Toward a Conceptual Data Model Design for the Report Digitalization Project in the CSR Department of PT ABC

Stenly Richard Pungus[1], Pauline Ira Tumbel[2], Gabriel Natalie Sumual[3*], Edson Yahuda Putra[4], Rolly Lontaan[5], Hizkia Eben Haezer[6], Jonathan Medellu[7]

[1]*Program of Information Systems, Faculty of Computer Science, Universitas Klabat*
[2]*Program of Informatics, Faculty of Computer Science, Universitas Klabat*

*Corresponding Author: s22210085@student.unklab.ac.id **

Abstract *- The increasing demand for digital transformation in the mining sector highlights the importance of developing efficient data management approaches, particularly in Corporate Social Responsibility (CSR) reporting. Current practices in mining companies often rely on conventional spreadsheet-based systems, which lead to inefficiencies such as inconsistent data formats, limited searchability, and difficulties in historical data tracking. This study explores a conceptual approach to data model design that can support the digitalization of CSR reporting in a mining company context, emphasizing efficiency, accuracy, and transparency. Adopting the Design Science Research (DSR) methodology, the study involves problem identification, formulation of solution objectives, and the development of an initial conceptual model using data modeling techniques. Data will be gathered through literature review, interviews, and observation of CSR reporting practices within the mining sector. This work aims to provide an early-stage foundation for future research and system development, offering both theoretical insights into data modeling for CSR digitalization and practical directions for subsequent implementation and evaluation.*

Keywords: Data Modeling, Digitalization, CSR Reporting, ERD, Mining Industry

I. INTRODUCTION

The rapid advancement of information technology has reshaped how organizations operate and interact with stakeholders. In particular, the shift from manual to digital systems has become a global imperative, affecting industries ranging from finance to healthcare. Although mining is traditionally considered a resource-driven sector, the integration of digital solutions into mining operations is increasingly recognized as essential, not only for operational efficiency but also for fostering accountability and sustainability (Danava, 2021). Putra et al. (2024) emphasized that digitalization functions not only as a medium for data storage but also as a means to enhance accessibility and transparency of information.

One of the key areas where digitalization plays a transformative role is Corporate Social Responsibility (CSR). Mining companies, especially those operating in developing regions, often face significant pressure from both government regulators and local communities to demonstrate that their operations generate positive social and environmental outcomes (Kereh, Supardjo, & Lintong, 2023). CSR programs are therefore not optional but a critical component of maintaining a company's social license to operate. However, the effectiveness of CSR initiatives depends largely on how well they are planned, documented, monitored, and reported.

In practice, CSR programs span multiple pillars, including health, education, environment, and infrastructure, and are often implemented across various communities (Pulukadang, 2018). Despite the breadth of these initiatives, the reporting process in many companies still relies heavily on Microsoft Excel spreadsheets. While Excel provides flexibility for small-scale data management, it becomes increasingly inefficient as data volume grows. Common problems include inconsistent formats, difficulties in retrieving historical data, and challenges in performing structured analysis. For example, if management wishes to identify all health-related programs conducted in a specific year, staff must manually search across multiple spreadsheets, leading to time-consuming and error-prone reporting (Kauwo, 2016).

These limitations underscore the urgent need for a more structured data management approach. Without a robust system, CSR reporting risks being fragmented and inconsistent, undermining both internal decision-making and external accountability. In contrast, a structured database supported by a carefully designed data model can reduce redundancy, ensure accuracy, and facilitate real-time analysis (Suryadi, 2019). This aligns with global trends in corporate governance, where transparency and evidence-based decision-making are increasingly demanded by stakeholders (Cristofel & Kurniawati, 2021).

