

RESUMOS INTELIGENTES: INTEGRANDO NOTÍCIAS E DADOS EPIDEMIOLÓGICOS PARA MELHORES DECISÕES EM SAÚDE

FRANKLIN SALES DE OLIVEIRA¹; DULCINÉIA ESTEVES SANTOS²; LUCAS
SUPERTI DA SILVA³; LUCAS FERREIRA SANTOS DE MELO⁴; BIANCA
CONRAD BOHM ⁵; BRENDA SALENAVE SANTANA⁶

¹Universidade Federal de Pelotas (UFPEL) – fsoliveira@inf.ufpel.edu.br

²Universidade Federal de Pelotas (UFPEL) – dulcinéiaesteveessantos@gmail.com

³Universidade Federal de Pelotas (UFPEL) – lsdsilva@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas (UFPEL) – lucas.fsm@gmail.com

⁵Universidade Federal de Pelotas (UFPEL) – biankabohm@hotmail.com

⁶Universidade Federal de Pelotas (UFPEL) – bssalenave@inf.ufpel.edu.br

1. DESCRIÇÃO DA INOVAÇÃO

Com o crescimento da internet, observamos também um aumento significativo na produção diária de textos jornalísticos online. Esse vasto volume de notícias pode ser esmagador, tornando mais difícil acompanhar e entender a evolução de eventos históricos (MAO *et al.* 2023). Segundo HAN (2017), essa sobrecarga informacional é característica da "sociedade de desempenho", em que o excesso de estímulos e a pressão por produtividade levam ao esgotamento mental e à incapacidade de processar adequadamente a enorme quantidade de dados com os quais somos constantemente bombardeados.

Para tanto, a proposta de inovação deste trabalho envolve uma ferramenta algorítmica que gera resumos de linhas do tempo com base em eventos descritos em textos jornalísticos. O objetivo é reduzir a grande quantidade de informações contidas nesses dados e automatizar a criação de resumos temporais que possam organizar informações sobre determinado assunto ao longo do tempo, facilitando sua análise posterior. Além disso, a ferramenta simplifica um processo que anteriormente demandava uma análise manual exaustiva, otimizando assim o tempo e o esforço necessários para a elaboração de resumos e acompanhamento da evolução da repercussão jornalística e social de determinados assuntos.

As motivações para o desenvolvimento dessa ferramenta decorrem do impacto negativo que o volume elevado de informações geradas continuamente pode ter na compreensão de eventos específicos. Esse excesso de dados frequentemente dificulta a análise dos acontecimentos que levaram a determinadas situações, prejudicando a capacidade dos pesquisadores e profissionais em interpretar informações relevantes. Ao oferecer uma solução automatizada e eficiente, a ferramenta busca mitigar esses desafios, promovendo uma análise mais clara e acessível dos eventos noticiados.

O sistema proposto utiliza técnicas avançadas de Processamento de Linguagem Natural (PLN), incluindo modelos de aprendizado profundo como *transformers* e a técnica *Timeline Summarization*, para identificar, organizar e sintetizar a descrição de eventos ao longo do tempo. Isso permite uma visualização simplificada e cronológica das principais notícias relacionadas a um determinado tema. A ferramenta opera de forma automatizada, processando grandes volumes de textos, e apresenta os eventos em uma estrutura temporal que facilita a compreensão e análise dos dados jornalísticos.

Com o objetivo de verificar a aplicabilidade da ferramenta, foi estabelecida uma parceria entre colaboradoras/res que pesquisam no Laboratório de Epidemiologia

Veterinária da Universidade Federal de Pelotas e a equipe de Computação composta por graduandos, sob orientação de uma professora, responsáveis pelo desenvolvimento da ferramenta. Essa colaboração visa não apenas a validação das funcionalidades, mas também promover a troca de conhecimentos e experiências entre as partes envolvidas. O estudo foi focado em notícias relacionadas à manifestação de casos de dengue e leptospirose, doenças já pesquisadas pelo grupo de epidemiologia.

Essa abordagem possibilita que os pesquisadores utilizem a tecnologia em um contexto prático e real, permitindo a aplicação direta da ferramenta em estudos relevantes sobre as doenças. Ao mesmo tempo, os pesquisadores fornecem *feedbacks* valiosos que podem ser utilizados para aprimorar a funcionalidade e a usabilidade da ferramenta. Essa troca de informações é essencial, pois permite que os desenvolvedores entendam as necessidades e desafios enfrentados na pesquisa em epidemiologia, garantindo que a ferramenta evolua de acordo com as demandas do campo.

Além disso, esse ambiente colaborativo não apenas contribui para o avanço das pesquisas em epidemiologia, mas também facilita a compreensão e a análise dos dados relacionados a essas doenças. A capacidade de resumir e organizar informações relevantes em tempo real permite que os pesquisadores se concentrem em desenvolver estratégias mais eficazes para a prevenção e controle das doenças, beneficiando assim a comunidade e potencialmente impactando a saúde pública.

