



DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS APLICADOS EM PROTEÇÃO E DEFESA CIVIL: UMA ANÁLISE DOS REGISTROS DE SOCORROS CONTRIBUINDO PARA A IDENTIFICAÇÃO DE DESASTRES

ARTIGO ORIGINAL

FERNANDES, Samir Batista ¹, SANTOS, Alexandre Luís Belchior dos ², SILVA, Rodrigo Werner da ³

FERNANDES, Samir Batista. SANTOS, Alexandre Luís Belchior dos. SILVA, Rodrigo Werner da. **Descoberta de conhecimento em banco de dados aplicados em proteção e defesa civil: uma análise dos registros de socorros contribuindo para a identificação de desastres.** Revista Científica Multidisciplinar Núcleo do Conhecimento. Ano. 07, Ed. 07, Vol. 06, pp. 05-20. Julho de 2022. ISSN: 2448-0959, Link de acesso: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>, DOI: 10.32749/nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil

RESUMO

Os desastres comprometem a capacidade local de resposta e não existem análises que possam sugerir, numa série histórica, os eventos anômalos baseados na contagem de ocorrências. Nesse contexto, a pergunta desta pesquisa é: a quantidade de registros de socorros em uma série histórica pode contribuir para a identificação de desastres por meio da técnica de Descoberta de Conhecimento em Banco de Dados (DCBD)? O objetivo do estudo é analisar a aplicabilidade da técnica de descoberta de conhecimento em banco de dados a fim de detectar anomalias

¹ Mestre. ORCID: 0000-0001-9976-5318.

² Orientador. Doutor. ORCID: 0000-0002-5314-3553.

³ Coorientador. Doutor. ORCID: 0000-0003-1874-9856.

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



associadas a ocorrências de desastres utilizando os registros dos eventos atendidos e as solicitações de reconhecimento de desastres no município de Duque de Caxias, no Estado do Rio de Janeiro entre o período de 2013 a 2020. Para aplicar o processo de DCDB foi utilizada a ferramenta *Power Bi* para detectar anomalias e para elaboração de uma árvore de decisão. A tarefa de classificação foi empregada junto com a coleção de algoritmos do programa *Waikato Environment for Knowledge Analysis* para avaliar o desempenho dos algoritmos. Após, foi realizada interpretação da avaliação dos resultados e gerado um grupo de regras. Ao concluir o ciclo foi possível obter conhecimento associado às anomalias dos registros de eventos e dos desastres por meio de regras que são úteis e aplicáveis em Proteção e Defesa Civil.

Palavras-chave: Defesa civil, Árvore de decisão, Mineração de dados, Power Bi.

1. INTRODUÇÃO

Os desastres são eventos adversos que comprometem a capacidade de resposta local, provocando graves danos e prejuízos. Em janeiro de 2011, o Estado do Rio de Janeiro foi acometido por um severo desastre na região Serrana que ceifou centenas de vidas. No Brasil, a análise de identificação de comprometimento do esgotamento da capacidade de responder um desastre é baseada em verificação documental, por meio de critérios normativos (BRASIL, 2020).

O evento mencionado gerou uma mobilização no cenário nacional e, com isso, deu início a convergência de esforços para a elaboração e promulgação da lei 12.608, de 10 de abril de 2012, que instituiu a Política Nacional de Proteção e Defesa Civil (BRASIL, 2012). Já no Estado do Rio de Janeiro, o decreto 46.935, de 12 de fevereiro de 2020, instituiu a Política Estadual de Proteção e Defesa Civil (PEPDEC) e reorganizou o Sistema Estadual de Proteção e Defesa Civil (SIEPDEC) (RIO DE JANEIRO, 2020).

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Além dos aspectos legais, o Brasil é signatário do Marco de Sendai, que é uma estratégia internacional para a Redução de Riscos de Desastres (RRD), e dos Objetivos de Desenvolvimento Sustentáveis (ODS), promovido pela Organizações das Nações Unidas (ONU). Em linhas gerais, a estratégia internacional visa uma mudança de paradigma, na qual as ações preventivas ganham mais ênfase do que as de resposta (NAÇÕES UNIDAS, 2015).

Ainda motivado pelo mega desastre na região serrana, em 2013 foi lançado o atlas brasileiro de desastres, com informações que detalham uma fragilidade sistêmica quanto ao registro e a integridade histórica de eventos próprios de defesa civil, bem como a ausência de unidades e campos padronizados para as informações declaradas nos documentos, a ausência de método de coleta sistêmica e armazenamento dos dados, as dificuldades que os responsáveis pela emissão dos documentos apresentavam na interpretação do tipo de desastre e dificuldades de consolidação, transparência e acesso aos dados (BRASIL, 2005).