Data modeling, and specifically the use of an Entity-Relationship Diagram (ERD), provides an effective solution to these challenges. By systematically identifying entities, attributes, and relationships, an ERD enables the design of a relational database that reflects the complexity of CSR operations (Pakpahan, 2023). Such a model not only streamlines reporting but also enhances scalability, ensuring that the system can accommodate future growth in program scope and data volume. Furthermore, integrating normalization principles into the model ensures that data is consistent, logically organized, and free from redundancy (Qiu, Wu, Ding, Xu, & Feng, 2016).

The purpose of this study is to design a conceptual and logical data model for CSR reporting in the mining sector. The study adopts the Design Science Research (DSR) methodology, which is well suited for developing practical artifacts grounded in academic rigor. Specifically, the research seeks to address three questions: (1) What is the current state of CSR data management? (2) How can a data model be designed to support the digitalization of CSR reports? and (3) How can the model be validated to ensure alignment with organizational needs?

By addressing these questions, this study contributes both theoretically and practically. From an academic perspective, it enriches the literature on CSR digitalization and data modeling, particularly within the mining industry, which has historically lagged behind other sectors in adopting digital tools. From a practical standpoint, it provides a structured foundation for digital transformation, enabling more accurate, transparent, and efficient CSR reporting. Ultimately, the study demonstrates that data modeling is not merely a technical exercise but a strategic enabler of corporate accountability and community engagement.

II. LITERATURE REVIEW

The literature review builds a theoretical foundation for CSR report digitalization.

1. **CSR in the Mining Sector**. Mining companies are expected to implement community programs in health, education, environment, and infrastructure as part of their Corporate Social Responsibility (CSR) commitments (Kereh, Supardjo, & Lintong, 2023).

2. **Data Modeling**. Data modeling provides the structure for organizing, storing, and managing information systematically. Conceptual, logical, and physical models each play a role in ensuring efficient and integrated data management (Pakpahan, 2023).

3. **Entity-Relationship Diagram (ERD)**. ERD is a conceptual tool for visualizing entities, attributes, and relationships in a database. It supports communication between developers and stakeholders and serves as the basis for relational schema design (Qiu, Wu, Ding, Xu, & Feng, 2016).

4. **Metadata**. Metadata describes and manages data, ensuring efficiency in retrieval and integration. It plays a crucial role in structured digital systems (Cristofel & Kurniawati, 2021).

5. **CSR and Digitalization**. CSR reflects a company's responsibility toward social and environmental sustainability. The digitalization of CSR reporting improves transparency, efficiency, and decision-making while reducing errors from manual reporting (Danava, 2021).

6. **Database Design**. Relational database design, supported by normalization, improves efficiency, reduces redundancy, and supports accurate reporting (Suryadi, 2019).

7. **Research Methodology**. Previous studies show that qualitative descriptive methods and Design Science Research (DSR) are effective for system design, enabling solutions that combine theory with organizational needs (Kauwo, 2016).

Several studies highlight the role of data modeling in ensuring system reliability and scalability. For example, normalization principles are widely discussed in database design

literature as a way to eliminate redundancy and ensure data integrity (Suryadi, 2019). Similarly, CSR digitalization has been shown to enhance corporate transparency, especially when supported by robust metadata management (Cristofel & Kurniawati, 2021).

The Entity-Relationship Diagram (ERD) remains a central tool in conceptual modeling. According to Qiu et al. (2016), ERDs provide a shared language between technical designers and organizational stakeholders, bridging the gap between abstract requirements and implementable systems. In the context of CSR, previous research emphasizes that a well-structured data model not only improves efficiency but also contributes to corporate reputation by facilitating clearer reporting (Kereh et al., 2023).

Based on these insights, this study develops a conceptual framework that integrates CSR data sources as inputs, digitalization and data modeling as processes, a structured data model as the output, and efficiency, accuracy, transparency, and future development as the expected outcomes.

Conceptual Framework.

This study proposes a simplified conceptual framework (Figure 1). CSR data sources serve as the input, which are processed through digitalization and data modeling techniques such as ERD and normalization. The process produces a structured data model as the output. Finally, the model is expected to lead to efficiency, accuracy, transparency, and provide a foundation for future system development.

