PROSIDING SNTTM XXIII 2025
Volume 23, No. 1, Tahun 2025, 9 Oktober 2025

e-ISSN 2623-0313
p-ISSN 3032-1972

Digital Transformation in Geothermal Power Plant Operations

Efrata Meliala[a], **Arfiandi Anugerah**[a,1]
[a]PT Geo Dipa Energi (Persero), Jakarta, 12740
[1]Email korespondensi: arfiandi@geodipa.co.id

Abstract. This paper explores the digital transformation of geothermal power plant operations through the Geothermal Operations Optimization, Reliability, and Efficiency using Machine Learning (GOOREMALE) framework. Geothermal operations have traditionally relied on manual data collection, periodic reporting, and expert-based analysis, which often delay decision-making and extend cycle times. These inefficiencies create risks of reduced output, higher operational costs, and slower responses to disturbances, underscoring the need for a structured, data-driven solution. The objective of this research is to evaluate whether integrating digitalization and machine learning can reduce the operational decision-making cycle time and enhance monitoring capabilities. The methodology adopts a multi-phase approach, beginning with the digitization of operational log sheets into a centralized database management system (DBMS), followed by the development of Business Intelligence dashboards for real-time visualization, and culminating in the application of machine learning models. Specifically, Locality Sensitive Hashing (LSH), OPTICS clustering, and Root Mean Square Error (RMSE) evaluation were applied to detect anomalies, forecast deviations, and provide decision support. The results show that GOOREMALE reduced the decision-making cycle time by approximately 50%, improved anomaly detection accuracy, and delivered real-time dashboards that enhanced situational awareness across operational teams. These outcomes confirm that digital transformation can significantly strengthen the timeliness and accuracy of geothermal operational decision-making. The main contribution of this research is the establishment of a practical and replicable framework for digital transformation in geothermal plant operations, while its boundary is limited to decision-making efficiency without yet covering long-term reliability, subsurface management, or predictive maintenance.

Keywords: geothermal plant, digital transformation, centralized database, decision-making, machine learning

Received: 30 September 2025; **Presented:** 9 October 2025; **Publication:** 9 March 2026
DOI: https://doi.org/10.71452/jctxg786

INTRODUCTION

Geothermal power plants are vital contributors to renewable energy development, but their operations are inherently complex, with fluctuating reservoir conditions, varying steam supply, and critical equipment dependencies. Traditionally, operational decision-making has been carried out through manual data collection, periodic reporting, and expert-based analysis [1]. While this approach has sustained operations for decades, it introduces significant delays in recognizing anomalies and responding to changes, thereby lengthening the decision-making cycle time. In a production environment where every hour of delay can affect output reliability and efficiency, accelerating the decision-making process has become a strategic imperative [2].

Recent research has demonstrated that digital transformation and machine learning (ML) can address these inefficiencies in the energy sector. For example, Ling et al. (2024) applied artificial neural networks (ANN) and support vector regression (SVR) to predict geothermal power plant performance with strong accuracy, while Siratovich et al. (2022) outlined opportunities for ML in anomaly detection, clustering, and optimization in geothermal operations [3], [4]. Okoroafor et al. (2022) highlighted two decades of ML applications in subsurface geothermal exploration, advancing techniques in seismic interpretation and reservoir property prediction [5]. Despite these advances, prior studies have largely focused on predictive modeling or exploration, rather than on developing a structured framework for operational decision-making. The gap lies in leveraging digitalization and ML not just for prediction, but specifically for reducing the time it takes to make operational decisions in daily plant management.

This study introduces the Geothermal Operations Optimization, Reliability, and Efficiency through Machine Learning (GOOREMALE) framework, which integrates Business Intelligence (BI) dashboards with ML techniques to automate anomaly detection and decision support. The central research problem is the inefficiency of traditional manual workflows, with the hypothesis that digital transformation supported by ML can significantly reduce the operational decision-making cycle time. The methodology follows a multi-phase approach, beginning with digitizing operational logs into centralized databases and dashboards, then embedding ML models to detect deviations and provide real-time insights.

The objective of this paper is to evaluate the impact of GOOREMALE on decision-making cycle time within geothermal operations, focusing exclusively on the operational dimension. The contribution of this research lies in empirically demonstrating that digitalization can cut cycle time by approximately 50%, establishing a practical framework for real-time decision-making. The scope of this study is deliberately bounded to operational decision-making processes, without extending into reliability and efficiency aspects, which are reserved for future phases of investigation.

Geothermal power plants play a strategic role in global renewable energy portfolios, but their operational performance depends on effective monitoring and rapid response to system changes [6]. Traditionally, operational data in geothermal facilities has been recorded manually or stored in siloed systems, limiting their usefulness for real-time decision-making [1], [7]. Omrani et al. (2024) highlight that Business Intelligence (BI) platforms provide a means to centralize data and support visualization, enabling faster access to operational insights. In recent years, BI dashboards and historian systems have increasingly been deployed in power plant environments to provide decision-makers with near real-time visibility of process conditions, although adoption in geothermal contexts has remained relatively limited [8].