Observada esta lacuna, o Governo Federal lançou a Plataforma chamada Sistema Integrado de Informações sobre Desastres (S2ID) e o governo Estadual do Rio de Janeiro lançou o Programa de Registro de Ocorrências em Defesa Civil (PRODEC), com objetivo de sanar as falhas apontadas no Atlas. Com o surgimento do S2ID, tornou-se possível incluir dados sobre as ocorrências de desastres. Entretanto, os registros de socorros de eventos não-desastres são desconsiderados na declaração de desastres (BRASIL, 2005).

Nesse contexto, a pergunta desta pesquisa é: a quantidade de registros de socorros em uma série histórica pode contribuir para a identificação de desastres por meio da técnica de Descoberta de Conhecimento em Banco de Dados (DCBD)? Para solucionar o problema, o objetivo do estudo foi analisar a aplicabilidade da técnica de descoberta de conhecimento em banco de dados a fim de detectar anomalias associadas a ocorrências de desastres utilizando os registros dos eventos atendidos

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



e as solicitações de reconhecimento de desastres no município de Duque de Caxias, no Estado do Rio de Janeiro entre o período de 2013 a 2020.

O que se espera por este artigo é conscientizar aos especialistas em Proteção e Defesa Civil e a todos os interessados nessa área de ciência de dados, que o conhecimento descoberto gere regras capazes de auxiliar na identificação de anomalias relacionadas com o aumento de registros de ocorrências e auxiliar na identificação do comprometimento de resposta local para facilitar os tomadores de decisão na identificação de um possível desastre.

Sendo assim, este foi estruturado da seguinte maneira: na seção 2, encontra-se o referencial teórico sobre o processo de DCBD, na seção 3 encontra-se a metodologia utilizada, na seção 4 os resultados e, na última seção, as considerações finais.

2. O PROCESSO DE DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS

O aumento do interesse por novas descobertas de conhecimento em bases de dados é motivado: pela redução dos custos envolvidos no armazenamento de dados; pelo potencial de geração e armazenamento de dados dos sistemas de computação; e pela necessidade de se obter informação útil nos dados que, até então, são desconhecidos pelas empresas e organizações (FAYYAD *et al.*, 1996).

A expressão “descoberta de conhecimento em bases de dados” foi introduzida no final dos anos 80, tendo, como propósito, identificar similaridades, padrões e conhecimentos a partir de dados, auxiliando a avaliação das informações a serem utilizadas por gestores tomadores de decisões (CABENA, 1998).

Segundo Fayyad *et al.* (1996), o processo de DCBD é definido por um processo não-trivial, oriundo da identificação de padrões novos, válidos, potencialmente úteis e compreensíveis a partir de uma base de dados. O processo, para o autor supracitado, é dito não-trivial, pois engloba um certo grau de dificuldade, considerando as

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



condições dos dados armazenados em suas bases, os quais podem ser volumosos, possuírem incoerências, serem esparsos, terem informações incompletas ou supérfluas. A descoberta encontrada durante o processo, também precisa ser válida para novos dados e confiável. Além disso, ao realizar descobertas deve-se entregar um resultado que seja útil e simples de ser compreendido pelas pessoas envolvidas no processo.

O processo de DCBD compreende várias etapas necessárias à obtenção de um conhecimento válido e útil, tais como a coleta e preparação dos dados indispensáveis ao alcance dos objetivos; a análise dos dados coletados com o auxílio de técnicas de mineração de dados; ou seja, a extração de padrões; a avaliação dos resultados e a assimilação do conhecimento obtido (FERNANDES, 2021).

3. METODOLOGIA

Nessa seção encontra-se a metodologia utilizada para a realização do artigo.

3.1 ESTRATÉGIA METODOLÓGICA

A estratégia metodológica utilizada neste artigo foi estabelecida por um método quantitativo, a partir de um estudo de caso apresentado segundo Fernandes (2021), onde foi proposto e apresentado como um pressuposto sobre o processo de DCBD, analisando a sua aplicabilidade para geração de conhecimento. Destacando-se a tarefa realizada, pelo algoritmo de classificação, no processo de construção de árvore de decisão.

No estudo de caso é contextualizada a descoberta de conhecimentos em banco de dados. Em seguida são contabilizados os registros de desastres do município de Duque de Caxias e analisadas as solicitações de homologação e/ou reconhecimento, de Situação de Emergência (SE) ou Estado de Calamidade Pública (ECP) do Município de Duque de Caxias – RJ, aplicando as etapas de Descoberta de Conhecimento em Banco de Dados (DCBD), para verificar se é possível descobrir

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



conhecimentos em registros de ocorrência em proteção e defesa civil, utilizando uma série histórica.

Ao término do processo, são interpretadas as avaliações do processo de mineração de dados e em seguida é gerado um conjunto de regras baseado no ciclo do processo de DCBD.

3.2 OBJETO DE ESTUDO ANALISADO

O objeto analisado, segundo Fernandes (2021), foram os atendimentos do município de Duque de Caxias, de 2013 a 2020, este com uma área total de 467,6 quilômetros quadrados, correspondentes a 6,2% da área da Região Metropolitana. Os limites municipais são: Miguel Pereira, Petrópolis, Magé, Rio de Janeiro, São João de Meriti, Belford Roxo e Nova Iguaçu.