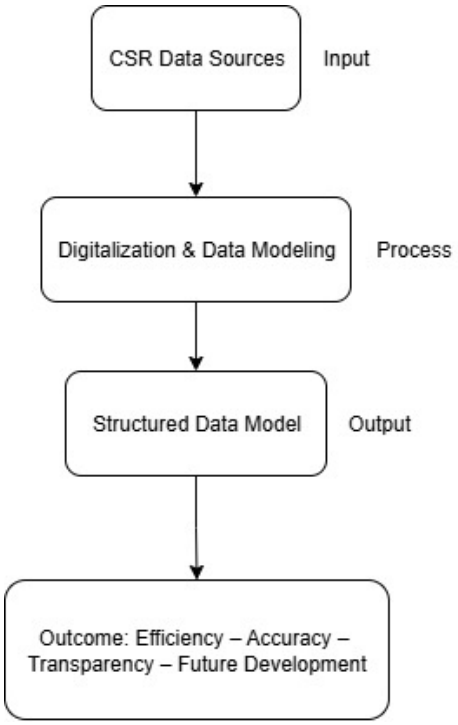

Figure 1. Conceptual Framework

III. METHODOLOGY

This research employs the Design Science Research (DSR) approach, which is widely recognized as a robust methodology for developing and evaluating artifacts that address real-world problems. The DSR framework consists of six iterative stages: problem identification, defining solution objectives, design and development, demonstration, evaluation, and communication. By following this cycle, the study ensures that the proposed data model is not only theoretically sound but also practically relevant to the needs of CSR reporting in the mining sector.

Research Design. The primary objective of this study is to create a conceptual and logical data model that addresses the inefficiencies of conventional spreadsheet-based CSR reporting systems. To achieve this, a qualitative descriptive method was adopted to capture practical challenges in managing CSR data and to translate them into structured requirements for system design. This approach emphasizes understanding the context and environment in which the artifact will be implemented, ensuring that the resulting model aligns with organizational practices and constraints.

Data Sources. Data for this study was obtained from both primary and secondary sources. Primary data included in-depth interviews with CSR staff and direct observation of reporting processes within a mining company's CSR department. These interactions provided insights into the practical difficulties of managing large volumes of CSR data using spreadsheets, such as duplication, formatting inconsistencies, and limited search functionality. Secondary data consisted of CSR reports, government compliance documents, and scholarly literature related to CSR reporting, data modeling, and digital transformation. Together, these data sources formed a comprehensive foundation for analyzing the current state of CSR reporting and designing an improved data model.

Sampling Strategy. A purposive sampling method was applied to ensure that the most representative and data-intensive CSR programs were analyzed. Health and education programs were chosen as the focus of model testing, as these areas involve a large number of beneficiaries, multiple performance indicators, and significant budget allocations. By concentrating on these pillars, the study was able to stress-test the proposed model under complex conditions, increasing its applicability across other CSR domains such as environment and infrastructure.

Procedures. The methodological procedures consisted of four main steps:

1. Analyzing existing CSR data – identifying redundancies, inconsistencies, and inefficiencies in spreadsheet-based reports.

2. Designing the ERD – constructing a conceptual and logical model that represents the key entities (programs, beneficiaries, locations, budgets, and indicators) and their relationships.

3. Validation with experts – conducting structured discussions with academic supervisors and industry practitioners to confirm the accuracy and relevance of the model.

4. Evaluation of efficiency, consistency, and scalability – assessing how well the model reduces redundancy, ensures data integrity, and supports potential expansion of CSR programs in the future.

Tools. The tools used in this research were selected to reflect both accessibility and practical implementation. Spreadsheet applications served as the primary tool for collecting and analyzing existing CSR data. For the modeling stage, ERD design software (such as draw.io or Lucidchart) was employed to construct the data model. A relational database management environment was also utilized to test the feasibility of implementing the ERD in practice, ensuring that the conceptual design could be translated into a working system.