Parallel to these developments, machine learning (ML) has emerged as a powerful tool for data-driven operational optimization. Ling et al. (2024) applied artificial neural networks (ANN) and support vector regression (SVR) for geothermal power plant performance prediction, achieving high levels of accuracy in forecasting operational parameters [3]. Siratovich et al. (2022) demonstrated opportunities for ML in anomaly detection and clustering, showing its potential to improve monitoring and reliability in geothermal plants [4]. Okoroafor et al. (2022) provided a broader review of ML applications in geothermal exploration, noting significant advances in seismic interpretation, well logging, and reservoir property prediction over the past two decades [5]. More recent studies in geothermal energy forecasting also highlight the use of regression and clustering techniques for temperature and flow prediction, underscoring the flexibility of ML approaches in handling complex geothermal datasets (e.g., Al-Fakih et al. 2025; Ibrahim et al. 2023; Gao et al. 2024) [9], [10], [11].

Despite these advances, most prior studies have focused either on predictive modeling of power plant outputs or on subsurface exploration. While such contributions are valuable, they do not directly address the operational challenge of accelerating decision-making in daily plant management. Specifically, there is a gap in integrating digitalization with ML within a structured framework that not only predicts performance but also reduces the cycle time for operational decisions. Addressing this gap, the GOOREMALE initiative contributes a combined BI and ML framework specifically designed to improve decision-making efficiency by centralizing data, automating anomaly detection, and providing real-time operational support.

METHODOLOGY

The methodology of this study is designed to evaluate how digital transformation, supported by machine learning (ML), can reduce the operational decision-making cycle time in geothermal power plant operations. The approach integrates structured workflow development, system architecture design, system implementation, and evaluation throughout the GOOREMALE implementation phase.

Research framework

The research framework adopts a structured and iterative process to address inefficiencies in geothermal operational decision-making. It begins with problem definition, research objectives, literature review, and gap analysis to establish the research context and identify shortcomings in current manual workflows. These insights are then translated into a roadmap for digital transformation, followed by system implementation and technology assessment to ensure that each stage delivers measurable improvements. The overall workflow of this research framework is illustrated in Figure 1.

A key decision point in the framework is the evaluation of prerequisites for machine learning integration, such as data readiness, system maturity, and organizational alignment [12], [13]. If these conditions are not met, the process loops back to refine earlier stages. Once readiness is achieved, machine learning models are embedded to support anomaly detection and accelerate decision-making, with outcomes evaluated based on reduced decision-making cycle time.

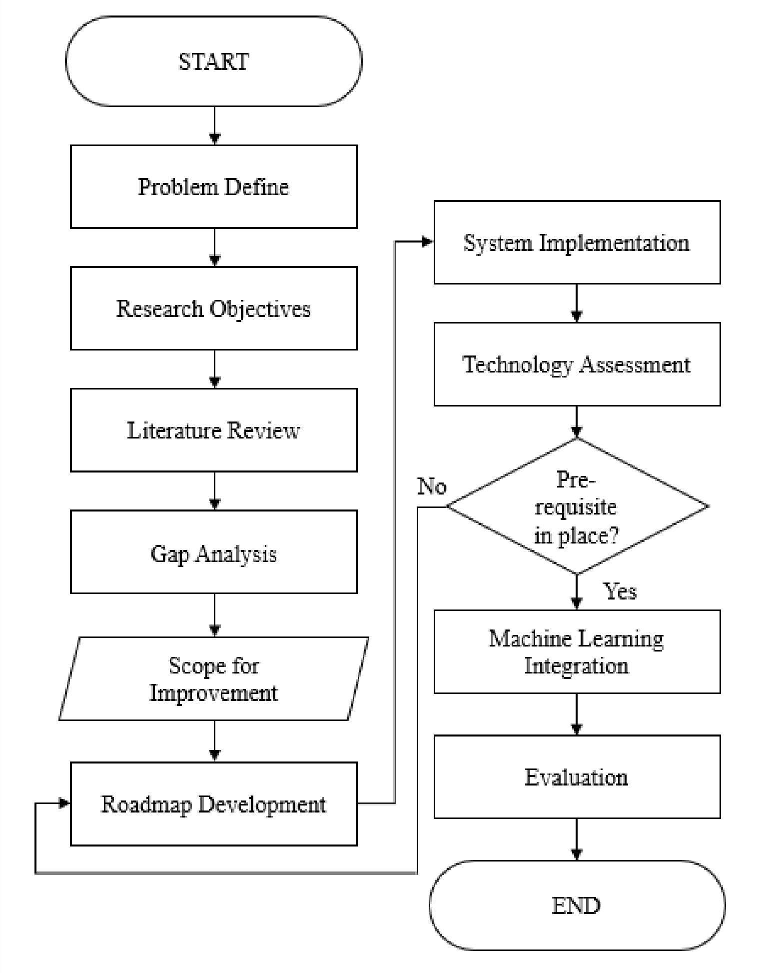


Figure 1. Staged workflow of the research framework integrating digitalization and machine learning

