Article

Development of MetaPolije: Cloud-Based Metabase GIS Data Analysis Platform

Rani Purbaningtyas 1*, Moh. Munih Dian Widianta 2 and Mochammad Rifki Ulil Albaab 3

1 Politeknik Negeri Jember; rpurbaningtyas@polije.ac.id
2 Politeknik Negeri Jember; munihdian@gmail.com
3 Politeknik Negeri Jember; mochrifki@polije.ac.id
* Correspondence: rpurbaningtyas@polije.ac.id

Abstract: This study aims to discuss the successful development of MetaPolije. MetaPolije is a platform used for cloud-based spatial-text data analysis utilizing the metabase library. MetaPolije as an alternative solution for users who need a medium for textual data analysis while also displaying precise location data. This platform is developed following the stages of agile method, which consists of several stages, from system requirement to testing. The platform has also undergone functional, validity, and performance testing techniques to ensure its efficiency. The test results have shown that the platform is capable of functioning well according to its intended purpose. Additionally, the platform is able to accurately display visualized data analysis results, including precise location points for each datum. Therefore, it can be concluded that the MetaPolije is a viable platform that meets the standards of functional, validity, and performance testing, making it a dependable and effective tool for data analysis. MetaPolije has been successfully developed by providing the basic features needed by users when conducting spatial text data analysis. The test results indicate that MetaPolije has met the quality testing standards for software based on each specified testing indicator, covering functional, validity, and performance aspects. These features can also be customized according to the users’ needs and purposes.

Keywords: metapolije, agile method, accurate result, precise location, data analysis tool

1. Introduction

In recent years, demand has increased for processing and visualizing analyzed data, particularly when it involves textual and geospatial data. This surge in demand can be attributed to the fact that users nowadays not only require accurate results from textual data analysis, but also precise location information to be displayed [1]. Moreover, the value of the displayed data is further enhanced when it is presented in real-time.

The advent of technology has brought about a massive influx of data from various sources including social media, sensor networks, and online platforms. This deluge of information has created a need for effective techniques to analyze and derive meaningful insights from the data [2]. Traditional methods of data analysis, which primarily focused on textual data, are no longer sufficient in meeting the evolving needs of users [3].

Textual data analysis involves extracting information from unstructured text, such as articles, social media posts, and customer reviews. While accurate textual analysis is crucial, it is equally important to consider the spatial aspect of the data. Geospatial data, which includes information related to geographic locations, provides valuable context to textual data analysis [4]. By combining textual and geospatial data, users can understand it more deeply analyzed information. This integration of textual and geospatial data...
allows for a more comprehensive analysis and interpretation of the information at hand [5].

Furthermore, the need for real-time data processing and visualization has become increasingly important [6]. Real-time data processing enables users to monitor and respond to changes as they happen, facilitating timely actions and interventions. For instance, in the case of disaster management, real-time analysis of social media data combined with geospatial information can help identify affected areas and allocate resources efficiently.

Visualizing analyzed data is another critical aspect of data analysis [7]. By presenting data in a visual format, users can quickly grasp complex patterns and relationships this may not be obvious in the raw data. Visualizations can take diagram and graphic formats, maps, or interactive dashboards, depending on the type of data and target audience. When combined with geospatial data, visualizations can provide a spatial context that enhances the understanding and interpretation of the analyzed information [8].

Several systems aimed at processing data analysis have been previously developed. However, this system only processes textual data. Additionally, it is only intended to solve specific topics or case studies. Therefore, if it were to be used for a different case study, the system would need to be redeveloped. This becomes ineffective for users who require a system with similar basic needs.

This research has successfully developed a cloud-based general metabase GIS data analysis platform called MetaPolije. This platform is highly customizable, allowing users to tailor it to their specific needs for data visualization. The system not only displays data visualizations based on textual analysis but also incorporates precise location-based visualizations using geospatial data. The development of this system followed the Agile method, ensuring a streamlined and efficient process [9].

One of the key features of this system is its API library, which is designed for data classification and clustering. This addition aims to improve the accuracy of data analysis results, ultimately enhancing the overall output of the system. By utilizing advanced algorithms and techniques, the system can provide users with more reliable and insightful data visualizations [10].

The cloud-based nature of this platform offers numerous advantages. Firstly, easily access your data anytime, anywhere as long as you have an internet connection. This flexibility allows users to analyze and visualize data conveniently, without being restricted by physical location or time constraints [11].

Furthermore, the cloud-based infrastructure enables seamless collaboration among multiple users. This is particularly beneficial for teams or organizations that need to work together on data analysis projects. The platform provides a centralized space where team members can share and access data, as well as collaborate on visualizations and analyses. This promotes efficiency, accuracy, and teamwork, ultimately leading to better outcomes [12].

The precision of the location-based visualizations is another noteworthy aspect of this platform. By leveraging geospatial data, the system can accurately display data points on a map, allowing users to gain deeper understanding of spatial patterns and relationships.