O município foi escolhido por sua série histórica, em comparação com os demais municípios que integram e utilizam a plataforma estadual para registros de ocorrências, contando com um banco de dados com mais de sete mil registros ao longo de quase uma década (FERNANDES, 2021).

Na Figura 1, é possível identificar a distribuição das contagens de todas as ocorrências ao longo do recorte temporal de 2013 a 2020, no município de Duque de Caxias – RJ.

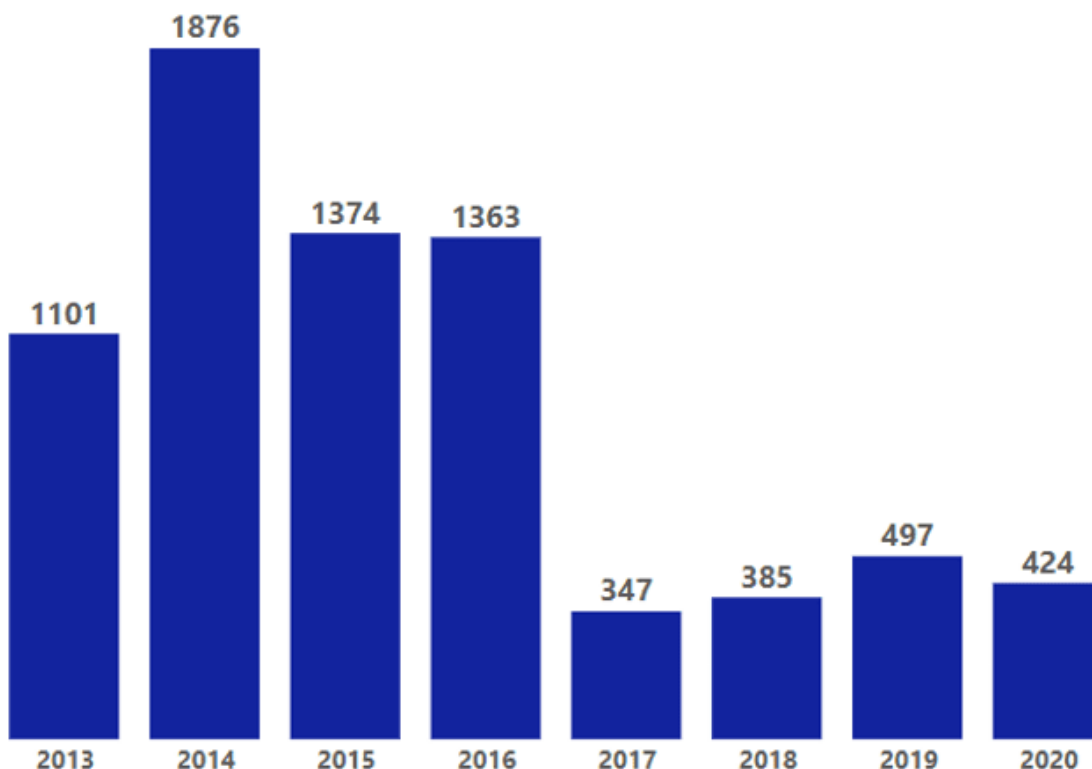
RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Figura 1: Frequência das ocorrências atendidas pela defesa civil entre os anos de 2013 a 2020.

FREQUENCIA DE OCORRENCIAS POR ANO - DUQUE DE CAXIAS/RJ



Fonte: Programa de gestão de ocorrências (2021)

3.3 FERRAMENTAS METODOLÓGICAS UTILIZADAS NO ESTUDO DE CASO

Nessa seção encontram-se as etapas que compõem o processo de DCBD.

3.3.1 O PROCESSO DE DESENVOLVIMENTO DE DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS (DCBD)

Para o desenvolvimento do processo de DCBD é fundamental o conhecimento do limite a ser trabalhado. A compreensão do limite, também chamado de domínio, não

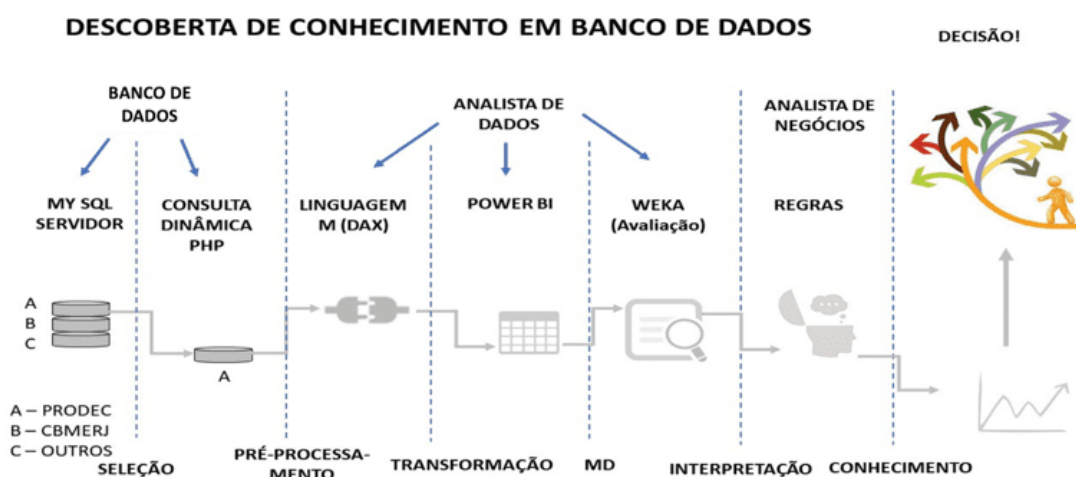
RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>

