

## APLICAÇÃO DE TÉCNICA DE MINERAÇÃO DE DADOS DE CLASSIFICAÇÃO EM REGISTROS DE ACIDENTES DE TRABALHO DO SETOR DE TRANSPORTE

Lucimara Ferreira da Silva (PIBIC/FA/CNPq/Uem), Gisllaine Camila Lapasini Leal (Orientadora), Edwin Vladmir Cardoza Galdamez (Co-Orientador), e-mail: ra108396@uem.br

Universidade Estadual de Maringá, Departamento de Engenharia de Produção/Maringá, PR.

**Área: Engenharia de Produção**  
**Subárea: Higiene e Segurança do Trabalho**

**Palavras-chave:** Saúde e Segurança do Trabalho; Setor de Transporte; Mineração de Dados

### Resumo:

A saúde e segurança do trabalho vêm sendo um assunto de destaque na sociedade. Afinal os acidentes de trabalho provocam um grande impacto sobre a produtividade e economia. No Brasil o panorama é preocupante, o maior número de registros está ligado à Comunicação de Acidentes de Trabalho (CAT), com abrangência nacional e acesso público. Dentre as maneiras de realizar uma análise está a mineração de dados, que permite operar com variados tipos, estruturas e dimensionalidade de dados. Esta pesquisa realizou um estudo com um conjunto de dados balanceados contendo 1.206 registros, disponíveis na CAT referentes ao setor de transporte. Com a finalidade de prever os óbitos no setor de transporte, esta pesquisa utilizou para a mineração de dados o algoritmo *Naïve Bayes*, o qual apresentou melhor capacidade preditiva, tendo como base os registros disponíveis na CAT referentes ao setor de transporte.

### Introdução

Os acidentes de trabalho são evitáveis e provocam um grande impacto sobre a produtividade e economia, além do abalo causado para o acidentado e seus familiares (SANTANA *et al.*, 2006). De acordo com os dados do Observatório Digital de Segurança e Saúde do Trabalho, o Brasil registra a cada 3 horas e 43 minutos uma morte por acidente de trabalho, tendo sido contabilizados 16.455 mortes entre 2012 e 2018.

No Brasil, uma grande parte dos custos diretos com acidente de trabalho são subsidiados pelo Ministério da Previdência Social, que por meio do Instituto Nacional de Seguridade Social (INSS), responsável por garantir o direito à

previdência social (SANTANA *et al.*, 2006). O acidente ocorrido é comunicado ao INSS por meio da CAT (Comunicação de Acidente do Trabalho), tendo as informações armazenados e disponibilizados em bancos de dados de acesso público (BATISTA *et al.*, 2019).

Diante do grande volume de dados produzidos referente a acidentes, doenças ocupacionais e óbitos, a mineração de dados (MD) é uma alternativa para atuar nessa área (HAN *et al.*, 2012). Através da mineração de dados e do aprendizado de máquina é possível converter um grande volume de dados em informações úteis, detectando padrões e tendências, possibilitando assim, prever e descobrir parâmetros que induzem as ocorrências analisadas (RUSO e STOJANOVIĆ, 2012).

Sendo assim, a aplicação de técnicas de mineração a dados ocupacionais, auxiliam na gestão das organizações, informando a inter-relações entre os dados, contribuindo para a tomada de decisão (BEVILACQUA *et al.*, 2008).

O objetivo do trabalho é apresentar a aplicação de uma técnica para mineração de dados de classificação em registros de acidentes de trabalho no setor de transportes. Para a pesquisa é utilizada a base de dados da CAT entre o período de julho a setembro de 2018 até julho a setembro de 2020, com um recorte no setor de transporte.

## Materiais e métodos

O estudo consiste em uma pesquisa exploratória, apresentando uma abordagem quantitativa, com fonte de pesquisa secundária. A realização deste estudo seguiu as seguintes etapas: planejamento; coleta e pré-processamento dos dados; aplicação de técnicas de mineração e por fim análise dos resultados.

Na etapa de planejamento foi desenvolvido uma contextualização do tema, buscando na literatura por estudos e referências na área de SST, setor de transporte e mineração de dados. Em seguida, na coleta e pré-processamento, utilizou-se o conjunto de dados com abertura de CAT, em razão de sua abrangência nacional, confiabilidade e fácil acesso.

