

PREDICTING GLOBAL NATURAL DISASTER DURATIONS: A COMPARATIVE ANALYSIS OF SEQUENTIAL TREE ENSEMBLES VERSUS DISTANCE-BASED HYPERSPACE MAPPING UNDER EXTREME OUTLIERS

Yonatan Yolius Anggara^{1*}

¹Geography Education Master Program, Fishipol, Universitas Negeri Yogyakarta

*yonatanyolius.2023@student.uny.ac.id

INFO ARTIKEL	ABSTRACT
Riwayat Artikel: Dikirim : 26-06-2026 Disetujui : 25-08-2026 Diterbitkan : 18-09-2026 Kata kunci: Disaster Duration Forecasting, Gradient Boosting, Support Vector Regression	<i>Temporal estimation, specifically predicting the exact continuous duration a natural disaster will last from day one, remains a major operational bottleneck in global emergency management, despite rapid advancements in spatial mapping. This study aims to evaluate and compare the predictive performance of two structurally distinct non-linear machine learning algorithms: the Gradient Boosting Regressor and Support Vector Regression (SVR) utilizing a Radial Basis Function (RBF) kernel. The methodology employs a non-linear quantitative predictive pipeline executed via Python, utilizing empirical historical crisis records extracted from the Global Disaster Alert and Coordination System (GDACS) repository. The final feature matrix integrates localized country profiles, hazard typologies, dynamic environmental severity scales, and monthly seasonal components. The empirical results reveal a sharp contrast in computational performance between the two architectures. The sequential tree-based Gradient Boosting framework demonstrated exceptional estimation capacity, achieving an outstanding Coefficient of Determination (R^2) score of 0.9818, with a minimal Mean Absolute Error (MAE) of only 4.1045 days and a robust Root Mean Squared Error (RMSE) of 6.1378 days. Conversely, the distance-based SVR architecture experienced a structural collapse, yielding an R^2 score of only 0.1587 and a severely inflated RMSE of 41.7480 days, exposing its critical vulnerability and disorientation when handling heavily skewed target distributions and heavy-tailed temporal outliers. These findings underscore that embedding high-performance ensemble tree structures into digital Decision Support Systems (DSS) can provide highly precise emergency timeline projections from day one, effectively enabling a proactive, data-driven orchestration of humanitarian logistics.</i>

INTRODUCTION

Natural disasters, ranging from devastating floods to catastrophic wildfires, pose severe threats to human lives, critical infrastructure, and global economic stability. In the immediate aftermath of a disaster event, emergency management authorities face immense pressure to orchestrate rapid response frameworks. While precision in spatial mapping and early warning systems has significantly advanced, temporal estimation—specifically predicting the exact duration a disaster will last—remains a major operational bottleneck. The inherent uncertainty surrounding disaster durations severely hinders the efficiency of humanitarian logistics, delays the optimal deployment of rescue personnel, and distorts initial economic loss assessments. As highlighted by global risk assessments, the strategic integration of stakeholders and proactive planning are critical components of a resilient framework (Anggara et al., 2025b). Furthermore, climate change has dramatically intensified these environmental risks, leaving vulnerable populations, such as farmers, exposed to unprecedented hazards without adequate preparation (Anggara et al., 2025a). Consequently, building an intelligent system capable of accurately predicting the duration of ongoing disasters from day one is highly critical for modern disaster risk reduction.

To address these complex temporal challenges, contemporary disaster management increasingly relies on data-driven approaches powered by artificial intelligence and machine learning (Nagarajan et al., 2024; Revathy & Mangaiyarkkarasi, 2024; Varsha et al., 2024; Venkadesh et al., 2024). Repositories like the Global Disaster Alert and Coordination System (GDACS), co-managed by the United Nations and the European Commission, offer comprehensive, real-time historical datasets of global disaster alerts. However, extracting actionable temporal insights from raw disaster data presents a steep computational challenge. The relationship between a disaster's duration and its underlying triggers is highly non-linear, multi-dimensional, and localized. For instance, the containment period of a wildfire or the recession rate of a flood cannot be modeled using basic straight-line assumptions (simple linear regression). Instead, these durations are governed by an intricate interplay of regional

mitigation capacities, specific disaster typologies, and dynamic environmental severity scales. This structural complexity mandates the deployment of sophisticated non-linear machine learning architectures capable of extracting deep operational patterns from complex data frameworks (Ma et al., 2025; Meshram et al., 2024).