Data collection

The primary data source consists of geothermal operational logs, including wellhead pressure, steam pressure, temperature, and flow rate, brine reinjection data, and turbine performance indicators. Traditional file sharing method was obviously not lean in a business process. To address its drawbacks, the digital initiative implemented a database management system (DBMS), data visualization, and an asset framework (AF).

A database management system, or DBMS, is software designed to assist in maintaining and utilizing large collections of data, and the need for such systems, as well as their use, is growing rapidly [14]. DBMS, built into this system, comprised a real-time historian database and an asset framework (AF) storage. Asset Framework is a platform to create a hierarchy or structure from collected data based on the site equipment hierarchy. AF is especially useful for creating a model (for analysis and forecasting), an event (data generated by a specific logic), and a notification. A robust DBMS, visualization, and diagnostic analytics architecture is shown in Figure 2 below. Its sources were OPC DA and Modbus Ethernet protocol, which were installed on the site server. The previous paper-based data was digitized and converted to real-time data, which was transferred to a DBMS server at the headquarters office. The process reduced the long cycle time for transferring data to the head office. After integrating the data, engineers can develop a user-friendly visualization and analytics tool that will help them make quick operational improvement decisions by leveraging trending data.

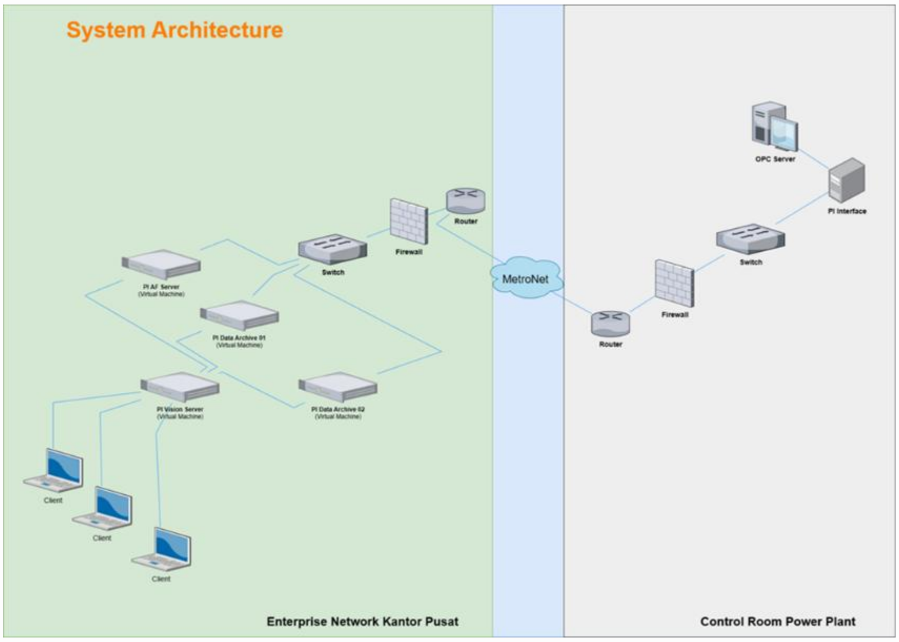


Figure 2. System architecture of the digital transformation initiative

Machine learning methods

The ML methodology focused on anomaly detection and decision acceleration. This phase leverages the historical and real-time data collected in earlier stages to train ML algorithms. By employing supervised or unsupervised learning methods, the system can autonomously identify deviations from normal operational patterns. These anomalies are crucial indicators of early-stage faults, sensor drift, or potential system failures, providing a data-driven foundation for proactive decision-making. This study employed supervised and unsupervised machine learning techniques to identify deviations in complex, high-dimensional operational datasets. Specifically, Locality Sensitive Hashing (LSH) was used to efficiently detect patterns and outliers in large-scale time-series data, while OPTICS (Ordering Points To Identify the Clustering Structure) was applied to uncover natural clustering structures without requiring a predefined number of clusters. In addition, Root Mean Square Error (RMSE) was used as a statistical measure to quantify deviations between actual sensor data and predicted operational conditions. By training on normal operation data, the ML models could flag abnormal states and support faster corrective actions.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (1)$$

Where n denotes the total number of data points considered in the evaluation, thereby defining the size of the dataset used for model assessment. The variable y_i corresponds to the actual or observed value at the i^{th} data point, reflecting the true measurement obtained from operational data. Conversely, $\hat{y}_i$ represents the predicted value generated by the machine learning model at the i^{th} data point.

