

INTISARI

Ketika pandemi COVID-19 terjadi, terdapat masalah tentang kebutuhan data dan metode analisis data. Terjadi kesenjangan besar antara peningkatan kebutuhan riset ilmiah untuk menganalisis data di berbagai bidang dalam kaitannya dengan pandemi dan kontribusi yang telah diberikan geovisual analitik. Masalah ini muncul karena data dalam bidang geospasial yang berbeda dapat memiliki karakteristik yang berbeda pula dan, dalam banyak kasus, sangat kompleks. Teknik analisis yang digunakan juga bervariasi untuk setiap jenis data, termasuk data COVID-19 untuk analitik sesuai fase *outbreak*. Data kesehatan juga tidak mudah untuk diperoleh dan diakses secara publik karena dikategorikan sebagai data yang konfidensial dan sensitif berdasarkan perlindungan data pribadi. Selain permasalahan terkait data dan metode, *dashboard* COVID-19 pada umumnya belum menerapkan geovisual analitik serta integrasi data non COVID-19, tetapi data penting lainnya terkait COVID-19 yang relevan. Dengan demikian dari segi analitik *dashboard* COVID-19 masih perlu ditingkatkan kemampuannya untuk memberikan dukungan dalam hal penalaran analitis berdasarkan data terkait yang dapat divisualisasikan guna mendapatkan *insights*.

Tujuan penelitian ini terkait penerapan dan pembuatan konsep serta desain model geovisual analitik eksplorasi *open-source datasets Emerging Infectious Diseases (EID) COVID-19*. Penerapan geovisual analitik meliputi integrasi data dan cara akses *open dataset* sebagai *input* data utama, serta desain geovisualisasi untuk memfasilitasi penalaran analitis. Selain itu, penelitian ini juga ditujukan untuk evaluasi kemampuan, ketersediaan, dan keandalan geovisual analitik dalam eksplorasi *open dataset*. Metode geovisual analitik eksplorasi *dataset* COVID-19 dilakukan dengan *platform* ArcGIS Insights dengan dukungan data sains R dan Python. Proses penerapan geovisual analitik diawali dengan menentukan tujuan, pengumpulan dan identifikasi *dataset*, pemilihan teknologi dan infrastruktur geospasial, dan menentukan model geovisualisasi serta interaksi antar muka pengguna. Metode dan teknik akses data, pengolahan data, dan integrasi data dilakukan melalui tahapan dan proses *data engineering*. Evaluasi dilakukan komparasi geovisualisasi dengan Financial Times (FT) *Visual Vocabulary*. Evaluasi ketersediaan dan keandalan didasarkan pada probabilitas dan *loading time* proses pada geovisual analitik.

Hasil desain dan konsep model yang dibangun dan diterapkan pada geovisual analitik yaitu dengan penerapan peta interaktif, grafik, dan bagan untuk geovisualisasi dan visualisasi data serta analisis spasial dengan fitur interaktif untuk interaksi pengguna. Geovisual analitik juga digunakan untuk mengintegrasikan berbagai *open-source datasets* dari *Open Application Programming Interface (Open API)* serta menggunakan konsep model *data engineering pipeline* untuk pengolahan data dan integrasi data dengan metode *Extraction Transform Load (ETL)* untuk akses dan eksplorasi *open dataset* seperti data epidemiologi, data mobilitas perjalanan, dan data demografi, serta perkembangan varian virus untuk memahami dampak dan perkembangan kondisi COVID-19. Hasil evaluasi kemampuan geovisual analitik dibandingkan dengan FT *Visual Vocabulary* diperoleh 5 (lima) kemampuan geovisualisasi. Dari evaluasi ketersediaan diperoleh nilai 0,92 (dari skala 0 s.d. 1), dan keandalan sebesar 60,38%. Geovisual analitik dapat digunakan untuk membantu menemukan pola, tren, dan *insights* dari setiap fase *outbreak* mulai dari deteksi awal hingga fase intervensi. Kemampuan utama dari geovisual analitik dalam mendukung geovisualisasi dan interaksi dapat menyederhanakan proses analisis. Geovisual analitik menjadi solusi alternatif dengan menggabungkan teknik geovisualisasi, interaksi, analisis data spasial untuk eksplorasi, dan interpretasi visual, serta membantu dalam penalaran analisis dari *open datasets*. Hal ini bermanfaat dalam konteks pandemi seperti COVID-19, karena peneliti, pembuat kebijakan, dan publik dapat menggunakan visualisasi sebagai dasar untuk mengelola pandemi.

Kata kunci: geovisual analitik, eksplorasi, *open dataset*, data sains, pandemi

ABSTRACT

When the COVID-19 pandemic occurred, there were issues with data needs and data analysis methods. There is a massive gap between the increasing need for scientific research to analyze data in different fields concerning the pandemic and the contribution that geovisual analytics has made. This problem arises because data in different geospatial fields can have different characteristics and, in many cases, are very complex. The analysis techniques vary for each data type, including COVID-19 data for analytics according to the outbreak phase. Health data is also not easy to obtain and access publicly as it is categorized as confidential and sensitive data under personal data protection. In addition to problems related existing COVID-19 dashboard, it lacks geovisual analytics and integration with non-COVID-19 data relevant to COVID-19. The dashboard needs improvement to provide analytical reasoning based on related data that can be visualized to get insights.

The purpose of this research is related to the application and development of concepts and design of geovisual analytics models for exploring open-source datasets of Emerging Infectious Diseases (EID) COVID-19. The development of geovisual analytics includes data integration and open dataset access methods as the primary data input, as well as geovisualization design to facilitate analytical reasoning. In addition, this research also aims to evaluate the capability, availability, and reliability of geovisual analytics in open dataset exploration. The geovisual analytics method of exploring the COVID-19 dataset was carried out with the ArcGIS Insights platform with R and Python data science. The process of applying geovisual analytics begins with determining objectives, collecting and identifying datasets, selecting geospatial technology and infrastructure, and determining geovisualization models and user interface interactions. Methods and techniques of data access, data processing, and data integration are carried out through data engineering stages and processes. Evaluation is compared with the Visual Vocabulary of the Financial Times (FT). Evaluation of availability and reliability is based on the probability and loading time of the analytical geovisual process.

The results of the design and concept model built and applied to geovisual analytics are by applying interactive maps, graphs, and charts for geovisualization and data visualization and spatial analysis with interactive features for user interaction. Geovisual analytics is also used to integrate various open-source datasets from the Open Application Programming Interface (Open API) and use the concept of the data engineering pipeline model for data processing and data integration with the Extraction Transform Load (ETL) method for access and exploration of open datasets such as epidemiological data, travel mobility data, and demographic data, as well as the development of virus variants to understand the impact and development of COVID-19 conditions. The results of the evaluation of analytical geovisual capabilities compared to the FT Visual Vocabulary obtained 5 (five) geovisualization capabilities. The availability evaluation obtained a value of 0.92 (from a scale of 0 to 1) and a reliability of 60.38%. Geovisual analytics can help find patterns, trends, and insights from each outbreak phase, from early detection to intervention. The critical capabilities of geovisual analytics in supporting geovisualization and interaction can simplify the analysis process. Geovisual analytics becomes an alternative solution by combining techniques of geovisualization, interaction, spatial data analysis for exploration, and visual interpretation, as well as assisting in analytical reasoning from open datasets. Geovisual analytics is beneficial in the context of a pandemic like COVID-19, as researchers, policymakers, and the public can use visualization as a basis for managing the pandemic.

Keywords: geovisual analytics, exploration, open dataset, data science, pandemic