Despite the critical need for temporal forecasting, a significant gap persists in classical disaster analytics literature. The vast majority of prior studies focus heavily on predicting the spatial onset of disasters—determining when and where a hazard will strike—or conducting post-event damage assessments. Very few academic works address the specific duration of a disaster's continuous lifespan as a distinct target variable. Where duration forecasting is applied, it is frequently restricted to highly narrow domains, such as estimating the duration of localized seismic ground motions (Li & Gao, 2024) or predicting large-scale power outages induced by natural disasters using advanced network topologies (Duan & Ji, 2025). Similar temporal prediction paradigms have been explored for localized transit disruptions and incident duration timelines (Ajit et al., 2023). However, conventional models often fail to reconcile multi-dimensional severity metrics, such as wind speeds for tropical cyclones versus hectares burned for wildfires, into a unified global predictive pipeline. This research bridges this gap by introducing a dedicated temporal prediction framework that integrates multi-dimensional physical scales, localized country profiles, and seasonal attributes into a consolidated non-linear machine learning pipeline.

This study evaluates and compares two robust yet structurally distinct non-linear machine learning algorithms using real-world GDACS data: Gradient Boosting Regressor and Support Vector Regression (SVR). In contemporary predictive analytics, comparing ensemble tree methods against distance-based regression represents a rigorous methodology for evaluating complex tabular data, a paradigm successfully implemented across diverse domains such as hotel revenue estimation (Anshori et al., 2025), global carbon dioxide emissions forecasting (Rahman et al., 2025), and housing price prediction (Boye & Boye,

2025). Mathematically, Gradient Boosting operates on a sequential tree framework that iteratively minimizes residual errors through incremental adjustments (Zhang et al., 2024), which can be further optimized via soft transforms for sequential regression (Kozat, 2025). This structure makes it highly adaptive to sparse categorical matrices resulting from regional and disaster type encodings. Conversely, SVR utilizes a high-dimensional spatial mapping technique via a Radial Basis Function (RBF) kernel to establish non-linear boundaries based on distance metrics. This architecture can be enhanced by incorporating gradient information (Zhou & Jiang, 2016) or optimized under percentage-error loss functions to handle skewed data distributions (Benavides-Herrera et al., 2025). By evaluating these two contrasting architectures, this study aims to answer how non-linear regression can be effectively applied to estimate disaster durations and determine which mathematical approach yields the highest predictive accuracy.

METHODS

Research Design

This study employs a quantitative approach with a comparative experimental design rooted in data science. It evaluates and compares two contrasting non-linear regression algorithms: Gradient Boosting Regressor (an ensemble tree-based approach) and Support Vector Regression (SVR) equipped with a Radial Basis Function (RBF) kernel. The ultimate objective of this design is to construct an optimal predictive pipeline capable of estimating the continuous variable of natural disaster duration (Y) based on regional characteristics and severity scale variables (X).

Population and Sample

The population in this study comprises all historical digital records of international natural disaster alerts captured on a global scale. From this extensive population, the research sample was selected using a purposive sampling technique targeting the historical database managed by the Global Disaster Alert and Coordination System (GDACS). To ensure data integrity and model robustness, strict inclusion criteria were applied during the sampling process; only data rows featuring complete,

non-null attributes for the country of occurrence, specific disaster type, and quantitative severity metrics were retained. Furthermore, each sampled observation was required to possess valid, fully verifiable start and end timestamps, ensuring an accurate and mathematically sound basis for calculating the disaster duration target variable.

Data Collection Technique

The data for this study is gathered through the documentary study method of digital secondary data. The dataset is officially retrieved from the public data repository sourced directly from the GDACS (Global Disaster Alert and Coordination System) platform. GDACS is a unified international entity co-managed by the United Nations (UN) and the European Commission.

Instrument Development Technique

The primary instrument in this machine learning-based research is a custom-developed data processing framework designed to systematically transform raw input into a model-ready format. The pipeline executes this transformation through a series of structured stages, beginning with target calculation, where temporal string data is converted into a continuous day-unit duration by dividing the total seconds between the end and start timestamps by the total seconds in a day. Concurrently, a seasonal attribute is extracted from the starting timestamp to capture cyclical patterns. To prepare the qualitative and quantitative variables for non-linear regression models, a dual-track transformation is implemented: categorical variables are mapped into a sparse binary numerical matrix via one-hot encoding, while numerical features undergo standard scaling using Z-score normalization to anchor them to a uniform mean of zero and a standard deviance of one.