é tratada como uma etapa, pois está vinculada à etapa de seleção de dados conforme apresentada na Figura 2. Esta etapa consistiu em escolher os dados que podem contribuir com o processo de DCBD. Na pesquisa foram atribuídas duas dimensões, ou seja, variáveis do conjunto de dados que auxiliaram na DCBD.

Os dados foram acessados em um banco de dados do tipo *My SQL* por meio de uma consulta dinâmica utilizando uma programação da plataforma *Power BI*, que é uma ferramenta de interface gráfica destinada para visualização de dados. Na plataforma é possível exibir relatórios, medidores de desempenho e realização de operações matemáticas simples. Ademais, ela possui alguns recursos, muito limitados, de aprendizado de máquina. Portanto, a funcionalidade do algoritmo utilizado pela plataforma escolhida foi o *Power Bi* da *Microsoft*, que somente permite escolher uma única variável para buscar a anomalia na série temporal.

Figura 2 - Método de obtenção de conhecimento em banco de dados.



Fonte: Autoria própria

Com base nas informações sobre os limites de estudo e origem dos dados, foram desenvolvidos testes, conforme as etapas demonstradas na figura 2, a fim de verificar a possibilidade de obtenção de padrões válidos mediante a construção de argumentos que são chamados de classes, tais como: dias analisados e classificados como

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



normais ou anômalos ou registros realizados. Os dias com anomalias foram aqueles em que houve uma quantidade de ocorrência significativa em relação a série histórica e os dias normais foram aqueles em que não houve uma quantidade significativa de ocorrência em relação a série histórica (FERNANDES, 2021).

Outro conjunto de dados selecionados foram os registros de solicitações realizadas pela Defesa Civil municipal de Duque de Caxias no S2ID com a finalidade de comprovar que houve um comprometimento da capacidade de resposta local e relacionar com os dias normais ou anômalos.

3.3.2 PRÉ-PROCESSAMENTO DOS DADOS

Segundo Fernandes (2021), o pré-processamento foi a etapa que modelou os dados selecionados na etapa anterior. A modelagem foi realizada por meio de comandos utilizando uma linguagem de programação dentro da plataforma *Power BI*. Os dados, na etapa de pré-processamento, foram avaliados e tratados segundo um processo de interação e de iteração até a etapa de transformação. Ainda, alguns dados foram substituídos ou descartados, pois apresentaram ruído.

Nessa etapa foi possível obter um roteiro para coletar dados e tratá-los na plataforma *Power BI*.

3.3.3. TRANSFORMAÇÃO E MINERAÇÃO DOS DADOS

Na transformação dos dados foi utilizado o *Power BI* para organizar os dados e transformá-los em uma tabela contendo colunas com: os dias das ocorrências, a contagem de ocorrências por dia e com os registros no S2ID. Cabe destacar que, após a programação, a transformação é feita de forma automática substituindo os dados do passado e para novos dados futuros.

Na mineração dos dados, o programa *Power BI* foi utilizado para realizar uma modelagem estatística que não requer treinos ou testes e entregam resultados que