Input–Process–Output Framework. The methodology can also be explained using the Input–Process–Output (IPO) framework. The input consisted of spreadsheet-based CSR data, including historical program reports and beneficiary records. The process involved multiple stages: data normalization, ERD design, validation, and expert feedback. The output was a structured data model that provides a conceptual foundation for future digitalization of CSR reporting. The IPO framework illustrates how raw CSR data is systematically transformed into a structured model, emphasizing the logical flow of the research.

By combining the DSR methodology with qualitative descriptive analysis, purposive sampling, and practical validation, this research ensures that the resulting conceptual data model is grounded in both theory and practice. The methodological rigor provides confidence that the model can serve as a reliable foundation for the digitalization of CSR reporting in the mining sector, while also being adaptable to future system development and application design.

IV. Results and Discussion

Normalization Result

Prior to designing the ERD, the CSR data originally stored in spreadsheets was normalized to eliminate redundancy and ensure consistency. The normalization process followed the standard database normalization rules, moving progressively from the first normal form (1NF) to the third normal form (3NF).

In the first stage, data was organized to ensure atomicity, meaning that each cell in the dataset contained only a single value. This resolved one of the key problems in the existing

spreadsheet-based system, where a single cell sometimes stored multiple pieces of information, such as the name of a beneficiary along with their location.

In the second stage, the process eliminated partial dependencies by ensuring that all attributes were fully dependent on the primary key. For example, in the original spreadsheet structure, program names were often repeated in rows where budget details or beneficiary data were recorded. This design created unnecessary duplication and increased the risk of inconsistencies. By restructuring the data into separate entities, such as health-related and education-related programs, each entity became more clearly defined, reducing redundancy.

Finally, in the third normal form, transitive dependencies were removed to ensure that non-key attributes were only dependent on primary keys. A clear example of this improvement can be seen in the treatment of location data. Initially, each program entry contained village and district information, which meant that when one location detail changed, it had to be updated across multiple rows, risking inconsistency. Through normalization, the Location entity was created as a standalone table, which was then connected to programs via relational keys. As a result, any updates to a location automatically propagate across all related records, ensuring data integrity.

Additionally, beneficiaries and indicators were modeled as independent entities rather than being embedded within program records. This separation allowed the database to handle many-to-many relationships more effectively. For instance, a single program could serve multiple categories of beneficiaries, while at the same time, one beneficiary group might be included in several different programs. Similarly, indicators were defined in a way that they could be reused across different programs, enabling standardized monitoring and evaluation.

The normalization step not only simplified the data structure but also established a scalable foundation for future expansion. As more CSR programs are implemented across multiple locations, the relational model ensures that new records can be integrated seamlessly without disrupting existing data. Moreover, the normalized structure supports advanced querying, making it possible to generate complex reports such as tracking the allocation of funds per village, comparing program outcomes by year, or analyzing beneficiary trends across multiple pillars of CSR.

Entity-Relationship Diagram

The proposed data model was designed using an Entity-Relationship Diagram (ERD) to represent the main entities and their relationships in CSR reporting. As emphasized by Pungus, Yahaya, Deraman, and Bakar (2019), data modeling is not merely about representation but about accurately capturing entities and the relationships among them. In this context, the ERD serves as a conceptual tool to ensure that CSR data is structured systematically and consistently. Figure 2 illustrates the ERD of the CSR program data model.

The ERD provides a conceptual map of how CSR data should be structured. The schema is designed to be applied across all CSR pillars, including health, education, environment, and infrastructure. Each program is represented as an entity and is linked to beneficiaries, locations, and indicators, ensuring that all dimensions of CSR initiatives are interconnected. This relational structure not only supports day-to-day reporting but also enables management to generate cross-cutting insights, such as comparing budget allocations across villages or evaluating program effectiveness through indicator trends.