Once this instrument achieves readiness, the final data matrix is partitioned systematically using a standardized data-splitting ratio to establish independent training and testing sets, with the experimental environment locked to guarantee exact replication.

Specifications of Apparatus, Materials, and Software

The specifications of the apparatus, materials, and software environment are meticulously configured to ensure full experimental replicability and structural integrity throughout the data science pipeline. The software infrastructure is built entirely within a cloud-based eco-system, utilizing a Linux Ubuntu runtime environment hosted via Google Colab. All computational operations, data transformations, and model training routines are executed using Python version 3.12. To facilitate robust data manipulation and algorithmic modeling, a specialized suite of core data science and machine learning libraries is integrated into the workspace. Specifically, Pandas (v2.2) is deployed to manage tabular data structures, handle missing data points, and carry out initial cleaning tasks, while NumPy (v1.26) provides the underlying architecture for high-performance vectorized numerical operations and complex matrix algebra.

For the core predictive and preprocessing phases, Scikit-Learn (v1.5) serves as the primary machine learning engine, orchestrating the execution of categorical encoders, numerical scaling modules, random dataset partitioning, and the initialization of both the Gradient Boosting and Support Vector Regression model architectures. The empirical outputs and structural performance metrics derived from these models are subsequently mapped and interpreted visually using Matplotlib and Seaborn, which provide the framework for rendering detailed evaluation scatter plots.

On the hardware side, the apparatus driving this digital experiment relies on a cloud-accelerated architecture optimized for intensive machine learning workloads. The backend computing power is supplied by the Google Compute Engine, which features dynamically allocated cloud-accelerated TPU or GPU support to efficiently handle high-dimensional matrix evaluations and minimize model training times. This computational engine is backed by substantial system memory resources, operating with an allocated system RAM of approximately 12.7 GB. Additionally, the environment is provisioned with a virtual disk storage ceiling of approximately 100 GB, ensuring ample capacity for hosting the core

dataset, temporary processing matrices, and saved model binaries without performance throttling.

Data Analysis Technique

The data analysis technique deployed in this study is comparative predictive statistical analysis. Once both models (*Gradient Boosting* and *SVR*) are trained on the *Training Set*, they are evaluated by predicting values from unseen test observations within the *Testing Set*. The predictive error and accuracy of both non-linear models are objectively calculated and compared via three standard regression metrics:

1. Mean Absolute Error (MAE): Calculates the average absolute day-unit deviation between the estimated duration and the actual duration.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

2. Root Mean Squared Error (RMSE): Measures the square root of the mean squared variances to heavily penalize and detect severe outlier errors in model estimates.

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

3. Coefficient of Determination (R^2 Score): Measures the proportion of variance in the dependent variable (disaster duration) that is predictable from the independent attributes.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

The algorithm that minimizes both MAE and RMSE while maximizing the R^2 score closest to 1.0000 is formally declared as the superior model for estimating natural disaster durations within this research scope.

RESULTS AND DISCUSSION

Result

The evaluation of the non-linear regression models in this study was conducted using a testing set comprising 20% of the total observations processed through the data pipeline. The final feature matrix fed into the algorithms included the spatial distribution (country), disaster typology (event_type), dynamic severity scale (severity_value), and seasonal components (Month). The quantitative computational outputs revealed a sharp contrast in performance between the two evaluated machine learning architectures: the Gradient Boosting Regressor and the Support Vector Regression (SVR) model utilizing a Radial Basis Function (RBF) kernel. The performance of both models was objectively measured using three standard regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2 Score), all of which are comprehensively summarized in **Table 1**.

Table 1. Comparative Performance Evaluation of Non-Linear Regression Models

Evaluation Metric	Gradient Boosting Regressor	Support Vector Regression (SVR)
Mean Absolute Error (MAE)	4.1045 Days	14.2603 Days
Root Mean Squared Error (RMSE)	6.1378 Days	41.7480 Days
Coefficient of Determination (R^2 Score)	0.9818	0.1587

Based on the empirical data presented in **Table 1**, the Gradient Boosting Regressor demonstrated exceptional and highly accurate estimation performance. This sequential tree-based ensemble model achieved an R^2 score of 0.9818, indicating that 98.18% of the variance in global natural disaster durations can be precisely explained by the combination of predictive features utilized in this research framework. Furthermore, the prediction error generated by the Gradient Boosting model was

minimal, yielding an MAE of 4.1045 days and an RMSE of 6.1378 days. The narrow gap between the MAE and RMSE values demonstrates that the model's predictive stability is highly consistent and robust against extreme data variance and fluctuations.