Locality sensitive hashing (LSH)

Locality Sensitive Hashing (LSH) is a dimensionality-reduction technique that efficiently clusters similar data points by hashing them into the same "bucket" with high probability. It is particularly useful for identifying outliers by evaluating how frequently a new data point falls into a bucket of previously observed normal behavior. LSH works by mapping similar data points into the same hash “bucket” using a series of hash functions that preserve similarity (e.g., cosine or Euclidean distance). This allows the system to group operational states that exhibit similar behavior (normal operation), while quickly identifying outliers (anomalies) that do not share characteristics with any known group [15], [16]. The working principles of LSH are depicted in Figure 3.

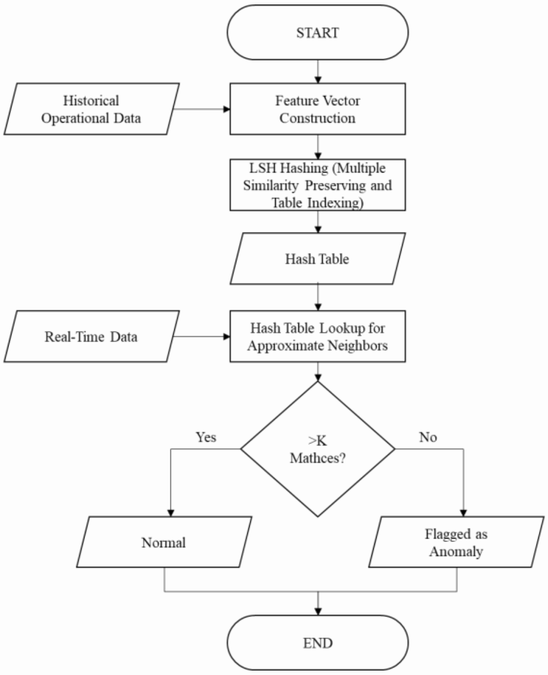


Figure 3. Working principles of LSH

Ordering points to identify the clustering structure (OPTICS)

OPTICS is a density-based clustering algorithm developed to identify underlying structures in large, noisy datasets without requiring the number of clusters to be defined beforehand. Unlike traditional algorithms like k-means, which assume spherical clusters and fixed counts, OPTICS can detect clusters of arbitrary shape and varying density—making it especially effective in industrial systems such as geothermal plants, where operational states can shift subtly or irregularly over time. OPTICS works by measuring how closely grouped data points are to one another in a multi-dimensional space [17], [18]. The working principles of OPTICS are depicted in Figure 4.

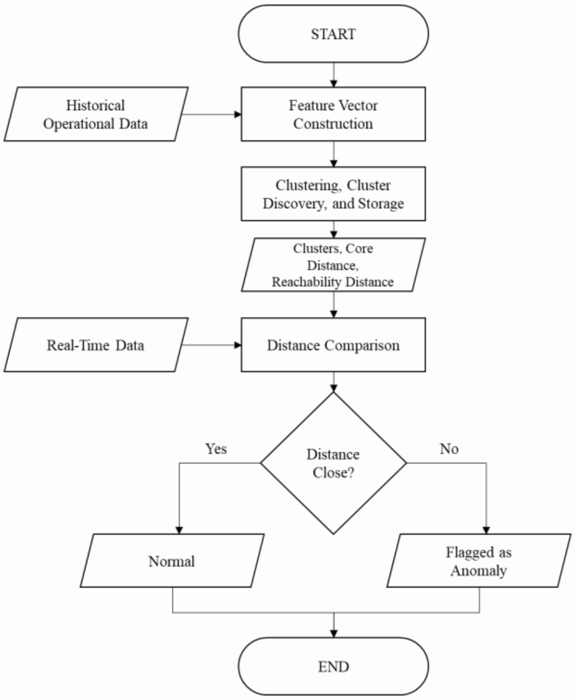


Figure 4. Working principles of OPTICS

RESULT AND DISCUSSION

Prior to digitalization, the communication protocol between local operators and headquarters engineers was exceedingly complex, with frequent issues of poor data quality. The use of a paper-based Operation Log Sheet led to unnecessary but necessary work. Several reports were supplied individually and in a scattered way from each field. As shown in Figure 5, there were various manual daily operational data, divided into three sections: steam field, power plant, and other reports. All reports were manually delivered to headquarters via the communicator applications or corporate mail. The entire traditional file-sharing workflow consumed about 14.5 hours each day. Late submission and data compilation triggered an increase in non-productive time and, consequently, weak decision-making, resulting in tremendous effort for engineers at headquarters.