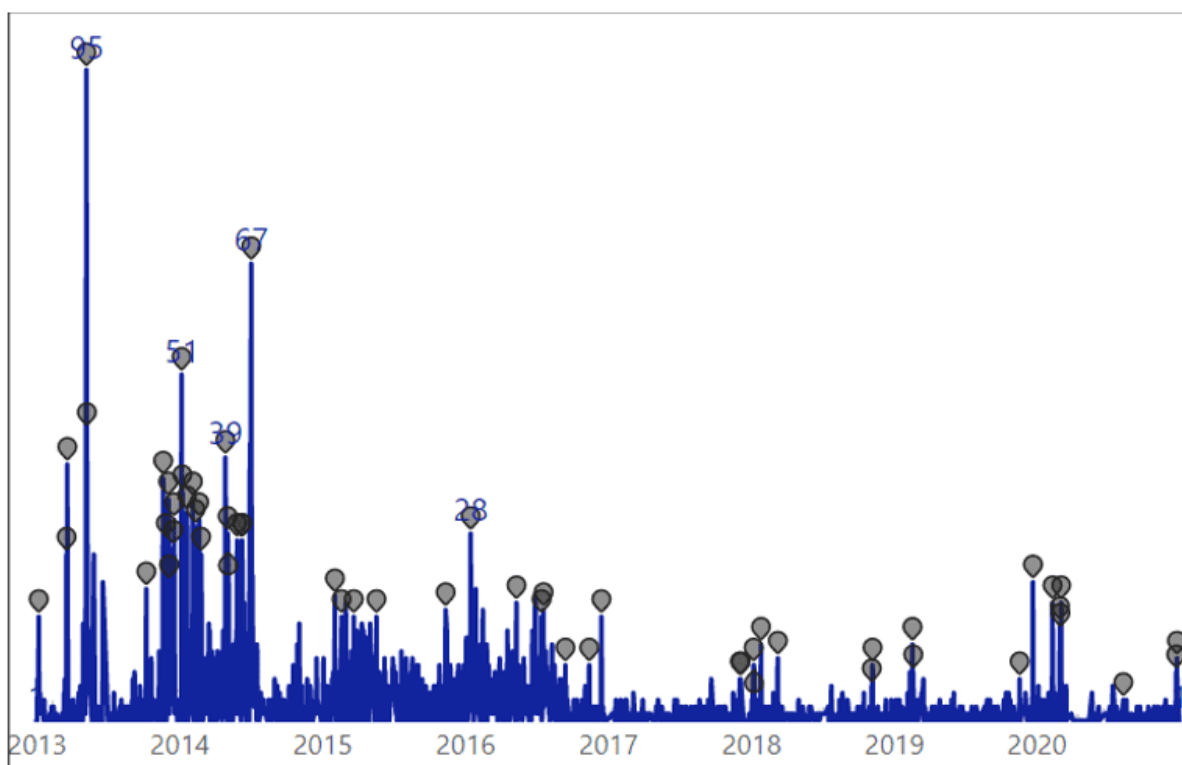
RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



podem ter algum interesse conforme demonstrado na figura 3, onde é representado graficamente os dias em que houve anomalias por meio dos círculos e, os demais dias, que são considerados normais, pelas retas em azul.

Figura 20 - Representação gráfica das anomalias geradas por meio da ferramenta *Power BI*.



Fonte: Autoria própria

Na tabela 1 são apresentadas as informações da existência de anomalia, ou seja, os dias em que houve um registro no S2ID com seu respectivo número de processo, tipo de registro, data e situação.

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Tabela 1 - Frequência dos registros no S2ID (Duque de Caxias) entre os anos de 2013 a 2020

Detectada anomalia	Processo	Tipo	data	Situação	
Sim	RJ-F-3301702-12100-20130318	Inundações	25/03/2013	Reconhecido	
Sim	RJ-F-3301702-12200-20130103	Enxurradas	25/03/2013	Reconhecido	
Não	RJ-F-3301702-23110-20130612	Incêndios industriais	03/09/2013	Registro	
Não	RJ-F-3301702-23120-20130902	Incêndios residenciais	02/10/2013	Registro	
Sim	RJ-F-3301702-12300-20130905	Alagamentos	29/10/2013	Registro	
Sim	RJ-F-3301702-23110-20131020	Incêndios industriais	01/11/2013	Registro	

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Sim	RJ-F- 3301702- 12100- 20131211	Inundações	23/12/2013	Registro
Não	RJ-F- 3301702- 22410- 20140220	Transporte de produtos perigosos rodoviário	25/02/2014	Registro
Não	RJ-F- 3301702- 11420- 20141208	Erosão de Margem Fluvial	10/12/2014	Não reconhecido
Sim	RJ-F- 3301702- 12300- 20150228	Alagamentos	09/03/2015	Registro
Sim	RJ-F- 3301702- 11420- 20150409	Erosão de Margem Fluvial	17/04/2015	Registro
Sim	RJ-F- 3301702- 13214- 20160102	Tempestade Local/Convectiva - Chuvas Intensas	05/01/2016	Registro
Sim	RJ-F- 3301702- 13215- 20151230	Tempestade Local/Convectiva – Vendaval	05/01/2016	Registro

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Sim	RJ-F- 3301702- 13214- 20160115	Tempestade Local/Convectiva - Chuvas Intensas	16/01/2016	Registro
Sim	RJ-F- 3301702- 13214- 20161113	Tempestade Local/Convectiva - Chuvas Intensas	23/11/2016	Registro
Sim	RJ-F- 3301702- 13214- 20161214	Tempestade Local/Convectiva - Chuvas Intensas	26/12/2016	Registro
Sim	RJ-F- 3301702- 13214- 20180126	Tempestade Local/Convectiva - Chuvas Intensas	27/01/2018	Registro
Sim	RJ-F- 3301702- 13214- 20181108	Tempestade Local/Convectiva - Chuvas Intensas	23/11/2018	Registro
Sim	RJ-F- 3301702- 13213- 20191223	Tempestade Local/Convectiva – Granizo	23/12/2019	Registro
Sim	RJ-F- 3301702- 13214- 20200210	Tempestade Local/Convectiva - Chuvas Intensas	18/02/2020	Registro

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Sim	RJ-F- 3301702- 11321- 20200301	Deslizamentos	02/03/2020	Registro
Sim	RJ-F- 3301702- 13214- 20201222	Tempestade Local/Convectiva - Chuvas Intensas	28/12/2020	Homologado pelo estado

Fonte: Autoria própria.