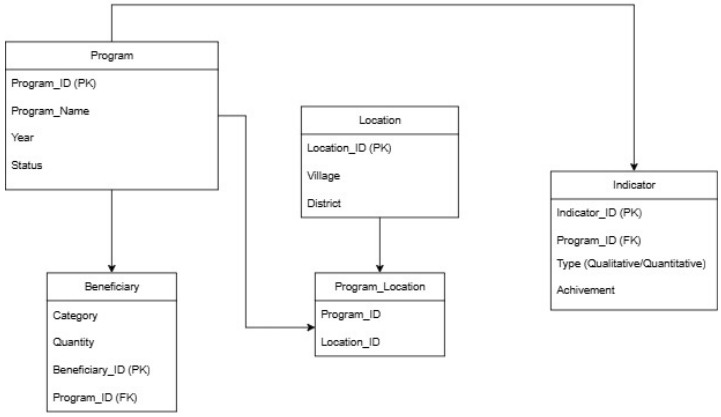

Figure 2. Entity-Relationship Diagram of the CSR Program Data Model.

This structure reduces redundancy by separating program, beneficiary, location, and indicator data into distinct entities, while maintaining integrity through the use of primary and foreign keys. For example, the *Program* entity captures key attributes such as program name, year, and status; the *Beneficiary* entity records categories and quantities; the *Location* entity stores geographical information; and the *Indicator* entity defines both qualitative and quantitative measures of achievement. The *Program_Location* table serves as a bridge entity, enabling a single program to be implemented across multiple locations.

By normalizing data in this way, the model ensures consistency, scalability, and efficient management of CSR information. Unlike spreadsheets, which are prone to duplication and human error, the relational database enforces data integrity rules, guaranteeing that CSR reporting remains reliable and consistent across reporting periods. Moreover, the structure allows for faster and more flexible query processing, enabling real-time insights that were previously difficult to obtain. In line with this, findings from Lontaan and Sinadia (2024) demonstrate that a well-designed system is capable of integrating diverse data—such as user profiles, schedules, and institutional records—into a unified framework, thereby reducing redundancy and improving consistency. This parallel reinforces the argument that structured data modeling in CSR reporting can deliver similar benefits by ensuring accuracy, transparency, and scalability.

The model also aligns with stakeholder demands for greater accountability and transparency. As highlighted by Mambu et al. (2022), the alignment between technology and organizational needs is essential to optimize performance and ensure the sustainability of the system. In this context, government regulators, community stakeholders, and corporate decision-makers can all benefit from clearer, more structured reports that reduce ambiguity and improve trust. Compared to existing literature, this study provides a unique contribution by focusing specifically on CSR data modeling in the mining sector, an area that has historically lacked structured digitalization initiatives (Kereh, Supardjo, & Lintong, 2023; Suryadi, 2019).

Interface Design (Dashboard Prototype)

To complement the proposed data model, a dashboard prototype was developed to illustrate how the structured data could be visualized and managed in practice. The dashboard provides several key functions, including program input, progress tracking, user management, and report generation. Figure Y shows the interface for the Health Program, which integrates a geographic map view of CSR activities in different locations.

It is important to note that this dashboard design was created as part of a collaborative academic project, while the focus of this study remained on the conceptual and logical data model. Therefore, the contribution of this research is limited to the data modeling stage, whereas the development of the full application, including dashboard functionalities, will be continued in future work by collaborating researchers.