The customization options offered by MetaPolije are also worth highlighting. Users can tailor the platform to suit their specific needs and preferences, ensuring that the visualizations and analyses generated align with their objectives. This flexibility empowers users to explore and present data in a way that is most meaningful and impactful to their particular field or industry. This feature is the highlight of MetaPolije. Its ability to be customized according to user needs sets MetaPolije apart from similar data analysis platforms.
2. Materials and Methods

This study aims to discuss the development of a cloud-based GIS data analysis platform using the metabase framework, following the Agile method.

The Agile method was chosen for the development stages of this platform due to its numerous advantages and suitability for this project [13]. The development stages of the cloud-based metabase GIS data analysis platform, in accordance with the Agile method [14], are as follows:

1. Planning and Requirement Gathering. The first stage of the Agile development process is the planning and requirement gathering phase. In this stage, the project team identifies the goals, objectives, and requirements of the platform [15]. We collaborate with stakeholders, such as users and clients, to understand their needs and expectations. This stage ensures that the platform’s development aligns with the desired outcomes and addresses the specific requirements of the users.

2. User Stories and Backlog Creation. Once the requirements are gathered, the project team creates user stories that briefly explain the functionality of the platform from the user's perspective [16]. These user stories help in defining the features and functionalities of the platform. The user stories are then prioritized and added to the backlog, which is a list of tasks to be completed during the development process. The backlog is continuously updated and adjusted throughout the development cycle.

3. Sprint Planning. In the Agile method, development is divided into iterations called sprints [17]. Each sprint typically lasts a set period of time (usually 2 to 4 weeks). Before the start of each sprint, the project team holds a sprint planning meeting. During this meeting, the team selects a set of user stories from the backlog to be developed and completed during the sprint. The team estimates the effort for each user story and assigns it to individual team members.

4. Development and Testing [18]. Once the sprint planning is complete, the development and testing phase begins. The team members work on their assigned user stories and develop the required functionalities of the platform. Continuous integration and testing is performed to ensure the quality and stability of the platform.

5. Sprint Review and Retrospective [19]. At the end of each sprint, there is a sprint review and retrospective. During a sprint review, the team demonstrates the completed functionality and gathers feedback. This feedback will help further develop and improve the platform. Retrospectives allow teams to reflect on the sprint, identify...
opportunities for improvement, and discuss strategies to improve the development process in future sprints.

6. Iterative Development and Continuous Improvement. The Agile method emphasizes iterative development and continuous improvement [20]. After each sprint, the development team incorporates the feedback received during the sprint review and retrospective meetings. They refine the platform’s functionalities, address any issues or bugs, and add new features based on the evolving requirements. This iterative approach ensures that the platform is continuously improved and aligned with the users' needs and expectations.

3. Results and Discussion

The process of system analysis involves several stages. Each stage plays a crucial role in ensuring that the data used for analysis is accurate, reliable, and accessible [21]. By following these steps, organizations can make informed decisions based on the insights gained from analyzing their data.

The first stage is planning and gathering requirement which divided into two sub phases as below:

The first sub phase is the needs analysis. In this phase, primary data is collected, which consists of location data, data categories, and complete profiles of each data based on available location points. Filtering is also done in this stage to remove invalid and biased data. The data that has been collected is shown in Figure 2.

![Figure 2. Primary data that has been collected](image)

Once the data is collected, a filtering process is performed to remove invalid or skewed data. This is important to ensure that the data used for analysis is reliable and unbiased. Invalid data can include data points that are missing or incomplete, while biased data can be influenced by personal opinions or preferences.

The second sub phase is data digitization, which involves converting the data collected in the previous stage into data stored in a PostgreSQL database with the following table relations:

![Figure 3. Data relation](image)
Once the data is filtered, the next step is to digitize it. The collected data is saved in a database and converted into a format that can be manipulated. In this case, the data is stored in a PostgreSQL database, which allows for efficient storage and retrieval of the data. The digitized data is then organized into tables with appropriate relations. This helps to establish a structured and organized database, making it easier to analyze and retrieve the data. The table relations ensure that the data is properly linked and can be accessed in a meaningful way.

The second stage is user stories and backlog creation. In this stage, the primary focus is on gathering accurate and relevant data. This involves collecting data from various sources, such as surveys, interviews, and existing databases. The data collected includes information about the location of the data, the categories it belongs to, and a complete profile of each data point. This stage as basic to describe sprint planning.

The next stage is development and testing. In this stage used to set up metabase. Metabase is a powerful data visualization and analysis tool. This allows users to create interactive dashboards and reports, making data easier to understand and interpret. In addition to metabase construction, built-in libraries for data classification and clustering are also implemented. These libraries provide advanced data analysis techniques such as machine learning algorithms for classifying and grouping data. This identifies patterns and relationships within your data, enabling deeper analysis and insight.

During this stage, a dashboard is also set up to visualize the results of the data analysis. This dashboard can be used to display the analysis results according to the user's needs. The menu options or icons displayed on the dashboard are as shown in Figure 4.