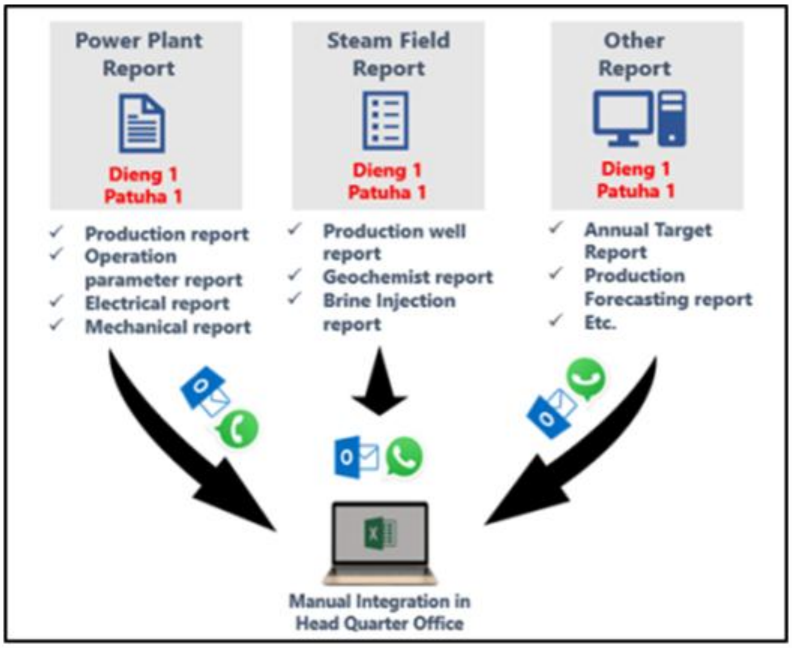


Figure 5. Traditional file-sharing methods

The digitalization initiative commenced with the conversion of operational log sheets into a centralized, online repository that functioned as a preliminary database management system (DBMS). This transition marked a fundamental shift from manual, siloed reporting to a structured, accessible data environment. On top of this DBMS, dashboards were developed using Business Intelligence (BI) tools, which served as the primary interface for data visualization and analysis. The dashboards enabled faster access to key operational parameters and trends, significantly reducing the time required for data retrieval and processing. The adoption of this digital approach resulted in a 50% reduction in the previous decision-making cycle time, shown in Figure 6. Additionally, this transformation yielded a soft benefit of approximately USD 2,245 per month, or around USD 26,940 annually.

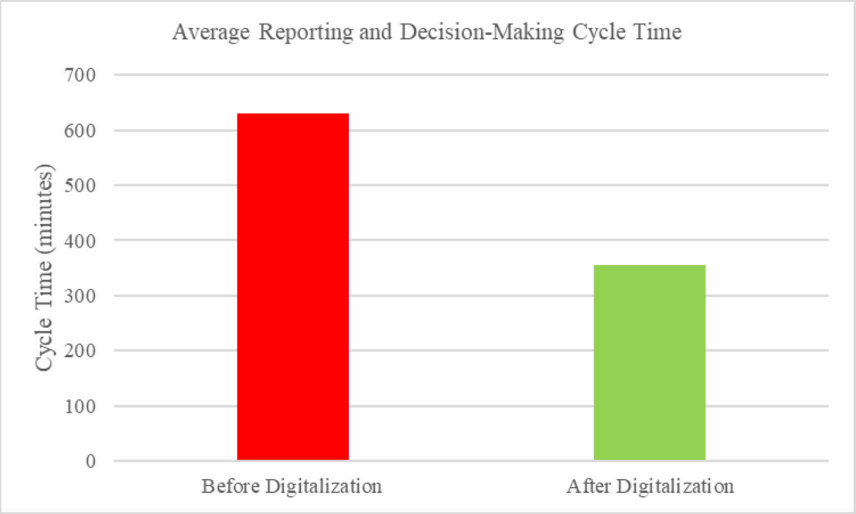


Figure 6. Reduction of decision-making cycle time

The ultimate outcome of integrating operational data into a robust database management system and machine learning was not only visualizing trending charts and other graphs but also providing deep insights for both operational and management levels. The real-time dashboards, as shown in Figure 7, enabled continuous monitoring of key operational parameters, such as power generation and specific steam consumption (SSC), allowing engineers and planners to track performance indicators in real time and respond to deviations without waiting for periodic reports.

$$\text{Power Generation (kW)} = \dot{m} \times (h_{\text{inlet}} - h_{\text{outlet}}) \tag{2}$$

$$\text{Specific Steam Consumption (SSC)} = \frac{\dot{m}}{\text{Power Generation}} \tag{3}$$

Where $\dot{m}$ represents mass flow rate (kg/s), h_{inlet} denotes enthalpy at turbine inlet (kJ/kg), and h_{outlet} represents enthalpy at turbine outlet. This shift from static reporting to dynamic visualization enhanced situational awareness and supported more agile operational control.

