57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

PSO-SciBERT: Otimização por Enxame de Partículas para Classificação Multimodal de Artigos Científicos sobre Eventos Climáticos Extremos

Augusto Magalhães Pinto de Mendonça[1]
augustompm@id.uff.br

Filipe Pessôa Sousa[2]
filipe.sousa@pos.ime.uerj.br

Igor Machado Coelho[1]
imcoelho@ic.uff.br

Mariza Ferro[1]
mariza@ic.uff.br

RESUMO

Eventos climáticos extremos representam um desafio global crescente, com impactos severos em diversas regiões, demandando ferramentas eficientes de otimização e análise do conhecimento científico. Este trabalho apresenta um sistema de classificação automática de artigos sobre desastres naturais, com duas contribuições: o conjunto AMCLIMA-BR, com 700 artigos classificados em três dimensões (tipo de desastre, fase de gerenciamento e paradigma de aprendizado), e uma abordagem de otimização via enxame de partículas (PSO) aplicada ao SciBERT para dados desbalanceados. Resultados demonstram melhorias significativas com PSO-SciBERT, atingindo ganhos de até 26,4% em acurácia balanceada, evidenciando robustez do modelo otimizado.

PALAVRAS CHAVE. Processamento de Linguagem Natural, Desastres Naturais, Metaheurísticas.

Tópicos: IA- PO e IA, AG&MA – PO na Agricultura, Meio Ambiente e Sustentabilidade

ABSTRACT

Extreme climate events represent a growing global challenge, with severe impacts across diverse regions, demanding efficient tools for optimization and analysis of scientific knowledge. This work presents an automatic classification system for articles on natural disasters, with two contributions: the AMCLIMA-BR dataset, comprising 700 articles classified across three dimensions (disaster type, management phase, and learning paradigm), and a particle swarm optimization (PSO) approach applied to SciBERT for imbalanced data. Results demonstrate significant improvements with PSO-SciBERT, achieving gains of up to 26.4% in balanced accuracy, evidencing the robustness of the optimized model.

KEYWORDS. Natural Language Processing. Natural Disasters. Metaheuristics.

Topics: AI- OR and AI, AG&MA – OR in Agriculture, Environment, and Sustainability

[1]Instituto de Computação - Universidade Federal Fluminense
Av. Gal. Milton Tavares de Souza, São Domingos – Niterói – RJ CEP: 24210-346

[2]CompMat - Instituto de Matemática e Estatística - Universidade do Estado do Rio de Janeiro
R. São Francisco Xavier, 524 - Maracanã, Rio de Janeiro - RJ, 20550-013

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

1. Introdução

Os eventos climáticos extremos representam um desafio global cada vez maior. Projeções indicam que cada grau de aquecimento global pode resultar em um aumento de 7% apenas nas precipitações extremas, intensificando os impactos e danos associados ao escalonamento da ocorrência desta e outras ameaças [WMO, 2025]. O relatório de síntese do IPCC (2023) confirma que "mudanças generalizadas e rápidas ocorreram na atmosfera, oceano, criosfera e biosfera" [IPCC, 2023]. O mesmo trabalho enfatiza que comunidades vulneráveis, que historicamente contribuíram menos para as mudanças climáticas, serão os mais afetados, como os países do Sul Global, incluindo o Brasil.

Este artigo desenvolve um sistema de classificação automática de artigos científicos relacionados a eventos climáticos extremos, utilizando o modelo SciBERT com técnicas de otimização. As duas contribuições principais são: (1) um novo conjunto de dados composto por 700 artigos científicos aplicados a ameaças climáticas, rotulado em três domínios complementares (tipo de desastre, fase de gerenciamento e paradigma de aprendizado de máquina aplicado), o corpus foi manualmente catalogado, e (2) e uma aplicação de otimização por enxame de partículas (PSO) para aprimorar o desempenho de um modelo de *deep learning* para classificação textual de conjuntos de dados desbalanceados.

Observa-se no Brasil uma intensificação dos eventos climáticos extremos, evidenciada em 2024 por recordes de temperatura e diversos desastres. Segundo o Cemaden, o Brasil enfrentou "a seca mais intensa já registrada", com cerca de 25% dos municípios afetados por seca severa e extrema [Ministério da Ciência, 2025]. Simultaneamente, o Rio Grande do Sul sofreu enchentes históricas causando 183 mortes e impactando $2,4$ milhões de pessoas. A ANA classificou estas enchentes como o maior desastre natural da história do estado, com chuvas de duração, intensidade e abrangência territorial sem precedentes. Estudos indicam que as mudanças climáticas tornaram estes eventos mais prováveis e intensos, com o Sul do Brasil projetando um aumento na frequência de cheias até cinco vezes maior de eventos antes raros [Agência Nacional de Águas e Saneamento Básico, 2024].

Apesar dos avanços científicos, persistem desafios no uso efetivo da informação disponível. [Sietsma et al., 2024] evidenciam que políticas de adaptação às mudanças climáticas são urgentes, mas o volume e variedade de dados dificultam decisões informadas. Esta realidade é crítica no contexto climático, onde localizar informações relevantes torna-se determinante para ações eficazes. A sobrecarga informacional, quando repositórios científicos acumulam conteúdo nem sempre aplicável, constitui barreira significativa. Conforme [Bawden e Robinson, 2009], "a sobrecarga de informação ocorre quando a informação recebida se torna um obstáculo em vez de uma ajuda, mesmo quando essa informação é potencialmente útil".