O principal diferencial da ferramenta proposta reside em sua habilidade de organizar eventos noticiados de forma cronológica, proporcionando uma visão clara e sequencial da evolução de um determinado assunto ao longo do tempo. Ao invés de apenas agrupar notícias, a ferramenta oferece um resumo que prioriza os eventos mais relevantes, o que contribui para a redução da sobrecarga informacional e possibilita uma análise mais direcionada e eficiente.

Outro aspecto inovador é a aplicação dessa técnica no monitoramento de casos epidemiológicos, permitindo a comparação entre notícias e dados epidemiológicos oficiais. Essa abordagem diferencia a ferramenta de outras soluções de monitoramento de mídia disponíveis, como as discutidas por (Pasquali et al., 2021). Ao integrar informações jornalísticas com dados epidemiológicos, a ferramenta não só facilita a compreensão do cenário atual, mas também auxilia na formulação de estratégias mais eficazes para a prevenção e controle de surtos, destacando sua relevância no contexto da saúde pública.

2. ANÁLISE DE MERCADO

A inovação proposta visa atender a um público diversificado que inclui pesquisadores e acadêmicos formados em áreas como Epidemiologia e Saúde Pública, os quais buscam ferramentas que otimizem seu tempo e aumentem a eficiência na interpretação de dados complexos. Esses profissionais necessitam de resumos claros e organizados de eventos relacionados a casos epidemiológicos, focando em informações relevantes e verificadas que embasam suas pesquisas e análises. Além disso, o público-alvo abrange profissionais de saúde pública, como agentes de saúde e gestores, que requerem informações atualizadas para a tomada de decisões em saúde coletiva. Essa categoria de usuários procura ferramentas que integrem dados de mídia e informações epidemiológicas, permitindo uma análise rápida e fundamentada dos riscos à saúde da população.

Nesse cenário competitivo, a solução enfrenta concorrentes como Google News e do NotebookLM¹, um assistente de pesquisa e escrita com tecnologia de IA que permite a análise de documentos, embora não se concentre especificamente em casos epidemiológicos. A ferramenta proposta se destaca por sua capacidade única de integrar dados jornalísticos e epidemiológicos, oferecendo uma solução que não apenas resume informações, mas também organiza eventos de forma cronológica e prioriza conteúdos relevantes. Isso proporciona aos usuários uma visão clara e concisa da evolução dos eventos relacionados à saúde, facilitando a análise e a interpretação dos dados.

Além disso, segundo a Mordor Intelligence², o mercado de análise de saúde está estimado em US\$ 46,5 bilhões em 2024, devendo atingir US\$ 130,49 bilhões até 2029, com um crescimento anual projetado (CAGR) de 22,92% durante o período de previsão. Este cenário de crescimento acentuado no mercado de saúde demonstra a necessidade crescente de ferramentas eficazes que possam lidar com a vasta quantidade de dados disponíveis e apoiar os profissionais em suas análises.

3. ESTRATÉGIA DE DESENVOLVIMENTO E IMPLEMENTAÇÃO

Atualmente, a ferramenta está em fase de desenvolvimento e aprimoramento de suas funcionalidades. O protótipo foi desenvolvido utilizando uma abordagem algorítmica orientada a objetos, composta por uma classe principal responsável por instanciar e invocar as demais classes encarregadas das funcionalidades de extração e manipulação de dados. O objetivo principal deste programa é automatizar o processo de coleta e armazenamento dos resultados em arquivos .json de maneira linear, isto é, conforme as notícias são apresentadas. Ao final do processo, o programa gera um arquivo que contém os títulos, datas e descrições extraídos das notícias, juntamente com os respectivos links de acesso.