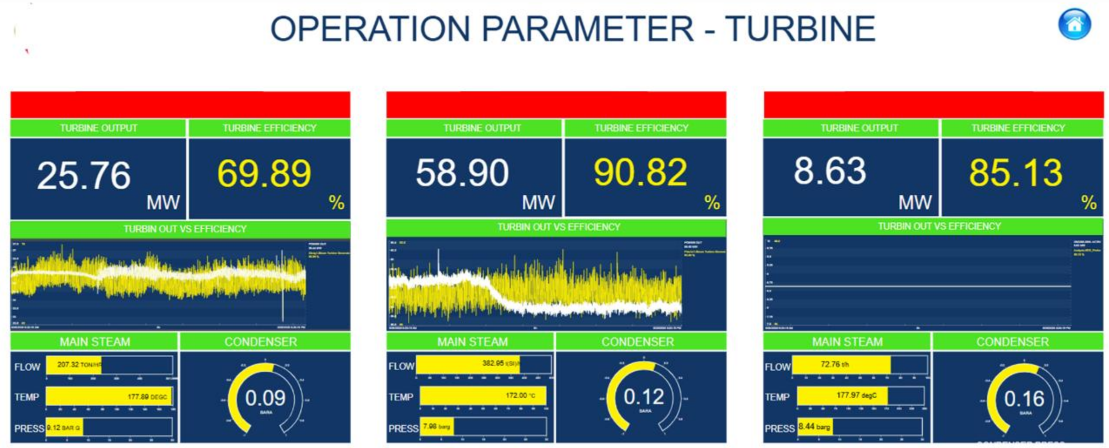
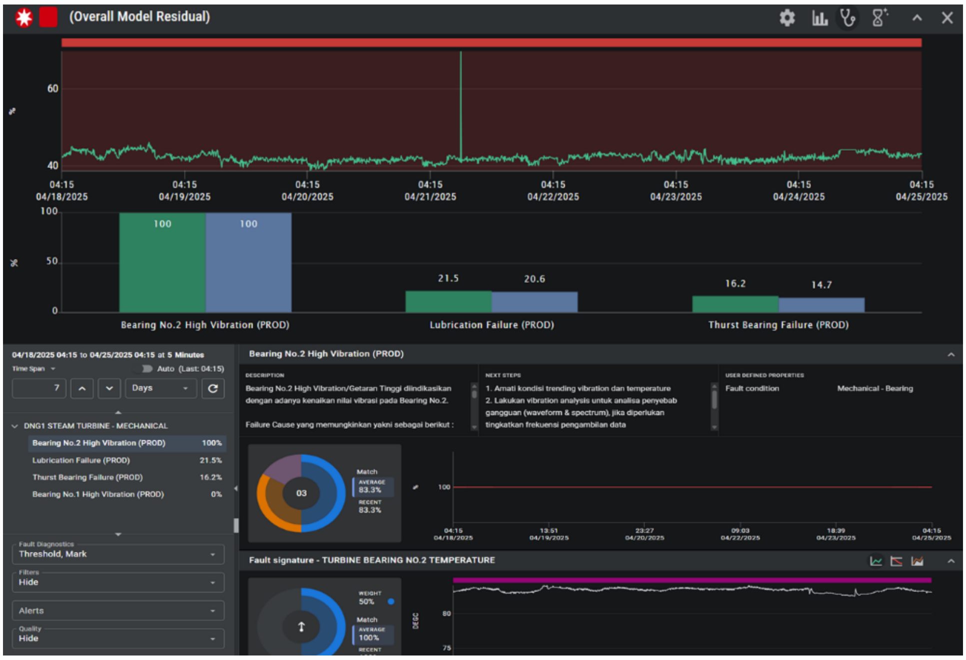


Figure 7. Turbine real-time operation parameter dashboard

In the upstream steam field, one recurring challenge is the formation of massive silica deposits, or scaling, which disrupts fluid flow and reduces system efficiency. To address this, a condition monitoring dashboard was implemented by integrating operational production data with geochemical brine analysis from the geochemistry team. This proactive approach prevented silica scaling in the brine line. As a result, the cleaning of the brine line—typically required due to severe scaling—was successfully avoided, preventing one day of lost production opportunity and resulting in estimated cost savings of approximately USD 19,500.

In addition, the integration of machine learning models provided an advanced layer of anomaly detection and diagnostic insights, as shown in Figure 8. As demonstrated in recent work (e.g., Abrasaldo et al. 2024; Azuma & Hashimoto. 2025), anomaly detection frameworks can flag deviations in steam production or equipment behavior hours before critical failure. These techniques rely on historical normal behavior to set thresholds or define clusters, enabling proactive maintenance rather than reactive responses [19], [20].