Diante deste quadro, surge a necessidade de ferramentas que organizem o conhecimento científico sobre desastres naturais. Este trabalho propõe o Corpus AMCLIMA-BR, novo conjunto de dados composto por 700 artigos sobre ameaças climáticas, classificados segundo critérios alinhados à legislação brasileira e padrões internacionais. O diferencial desta base é sua categorização em três dimensões complementares, integrando taxonomias oficiais reconhecidas e padronizações técnicas estabelecidas, proporcionando uma estrutura classificatória coerente com marcos regulatórios nacionais e diretrizes internacionais para gestão de desastres.

Durante a classificação, observou-se um desequilíbrio na distribuição entre categorias, refletindo a realidade da produção científica, onde certos tipos de desastres recebem maior atenção. Tal característica tornou a classificação com Aprendizado de Máquina desafiadora, exigindo modelos complexos com múltiplos hiperparâmetros a serem ajustados.

A segunda contribuição deste trabalho é a aplicação de otimização por enxame de partículas

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

(PSO) [Kennedy e Eberhart, 1995] para aprimorar o modelo SciBERT [Beltagy et al., 2019], através de *fine-tuning* para classificação de textos desbalanceados. O SciBERT foi escolhido por ser pré-treinado em texto científico e já demonstrar excelentes resultados em classificação textual [Inayah et al., 2023]. Nossa abordagem emprega PSO para otimizar automaticamente hiperparâmetros e balanceadores, explorando sistematicamente um espaço de configurações compatível com otimização em variáveis discretas e categóricas deste problema.

Este trabalho está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico, abordando conceitos fundamentais sobre processamento de linguagem natural para textos científicos e técnicas de otimização por enxame de partículas. A Seção 3 detalha a construção da base de dados AMCLIMA-BR e a implementação do modelo SciBERT otimizado via PSO. Na Seção 4, são apresentados e discutidos os resultados, comparando o desempenho do modelo proposto com abordagens *baseline*. Por fim, a Seção 5 traz as conclusões, limitações e perspectivas para trabalhos futuros.

2. Referencial Teórico

Nesta seção são apresentados os conceitos básicos de Processamento de Linguagem Natural (PLN) e otimização de modelos com metaheurísticas que fundamentam este bem como alguns trabalhos relacionados.

2.1. Processamento de Linguagem Natural

O PLN é uma área da inteligência artificial que permite a interação entre computadores e a linguagem humana [Caseli e Nunes, 2024], com aplicações significativas, entre outros, em textos científicos. A análise computacional de artigos científicos apresenta desafios particulares que a diferencia do processamento de textos convencionais. A terminologia especializada de cada área do conhecimento exige modelos capazes de interpretar vocabulários restritos e contextuais. Soma-se a isso a alta densidade informacional dos textos científicos, onde cada sentença frequentemente contém conceitos interrelacionados, demandando técnicas avançadas para identificação e classificação precisa do conteúdo.

A classificação automática de artigos científicos tem avançado com a aplicação de técnicas de Aprendizado de Máquina (AM) e PLN. Um trabalho relevante nesta área foi conduzido por [Yesenia e Segarra-Faggioni, 2021], onde os autores investigaram a classificação de artigos científicos publicados na base Scopus de acordo com linhas de pesquisa universitárias. Os autores aplicaram algoritmos *K-Nearest Neighbors* (K-NN) e Análise Discriminante Linear (LDA) a um corpus de 596 artigos publicados entre 2003 e 2017. O modelo proposto, que extraiu informações relevantes dos resumos, serviu de base para este estudo que também utilizou o Scopus para buscar os 700 artigos que foram manualmente classificados.

2.2. Modelos de Classificação de Texto

A classificação textual constitui um componente essencial em diversas aplicações, desde detecção de spam até análise de sentimentos. Considerando o crescente volume de documentos, métodos de Aprendizado de Máquina cada vez mais avançados são necessários. [Hsu, 2020] conduziu uma avaliação abrangente do desempenho de múltiplos modelos supervisionados para classificação textual. O estudo comparou algoritmos como regressão logística, árvores de decisão, máquinas de vetores de suporte (SVM), AdaBoost, *random forest*, Naive Bayes multinomial, *perceptrons* multicamadas e *gradient boosting*. Os resultados indicaram que SVM, regressão logística e *perceptrons* multicamadas apresentaram melhor desempenho geral. O autor destacou que modelos que generalizam bem em espaços de características de dimensões maiores têm melhor poder preditivo na classificação de texto.

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

Para conjuntos de dados pequenos, comuns devido ao custo de anotação, [Riekert et al., 2021] investigaram como classificadores de texto devem ser projetados para alcançar alta precisão com pouca informação. Os autores avaliaram o dimensionamento de conjuntos de treinamento, empregando um *design* fatorial de quatro vias com 32 combinações, utilizando subconjuntos de conjuntos públicos. Outro desafio na classificação textual é o desequilíbrio de classes, tratado neste trabalho. [Chen et al., 2024] apresentam uma revisão sobre aprendizado com dados desbalanceados, destacando que distribuições desequilibradas restringem a utilidade prática de modelos. Os autores exploram estratégias incluindo técnicas de *oversampling*, *undersampling*, métodos híbridos e ponderação de classes que atribui maior importância a exemplos de classes menos representadas durante o treinamento.