Com o *Power BI* foi gerado uma árvore de decisão, que é uma forma simples e eficaz de representar o conhecimento, na qual o conjunto de dados são identificados em classes. Na figura 4, os conjuntos de dados foram agrupados em duas classes: dias analisados e registros ou SE (FERNANDES, 2021).

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>

Figura 4 - Árvore de decisão gerada pelo algoritmo da *Microsoft*.



Fonte: Autoria própria.

Observando a figura 4 e contabilizando os dias com ou sem anomalias e os dias anômalos com ou sem registros (conforme a tabela 1) foi possível gerar as possíveis regras que necessitam ser validadas e interpretadas:

- Existem 18 dias em que houve detecção de anomalias e, nesse período, houve pelo menos um registro no sistema integrado de informações de desastres ou uma solicitação para reconhecimento do desastre;
- Houve 1 dia em que não houve detecção de anomalia e foi realizado registro, porém como houve apenas uma única ocorrência o dia foi interpretado como normal. Cabe destacar que essa solicitação foi recusada pelo Governo Federal;
- Em 141 dias foram detectadas anomalias e não houve registro ou pedido de reconhecimento, exceto em dois dias que não foram detectadas anomalias, mas foi realizado sem considerar a quantidade de ocorrências, por isso esses

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



dois dias foram considerados anômalos contabilizando um total de 143 dias anômalos;

- Por fim, os 1788 registros de dias que não foram detectadas anomalias e não houve registro e nem anomalia, encerram a contabilização da matriz.

3.4 TÉCNICAS DE AVALIAÇÃO DE DESEMPENHO DOS ALGORITMOS – AVALIAÇÃO DA MINERAÇÃO DE DADOS

Na etapa de avaliação da mineração dos dados foram utilizados os valores das classes, ou seja, as previsões realizadas por meio do algoritmo da Microsoft na matriz de confusão, para representar o número de previsões corretas e incorretas. E, em seguida, utilizou-se os algoritmos da plataforma *Waikato Environment for Knowledge Analysis (WEKA)* para avaliar e comparar o desempenho entre os algoritmos, segundo Fernandes (2021).

A Tabela 2 representa a matriz de exemplo e foi elaborada para classificar os modelos em classes e quantificar a aproximação entre o valor previsto e o valor real. Cada caso de classe é contabilizado e o total é representado na matriz.

Tabela 2 - Matriz de confusão - categorias observadas e categorias previstas

	Normal	Anomalia
Anomalia	Verdadeiro positivo	Falso positivo
Normal	Falso negativo	Negativo verdadeiro

Fonte: Adaptada de Bruce (2019).

O verdadeiro positivo e o negativo verdadeiro referem-se aos valores que foram corretamente classificados. Já os Positivos falsos e negativos falsos são os dados classificados de forma incorreta.

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Tabela 3 - Matriz de confusão gerada para representar as classificações de cada rótulo.

	Anomalia	Normal	Não classificado
Anomalia	18	1	1
Normal	141	1788	2
Total	159	1789	3

Fonte: Autoria própria.

Na Tabela 3 são apresentados os valores da quantidade de instâncias que foram avaliados corretamente e os que foram rotulados incorretamente, possibilitando, assim, contabilizar a taxa da precisão de positivos e negativos, a especificidade dos positivos e negativos, a média harmônica entre a precisão e a especificidade dos valores positivos e negativos e o valor da área sob a curva *Receiver Operating Characteristic* (ROC).

Cabe destacar que para avaliar o conjunto de dados modelados por algoritmo do *Power BI* foi utilizado a matriz de confusão e, para avaliar o conjunto de dados dos algoritmos *D. Stump*, *Forest random*, *J.48*, *Random tree* e *RepTree* foi utilizado a plataforma *WEKA* por meio da técnica de classificação usando um processo de treino-teste supervisionado com a validação cruzada.

Os resultados da avaliação dos algoritmos da plataforma *WEKA* e do *POWER BI* estão representados na tabela 4 com seus respectivos resultados

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Tabela 4: Avaliação dos algoritmos empregados para classificar o conjunto de dados.