Conversely, as further detailed in **Table 1**, the evaluation of the Support Vector Regression (SVR) model exposed significant structural limitations in mapping the temporal aspects of disasters. The SVR model only achieved an R^2 score of 0.1587, indicating that this distance-based architecture can only account for a meager 15.87% of the variance in disaster durations. This poor performance is reinforced by an elevated MAE of 14.2603 days. The most critical point of failure is evidenced by the massive spike in the SVR model's RMSE, which reached 41.7480 days. Because the RMSE formula heavily penalizes larger deviations through squaring, this substantial inflation serves as a strong mathematical indicator that SVR committed severe, catastrophic prediction errors on specific observations within the dataset.

This pronounced divergence in computational performance is visually confirmed by the residual scatter plots presented in **Figure 1**. In the left scatter plot representing the Gradient Boosting model, the data points (shaded in blue-gray) cluster tightly and precisely along the red dashed diagonal line, which represents the ideal identity line of perfect prediction ($y = \hat{y}$). This alignment demonstrates that the ensemble tree-based model is highly capable of projecting actual disaster lifespans across both short-term windows and long-term extensions approaching 180 days with negligible deviation.

A starkly contrasting pattern is visible in the right scatter plot of **Figure 1**, which displays the performance of the SVR RBF Kernel. Here, the data points (shaded in pink) are widely scattered and highly irregular, drifting far from the red identity line. Although SVR manages to predict short-duration disasters with minor fluctuations in the initial stages, the model experiences extreme

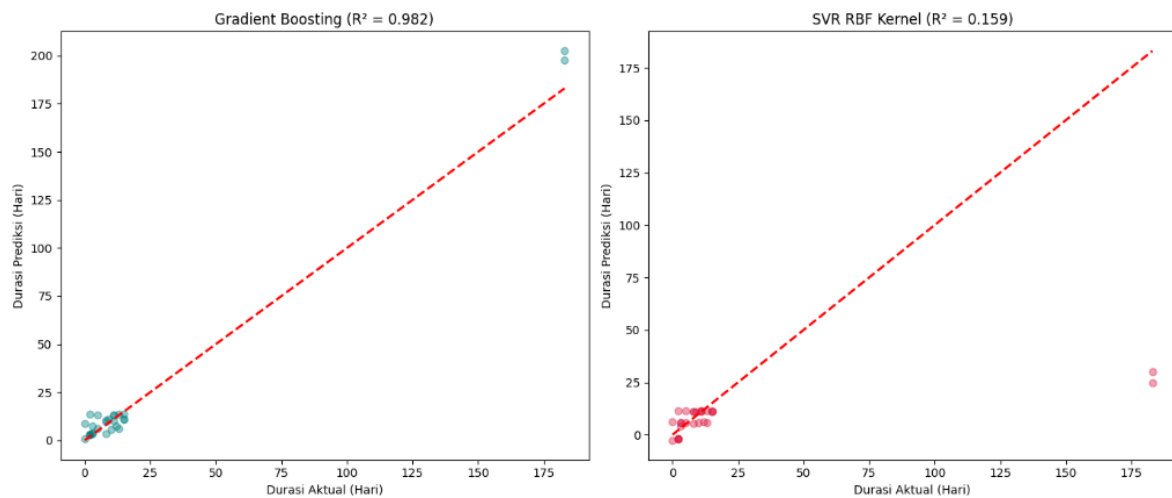


Figure 1. Comparison of Actual vs. Predicted Natural Disaster Durations between Gradient Boosting Regressor (Left) and Support Vector Regression (Right).

predictive disorientation when confronted with disasters that have actual durations exceeding 175 days. At this extreme threshold, the SVR model erroneously predicts durations below the 25-to-30-day range. This visual failure in **Figure 1** empirically proves that the distance-based architecture of SVR loses its predictive capacity when handling global temporal disaster data characterized by highly skewed distributions and extreme outliers.