2.3. Evolução dos Modelos de Linguagem

A evolução do PLN foi marcada pelos modelos baseados na arquitetura *Transformer*. [Vaswani et al., 2017] propuseram uma arquitetura inteiramente baseada em mecanismos de atenção, eliminando recorrência e convoluções. Diferentemente de modelos sequenciais anteriores, o *Transformer* facilita a paralelização, servindo como base para o BERT (*Bidirectional Encoder Representations from Transformers*) [Devlin et al., 2019]. O BERT foi projetado para pré-treinar representações bidirecionais profundas, condicionando simultaneamente nos contextos à esquerda e à direita. Sua principal vantagem é a capacidade de *fine-tuning* para tarefas específicas com apenas uma camada adicional, alcançando resultados estado da arte em múltiplas tarefas de PLN.

Para domínios científicos, [Beltagy et al., 2019] desenvolveram o SciBERT, treinado em $1,14$ milhão de artigos científicos. Os autores ressaltam que dados anotados para PLN científico são desafiadores e dispendiosos, motivando um modelo pré-treinado específico. O SciBERT demonstrou melhorias significativas em tarefas como classificação de sentenças e análise de dependências, sendo por isso selecionado como base para classificação do conjunto AMCLIMA-BR.

2.4. Otimização de Modelos com Metaheurísticas

A correta aplicação de técnicas de avaliação e seleção de modelos é crucial tanto em pesquisa quanto na indústria. Raschka [2018] revisa estas abordagens, destacando que o método *holdout* não é recomendado para conjuntos pequenos, sugerindo alternativas como *bootstrap* para estimar incertezas e diversas técnicas de validação cruzada como *leave-one-out* e *k-fold*, com análise do trade-off viés-variância. Para comparação com dados limitados, recomenda métodos como o teste F combinado $5x2$ *folds* e validação cruzada aninhada, que proporcionam avaliações mais robustas.

Entre as técnicas de otimização aplicáveis, destaca-se a Otimização por Enxame de Partículas (PSO), introduzida por [Kennedy e Eberhart, 1995], desenvolvida a partir de modelos sociais simplificados inspirados no comportamento coletivo de pássaros e peixes. Suas principais vantagens incluem simplicidade conceitual, implementação descomplicada e eficiência computacional. As aplicações de PSO em AM têm demonstrado resultados promissores, como evidenciado por [Dongoran et al., 2018], que comprovaram sua eficácia na otimização de pesos de características em classificadores baseados em *Learning Vector Quantization*, resultando em melhorias significativas no desempenho, capacidade que pode ser estendida para múltiplos aspectos de modelos complexos.

3. Metodologia

Nesta seção são apresentados os passos metodológicos adotados para a construção do corpus de teste AMCLIMA-BR, detalhes a aplicação da otimização PSO e sua utilização no ajuste do modelo SciBERT, incluindo detalhes dos hiperparâmetros e seus valores.

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

3.1. Construção do Corpus AMCLIMA-BR

O corpus AMCLIMA-BR foi desenvolvido por meio de um processo metodológico rigoroso, desde a seleção de fontes até a classificação multidomínio dos artigos. A base Scopus foi selecionada como fonte primária após análises comparativas com outras plataformas, demonstrando menor redundância e maior completude de metadados, reduzindo significativamente o retrabalho no processamento posterior.

O processo de coleta seguiu duas etapas: inicialmente, uma busca estruturada abrangendo os últimos 10 anos (2015-2025) resultou em aproximadamente 500 artigos. Esta busca integrou termos de três domínios principais: aprendizado de máquina, tipos de desastres naturais e fases de gerenciamento. Posteriormente, para categorias com representatividade inferior a 20 artigos, realizou-se uma busca complementar expandindo o horizonte temporal para 20 anos (2005-2025), acrescentando mais 200 documentos ao corpus final.

Os 700 artigos científicos coletados foram classificados manualmente em três dimensões complementares, com cada classificação sendo realizada em consenso por pelo menos dois autores especialistas na área. Este processo assegurou consistência nas anotações e mitigou possíveis vieses individuais de interpretação.

A primeira dimensão classifica os eventos segundo o tipo de ameaça, seguindo o segundo nível da Classificação e Codificação Brasileira de Desastres (COBRADE) [Secretaria Nacional de Proteção e Defesa Civil, 2024], que padroniza a taxonomia de desastres. Esta classificação facilita a integração com sistemas oficiais de monitoramento e permite comparabilidade com registros históricos nacionais.

A segunda dimensão refere-se à fase de gerenciamento do desastre, estruturada conforme adaptação do Plano Nacional de Proteção e Defesa Civil, em consonância com a Lei 12.608/2012 e com o Marco de Sendai para a Redução do Risco de Desastres 2015-2030 [United Nations Office for Disaster Risk Reduction, 2015]. Esta abordagem permite identificar pesquisas relevantes para cada etapa do ciclo de gestão, otimizando o acesso à informação para tomadores de decisão. As categorias prevenção e mitigação foram agrupadas, segundo sua descrição; a primeira trata de "evitar riscos de desastres" e a segunda "medidas destinadas a reduzir, limitar ou evitar o risco de desastre", sendo similar demais para classificação manual ou sistemática.

A terceira dimensão, focada no paradigma de AM aplicado, baseia-se na hierarquia do ACM Computing Classification System (CCS) em seu caminho "Computing methodologies → Machine learning → Learning paradigms" [Association for Computing Machinery, 2012], que possui quatro classificações principais, às quais adicionamos a categoria "Não Classificado" para completude taxonômica. Esta classificação permite identificar quais abordagens de aprendizado têm sido aplicadas a cada tipo de desastre e fase de gestão.

Na Tabela 1 é apresentada a distribuição de artigos no corpus AMCLIMA-BR nas três dimensões de classificação.