![Figure 4. Dashboard menu icon](image)

When users access this cloud-based GIS data analysis platform, they only need to select one of the available menus/icons on the dashboard. For example, if a user selects the "crafts" menu, the system will display the locations of the points that fall into the crafts category as depicted in Figure 5.
The displayed data includes the profiles of each data point. Additionally, the data is real-time, and the time required to access and display the data analysis results is only around 1.46 seconds. This is made possible because the platform for this GIS data analysis is cloud-based. Therefore, the resources for data storage and access are extensive and fast [22].

After each sub module had developed than the next step is platform testing. Platform testing is an important step in any software or application development process. This ensures that the platform works as intended, meets user needs, and operates efficiently. For MetaPolije, testing focused on his three main areas: functional testing, validity testing, and performance testing [23].

Functional testing is essential to ensure that the platform functions properly and meets the needs of the users [24]. This includes testing various features of the platform to ensure they work as expected. This includes testing data analysis features, visualization tools, and user interfaces.

Validity testing, on the other hand, assesses the accuracy and reliability of the data analysis results displayed by the platform [25]. This testing ensures that the platform provides accurate and valid information to the users. It involves comparing the results obtained from the platform with known data sets or benchmarks to verify their accuracy.

Performance testing evaluates the performance and responsiveness of the platform, particularly in terms of the time required for users to access and interact with it [26]. This testing helps identify any performance bottlenecks or issues that may affect the user experience.

Table 1 depicted the result of MetaPolije platform testing based on functionality, validity and performance.

Table 1. Platform testing result

<table>
<thead>
<tr>
<th>Testing area</th>
<th>Description</th>
<th>Testing Method</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functionality</td>
<td>Test all feature</td>
<td>Manual direct testing</td>
<td>All features are functioning properly.</td>
</tr>
<tr>
<td>Validity</td>
<td>Test accuracy and reliability data</td>
<td>Manual direct testing</td>
<td>Platform provides accurate and valid information</td>
</tr>
<tr>
<td>Performance</td>
<td>Platform speed test</td>
<td>tools.keycdn.com</td>
<td>Access time only 1.46 seconds</td>
</tr>
</tbody>
</table>

Figure 5. Display location based on category
Through functional testing, it was determined that the platform met the functional requirements and operated smoothly. The validity testing conducted on the metabase GIS data analysis platform showed that the results were consistent and reliable. The performance testing conducted on the platform indicated that it performed efficiently and provided a smooth user experience.

Based on the above test results, it can be stated that the developed MetaPolije is deemed suitable in terms of functional, validity, and performance aspects of the system. Overall, the testing results indicate that the developed cloud-based metabase GIS data analysis platform is suitable for use. It meets the functional requirements, provides accurate and valid data analysis results, and performs efficiently. The testing process ensures that the platform is reliable and meets the needs of the users.

4. Conclusions

This study aims to discuss the successful development of a cloud-based metabase GIS data analysis platform called MetaPolije, which follows the Agile method system development life cycle. The Agile method consists of several stages. The platform has also undergone functional, validity, and performance testing techniques to ensure its efficiency. The test results have shown that the platform is capable of functioning well according to its intended purpose. Additionally, the platform is able to accurately display visualized data analysis results, including precise location points for each data. The time required to access and display the data analysis results is also remarkably short, taking only 1.46 seconds. Therefore, it can be concluded that the MetaPolije is a viable platform that meets the standards of functional, validity, and performance testing, making it a dependable and effective tool for data analysis. For the next stage, MetaPolije can be implemented according to user needs. This is because the features in MetaPolije are basic features needed when conducting spatial-text data analysis. Category sorting can be adjusted to the user’s objectives.

5. Patents

The copyright of Metapolije platform is registered at the Directorate General of Intellectual Property (DJKI) of the Ministry of Law and Human Rights with registration number ECO02023117464 and recording number 000550419.

Supplementary Materials: The following supporting information can be downloaded at:
https://docs.google.com/spreadsheets/d/17lAW5zXpmzHfhiSOGRMY4TMFn0OVu0A6s/edit?usp=s haring&ouid=10315231873925438998&rtpof=true&sd=true


Funding: This research was funded by Politeknik Negeri Jember, grant number 871/PL17.4/PG/2023.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Available in
https://docs.google.com/spreadsheets/d/17lAW5zXpmzHfhiSOGRMY4TMFn0OVu0A6s/edit?usp=s haring&ouid=10315231873925438998&rtpof=true&sd=true.
Acknowledgments: Acknowledgments are extended to Politeknik Negeri Jember for providing support to this research through the Basic Research scheme funded by PNBP 2023 with contract number 871/PL17.4/P/G/2023. We express our gratitude for their assistance and contribution to the successful completion of this study. Their financial support has been instrumental in facilitating the research process and enabling us to achieve our research objectives. We are thankful for their commitment to promoting and supporting scientific endeavors.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References


