

APLICAÇÃO DA TÉCNICA DE MINERAÇÃO DE CLUSTERING NOS REGISTROS DE ACIDENTES DE TRABALHO NO PERÍODO DE 2021 – BRASIL

Millena Jussani da Cruz (PIBIC/CNPq/FA/Uem), Edwin Vladimir Cardoza Galdamez (Orientador), E-mail: evcgaldamez@uem.br.

Universidade Estadual de Maringá, Centro de Tecnologia, Maringá-PR

Área e subárea do conhecimento: Engenharia de Produção / Gerência de produção

Palavras-chave: Mineração de *Clustering*; Acidentes de Trabalho; Clusterização de Dados.

RESUMO

Este projeto de iniciação científica visa aplicar técnicas de mineração de cluster para registros de acidentes de trabalho no Brasil em 2021. A proposta visa identificar padrões e agrupamentos em dados de acidentes para entender melhor as circunstâncias e fatores por trás desses acidentes. A análise de agrupamento identificará grupos com características semelhantes, fornecendo pontos relevantes para prevenção e mitigação de incidentes. Ao desvendar áreas de alto risco e condições recorrentes, esta pesquisa visa contribuir para a promoção de ambientes de trabalho mais seguros e saudáveis. A abordagem envolverá a aplicação de técnicas de coleta de dados, pré-processamento e agrupamento. Através desta abordagem, espera-se que subsídios valiosos sejam fornecidos para melhorar a segurança ocupacional e desenvolver estratégias eficazes de prevenção.

INTRODUÇÃO

O crescente interesse pela segurança e proteção da saúde no trabalho tem levado a uma procura constante de métodos mais eficazes de análise e compreensão dos acidentes de trabalho. No contexto do Brasil, onde a diversidade de indústrias e o diversificado cenário econômico influenciam diretamente os registros de acidentes, a aplicação de técnicas avançadas, como a mineração em cluster, é uma abordagem promissora para identificar

padrões ocultos e relações entre diferentes variáveis. Neste relatório, examinamos a aplicação da técnica de agrupamento de Mineração de Dados nos registros de acidentes de trabalho no período de 2021 com o objetivo não só de identificar agrupamentos significativos, mas também de obter conhecimentos que possam levar à formulação de medidas preventivas mais direcionadas e eficazes. Espera-se que, ao compreender as características internas dos acidentes e suas inter-relações, contribua para a redução da ocorrência de acidentes e, portanto, para a melhoria das condições de trabalho em todo o estado. Neste relatório descrevemos a metodologia utilizada, os dados utilizados, os resultados obtidos e as implicações práticas dessas informações no contexto da segurança do trabalho no Brasil.

MATERIAIS E MÉTODOS

A base de dados usada neste estudo foi disponibilizada pelo portal de Comunicação de Acidente de Trabalho (CAT). Foram selecionadas planilhas de 2020 a 2021, que são disponibilizadas de três em três meses, por isso, foi necessário juntar os arquivos para realizar o pré-processamento dos dados. A primeira ação antes de enviar o Corpus para Orange Data Mining foi excluir colunas vazias e não classificadas e padronizar os atributos.

Após incorporar o Corpus ao software, foi feito o pré-processamento que consiste nas etapas de Tokenização que extrai unidades mínimas de texto, remoção de Stopwords, palavras que aparecem muito na massa textual e Normalização que se baseia no agrupamento de tokens com padrões semelhantes. Após estas etapas, a base de dados pré-processada foi enviada para um recipiente de tokens, em que a sequência e a conexão entre os tokens não possui relevância para o sistema.

RESULTADOS E DISCUSSÃO

Foi empregada a técnica de Clusterização Hierárquica usando como medida a distância Jaccard, que mensura a diferença entre conjuntos de amostras para encontrar os clusters dos dados. A escolha desse modelo teve como objetivo identificar documentos de conteúdo similar e armazená-los em um mesmo aglomerado. A identificação de documentos similares dá-se pela quantidade de palavras iguais que eles contêm.