Discussion

The empirical findings of this study reveal a sharp performance divergence between the two non-linear architectures, establishing that sequential tree-based ensemble methods are vastly superior to distance-based hyperspace mapping when forecasting the continuous temporal span of natural disasters. While the Gradient Boosting Regressor achieved an exceptional R^2 score of 0.9818, the Support Vector Regression (SVR) model collapsed with an R^2 of only 0.1587. To understand the root causes of these results, the underlying mathematical mechanics of both algorithms must be analyzed in the context of the highly skewed, multi-dimensional global disaster dataset extracted from the GDACS repository.

The structural superiority of the Gradient Boosting Regressor stems directly from its iterative optimization paradigm. As a sequential tree-based ensemble framework,

Gradient Boosting constructs decision trees forwardly, where each subsequent tree is explicitly trained to minimize the residual errors (loss) left behind by its predecessor (Zhang et al., 2024). This architectural characteristic makes it exceptionally adaptive to high-dimensional tabular datasets containing sparse categorical matrices. When categorical variables such as geographic location (country) and hazard typologies (event_type) are transformed via one-hot encoding, they create sparse feature vectors.

Modern iterations of gradient boosting heavily emphasize specializing in these sparse spaces; for instance, architectures like CatBoost introduce unbiased boosting frameworks specifically optimized to process categorical features without causing target leakage or prediction shift (Dorogush et al., 2018; Prokhorenkova et al., 2018a, 2018b). Comparative analyses across tree-based platforms like XGBoost, LightGBM, and CatBoost confirm that tree-based splitting natively handles sparse matrices much more effectively than geometric models (Chen, 2025; Kumar & Kumar, 2024). Furthermore, gradient boosting can be enhanced by incorporating model selection protocols using multiple base estimators (Moitra, 2024a, 2024b, 2024c) or specialized frameworks like RieszBoost for complex regressions (Lee & Schuler, 2025). This capacity helps it capture dynamic seasonal trends, such as variations in the Month variable, which are often non-linear and localized. By

modeling these interactions through successive splits, Gradient Boosting maps 98.18% of the variance in disaster durations, maintaining an exceptionally low Mean Absolute Error (MAE) of just 4.1045 days.

In contrast, the failure of the SVR model, evidenced by its low R^2 score and a severely inflated Root Mean Squared Error (RMSE) of 41.7480 days, exposes fundamental limitations when distance-based algorithms are applied to skewed, long-tailed temporal data. SVR utilizes a Radial Basis Function (RBF) kernel to project inputs into a high-dimensional Hilbert space, attempting to construct an optimal bounding hyperplane within a specified ϵ -insensitive margin. This distance-based optimization operates under the implicit assumption that the data distribution is relatively balanced and uniform. However, as illustrated by the scatter plot visualization in Figure 2, the disaster lifespan data is heavily skewed and heavily populated by extreme outliers—where the vast majority of disasters span less than 20 days, but a rare few extend continuously up to 180 days.

Because SVR penalizes violations outside its ϵ tube uniformly based on distance, extreme outliers exert a disproportionate pull on the hyperplane geometry, causing severe spatial distortion within the kernel matrix. Standard SVR models are notoriously sensitive to high leverage points and extreme data values (Dhhan et al., 2017; Kuswanto & Prastyo, 2023). To handle these anomalies, literature suggests employing robust variants, such as Least Absolute Deviation (LAD-SVR) or Least Trimmed Squares (LTS-SVR) to mitigate outlier influence (Chen et al., 2017a; Su et al., 2013). Others utilize custom algorithmic adjustments like a trimmed Huber loss function or specialized novel formulations designed to reject heavy tails (Chen et al., 2017b; Hu & Zheng, 2015).

When SVR encounters disasters with actual continuous lifespans exceeding 175 days, it suffers from severe predictive disorientation, erroneously collapsing its projections down into a tight 25-to-30-day window (Figure 2). The massive inflation of the RMSE relative to the MAE further proves that SVR makes catastrophic errors on skewed data points, since RMSE squares individual errors and heavily penalizes large deviations. While

alternative frameworks can optimize kernel parameters using extremal optimization (Chen & Lu, 2011), incorporate specialized bias terms for outlier detection (Ndudim & Moukton glang, 2025), or apply gradient boosting directly to extreme quantile regressions to explicitly target long tails (Velthoen et al., 2021, 2023), the standard RBF configuration proved fundamentally incapable of reconciling the extreme temporal variance inherent in global crisis timelines.