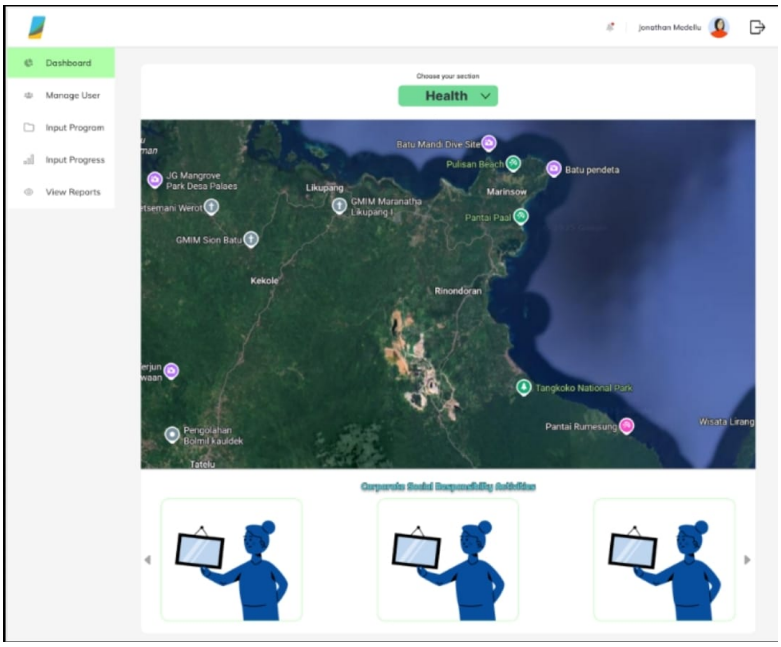

Figure 3. Dashboard prototype for CSR program monitoring (designed by collaborating group).

The dashboard prototype provides a tangible glimpse into how the data model could be operationalized. Designed in parallel with this study, the dashboard includes features for inputting new programs, visualizing data on a geographic map, and generating reports based on selected indicators. Although the focus of this research is limited to the conceptual data model, the dashboard illustrates its potential applicability in practice and highlights the importance of structured evaluation. As noted by Sondakh, Pungus, and Putra (2022), structured evaluation offers not only a snapshot of current capabilities but also a foundation for continuous improvement. In a similar way, the proposed CSR data model provides a conceptual basis that can be further developed into a comprehensive digital system in future studies.

It is emphasized that the full development of the application will be undertaken in future research. The separation of responsibilities ensures clarity in academic contributions: this paper provides the conceptual and logical foundation, while the next stage of research will focus on technical implementation and user interface design.

V. CONCLUSION

This study proposes the design of a conceptual and logical data model to support the digitalization of CSR reporting in the mining sector. The model addresses inefficiencies inherent in spreadsheet-based systems by applying normalization principles and representing entities through an Entity-Relationship Diagram (ERD). The resulting structure enhances efficiency, accuracy, and transparency while providing a scalable foundation for digital transformation in CSR reporting.

Although the scope of this research is limited to data modeling, the study contributes both theoretically and practically. Theoretically, it demonstrates the applicability of design science research and data modeling techniques in the context of CSR digitalization, particularly in industries that have historically lagged in adopting digital tools. Practically, it provides a structured conceptual framework that organizations can adopt as a foundation for future system development.

Future studies may extend this research to the full implementation of a CSR information system, focusing on application design, user interface usability, and real-time reporting. Additional research could also explore integration with advanced technologies such as big data analytics and machine learning, thereby maximizing the impact of CSR initiatives on corporate accountability and community outcomes.

AUTHORS' CONTRIBUTIONS

S.R. Pungus provided the original idea and his earlier research served as the primary reference for this study. P.I. Tumbel contributed to conceptualization, data collection, and drafting. G.N. Sumual contributed to methodology, validation, and revisions. H.E. Haezer and J. Medellu contributed to the design of the user interface and will continue the development of the application in future stages of the project. E.Y. Putra and R. Lontaan supervised the research and provided academic guidance. All authors reviewed and approved the final manuscript.

ACKNOWLEDGEMENT

The authors would like to thank the CSR department of a mining company for their cooperation and the academic supervisors for their valuable guidance throughout this research.