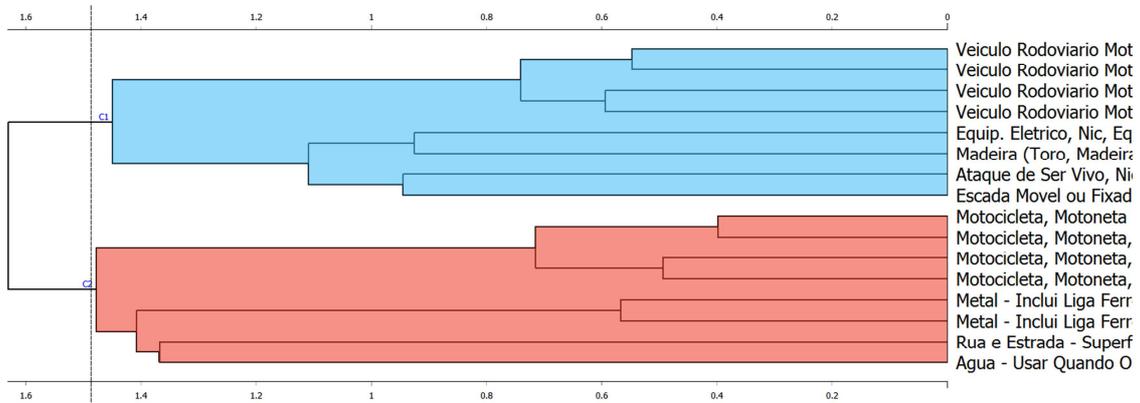


Figura 1. Clusters encontrados

Dos atributos existentes na planilha, foram selecionadas 5 classes de áreas de trabalho para fazer a análise de clusters, isso foi feito devido ao tamanho da planilha original (6.815.744 bytes) que não foi aceita pelo software devido a incapacidade de armazenamento de dados. No entanto, o uso dessas classes provou a eficiência do algoritmo em realizar a transformação. A separação foi feita de acordo com o número de características parecidas entre cada dado. A variável usada nessa análise foi “Agente Causador do Acidente”.

CONCLUSÃO

A partir do objetivo proposto pelo projeto, foi possível encontrar grandes grupos divididos em 2 clusters a partir da Clusterização Hierárquica. Cada grupo, analisado separadamente possuem características bem definidas, como por exemplo a variável sexo onde o sexo masculino é predominante. As idades variam entre 28 a 56 anos, e a maioria dos acidentes não resultaram em óbito. E em relação a disposição geográfica, o local que mais apresenta ocorrências é o município do Maranhão e em seguida Rondônia.

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador Edwin Vladimir Cardoza Galdamez pela orientação acadêmica, apoio e confiança. Aos meus familiares pelo apoio incondicional, e a Fundação Araucária pelo apoio financeiro que possibilitou que eu pudesse me dedicar integralmente ao projeto.

REFERÊNCIAS

FAYYAD , Usama. Descoberta de Conhecimento e Mineração de Dados: Rumo a uma estrutura unificadora. *In*: PIATETSKY-SHAPIRO , Gregory. **Problemas de Aplicação**. [S. l.: s. n.], 1996. cap. 2, p. 82-88

AMORIM , Thiago. Limpeza dos Dados (Data Preparation). *In*: AMORIM , Thiago. **Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de bases de dados**. 2006. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Ciência da Computação, Universidade Federal de Pernambuco, [S. l.], 2006

CANTAVE , Jerry. ACIDENTES DE TRABALHO NO BRASIL: ANÁLISE TEMPORAL DE 2012 A 2021. **INSTITUTO LATINO-AMERICANO DE CIÊNCIAS DA VIDA E DA NATUREZA (ILACVN)**, [s. l.], 2022.

JESUS, Liandra dos Santos. **APLICAÇÃO DE MINERAÇÃO DE DADOS EMPREGANDO CLUSTERIZAÇÃO EM DADOS NACIONAIS DE ACIDENTES OCUPACIONAIS**. *In*: JESUS, Liandra dos Santos. TCC (SUPERIOR) - MESTRANDA, [S. l.], 2021.