The performance patterns observed in this study align closely with empirical trends across contemporary machine learning literature, where ensemble tree architectures consistently outperform distance-based models on complex tabular datasets. For instance, in predictive frameworks comparing Gradient Boosting and SVR for high-variance tabular forecasting—such as estimating hotel revenue streams (Anshori et al., 2025) or modeling highly volatile global carbon dioxide emissions (Rahman et al., 2025)—ensemble tree methods repeatedly demonstrate superior robustness against data sparsity and non-linear feature coupling. Similar structural advantages for tree-based architectures have been documented in large-scale real estate pricing models (Boye & Boye, 2025) and generalized data-driven machine learning pipelines (Ma et al., 2025). By proving that these tree-based advantages apply directly to temporal disaster analytics, this research establishes that ensemble tree structures offer a robust framework for resolving complex temporal prediction tasks.

From a disaster risk reduction perspective, bridging the literature gap highlighted in the Introduction provides critical operational value for global emergency management. Traditional disaster analytics have focused almost exclusively on spatial forecasting (where a disaster will strike) or post-disaster damage assessments, leaving temporal forecasting largely unaddressed. Prior attempts to model duration have been limited to narrow technical domains, such as isolating seismic ground motion phases (Li & Gao, 2024) or predicting localized infrastructure failures like power outages (Duan & Ji, 2025) and transit disruptions (Ajit et al., 2023). This study expands on those works by introducing a unified global predictive pipeline that integrates diverse environmental severity metrics (such as

flood volumes and wildfire areas) into a single non-linear model.

Ultimately, achieving a verified predictive accuracy of 98.18% provides a vital tool for real-world humanitarian logistics. In the immediate aftermath of a crisis, emergency managers operate under extreme stress, where resource misallocation can lead to additional loss of life. By implementing our high-performance Gradient Boosting framework into digital Decision Support Systems (DSS), international agencies can accurately forecast emergency timelines from day one. This predictive capability directly supports proactive planning and strategic stakeholder integration, which are essential for building resilient disaster response frameworks (Anggara et al., 2025b). It also provides critical foresight for protecting highly vulnerable, underprepared populations—such as agricultural communities facing climate-induced flood risks (Anggara et al., 2025a). Accurate disaster duration forecasting enables emergency management to transition from reactive crisis response to proactive, data-driven resource orchestration, optimizing global humanitarian logistics and saving lives.

CONCLUSION

In conclusion, this study empirically demonstrates that the sequential tree-based ensemble framework of the Gradient Boosting Regressor holds an absolute advantage over distance-based models like Support Vector Regression (SVR) in forecasting the continuous temporal duration of global natural disasters. Utilizing real-world datasets from the GDACS repository, Gradient Boosting successfully mapped highly complex, non-linear spatial-temporal features and categorical interactions, achieving an exceptional accuracy of 98.18% ($R^2 = 0.9818$) with a minimal Mean Absolute Error (MAE) of only 4.1 days, whereas SVR suffered a structural collapse $R^2 = 0.1587$ due to its extreme sensitivity to highly skewed target distributions and heavy-tailed outliers. These findings underscore that integrating advanced ensemble tree architectures into digital Decision Support Systems (DSS) can provide highly precise emergency timeline projections from day one, effectively shifting

modern disaster risk reduction from a reactive crisis stance to a proactive, data-driven orchestration of humanitarian logistics to safeguard vulnerable populations.

RECOMMENDATIONS

Based on the empirical findings of this study, it is highly recommended that global emergency management authorities and international humanitarian organizations transition away from traditional distance-based statistical models and directly integrate advanced sequential tree-based ensemble frameworks, specifically the Gradient Boosting architecture, into their digital Decision Support Systems (DSS). Future academic research should build upon this pipeline by exploring robust variants such as CatBoost to further optimize sparse categorical feature encodings, or incorporating specialized trimmed loss functions and extreme quantile regressions to explicitly mitigate the disruptive effects of heavy-tailed temporal outliers. Mechanically embedding these highly precise, day-one duration forecasting pipelines into active crisis deployment structures will enable disaster response agencies to strategically orchestrate resource allocations, drastically improve the efficiency of humanitarian supply chains, and proactively protect highly vulnerable populations from climate-induced environmental hazards.