REFERENCES

Cristofel, C., & Kurniawati, K. (2021). Pengaruh enterprise risk management, corporate social responsibility dan kepemilikan institusional terhadap nilai perusahaan. *Jurnal Akuntansi Bisnis, 14*(1), 1–12. https://doi.org/10.30813/jab.v14i1.2468

Danava, D. (2021). Menuju era digitalisasi pertambangan. *Prosiding Temu Profesi Tahunan PERHAPI*, 0–1.

Dirgantara, M. R., Syahputri, S., & Hasibuan, A. (2023). Pengenalan database management system (DBMS). *Jurnal Ilmiah Multidisipline, 1*(6), 300–301. https://doi.org/10.5281/zenodo.8123019

Kauwo, F. M. (2016). Analysis of the process of knowledge management at PT Meares Soputan Mining Likupang. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi, 4*(2), 266–272. https://doi.org/10.35794/emba.4.2.2016.12969

Kereh, V. M., Supardjo, S., & Lintong, S. (2023). Dampak pertambangan emas PT MSM Bitung Minahasa Utara terhadap kawasan permukiman dan sosial ekonomi masyarakat. *Sabua: Jurnal Lingkungan Binaan dan Arsitektur, 12*(1), 30–39. https://doi.org/10.35793/sabua.v12i1.48814

Lontaan, R. J., & Sinadia, A. R. (2024). Design and development of a web-based high school information system. *CogITo Smart Journal, 10*(2), 593–606. https://doi.org/10.31154/cogito.v10i2.745.593-606

Mambu, J. Y., Lontaan, R. J., Lompoliu, E., Salindeho, J., & Sambul, J. (2022). IT governance capability level identification of COBIT 2019 at the RSUP Prof. Dr. R.D. Kandou, Manado, North Sulawesi. *semanTIK, 8*(2), 121–128. https://doi.org/10.55679/semantik.v8i2.28547

Pakpahan, C. (2023). Memahami pemodelan data: Ulasan jurnal. Universitas Sumatera Utara.

Pulukadang, U. R. (2018). Studi sequence penambangan PT Meares Soputan Mining dalam mencapai produksi 347.850 ton bijih tertambang pada semester 2 di Pit Toka, Tokatindung Gold Mine Project, Likupang, Sulawesi Utara, Indonesia [Undergraduate thesis, Universitas Sriwijaya].

Pungus, S. R., Poluan, S. E. R., & Rantepadang, A. R. (2016). Aplikasi informasi restoran berbasis web services di Kota Manado. *CogITo Smart Journal, 2*(1), 14–27. https://doi.org/10.31154/cogito.v2i1.11.14-27

Pungus, S. R., Yahaya, J., Deraman, A., & Bakar, N. H. (2019). A data modeling conceptual framework for ubiquitous computing based on context awareness. *International Journal of Electrical and Computer Engineering (IJECE), 9*(6), 5495–5501. https://doi.org/10.11591/ijece.v9i6.pp5495-5501

Putra, E., Haerani, C., Sidomulyo, R., Rondonuwu, V., Golung, C., & Stenly, A. (2024). Digitalisasi museum berbasis web dengan AR dan QR Code (studi kasus pada Museum Sulawesi Utara). *Jurnal Teknik Elektro dan Komputer TRIAC, 11*(1). Universitas Klabat. https://journal.trunojoyo.ac.id/triac

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing, 2016*(1), 67. https://doi.org/10.1186/s13634-016-0355-x

Sondakh, D. E., Pungus, S. R., & Putra, E. Y. (2022). Indonesian undergraduate students' perception of their computational thinking ability. *CogITo Smart Journal, 8*(1), 68–80. https://doi.org/10.31154/cogito.v8i1.387

Suryadi, S. (2019). Implementasi normalisasi dalam perancangan database relational. *U-NET Jurnal Teknik Informatika, 3*(2), 20–26. https://doi.org/10.52332/u-net.v3i2.7