas podem ser particularmente úteis em cenários onde os dados são complexos e não lineares, como em setores com alta variabilidade nas condições de trabalho.

Além disso, a análise temporal dos acidentes de trabalho, utilizando séries temporais e modelos preditivos, pode fornecer insights valiosos sobre como os padrões de risco evoluem ao longo do tempo. Estudos que explorem a sazonalidade dos acidentes, por exemplo, podem ajudar a identificar períodos do ano em que os riscos são maiores e a planejar intervenções preventivas de forma mais eficaz.

A integração de dados de saúde ocupacional, como registros médicos e relatórios de saúde dos trabalhadores, também pode enriquecer as análises futuras (Mendes & Wünsch, 2007; Vitrano et al., 2023). A combinação desses dados com informações de acidentes pode ajudar a identificar correlações entre condições de saúde e a probabilidade de acidentes, permitindo a implementação de programas de saúde ocupacional mais eficazes.

Finalmente, é importante considerar a replicação deste estudo em outros países ou contextos, para verificar a generalização dos resultados e adaptar as metodologias a diferentes realidades. A segurança no trabalho é uma questão global, e estudos comparativos que explorem as diferenças e semelhanças nos padrões de acidentes entre diferentes países podem contribuir para o desenvolvimento de estratégias de segurança mais universalmente aplicáveis.

7. Conclusões

Este estudo avançou significativamente na compreensão dos padrões de acidentes de trabalho no Brasil, utilizando uma combinação inovadora de machine learning e análise espacial para identificar hotspots e propor intervenções direcionadas. Os resultados destacam a importância de uma abordagem baseada em dados para a gestão da segurança no trabalho, que não só identifica os riscos existentes, mas também prevê onde e quando os acidentes são mais prováveis de ocorrer.

As implicações práticas deste estudo são vastas, desde a implementação de sistemas de monitoramento contínuo até a formulação de políticas públicas mais eficazes e adaptadas às realidades regionais e setoriais. No entanto, a pesquisa também abriu novas questões e áreas para exploração futura, que podem aprofundar ainda mais o entendimento dos fatores que contribuem para os acidentes de trabalho e ajudar a desenvolver soluções ainda mais eficazes.

Em última análise, o trabalho realizado aqui representa um passo importante na direção de ambientes de trabalho mais seguros e saudáveis, e estabelece uma base sólida para futuras pesquisas e intervenções na área de segurança do trabalho. A continuação deste tipo de pesquisa é essencial para reduzir os índices de acidentes e promover uma cultura de segurança em todos os setores da economia brasileira.

Contribuições dos Autores
R. H. S. G.; W. M. S.: Curadoria de Dados, Redação - Preparação do Rascunho Original; Edição, Revisão e Edição. Todos os autores leram e aprovaram o manuscrito final.

Conflitos de Interesses
Os autores declaram que não têm interesses conflitantes.

Referências

Abdullah, K. H., & Sofyan, D. (2023). Machine learning in safety and health research: a scientometric analysis. *International Journal of Information Science and Management*, *21*(1). https://doi.org/https://doi.org/10.22034/ijism.2022.1977763.0

Ackermann, M. R., Blömer, J., Kuntze, D., & Sohler, C. (2014). Analysis of Agglomerative Clustering. *Algorithmica*, *69*(1), 184–215. https://doi.org/10.1007/s00453-012-9717-4

Akalonu, G., Nwaogazie, I., & Ugwoha, E. (2017). Evaluation of workplace safety culture implementation and practice using agglomerative hierarchy clustering. *Archives of Current Research International*, *10*(2), 1–15. https://doi.org/10.9734/ACRI/2017/36464

Akay, A. O., Akgul, M., Esin, A. İ., Demir, M., Şenturk, N., & Özturk, T. (2021). Evaluation of occupational accidents in forestry in Europe and Turkey by k-means clustering analysis. *Turkish Journal of Agriculture and Forestry*, *45*(4), 495–509. https://doi.org/10.3906/tar-2010-55

Babalola, A., Manu, P., Cheung, C., Yunusa-Kaltungo, A., & Bartolo, P. (2023). Applications of immersive technologies for occupational safety and health training and education: A systematic review. *Safety Science*, *166*, 106214. https://doi.org/10.1016/j.ssci.2023.106214