Tabela 1: Distribuição de artigos no conjunto de dados AMCLIMA-BR nas três dimensões de classificação

Tipo de Desastre			**Fase de Gerenciamento**			**Paradigma de Aprendizado**		
Categoria	*N*	**%**	**Categoria**	*N*	**%**	**Categoria**	*N*	**%**
Geológico	267	38,14	Preparação	321	45,86	Aprendizado Supervisionado	545	77,86
Hidrológico	192	27,43	Prevenção	191	27,29	Aprendizado Multi-Tarefa	49	7,00
Meteorológico	78	11,15	Resposta	124	17,71	Aprendizado Não-Supervisionado	44	6,29
Climatológico	65	9,29	Recuperação	47	6,71	Não Classificado	38	5,43
Não Classificado	98	14,00	Não Classificado	17	2,43	Aprendizado por Reforço	24	3,43
Total geral			**700 artigos — 100%**					

A análise da distribuição de artigos revelou pronunciado desbalanceamento nas três di-

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

mensões (Tabela 1), refletindo tendências na literatura científica. Na dimensão de paradigmas de aprendizado, o "Aprendizado Supervisionado" domina com 77,86% dos artigos, enquanto na dimensão de tipos de desastre, "Geológico" (38,14%) e "Hidrológico" (27,43%) são predominantes. Quanto às fases de gerenciamento, "Preparação" representa 45,86% do corpus. Este desbalanceamento natural constitui um desafio significativo para algoritmos de classificação automática, motivando a segunda contribuição.

3.2. Implementação do PSO-SciBERT

A segunda contribuição principal deste trabalho consiste na aplicação sistemática do PSO para aprimorar o desempenho do modelo SciBERT na classificação de texto científico com classes desbalanceadas. O SciBERT é uma arquitetura de *deep learning* baseada em *Transformer* com 12 camadas codificadoras, 768 neurônios nas camadas ocultas e 12 cabeças de atenção em sua configuração original ($allenai/scibert_scivocab_uncased$).

Nossa implementação utiliza o modelo pré-treinado como base e adiciona camadas específicas para classificação: uma camada de *pooling* para o token [CLS], um conjunto de camadas de *dropout* aplicadas em paralelo para maior robustez, e uma camada final de projeção linear para as classes-alvo. O PSO foi escolhido por sua capacidade de otimizar tanto parâmetros discretos quanto contínuos simultaneamente, sem exigir gradientes da função objetivo. Um ponto importante é que esta versão atual gera três modelos, um otimizado para cada domínio.

A implementação do PSO-SciBERT explora um espaço multidimensional de 8 parâmetros-chave. Os intervalos foram definidos através de experimentação preliminar e análise da literatura, com margens para evitar soluções de borda. Devido à intensidade computacional da otimização, todos os parâmetros contínuos foram discretizados em conjuntos de valores pré-selecionados, como ilustrado na Tabela 2.

Tabela 2: Espaço de busca otimizado para o PSO-SciBERT

Hiperparâmetro	**Valores**
Estratégias de balanceamento	None, Random Oversampling, SMOTE, Combinada
Taxa de aprendizado (LR)	2.0×10^{-5}, 2.5×10^{-5}, 3.0×10^{-5}, 4.0×10^{-5}, 5.0×10^{-5}
Dropout	0.20, 0.25, 0.30, 0.35, 0.40, 0.45
Decaimento de peso (WD)	1.0×10^{-5}, 2.0×10^{-5}, 5.0×10^{-5}, 1.0×10^{-4}, 2.0×10^{-4}
Camadas descongeladas	4, 5, 6, 7
Cabeças de atenção (CA)	4, 8, 12
Scheduler	cosine_warmup, one_cycle, cyclic
Gamma da Focal Loss (γ)	1.5, 2.0, 2.5

A estratégias de balanceamento Combinada se refere à junção em partes iguais das técnicas de *Random Undersampling* nas classes majoritárias com SMOTE nas classes minoritárias. Adicionalmente, os balanceadores foram apenas aplicados aos dados de treinamento.

Os valores finais encontrados pelo PSO-SciBERT, dentro do escopo computacional para este trabalho, para cada subtarefa são apresentados na Tabela 3. Parâmetros não listados em uma subtarefa mantiveram seu valor padrão do SciBERT.

Tabela 3: Hiperparâmetros otimizados pelo PSO-SciBERT para cada subtarefa

Subtarefa	**Estratégia balanceamento**	**LR**	**Dropout**	**WD**	γ	**Camadas descong.**	**Scheduler**	**CA**
`Tipo de Desastre`	`Combinada`	2.5×10^{-5}	0.35	5.0×10^{-5}	2.0	4	`one_cycle`	–
`Fase de Gerenciamento`	`Random Oversampling`	4.0×10^{-5}	0.30	2.0×10^{-5}	–	–	`cyclic`	4
`Paradigma de Aprendizado`	`Random Oversampling`	3.0×10^{-5}	0.40	1.0×10^{-4}	2.5	6	`cosine_warmup`	8

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

O PSO utiliza um enxame de partículas que navegam no espaço de busca guiadas por suas melhores posições históricas e pela melhor global. Nossa adaptação mapeia posições contínuas para índices discretos dos hiperparâmetros, permitindo otimização em espaços mistos. Utilizamos 12 partículas por 30 gerações, configuração determinada empiricamente: testes preliminares mostraram que esse dimensionamento encontra convergência consistente (tipicamente em 15-20 gerações) com custo computacional viável (4h por domínio). Cada partícula representa uma configuração completa do modelo e é avaliada por meio de validação cruzada estratificada com 2 *folds*, escolha motivada pela necessidade de conter o custo de treinamento em cenários de aprendizado com dados desbalanceados. A função objetivo $\mathcal{F}(\boldsymbol{\theta})$ utilizada para avaliação de cada configuração é definida pela média da acurácia balanceada em todos os *folds*, como apresentado na Equação 1. Para equilibrar exploração e tempo computacional, adotamos *early stopping* sempre que não ocorre uma melhoria mínima de $0{,}25\%$ na função-objetivo durante 5 gerações consecutivas.