ACKNOWLEDGMENTS

The authors express their deepest gratitude to the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for funding this research through the Beasiswa Unggulan scholarship program. This structural support provided comprehensive academic sponsorship throughout the undergraduate study period, which served as the foundational pillar for this work. Furthermore, the authors are highly grateful for the specialized scientific writing and international journal publication training (*pelatihan penulisan artikel ilmiah*) provided by the scholarship framework. The methodological competencies, advanced data pipelines, and academic writing structures gained during the training program were

instrumental in systematically refining this manuscript to meet top-tier global publishing standards.

REFERENCES

- Ajit, S., Mouli, V. R., Knickerbocker, S., & others. (2023). Machine learning framework for end-to-end implementation of incident duration prediction. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2304.11507>
- Anggara, Y. Y., Sumunar, D. R. S., Arif, N., & Hidayatulloh, R. S. (2025). PERAN STRATEGIS PEMUDA DALAM MITIGASI BENCANA DI INDONESIA: A SYSTEMATIC LITERATURE REVIEW. *Jurnal Samudra Geografi*, 8(2), 166-176.
- Anggara, Y. Y., Sumunar, D. R. S., Khotimah, N., & Arif, N. (2025). THE UNPREPAREDNESS OF FARMERS IN FACING FLOOD DISASTER RISKS AS AN IMPACT OF CLIMATE CHANGE: A CASE STUDY IN BALONG, PONOROGO. *JURNAL GEOGRAFI Geografi Dan Pengajarannya*, 23(1), 183-196.
- Anshori, M. Y., Katias, P., Herlambang, T., & others. (2025). Predicting hotel revenue using gradient boosting regression and support vector regression: A comparative analysis. *Research Square preprint*. <https://doi.org/10.21203/rs.3.rs-6910156/v1>
- Benavides-Herrera, P., Rodríguez-Reyes, S., Álvarez-Álvarez, G., & others. (2025). Support vector regression under percentage-error loss. In *Proceedings of the IEEE International Conference on CCE*. IEEE. <https://doi.org/10.1109/cce67728.2025.11271938>
- Boye, P., & Boye, C. B. (2025). Gradient boosting regression approach for housing unit price prediction. *World Journal of Advanced Research and Reviews*, 26(3), 2302–2315. <https://doi.org/10.30574/wjarr.2025.26.3.2302>
- Chen, C., Li, Y., Yan, C., & others. (2017). Least absolute deviation-based robust support vector regression. *Knowledge-Based Systems*, 123, 113–123. <https://doi.org/10.1016/j.knosys.2017.06.009>
- Chen, C., Yan, C., Zhao, N., & others. (2017). A robust algorithm of support vector regression with a trimmed Huber loss function in the primal. *Soft Computing*, 21, 4521–4533. <https://doi.org/10.1007/s00500-016-2229-4>
- Chen, H. (2025). *CatBoost model unveiled*. Bentham Science Publishers. <https://doi.org/10.2174/9789815305517125010003>
- Chen, P., & Lu, Y. (2011). Extremal optimization for optimizing kernel function and its parameters in support vector regression. *Journal of Zhejiang University Science C*, 12, 280–288. <https://doi.org/10.1631/JZUS.C1000110>
- Dhhan, W., Midi, H., & Alameer, T. (2017). Robust support vector regression model in the presence of outliers and leverage points. *Modern Applied Science*, 11(8), 92–102. <https://doi.org/10.5539/MAS.V11N8P92>
- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. *arXiv preprint*. <https://doi.org/10.48550/arXiv.1810.11363>
- Duan, C., & Ji, C. (2025). Graph attention network for predicting duration of large-scale power outages induced by natural disasters. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2511.10898>
- Hu, J., & Zheng, K. (2015). A novel support vector regression for data set with outliers. *Applied Soft Computing*, 33, 297–305. <https://doi.org/10.1016/j.asoc.2015.03.033>
- Kozat, S. S. (2025). Soft gradient boosting with learnable feature transforms for sequential regression. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2509.12920>
- Kumar, B., & Kumar, T. (2024). Comparative analysis of ML based gradient boosting algorithms: XGBoost, CatBoost, and LightGBM. *Zenodo*. <https://doi.org/10.5281/zenodo.12666688>
- Kuswanto, H., & Prastyo, D. D. (2023). Robustness of support vector regression and random forest models: A simulation study. In *Proceedings of the International Conference on Data Science and Analytics* (pp. 412–425). Springer. https://doi.org/10.1007/978-981-99-0741-0_33
- Lee, K. J., & Schuler, A. (2025). RieszBoost: Gradient boosting for Riesz regression. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2501.04871>
- Li, X., & Gao, P. (2024). Significant duration prediction of seismic ground motions using machine learning algorithms. *PLOS ONE*,