04/18/2025 04:15 to 04/25/2025 04:15 at 5 Minutes

DNG1 STEAM TURBINE - MECHANICAL

Bearing No.2 High Vibration (PROD)	100%
Lubrication Failure (PROD)	21.5%
Thurst Bearing Failure (PROD)	16.2%
Bearing No.1 High Vibration (PROD)	0%

Fault Diagnostics: Threshold, Mark
Filters: Hide
Alerts
Quality: Hide

Bearing No.2 High Vibration (PROD)

DESCRIPTION
Bearing No.2 High Vibration/Getaran Tinggi diindikasikan dengan adanya kenaikan nilai vibrasi pada Bearing No.2.

Failure Cause yang memungkinkan yakni sebagai berikut :

NEXT STEPS
1. Amati kondisi trending vibration dan temperature
2. Lakukan vibration analysis untuk analisa penyebab gangguan (waveform & spectrum), jika diperlukan tingkatkan frekuensi pengambilan data

USER DEFINED PROPERTIES
Fault condition: Mechanical - Bearing

Fault signature - TURBINE BEARING NO.2 TEMPERATURE

Figure 8. Machine learning integration and diagnostic insights

From an operational perspective, integrating real-time dashboards proved a key enabler in bridging the gap between raw data and actionable insights. Engineers and operators reported that visualisation through BI interfaces reduced the time required to interpret system performance trends, while ML-driven alerts provided contextual decision support. This change reflects a shift from reactive to proactive management of geothermal plant operations, in line with broader digital transformation trends in the energy sector [5].

CONCLUSION

This study examined the implementation of the Geothermal Operations Optimization, Reliability, and Efficiency through Machine Learning (GOOREMALE) framework to address delays in operational decision-making caused by manual processes in geothermal power plant management. The results show that digitalizing operational data through a database management system (DBMS) and using Business Intelligence dashboards significantly improved information accessibility and transparency in decision-making. In addition, this digital foundation enabled real-time collaboration between field engineers and headquarters, enabling operational insights to be shared and acted on more quickly across organizational levels.

The integration of machine learning methods, particularly Locality Sensitive Hashing (LSH), OPTICS clustering, and Root Mean Square Error (RMSE) evaluation, enabled faster, more accurate anomaly detection. This implementation successfully reduced decision-making cycle time by approximately 50%, achieved cost savings, and enhanced real-time monitoring of operational performance. Moreover, the use of ML-driven anomaly detection strengthened the

ability of operational teams to anticipate risks and proactively
intervene before underperformance or failures occur.

The contribution of this research lies in providing a practical, replicable framework for digital transformation in geothermal plant operations. The research boundary is limited to operational decision-making and cycle-time efficiency, without extending to long-term reliability, maintenance cost optimization, or subsurface management. Future research is recommended to integrate predictive reliability models, such as RAM analysis, long-term optimization strategies, and preparations for the incorporation of additional generating units.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to all parties who have contributed to the completion of this research. Special gratitude is extended to the management and staff of the company for providing access to operational data and valuable insights that served as the foundation for this study. The constructive feedback and technical guidance from engineers, operators, and subject-matter experts were instrumental in shaping the methodology and validating the findings. Appreciation is also given to academic advisors and colleagues for their continuous encouragement, critical reviews, and suggestions throughout the research process. Their support not only strengthened the analytical framework but also ensured that the research remained aligned with both academic rigor and practical relevance.

AUTHORS' CONTRIBUTION

Author 1 was primarily responsible for conceptualizing the research framework, developing the methodology, and implementing the GOOREMALE system in the geothermal power plant context. Author 1 also conducted the data analysis, prepared the figures and tables, and drafted the initial version of the manuscript. Author 2 contributed to the research design, provided technical validation of the machine learning models, and offered critical input during the interpretation of results. Author 2 also revised the manuscript, improved the structure and clarity, and ensured alignment with academic and industry standards. Both authors reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

REFERENCES

[1] M. Krieger, K. A. Kurek, and M. Brommer, "Global geothermal industry data collection: A systematic review," *Geothermics*, vol. 104, p. 102457, Sep. 2022, doi: 10.1016/j.geothermics.2022.102457.

[2] F. V. Hackstein and R. Madlener, "Sustainable operation of geothermal power plants: why economics matters," *Geotherm. Energy*, vol. 9, no. 1, p. 10, Mar. 2021, doi: 10.1186/s40517-021-00183-2.

[3] W. Ling, Y. Liu, R. Young, T. T. Cladouhos, and B. Jafarpour, "Efficient data-driven models for prediction and optimization of geothermal power plant operations," *Geothermics*, vol. 119, Feb. 2024, doi: 10.1016/j.geothermics.2024.102924.