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{1}{K} \sum_{k=1}^{K} Acc_{bal}^{(k)}(\boldsymbol{\theta}) \tag{1}$$

onde $Acc_{bal}^{(k)}(\boldsymbol{\theta})$ é a acurácia balanceada obtida no k-ésimo fold com os parâmetros $\boldsymbol{\theta}$, e $K = 2$ é o número de folds na validação cruzada utilizada durante a otimização.

Para a avaliação final, após identificar os hiperparâmetros ótimos, aplicamos validação cruzada estratificada 5-fold independente. Cada fold usa 80% dos dados para treino e 20% para teste, preservando rigorosamente a proporção de classes; o bloco de validação (20%) é também empregado para *early stopping*. Esta avaliação foi realizada separadamente para cada tarefa de classificação (*Tipo de Desastre*, *Paradigma de Aprendizado* e *Fase de Gerenciamento*). Todas as partições foram geradas com a semente `numpy.random.seed(42)` para garantir reprodutibilidade. Reportamos a *média* $\pm$ *desvio-padrão* das métricas obtidas nos conjuntos de teste.

Os resultados demonstraram que diferentes configurações de balanceamento são mais efetivas para cada domínio: a estratégia combinada para Tipo de Desastre e *Random Oversampling* tanto para Fase de Gerenciamento quanto para Paradigma de Aprendizado. O modelo implementa um esquema de *fine-tuning* com congelamento seletivo das camadas iniciais, Cabeças de Atenção (*Attention Heads*) e *Scheduler* cujos tipos e números exatos são determinados pelo PSO, melhorando o desempenho quanto à acurácia balanceada.

O processo de otimização foi desenvolvido em Python e executado em um sistema Intel i7 13700 5.2GHz Max com uma GPU NVIDIA RTX 4090 com 24GB de memória, demandando cerca de 4 horas por domínio, totalizando aproximadamente 12 horas para os três domínios de classificação. A implementação completa do PSO-SciBERT utilizou PyTorch e Hugging Face Transformers [Wolf et al., 2020], com código customizado para balanceamento de classes, *focal loss* adaptativa por dificuldade de amostra, e otimização automática. Incorporamos também uma função de perda *Focal Loss* com ajuste gamma por classe, permitindo penalização diferenciada para classes minoritárias. O algoritmo implementa mecanismos de *checkpointing* para resiliência contra falhas, permitindo retomar a otimização a partir do último estado salvo.

4. Resultados

Nesta seção, apresentamos os resultados dos experimentos de classificação automática de artigos científicos, comparando o desempenho do SciBERT padrão com nossa abordagem PSO-SciBERT para as três dimensões de classificação.

4.1. Avaliação Comparativa

A Tabela 4 apresenta o desempenho comparativo entre o SciBERT com configurações padrão e o modelo otimizado via PSO, para cada uma das três dimensões de classificação.

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

Tabela 4: Desempenho comparativo entre SciBERT-padrão e PSO-SciBERT (resultados médios ± DP em 5-fold CV estratificada; hiperparâmetros obtidos via PSO com 2-fold CV)

Métrica	*Tipo de Desastre*			*Fase de Gerenciamento*			*Paradigma de Aprendizado*		
	SciBERT	**PSO-SciBERT**	**Melhoria**	**SciBERT**	**PSO-SciBERT**	**Melhoria**	**SciBERT**	**PSO-SciBERT**	**Melhoria**
Bal. Acc.	0,751 ± 0,13	**0,882 ± 0,02**	*+17,4%*	0,705 ± 0,04	**0,731 ± 0,13**	*+3,6%*	0,574 ± 0,31	**0,725 ± 0,09**	*+26,4%*
F1-Score	0,778 ± 0,09	**0,895 ± 0,02**	*+15,0%*	0,755 ± 0,03	**0,803 ± 0,02**	*+6,4%*	0,773 ± 0,15	**0,885 ± 0,03**	*+14,5%*
F2-Score	0,787 ± 0,09	**0,895 ± 0,02**	*+13,6%*	0,752 ± 0,04	**0,805 ± 0,02**	*+7,1%*	0,770 ± 0,15	**0,888 ± 0,03**	*+15,3%*
MCC	0,746 ± 0,09	**0,861 ± 0,02**	*+15,5%*	0,646 ± 0,05	**0,715 ± 0,02**	*+10,8%*	0,434 ± 0,36	**0,699 ± 0,08**	*+61,3%*