Análise	D. Stump	Forest random	J.48	Power BI	RandomTree	RepTree
Classificado corretamente	0,94	0,99	0,98	0,98	0,99	0,98
Classificado incorretamente	0,06	0,13	0,16	0,02	0,13	0,20
F. measure VF	0,97	0,99	0,99	0,54	0,99	0,99
F. Measure VP	0,05	0,88	0,84	0,10	0,88	0,80
Negativos Falsos	1,00	1,00	1,00	0,99	1,00	1,00
Negativos verdadeiros	0,97	0,19	0,25	0,00	0,16	0,30
Positivos falsos	0,00	0,00	0,00	0,00	0,00	0,00
Positivos verdadeiros	0,03	0,81	0,75	1,00	0,84	0,70
Precisão VF	0,94	0,99	0,99	1,00	0,99	0,98
Precisão VP	0,33	0,96	0,96	0,95	0,93	0,93
ROC negativos	0,78	0,98	0,98		0,91	0,91
ROC verdadeiros	0,78	0,98	0,81		0,91	0,91
Sensibilidade FP	1,00	1,00	1,00	0,89	1,00	1,00
Sensibilidade VP	0,03	0,81	0,75	0,11	0,84	0,70

Fonte: Autoria própria.

4. RESULTADOS

Nessa seção foram analisados e discutidos os resultados da avaliação do algoritmo e discutido o desempenho deles quando comparado uns com os outros. Pela avaliação de desempenho dos algoritmos, na mineração de dados apresentaram-se resultados distintos conforme a Tabela 4. De modo geral, os dados apresentaram bons desempenhos e foram capazes de prever bem os resultados.

Bases binárias, ou seja, aquelas que possuem classe com duas opções de classificação oferecem uma modelagem mais assertiva para as previsões positivas e negativas. Nesse caso, as respostas para um dia ser classificado como anômalo foram sim ou não e, se houve registro, a opção foi sim ou não.

O algoritmo realizou uma tarefa de classificação que foi positiva, quando houve acerto na previsão dos dias analisados e dos registros realizados corretamente. Houve 3 previsões incorretas, ou seja, o algoritmo previu diferente do que de fato ocorreu.

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Analisando as previsões incorretas temos o processo RJ-F-3301702-11420-20141208, o RJ-F-3301702-23110-20130612 e o RJ-F-3301702-23120-20130902, entretanto estes erros não comprometeram o bom desempenho preditivo dos algoritmos.

Na comparação entre os algoritmos, observando os valores no interior da tabela 4, é possível avaliar sua performance. Todos os algoritmos foram excelentes para a classificação dos dias analisados, ou seja, para identificar se houve ou não anomalia.

Entretanto, a análise da contabilização pura de acertos e erros não é o suficiente para análise de predição, pois dentro da classe dos dias analisados e dos registros realizados foi importante destacar qual classe mais contribui para identificar que houve um desastre. Nesse caso, a sensibilidade e a especificidade são bem efetivas para achar uma classe específica de previsões.

Classificar um dia analisado como normal, implica que não houve desastre comprometendo a capacidade de resposta local ou que houve desastre comprometendo a capacidade de resposta local, porém não possuem relação com a contagem de ocorrências em ambos os casos. É possível que o município possa analisar, por outros critérios além da contagem de ocorrência, o motivo do comprometimento da capacidade de resposta local. Nesse caso, deve-se ser avaliado outros fatores, que não foram o escopo dessa pesquisa.

Classificar um dia analisado como anômalo, implica que não houve desastre ou que houve desastre comprometendo a capacidade resposta local considerando a existência de um acréscimo na quantidade de contagem de ocorrências.

Medidas estatísticas como a sensibilidade e precisão, ajudam a escolher melhor o algoritmo que mais contribui com a sua capacidade preditiva para uma dada classe específica. A tabela 4, com sua frequência de classificações para as métricas de sensibilidade, permitiu visualizar melhor o desempenho de cada algoritmo. É possível,

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



ainda, observar que os algoritmos são excelentes, pois possuem valores acima de 0,9, exceto o *Power BI* que é bom, pois possui valor entre 0,8 até 0,9.

Na análise de desempenho, é levado em consideração a área sob a curva ROC para atribuir um certo grau de confiabilidade no modelo, portanto, observando a tabela 4, os algoritmos que possuem área sob a curva ROC superior a 90% são confiáveis, tais como os algoritmos *Random Tree*, *Rep Tree* e *Forest random*. Os algoritmos *D. Stump* e *J.48* tiveram desempenho menor, porém bons. Em todos os algoritmos foi possível construir um modelo de árvore de decisão semelhante ao construído no *Power BI* no que diz respeito às classes, dessa forma, quanto ao algoritmo do *Power BI*, cumpre destacar que não foi avaliado o desempenho sob a curva ROC.

Foi identificado que todos os algoritmos demonstraram, exceto o *D. Stump* e o *J.48*, uma performance que pode ser reproduzida em novos conjuntos de dados, ou seja, o modelo apresentado pelo desempenho do algoritmo foi excelente. Vale destacar que não foi realizada a medida sob a curva ROC para o algoritmo do *Power BI*, portanto, não foi possível verificar a performance da classificação.

É importante considerar que o dia que foi normal e com registro ou decretação do desastre não foi considerado como uma classe distinta, pois todos os algoritmos modelaram ausência dessa nova possibilidade, ou seja, não houve contribuição com ganho de informação suficiente para ser criada uma nova classe.