[4] P. Siratovich *et al.*, "GOOML - Finding Optimization Opportunities for Geothermal Operations," *Proc. 47 Th Workshop Geotherm. Reserv. Eng.*, vol. 47, 2022.

[5] E. R. Okoroafor, C. M. Smith, K. I. Ochie, C. J. Nwosu, H. Gudmundsdottir, and M. (Jabs) Aljubran, "Machine learning in subsurface geothermal energy: Two decades in review," *Geothermics*, vol. 102, p. 102401, Jun. 2022, doi: 10.1016/j.geothermics.2022.102401.

[6] A. Alzayani, A. Al-Qasim, P. Birkle, and A. Al-Muhaideb, "An Overview of Monitoring and Surveillance Technologies for Geothermal Systems," in *ADIPEC*, Abu Dhabi, UAE: SPE, Oct. 2023, p. D041S124R002. doi: 10.2118/216945-MS.

[7] J. Prihantoro, H. Taufiqurrohman, Suyanto, Cahyadi, A. Larasati, and A. Nurrohim, "Implementation SCADA Web Client For Monitoring Geothermal Power Plant 3 MW Kamojang - Indonesia," in *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, Yogyakarta, Indonesia: IEEE, Nov. 2022, pp. 270–275. doi: 10.1109/ICE3IS56585.2022.10010311.

[8] P. S. Omrani *et al.*, "Digital Twin of Geothermal Assets Assisting the Production and Operational Decisions," *GeoTHERM J.*, vol. 2, Dec. 2024, Accessed: Sep. 13, 2025. [Online]. Available: https://www.geotherm-journal.com/index.php/gtjw/article/view/216

[9] A. Al-Fakih, A. Al-khudafi, A. Koeshidayatullah, S. Kaka, and A. Al-Gathe, "Forecasting geothermal temperature in western Yemen with Bayesian-optimized machine learning regression models," *Geotherm. Energy*, vol. 13, no. 1, p. 4, Jan. 2025, doi: 10.1186/s40517-024-00324-3.

[10] B. Ibrahim, J. O. Konduah, and I. Ahenkorah, "Predicting reservoir temperature of geothermal systems in Western Anatolia, Turkey: A focus on predictive performance and explainability of machine learning models," *Geothermics*,

vol. 112, p. 102727, Jul. 2023, doi: 10.1016/j.geothermics.2023.102727.

[11] W. Gao and J. Zhao, “Prediction of geothermal temperature field by multi-attribute neural network,” *Geotherm. Energy*, vol. 12, no. 1, p. 22, Jun. 2024, doi: 10.1186/s40517-024-00300-x.

[12] M. M. John, H. H. Olsson, and J. Bosch, “An empirical guide to MLOps adoption: Framework, maturity model and taxonomy,” *Inf. Softw. Technol.*, vol. 183, p. 107725, Jul. 2025, doi: 10.1016/j.infsof.2025.107725.

[13] M. Zarour, H. Alzabut, and K. T. Al-Sarayreh, “MLOps best practices, challenges and maturity models: A systematic literature review,” *Inf. Softw. Technol.*, vol. 183, p. 107733, Jul. 2025, doi: 10.1016/j.infsof.2025.107733.

[14] M. T. Özsu and P. Valduriez, “Database Integration—Multidatabase Systems,” in *Principles of Distributed Database Systems*, M. T. Özsu and P. Valduriez, Eds., Cham: Springer International Publishing, 2020, pp. 281–347. doi: 10.1007/978-3-030-26253-2_7.

[15] J. Wang, T. Zhang, jingkuan song, N. Sebe, and H. T. Shen, “A Survey on Learning to Hash,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 769–790, Apr. 2018, doi: 10.1109/TPAMI.2017.2699960.

[16] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*, New internat. edition. in Always learning. Harlow: Pearson, 2014.

[17] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “OPTICS: ordering points to identify the clustering structure,” *SIGMOD Rec*, vol. 28, no. 2, pp. 49–60, Jun. 1999, doi: 10.1145/304181.304187.

[18] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek, “Density-based clustering,” *WIREs Data Min. Knowl. Discov.*, vol. 1, no. 3, pp. 231–240, 2011, doi: 10.1002/widm.30.

[19] K. Azuma and Y. Hashimoto, “Anomaly Detection in Geothermal Steam Production Time Series Using Singular Spectrum Analysis,” *Eng. Proc.*, vol. 107, no. 1, p. 24, 2025, doi: 10.3390/engproc2025107024.

[20] P. Michael B. Abrasaldo, S. J. Zarrouk, A. Mudie, J. Cen, C. Siega, and A. W. Kempa-Liehr, “Detection of abnormal operation in geothermal binary plant feed pumps using time-series analytics,” *Expert Syst. Appl.*, vol. 247, p. 123305, Aug. 2024, doi: 10.1016/j.eswa.2024.123305.