Como pode ser observado na Figura 1, a estratégia PSO-SciBERT supera o SciBERT-padrão em todas as métricas avaliadas. No domínio *Tipo de Desastre*, o F_2*-score* aumentou em $13,64\%$ (de $0,7873$ para $0,8947$), indicando maior sensibilidade às classes minoritárias. O F_1 *médio* elevou-se em $15,02\%$ (de $0,7784$ para $0,8953$). Na *Acurácia* e na *Acurácia Balanceada*, observamos melhorias de $11,98\%$ (de $0,8000$ para $0,8958$) e $17,35\%$ (de $0,7512$ para $0,8815$), evidenciando melhor equilíbrio entre classes majoritárias e minoritárias. O *coeficiente de correlação de Matthews* (*MCC*)[1] aumentou de $0,7456$ para $0,8613$ (incremento de $15,52\%$), confirmando redução significativa de falsos positivos e falsos negativos. Isso atesta que o PSO não só melhora o acerto global mas também reduz decisivamente falsos positivos e negativos. Em termos de *Tempo de Execução*, observamos uma correlação entre o grau de desbalanceamento e o custo computacional. O modelo PSO-SciBERT para *Tipo de Desastre* requer 393 segundos ($78,6s$ por *fold*), 5 vezes mais que o modelo base, enquanto o domínio *Paradigma de Aprendizado*, com desequilíbrio mais severo, demandou 948 segundos, 12 vezes mais tempo. Este aumento é atribuído principalmente ao mecanismo de *ensemble* de *dropouts* e número maior de épocas necessárias para aprendizado eficaz das classes minoritárias.

Como pode ser observado também na Figura 2, revela um desempenho de excelência em "Aprendizado por Reforço", com 100% de acertos. Entretanto, este resultado deve ser interpretado com cautela devido ao reduzido tamanho amostral ($n = 24$), que limita sua significância estatística. Tal fenômeno, comum em conjuntos de dados desbalanceados, está sujeito ao viés de amostra pequena, onde a variância do estimador de desempenho aumenta inversamente ao tamanho da classe minoritária, podendo levar a conclusões superestimadas sobre a capacidade de generalização do modelo para esta categoria.

"Aprendizado Supervisionado" atinge 92 % de precisão, sofrendo pequenas confusões de 3 % sendo classificados erroneamente como "Aprendizado Não Supervisionado" e 3 % como "Não Classificado". O "Aprendizado Não Supervisionado" registra 89 % de acerto, mas 11 % de seus exemplos foram classificados como "Não Classificado", e o "Aprendizado Multi-tarefa" alcança 80 % de acertos, com 20 % dos casos erroneamente classificados com do paradigma supervisionado. Já a categoria "Não Classificado" obteve apenas 57 % de acertos, classificando 29 % dos exemplos como "Aprendizado Supervisionado" e 14 % como "Aprendizado por Reforço".

No domínio Tipo de Desastre, observa-se discriminação perfeita da classe "Climatológico" (100 %) e elevado desempenho em "Hidrológico" (95 %) e "Geológico" (94 %), que apresentam margens de erro de apenas 2–3 % entre si. A categoria "Meteorológico" atinge 80 % de acerto, mas confunde 13 % dos exemplos com "Hidrológico" e 7 % com "Não Classificado". Por sua vez, "Não Classificado" obtém 85 % de acertos, alocando 10 % dos textos incorretamente como "Meteorológico" e 5 % como "Geológico".

Finalmente, na Fase de Gerenciamento, o rótulo "Não Classificado" alcança 100 % de

[1] Disponível em `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.matthews_corrcoef.html`

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

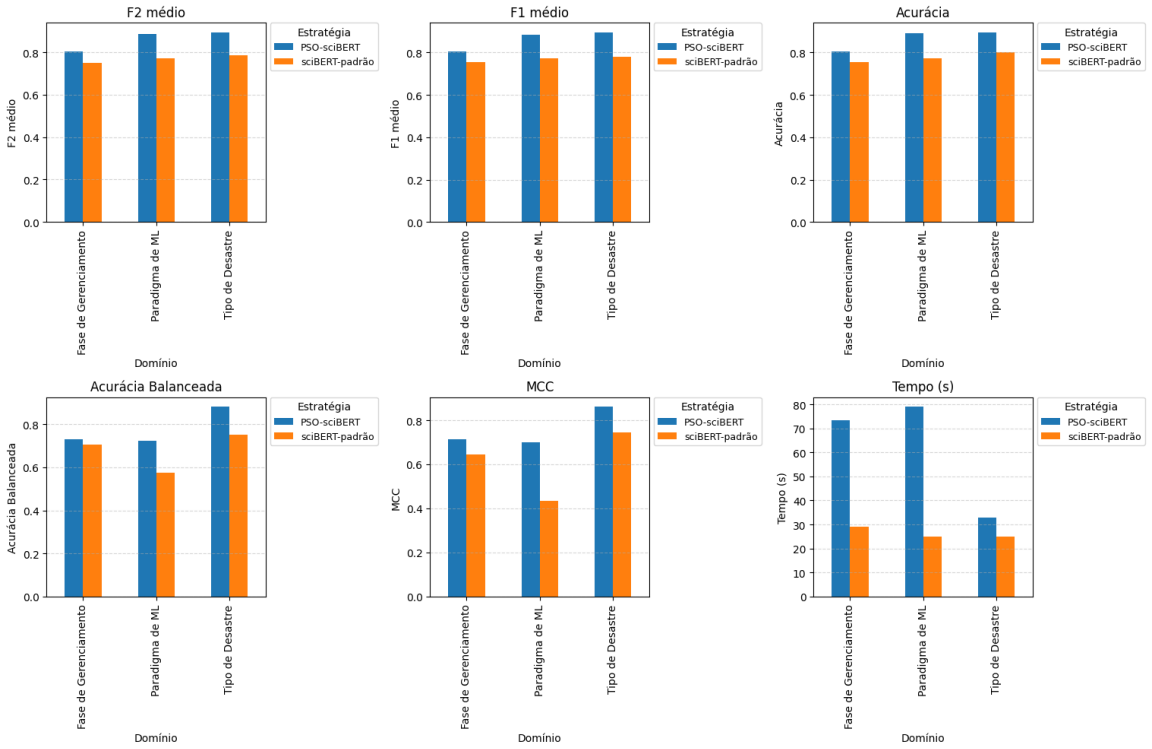

Figura 1: Comparação das métricas F_2 *médio*, F_1 *médio*, Acurácia, Acurácia Balanceada, MCC e tempo de execução (s) entre as estratégias **PSO-sciBERT** e **sciBERT-padrão**, avaliadas nos domínios *Fase de Gerenciamento*, *Paradigma de Aprendizado* e *Tipo de Desastre*.