Carvalho, C. A. da S., Silva, J. C. da, Lima, J. L. L. P. C. de, & Brum, S. da S. (2020). Saúde e Segurança no Trabalho: um relato dos números de acidentes do trabalho e doenças ocupacionais no Brasil (2012-2018) / Health and Safety at Work: a portrait of occupational accident and disease numbers in Brazil (2012-2018). *Brazilian Journal of Business*, *2*(3), 2909–2926. https://doi.org/10.34140/bjbv2n3-070

Cunha, H. D., Silva , A. D. da, Martins, B. B., Guedes, B. S., Nunes, I. M., Maranhão, M. R. de A., & Conforto, M. do N. F. (2024). Detection of slums in Rio de Janeiro through satellite images. *Dataset Reports*, *3*(1), 107–113. https://doi.org/10.58951/dataset.2024.019

Deng, D. (2020). DBSCAN Clustering Algorithm Based on Density. *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, 949–953. https://doi.org/10.1109/IFEEA51475.2020.00199

Deng, F., Gu, W., Zeng, W., Zhang, Z., & Wang, F. (2020). Hazardous Chemical Accident Prevention Based on K-Means Clustering Analysis of Incident Information. *IEEE Access*, *8*, 180171–180183. https://doi.org/10.1109/ACCESS.2020.3028235

Filgueiras, V. A. (2017). Saúde e segurança do trabalho no Brasil. In V. A. Filgueiras (Ed.), *Saúde e Segurança do Trabalho no Brasil*. Gráfica Movimento. pp. 19–78.

Jaafar, M. H., Arifin, K., Aiyub, K., Razman, M. R., Ishak, M. I. S., & Samsurijan, M. S. (2018). Occupational safety and health management in the construction industry: a review. *International Journal of Occupational Safety and Ergonomics*, *24*(4), 493–506. https://doi.org/10.1080/10803548.2017.1366129

Jozan, M. M. B., Ghorbani, B. D., Khalid, M. S., Lotfata, A., & Tabesh, H. (2023). Impact assessment of e-trainings in occupational safety and health: a literature review. *BMC Public Health*, *23*(1), 1187. https://doi.org/10.1186/s12889-023-16114-8

Kemajl, Z., Stojance, M., Gzim, I., & Ledi, M. L. (2024). Comprehensive analysis of the mining accident forecasting and risk assessment methodologies: Case study – Stanterg Mine. *Mining of Mineral Deposits*, *18*(2), 11–17. https://doi.org/10.33271/mining18.02.011

Konzen, I. G. do N. C., Souto, A. B., Konzen, M. R., & Neto, J. M. da S. (2023). Segurança no trabalho: motivos que levam o trabalhador da construção civil a deixar de utilizar do EPIs. *Revista de Gestão e Secretariado (Management and Administrative Professional Review)*, *14*(6), 8875–8896. https://doi.org/10.7769/gesec.v14i6.2271

Liang, C.-J., & Cheng, M. H. (2023). Trends in Robotics Research in Occupational Safety and Health: A Scientometric Analysis and Review. *International Journal of Environmental Research and Public Health*, *20*(10), 5904. https://doi.org/10.3390/ijerph20105904

Lima, M. D. F. de, Silva, M. C., França, C. D. V., Santos, W. B. dos, Costa, V. A., Costa, W. L. M. da, Segundo, J. de J. P., & Teixeira, M. R. (2023). Análise sobre a segurança do trabalho em empreendimentos de construção civil: uma revisão sistemática em diferentes categorias de canteiros de obras. *Contribuciones a Las Ciencias Sociales*, *16*(9), 18314–18328. https://doi.org/10.55905/revconv.16n.9-266

Mendes, J. M. R., & Wünsch, D. S. (2007). Elementos para uma nova cultura em segurança e saúde no trabalho. *Revista Brasileira de Saúde Ocupacional*, *32*(115), 153–163. https://doi.org/10.1590/S0303-76572007000100014

Mendonça, M. F. S. de, Silva, A. P. de S. C., & Castro, C. C. L. de. (2017). Análise espacial dos acidentes de trânsito urbano atendidos pelo Serviço de Atendimento Móvel de Urgência: um recorte no espaço e no tempo. *Revista Brasileira de Epidemiologia*, *20*(4), 727–741. https://doi.org/10.1590/1980-5497201700040014

Menezes, M. N., & Dal Magro, M. L. P. (2023). Impactos psicossociais dos acidentes de trabalho graves: um olhar sobre os trabalhadores acompanhados pelo Centro de Referência em Saúde do Trabalhador. *Revista Jurídica Trabalho e Desenvolvimento Humano*, *6*, 1–30. https://doi.org/https://doi.org/10.33239/rjtdh.v6.152

Ministério da Previdência Social. (2024). *Anuário Estatístico de Acidentes do Trabalho – AEAT*. Dados Estatísticos – Saúde e Segurança do Trabalhador. Acesso em 17 de outubro de 2024. Disponível em: <https://www.gov.br/previdencia/pt-br/assuntos/previdencia-social/saude-e-seguranca-do-trabalhador/acidente_trabalho_incapacidade>.

Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foiadelli, F. (2023). K-means and alternative clustering methods in modern power systems. *IEEE Access*, *11*, 119596–119633. https://doi.org/10.1109/ACCESS.2023.3327640

Mutlu, N. G., Altuntas, S., & Dereli, T. (2023). The evaluation of occupational accident with sequential pattern mining. *Safety Science*, *166*, 106212. https://doi.org/10.1016/j.ssci.2023.106212

Oliveira, O. J. de, Oliveira, A. B. de, & Almeida, R. A. de. (2010). Gestão da segurança e saúde no trabalho em empresas produtoras de baterias automotivas: um estudo para identificar boas práticas. *Production*, *20*(3), 481–490. https://doi.org/10.1590/S0103-65132010005000029

Peinado, H. S. (2019). *Segurança e Saúde do Trabalho na Indústria da Construção Civil*. Editora Scienza. https://doi.org/10.26626/978-85-5953-048-3.2019B0001

Queiroz, M. T. A., Queiroz, F. A., & Queiroz, V. A. (2023). Ocorrência de acidentes de

trabalho na Região do Vale do Aço, MG, Brasil. *Sistemas & Gestão*, *18*(1). http://dx.doi.org/10.20985/1980-5160.2023.v18n1.1855

Rosa, R. (2011). Análise espacial em geografia. *Revista da ANPEGE*, *07*(01), 275–289. https://doi.org/10.5418/RA2011.0701.0023

Schwambach, G. C. dos S., Sott , M. K., & Schwambach, R. E. (2024). Wearable devices and workplace productivity: a bibliometric analysis of their integration into professional environments. *Dataset Reports*, 3(1), 101–106. https://doi.org/10.58951/dataset.2024.018

Silva, D. da, Lopes, E. L., & Braga Junior, S. S. (2014). Pesquisa quantitativa: elementos, paradigmas e definições. *Revista de Gestão e Secretariado*, *5*(1), 01–18. https://doi.org/10.7769/gesec.v5i1.297

Simonelli, A. P., Jackson Filho, J. M., Vilela, R. A. G., & Almeida, I. M. de. (2016). Influência da segurança comportamental nas práticas e modelos de prevenção de acidentes do trabalho: revisão sistemática da literatura. *Saúde e Sociedade*, *25*(2), 463–478. https://doi.org/10.1590/S0104-12902016147495

Soares, D. de C. (2019). Análise espacial exploratória dos acidentes de trabalho no Brasil. *Boletim Científico Escola Superior do Ministério Público da União*, *53*, 205–232. https://escola.mpu.mp.br/publicacoescientificas/index.php/boletim/article/view/507

Sousa, A. do R. F. de, & Rodolpho, D. (2020). Importância da segurança do trabalho na produção industrial. *Revista Interface Tecnológica*, *17*(2), 817–824. https://doi.org/10.31510/infa.v17i2.1008

Souza, D. F. de, Martins, W. A., Martinho, E., & Santos, S. R. (2023). An analysis of accidents of electrical origin in Brazil between 2016 and 2021. *IEEE Transactions on Industry Applications*, *59*(3), 3151–3160. https://doi.org/10.1109/TIA.2023.3241138

Stewart , G., & Al-Khassaweneh, M. (2022). An implementation of the HDBSCAN* clustering algorithm. *Applied Sciences*, *12*(5), 2405. https://doi.org/10.3390/app12052405

Tokuda, E. K., Comin, C. H., & Costa, L. da F. (2022). Revisiting agglomerative clustering. *Physica A: Statistical Mechanics and its Applications*, *585*, 126433. https://doi.org/10.1016/j.physa.2021.126433

Vitrano, G., Micheli, G. J. L., Guglielmi, A., De Merich, D., Pellicci, M., Urso, D., & Ipsen, C. (2023). Sustainable occupational safety and health interventions: A study on the factors for an effective design. *Safety Science*, *166*, 106249. https://doi.org/10.1016/j.ssci.2023.106249

DATASET REPORTS

journals.royaldataset.com/dr