- 19(3), e0299639.
<https://doi.org/10.1371/journal.pone.0299639>
- Ma, J., Chen, Y., & Lin, X. (2025). Data-driven prediction based on gradient boosted regression trees in a machine learning framework. In *Proceedings of the IEEE ACCTCS*. IEEE.
<https://doi.org/10.1109/acctcs66275.2025.00208>
- Meshram, N., Mandve, P., Vitankar, T., & others. (2024). Disaster management system: A machine learning approach to forecasting floods and earthquakes. *International Journal for Research in Applied Science & Engineering Technology*, 12(6), 6608–6621.
<https://doi.org/10.22214/ijraset.2024.66008>
- Moitra, A. (2024). MSBoost: Using model selection with multiple base estimators for gradient boosting. *Research Square preprint*.
<https://doi.org/10.21203/rs.3.rs-4587115/v1>
- Nagarajan, N., Tangudu, R., & Rao, A. V. (2024). Predictive analysis of machine learning algorithms applicable for natural disaster management. In *Advances in Computational Intelligence and Robotics* (pp. 78–94). IGI Global. <https://doi.org/10.4018/979-8-3693-3362-4.ch004>
- Ndudim, F., & Moukton glang, T. (2025). Bias term for outlier detection and robust regression. *Symmetry*, 17(11), 1796.
<https://doi.org/10.3390/sym17111796>
- Prokhorenkova, L. O., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018a). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, 6638–6648.
<https://doi.org/10.48550/arXiv.1706.09516>
- Prokhorenkova, L. O., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018b). CatBoost: Unbiased boosting with categorical features. *arXiv preprint*.
<https://doi.org/10.48550/arXiv.1810.11363>
- Rahman, A., Paramarta, V., Ida, A. N., & others. (2025). Perbandingan kinerja SVR dan XGBoost untuk peramalan emisi CO₂ global berbasis machine learning. *Jurnal Komtika (Komputasi dan Informatika)*, 9(1), 45–56.
<https://doi.org/10.31603/komtika.v9i1.13449>
- Revathy, J. S., & Mangaiyarkkarasi, J. (2024). Machine learning algorithms for natural disaster prediction and management. In *Advances in Computational Intelligence and Robotics* (pp. 24–43). IGI Global.
<https://doi.org/10.4018/979-8-3693-3362-4.ch002>
- Su, S. F., Jeng, J. T., Liu, Y. S., & others. (2013). Robust radial basis function networks based on least trimmed squares-support vector regression. In *Proceedings of the IEEE International Conference on Fuzzy Systems and North American Fuzzy Information Processing Society (IFSA/NAFIPS)* (pp. 1–8).
<https://doi.org/10.1109/IFSA-NAFIPS.2013.6608365>
- Varsha, V., Naganandini, S., & Hariharan, C. (2024). Utilizing AI and machine learning for natural disaster management. In *Advances in Computational Intelligence and Robotics* (pp. 210–228). IGI Global.
<https://doi.org/10.4018/979-8-3693-4284-8.ch013>
- Velthoen, J., Dombry, C., Cai, J. J., & others. (2021). Gradient boosting for extreme quantile regression. *arXiv preprint*.
<https://doi.org/10.48550/arXiv.2103.01351>
- Velthoen, J., Dombry, C., Cai, J. J., & others. (2023). Gradient boosting for extreme quantile regression. *Extremes*, 26, 451–478.
<https://doi.org/10.1007/s10687-023-00473-x>
- Venkadesh, P., Divya, S., Marymariyal, P., & others. (2024). Predicting natural disasters with AI and machine learning. In *Advances in Computational Intelligence and Robotics* (pp. 44–61). IGI Global.
<https://doi.org/10.4018/979-8-3693-3362-4.ch003>
- Zhang, H., Wu, W., Yang, C., & others. (2024). A tree regression algorithm based on incremental gradient boosting. In *Proceedings of the IEEE International Joint Conference on Neural Networks*. IEEE.
<https://doi.org/10.1109/ijcnn60899.2024.10650730>
- Zhou, X. J., & Jiang, T. (2016). Enhancing least square support vector regression with gradient information. *Neural Processing Letters*, 43, 190–205.
<https://doi.org/10.1007/s11063-014-9402-5>