precisão, enquanto "Preparação" e "Prevenção" apresentam 86 % e 74 % de acerto, respectivamente, com 12 % de exemplos de "Preparação" confundidos com "Prevenção" e 24 % de "Prevenção" lançados em "Preparação". A etapa "Resposta" soma 92 % de acertos, com apenas 4 % de confusão para "Recuperação", ao passo que "Recuperação" registra 78 % de precisão, dividindo igualmente 11 % dos casos entre "Prevenção" e "Resposta". Esses resultados indicam que o modelo domina bem categorias conceitualmente distantes, mas encontra maior dificuldade na distinção de classes vizinhas, especialmente nas fases de gerenciamento e nos subtipos de eventos climáticos.

Os resultados demonstram que a otimização via PSO proporcionou melhorias em todas as métricas avaliadas. Conforme demonstrado na Tabela 4, a dimensão *Paradigma de Aprendizado* apresentou o maior ganho em acurácia balanceada (+26, 4%) e MCC (+61, 3%), indicando que esta classificação se beneficiou da otimização de hiperparâmetros e das estratégias de balanceamento. A dimensão *Tipo de Desastre* também apresentou ganhos, com aumento de $17,4\%$ na acurácia balanceada, enquanto a dimensão *Fase de Gerenciamento* mostrou melhorias mais modestas. Por fim, o MCC para Paradigma de Aprendizado apresentou melhoria excepcional de $61,3\%$, confirmando a eficácia da abordagem proposta em lidar com o significativo desbalanceamento de classes neste domínio.

4.2. Análise dos Parâmetros Otimizados

A otimização PSO identificou configurações específicas para cada domínio, destacando-se a estratégia de balanceamento combinada para *Tipo de Desastre* e *random oversampling* para *Paradigma de Aprendizado* e *Fase de Gerenciamento*. Nos hiperparâmetros, observamos taxas de aprendizado menores (2-4e-5) em todos os domínios, valores de gamma da *Focal Loss* ajustados à

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

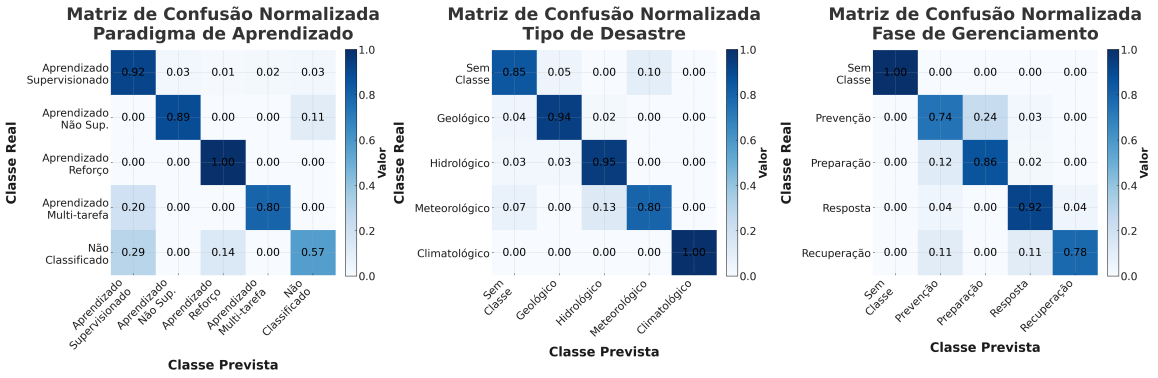

Figura 2: Matrizes de confusão normalizadas: (a) paradigma de aprendizado de máquina, (b) tipo de desastre e (c) fase de gerenciamento.

dificuldade de classificação de cada domínio (1,5-2,5), e congelamento consistente de 4-6 camadas do *transformer*.

A redução significativa no desvio padrão das métricas, especialmente nos domínios *Tipo de Desastre* ($\pm 0,0172$ vs. $\pm 0,1285$ na acurácia balanceada) e *Paradigma de Aprendizado* ($\pm 0,0888$ vs. $\pm 0,3055$), evidencia que os modelos otimizados apresentam maior estabilidade e robustez entre diferentes divisões dos dados. Esta estabilidade é relevante para aplicações práticas, onde a consistência do desempenho é tão importante quanto os valores absolutos das métricas, especialmente no contexto de classes desbalanceadas típicas de conjuntos de dados reais como o AMCLIMA-BR.

5. Conclusão

Este trabalho apresentou duas contribuições principais: o conjunto de dados original, AMCLIMA-BR, com 700 artigos científicos classificados em três dimensões complementares, e uma abordagem baseada em otimização por enxame de partículas para aprimorar o modelo SciBERT em contextos de desbalanceamento de classes.