A lógica aplicada na avaliação de desempenho dos algoritmos permitiu analisar a aplicabilidade da técnica de DCBD, aplicado à proteção e defesa civil a fim de detectar anomalias associadas a ocorrências de desastres. Ademais, a quantidade de registros de socorros em uma série histórica pode contribuir para a identificação de desastres por meio da técnica de DCBD.

Entretanto, cabe destacar que existem os fatores dos danos e prejuízos que podem ser significativos, as ameaças complexas e pôr fim a vulnerabilidade que também contribuem para o surgimento dos desastres. Todos esses fatores contribuem, em

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



separado ou juntos, para a ocorrência de um desastre sem que necessariamente aconteça um aumento da quantidade de ocorrências e sugestivo esgotamento da capacidade de resposta.

5. CONSIDERAÇÕES FINAIS

Portanto, pelas análises proferidas nas avaliações satisfatórias de desempenho dos algoritmos, obtidos pela técnica de DCBD, a partir da árvore de decisão gerada, com relação aos dados relativos à Defesa Civil de Duque de Caxias, conclui-se que:

- Se o dia for normal, então não existe comprometimento da capacidade local de resposta.
- Se o dia for anômalo e não houver registro ou decretação do desastre, então não existe comprometimento da capacidade local de resposta.
- Se o dia for anômalo e houver registro ou decretação do desastre, então existe comprometimento da capacidade local de resposta.

Desta forma, após a seleção dos dados, pré-processamento, transformação, mineração, avaliação dos dados e interpretação da avaliação e a generalização dos dados é possível afirmar que a quantidade de registros de socorros em uma série histórica contribui para a identificação de desastres por meio da técnica de descoberta de conhecimento em banco de dados gerando um conjunto de três regras que têm o potencial de serem úteis e são novas, respondendo, assim, a questão norteadora deste estudo.

O conjunto de regras obtidas constitui a principal descoberta oriunda do processo que foram submetidas a todas as etapas. Estas regras podem contribuir na construção de uma identificação de desastre de forma objetiva por meio do emprego de inteligência artificial e não esgotam as possibilidades de outras análises que venham a contribuir no detalhamento para determinação de desastres e possam contribuir de forma significativa para a redução de riscos e desastres.

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



Portanto, cabe destacar a adoção dos gestores, principalmente em Proteção e Defesa Civil, cada vez mais uma nova perspectiva e aceitação de uma governança facilitada por instrumentos tecnológicos de suporte à decisão. A metodologia ora apresentada neste artigo poderá servir de base para que outros trabalhos futuros possam ser idealizados no universo da ciência de dados.

REFERÊNCIAS

BRASIL. Dispõe sobre o Sistema Nacional de Defesa Civil - SINDEC e o Conselho Nacional de Defesa Civil, e dá outras providências. **Câmara legislativa**, 2005. <<https://www2.camara.leg.br/legin/fed/decret/2005/decreto-5376-17-fevereiro-2005-535733-publicacaooriginal-25097-pe.html>>. Acesso em: 2 maio 2021.

BRASIL. **Atlas brasileiro de desastres naturais: 1991 a 2012**. Centro Universitário de Estudos e Pesquisas da Universidade Federal de Santa Catarina. Florianópolis, p. 120. 2013.

BRASIL. **Instrução normativa 36**. Ministério do Desenvolvimento Regional. Brasília. 2020.

CABENA, Peter. *et al. Discovering Data Mining from Concept to Implementation*. New Jersey, 1998.

FAYYAD Usama. *et al. From Data Mining to Knowledge Discovery in Databases*. *AI Magazine*. American Association for Artificial Intelligence. Volume 17, 1996.

FERNANDES, Samir Batista. **Descoberta de conhecimento em banco de dados aplicados em defesa civil utilizando árvore de decisão**. Dissertação (mestrado profissional) -Universidade Federal Fluminense, Niterói, 2021.

NAÇÕES UNIDAS. **Oficina das Nações Unidas para a redução dos riscos de desastres**. Site da Oficina das Nações Unidas para a redução dos riscos de

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>



MULTIDISCIPLINARY SCIENTIFIC JOURNAL

**NÚCLEO DO
CONHECIMENTO**

REVISTA CIENTÍFICA MULTIDISCIPLINAR NÚCLEO DO
CONHECIMENTO ISSN: 2448-0959

<https://www.nucleodoconhecimento.com.br>

desastres,

2015.<https://www.unisdr.org/files/43291_63575sendaiframeworkportunofficialf%5B1%5D.pdf>. Acesso em: 08 nov. 2021.

RIO DE JANEIRO. **Decreto 43.599**, 2020. Acesso em 08 nov 2021:
<<http://defesacivil.rj.gov.br/index.php/legislacao-esdec/decretos-esdec>>.

Enviado: Maio, 2022.

Aprovado: Julho, 2022.

RC: 122547

Disponível em: <https://www.nucleodoconhecimento.com.br/tecnologia/protecao-e-defesa-civil>