Após a coleta e armazenamento dos dados, é aplicado um modelo de resumo chamado BERT Extractive Summarizer (MILLER, 2019), projetado para gerar resumos concisos dos textos completos. Esse modelo utiliza técnicas de aprendizado profundo para identificar e extrair as partes mais relevantes das notícias, assegurando que as informações essenciais sejam preservadas. Os resumos gerados são, então, adicionados ao arquivo .json, juntamente com as informações coletadas no passo anterior. Dessa forma, o sistema não apenas organiza os dados originais, mas também proporciona um recurso valioso para a análise posterior, facilitando a compreensão rápida e eficiente dos conteúdos jornalísticos coletados. A **Figura 1** ilustra os resultados parciais do protótipo em um arquivo .json.

```

url: "https://g1.globo.com/bus-63556742d68a00354can441e"
conteudo: <html><head><meta name="...noscript"></head></html>
link: "https://g1.globo.com/sp/l_nor-dengue-em-2024.html"
texto_completo: "A cidade deGuareí (SP)confirmou aprimeira morte por dengue em 2024. Segundo a Vigilância Epidemiológica,a vítima é um homem de 40 anos. Ele foi hospitalizado em 12 de abril e morreu seis dias depois, porém, o resultado positivo para a dengue foi divulgado somente nesta quarta-feira (14). Participe do canal do g1 Itapetininga e Região no WhatsApp A Prefeitura de Guareí divulgou, ainda, quea vacinação contra a dengue está com baixa adesão no município. Podem se vacinarcrianças e adolescentes de 10 a 14 anos. Basta os pais ou responsáveis acompanharem o menor até uma das Unidades Básicas de Saúde, com documentos pessoais. Região De acordo com o painel de monitoramento da dengue do estado,outras duas mortes foram confirmadas na região de Itapetininga, emÁguas de Santa Bárbara eParanapanema. As prefeituras desses municípios não divulgaram detalhes dos óbitos até a publicação desta reportagem. Na região de Itapetininga, em 2024 foram registrados46.390 casos de dengue e 29 pessoas morreram em decorrência da doença. O que é a dengue? A dengue faz parte de um grupo de doenças denominadas arboviroses. O vírus é transmitido pela picada da fêmea do mosquito Aedes aegypti e possui quatro sorotipos diferentes: DENW-1, DENW-2, DENW-3 e DENW-4 – todos podem causar as diferentes formas da doença. Todas as faixas etárias são igualmente suscetíveis à doença, porém,as pessoas mais velhas e aquelas que possuem doenças crônicas, como diabetes e hipertensão arterial, têm maior risco de evoluir para casos graves e outras complicações que podem levar à morte. Veja quais são os principais sintomas da dengue: Febre;Dor no corpo e articulações;Dor atrás dos olhos;Mal-estar;Falta de apetite;Dor de cabeça;Manchas vermelhas no corpo. Veja mais notícias no1 Itapetininga e Região VÍDEOS: assista às reportagens da TV TEM"
resumo: "A cidade deGuareí (SP)confirmou aprimeira morte por dengue em 2024. A dengue faz parte de um grupo de doenças denominadas arboviroses."

```

Figura 1: arquivo .json com as respectivas células resultantes da execução.

¹ Para saber mais sobre a ferramenta: <https://notebooklm.google.com/>

² Para saber mais acesse: [Mordor Intelligence](https://www.mordorintelligence.com/). Acessado em 09 out. 2024

4. RESULTADOS ESPERADOS E IMPACTO

A inovação proposta não apenas otimiza a coleta e análise de dados jornalísticos e epidemiológicos, mas também tem o potencial de incidir no setor de tomadas de decisões na área de saúde pública e em órgãos de pesquisa epidemiológica. Ao fornecer resumos claros e organizados sobre casos epidemiológicos, a ferramenta facilita a disseminação de informações relevantes e atualizadas para profissionais de saúde pública, pesquisadores e profissionais da mídia. Isso pode resultar em decisões mais informadas e ágeis na gestão de crises de saúde, como surtos de doenças, promovendo uma resposta mais eficaz e, conseqüentemente, protegendo a saúde da população.

A visão de futuro para esta inovação é torná-la uma solução multimodal que não só abrange a análise de notícias, mas também incorpora resumos de artigos científicos e outras fontes de informação relevantes. Essa evolução permitirá que a ferramenta se torne um recurso abrangente para a pesquisa em saúde pública, promovendo a interdisciplinaridade e a colaboração entre diferentes áreas do conhecimento.

5. CONCLUSÕES

A inovação proposta representa uma ferramenta algorítmica avançada que integra dados jornalísticos e epidemiológicos para fornecer resumos claros e organizados sobre casos epidemiológicos. Ao automatizar a coleta e análise de informações, a ferramenta não apenas otimiza o trabalho de pesquisadores e profissionais de saúde pública, mas também promove uma gestão mais eficiente e informada em situações de crise. O impacto social da inovação é significativo, pois facilita a disseminação de dados relevantes e atualizados, resultando em decisões mais rápidas e fundamentadas que podem salvar vidas e melhorar a saúde da população.

Convidamos *stakeholders*, investidores e parceiros a se juntarem a nós para aprimorar e implementar esta ferramenta inovadora. Sua colaboração pode transformar a gestão de informações em saúde pública e gerar um impacto positivo na sociedade. Vamos juntos construir um futuro mais informado e decisões inteligentes na área de saúde.

6. REFERÊNCIAS BIBLIOGRÁFICAS

HAN, Byung-Chul. **Sociedade do cansaço**. Editora Vozes Limitada, 2015.

MILLER, Derek. Leveraging BERT for extractive text summarization on lectures. **arXiv preprint arXiv:1906.04165**, 2019.

McCreadie, R.; Santos, R. L. T.; Macdonald, C.; Ounis, I. Explicit diversification of event aspects for temporal summarization. **ACM Trans. Inf. Syst.**, 36(3), 2018.

Pasquali, A.; Campos, R.; Ribeiro, A.; Santana, B.; Jorge, A.; Jatowt, A. Tls-covid19: A new annotated corpus for timeline summarization. In **European Conference on Information Retrieval, Springer**. pages 497–512, (2021).