Os resultados demonstram a eficácia da abordagem proposta, com o PSO-SciBERT superando o modelo padrão em todas as dimensões avaliadas. As melhorias mais expressivas ocorreram para "Paradigma de Aprendizado" (+26,4% em acurácia balanceada) e "Tipo de Desastre" (+17,4%). A especificidade das estratégias de balanceamento para cada domínio evidenciou a importância de considerar as características particulares de cada dimensão: abordagem combinada para "Tipo de Desastre" e random oversampling para as demais dimensões. É importante ressaltar que o tempo de execução do SciBERT padrão (15 segundos por *fold*) oferece vantagem para aplicações com restrições computacionais onde ganhos moderados de acurácia são aceitáveis.

A redução no desvio padrão das métricas indica maior robustez do modelo otimizado, característica fundamental para aplicações reais. Entre as limitações, destacamos o alto custo computacional do processo de otimização.

Como trabalhos futuros, pretendemos: desenvolver um modelo único capaz de classificar simultaneamente as três dimensões, aproveitando suas inter-relações; expandir o corpus com artigos recentes, possibilitando uma classificação automática; e implementar técnicas de transferência de aprendizado entre domínios.

O *dataset* AMCLIMA-BR está disponível para colaboração e expansão em: `https://github.com/augustompm/PSO-SciBERT`. Os resultados deste trabalho têm aplicações potenciais para a gestão de conhecimento de desastres naturais, em um cenário onde a rápida localização de informações e métodos relevantes podem impactar as ações de pesquisa em eventos climáticos extremos.

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior –Brasil (CAPES) – Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) processo (PQ 2021 315059/2021-4) e da Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) processo (JCNE 2021 E-26/201.451/2022).

Referências

Agência Nacional de Águas e Saneamento Básico (2024). Estudo aponta que enchentes de 2024 foram maior desastre natural da história do rs e sugere caminhos para futuro com eventos extremos mais frequentes. https://www.gov.br/ana/pt-br/assuntos/noticias-e-eventos/noticias/estudo-aponta-que-enchentes-de-2024-foram-maior-desastre-natural-da-historia-do-rs-e-sugere-caminhos-para-futuro-com-eventos-extremos-mais-frequentes. Acessado: 2025-04-30.

Association for Computing Machinery (2012). The 2012 acm computing classification system. https://www.acm.org/publications/class-2012. Acessado: 2024-10-02.

Bawden, D. e Robinson, L. (2009). The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35:180–191.

Beltagy, I., Lo, K., e Cohan, A. (2019). Scibert: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, p. 3615–3620, Hong Kong, China. Association for Computational Linguistics.

Caseli, H. M. e Nunes, M. G. V. (2024). *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. BPLN, São Carlos. Disponível em: brasileiraspln.com/livro-pln.

Chen, W., Yang, K., Yu, Z., Shi, Y., e Chen, C. L. P. (2024). A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, 57:137.

Devlin, J., Chang, M.-W., Lee, K., e Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. https://arxiv.org/abs/1810.04805. Acessado: 2024-10-28.

Dongoran, A., Rahmadani, S., Zarlis, M., e Zakarias (2018). Feature weighting using particle swarm optimization for learning vector quantization classifier. *Journal of Physics: Conference Series*, 978:012032.

Hsu, B.-M. (2020). Comparison of supervised classification models on textual data. *Mathematics*, 8(5):851.

Inayah, N., Liebenlito, M., Fitriyati, N., Manaqib, M., e Aryanti, N. (2023). Relation classification in scientific article abstracts using scibert with entity marker. In *2023 11th International Conference on Cyber and IT Service Management (CITSM)*, p. 1–5, New York. IEEE.

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br

IPCC (2023). Summary for policymakers: Climate change 2023: Synthesis report. https://www.ipcc.ch/report/ar6/syr/resources/spm-headline-statements. Acessado: 2025-04-08, In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1-34, doi: 10.59327/IPCC/AR6-9789291691647.001.

Kennedy, J. e Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, p. 1942–1948, New York. IEEE.

Ministério da Ciência, T. e. I. (2025). "a era dos extremos já chegou ao brasil", avalia pesquisador do cemaden. https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2025/01/201ca-era-dos-extremos-ja-chegou-ao-brasil201d-avalia-pesquisador-do-cemaden. Acessado: 2025-03-15.

Raschka, S. (2018). Model evaluation, model selection, and algorithm selection in machine learning. *CoRR*, abs/1811.12808:1811.12808.

Riekert, M., Riekert, M., e Klein, A. (2021). Simple baseline machine learning text classifiers for small datasets. *SN Computer Science*, 2:178.

Secretaria Nacional de Proteção e Defesa Civil (2024). Publicações: Proteção e defesa civil - sedec. https://www.gov.br/mdr/pt-br/centrais-de-conteudo/publicacoes/protecao-e-defesa-civil-sedec. Acessado: 2024-11-02.

Sietsma, A. J. et al. (2024). Machine learning evidence map reveals global differences in adaptation action. *One Earth*, 7:280–292.

United Nations Office for Disaster Risk Reduction (2015). Sendai framework for disaster risk reduction 2015-2030. https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030. Acessado: 2025-01-08.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., e Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, p. 6000–6010. Curran Associates Inc.

WMO (2025). Extreme weather. https://wmo.int/topics/extreme-weather. Acessado: 2025-04-08.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., e Rush, A. (2020). Transformers: State-of-the-art natural language processing. In Liu, Q. e Schlangen, D., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, p. 38–45, Online. Association for Computational Linguistics.

Yesenia, O. e Segarra-Faggioni, V. (2021). Automatic classification of research papers using machine learning approaches and natural language processing. In *Information Technology and Systems*, p. 80–87. Springer International Publishing.

DOI: 10.59254/sbpo-2025-212251
https://proceedings.science/p/212251?lang=pt-br