Na etapa de mineração foi realizado um balanceamento entre os dados de óbito e não óbito, deixando a quantidade de ocorrências equilibrada, resultando ao final 1.206 registros, composto por 50% de ocorrências com óbitos e 50% sem óbito, sendo aleatória a seleção dos dados sem óbito. Para aplicação da mineração de dados foi utilizado o ambiente do Jupyter Notebook, com modelos desenvolvidos na linguagem Python. O conjunto de dados balanceado foi processado utilizando o algoritmo *Naive Bayes* (NB), buscando prever a ocorrência de óbitos a partir de variáveis registrados com a abertura da CAT.

Os resultados da técnica foram analisados a partir de cinco métricas: acurácia, precisão, *recall*, F1 score e curva ROC/AUC (*Receiver Operating Characteristic/Area Under Curve*). Baseando-se nos resultados encontrados, algumas considerações foram realizadas ao final da pesquisa.

## Resultados e Discussão

Com o objetivo de prever a ocorrência de óbitos, dentro dos nove atributos considerados o óbito foi a variável dependente e os demais atributos foram variáveis independentes. Por meio da função *train\_test\_split* do *scikit-learn*, o conjunto de dados foi dividido em 70% treino e 30% teste. Assim, o conjunto de treino se caracterizava por 844 instâncias, enquanto o conjunto de teste apresentava apenas 362 instâncias.

Os resultados apresentados na Tabela 1 são os resultados do conjunto de dados executado no algoritmo *Naive Bayes* (NB).

**Tabela 1.** Resultados das métricas e tempo computacional da técnica

Técnica	Acurácia	Precisão	Recall	F1 score	ROC/AUC	Tempo (s)
NB	0,8912	0,8665	0,9194	0,8919	0,9505	5,7622

Fonte: Autoria própria (2021)

A técnica *Naive Bayes*, apresentou os melhores resultados nas métricas *Recall* e ROC/AUC. Porém, averiguar o desempenho das técnicas de mineração apenas por meio das métricas não é satisfatório para entender seu funcionamento. Para compor a etapa de explicação foi utilizado um algoritmo de inteligência artificial, *SHapley Additive exPlanations* (SHAP), que tem como objetivo explicar as saídas de um modelo de aprendizagem de máquinas.

Procedendo com a técnica ensemble, *Naive Bayes* (NB), a variável natureza da lesão descreve a maior relevância para previsão do óbito, seguida pela parte do corpo atingida, destacando que o óbito do trabalhador é diretamente impactado pelo tipo da lesão sofrida e pela região do seu corpo que é atingido.

Em compensação, o tipo de acidente (típico, trajeto ou doença), não apresenta relevância para a previsão de óbitos no setor de transporte, assim como o sexo do trabalhador, o estado, idade e CBO do empregador.

## Conclusões

A utilização de técnicas de mineração de dados vinculados a dados de saúde e segurança do trabalho, resultam em informações que podem servir de subsídio para tomada de decisões tanto para organizações públicas

quanto privadas. Os resultados também podem indicar quais fatores mais influenciam na ocorrência de óbito, auxiliando gestores na implantação de programas de prevenção, atuando na redução de ocorrências para o setor de transporte.

## Agradecimentos

Agradecimento à CAPES e a Fundação Araucária pela oportunidade de desenvolvimento da pesquisa.

## Referências

BATISTA, A. G.; SANTANA, V. S.; FERRITE, S. Registro de dados sobre acidentes de trabalho fatais em sistemas de informação no Brasil. **Revista Ciência e Saúde Coletiva**, v.24, p. 693-704, 2019.

BEVILACQUA, M., CIARAPICA, F.E., GIACCHETTA, G. Industrial and occupational ergonomics in the petrochemical process industry: A regression trees approach. *Accident Analysis and Prevention*, 40, 1468–1479, 2008.

HAN, J; KAMBER, M.; PEI, J. *Data Mining Concepts and Techniques*. **Morgan Kaufmann Publishers**. Third Edition, 2012.

RUSO, J.; STOJANOVIĆ, V. Occupational health and safety using data mining. **International Journal for Quality Research**, v.6, n.4, p. 355-364, 2012.

SANTANA, V.S.; ARAÚJO-FILHO, J.B.; ALBUQUERQUE, P. B. O.; BARBOSA A. B. Acidentes de trabalho: custos previdenciários e dias de trabalho perdidos. **Revista Saúde Pública**. 2006. Disponível em:<[http://www.scielo.br/scielo.php?pid=S0034-89102006000700007&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S0034-89102006000700007&script=sci_arttext)>. Acesso em: 24 nov. 2020.