



QUALIDADE DE DADOS E INDICADORES ALTERNATIVOS SOBRE EPIDEMIAS

José Gabriel De Souza Silva¹
Tales Paiva Nogueira²

RESUMO

Este trabalho visa descrever resumidamente a realização da pesquisa focada no plano de trabalho intitulado “Qualidade de Dados” ligado ao projeto “Qualidade de dados e indicadores alternativos sobre epidemias”. No contexto da engenharia de dados, garantir dados de alta qualidade é crucial para o funcionamento eficaz de aplicativos, análises precisas e processos de tomada de decisão confiáveis. A qualidade dos dados pode ser descrita por meio de vários atributos, cada um destacando uma dimensão específica que afeta a utilidade geral, a precisão e o valor dos dados. Manter alta qualidade de dados em todos esses atributos envolve a implementação de práticas robustas de governança de dados, uso de ferramentas modernas de gerenciamento de dados e monitoramento contínuo e correção de problemas de qualidade de dados. Isso garante que os esforços de engenharia de dados suportem efetivamente os objetivos organizacionais, fornecendo dados de alta qualidade que sejam adequados para o uso pretendido. No caso específico do projeto em questão, dados sobre a casos confirmados e óbitos ocasionados pela pandemia de Covid-19 foram o alvo das análises. O trabalho consistiu no estudo aprofundado sobre qualidade de dados e análise da aplicação de métricas de qualidade para dados sobre a Covid, com o intuito de incrementar um website informativo sobre a pandemia, voltado à disponibilização de informações atualizadas para pesquisadores e sociedade. O estudo focou-se na análise sistemática de bibliotecas (módulos de software reutilizáveis) de qualidade de dados, buscando responder à necessidade de avaliar ferramentas computacionais que atuam em etapas fundamentais dos processos de análise, garantindo a confiabilidade dos dados utilizados em estudos científicos. A metodologia consistiu em duas etapas principais: (i) a busca e seleção de bibliotecas relacionadas à qualidade de dados em bases acadêmicas e repositórios de software, especificamente Google Scholar e GitHub, com critérios de inclusão voltados à atualidade, relevância e disponibilidade de código, onde as bibliotecas escolhidas foram gg-great-expectations, pysage e deequ; e (ii) a mensuração da qualidade dessas bibliotecas a partir de métricas como completude, consistência, validade, pontualidade, uniformidade e duplicidade. Essa abordagem permitiu uma avaliação comparativa sistemática, baseada em parâmetros reconhecidos pela literatura. Os resultados preliminares demonstraram que, assim como ocorre em trabalhos que comparam programas de análise de dados, a avaliação crítica de módulos específicos de qualidade de dados pode fornecer subsídios relevantes tanto para a comunidade acadêmica quanto para desenvolvedores, auxiliando na escolha de ferramentas adequadas a diferentes cenários de pesquisa. Conclui-se que o projeto possibilitou a geração de conhecimento aplicado, com potencial de contribuir para a consolidação de boas práticas em Ciência de Dados. Agradeço à Unilab pelo financiamento da pesquisa intitulada Qualidade de dados e indicadores alternativos sobre epidemias, executada entre outubro de 2024 e setembro de 2025, por meio do Programa Institucional de Bolsas de Iniciação Científica (Pibic).

Palavras-chave: qualidade de dados; bibliotecas; análise comparativa; epidemias.

Palavras-chave: qualidade de dados; bibliotecas; análise comparativa; epidemias.

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável, Discente, jgabrielsilva@aluno.unilab.edu.br¹

Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB), Instituto de Engenharias e Desenvolvimento Sustentável, Docente, tales@unilab.edu.br²