

Two-Stage Machine Learning: Clustering-Based Feature Engineering to Improve the Accuracy of Support Vector Machines in Flood Loss Prediction in Gorontalo Province

Rampi Yusuf¹, Hermila A.², Mario W. Yani³, Nikmasari Pakaya⁴, Muchlis Polin⁵, Rahmat Taufik R.L Bau⁶

¹ Informatic Engineering, Gorontalo State University, Gorontalo, Indonesia

Article Info

Article history:

Received June 09, 2018

Revised November 20, 2018

Accepted December 11, 2018

Keywords:

Two-Stage Machine Learning;

K-Means Clustering;

Support Vector Machine;

Flood Loss Prediction;

Flood Disaster Mitigation;

ABSTRACT

Floods are hydrological disasters that cause significant economic losses and require rapid and accurate damage estimates to support post-disaster decision-making. However, the process of assessing damage in the field by relevant agencies often takes months, thereby hindering recovery efforts. This study proposes a two-stage machine learning approach to improve the accuracy of floods loss predictions by integrating unsupervised and supervised learning methods. In the first stage, K-Means Clustering is used to group 294 flood events in Gorontalo Province based on rainfall intensity, number of victims, number of affected houses, and infrastructure damage. The cluster labels are then used as additional features in the second stage to train a Support Vector Machine (SVM) model to predict economic loss values. Model evaluation is performed using RMSE and R² metrics with an 80:20 data split. The results show that adding cluster features significantly improves model performance, with a 4.27% decrease in RMSE and a 3.43% increase in R² compared to the baseline model without them. These findings prove that clustering-based feature engineering can enrich data representation and overcome the heterogeneity of loss patterns that cannot be captured by basic features alone. This approach provides theoretical contributions to the development of predictive models for disaster losses and practical contributions to rapid loss estimation for flood management policies. Further research is recommended to expand hydrological features and evaluate alternative ensemble-based models to improve prediction quality.

Copyright © 2019 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Hermila A,

Computer Engineering,

Gorontalo State University,

Jl. Jend. Sudirman No.6, Dulalowo Tim., Kec. Kota Tengah, Kota Gorontalo, Gorontalo 96128

Email: hermila@ung.ac.id

1. INTRODUCTION

Floods are among the most frequent and destructive natural disasters worldwide, causing loss of life, property damage, and significant economic disruption. This is especially true in tropical countries such as Indonesia, where high rainfall, topographical factors, and land-use changes exacerbate the risk of flooding [1], [2], [3]. The province of Gorontalo is a prime example of this vulnerability, with 169 flood events between 2010 and 2019, affecting more than 1.2 million people and potentially causing hundreds of billions of rupiah in economic losses each year [4]. The protracted process of damage assessment and post-disaster fund

allocation, which can take up to 90 days according to the National Disaster Management Agency (BNPB), further exacerbates economic losses by delaying recovery [5]. This underscores the urgent need for rapid, accurate estimates of flood damage to enable targeted, efficient mitigation responses.

In this context, *machine learning* (ML) offers promising tools for predictive modeling. Among supervised learning algorithms, *the Support Vector Machine* (SVM) has demonstrated high effectiveness in disaster-related prediction tasks, achieving up to 95% accuracy in identifying flood-prone areas [6]. Simultaneously, unsupervised learning methods such as *K-Means Clustering* have been successfully applied to group areas with similar disaster risk characteristics [7]. However, conventional predictive models often treat all flood events uniformly, ignoring the inherent diversity in scale, impact, and triggering factors. The same approach to all of these can limit prediction accuracy, as a single model may not sufficiently capture the distinctive patterns of different types of floods.

Recent research indicates that a hybrid approach, integrating unsupervised and supervised learning, can improve predictive performance. Recent studies suggest that Clustering can reveal hidden patterns in data, thereby enriching the feature space for subsequent predictive models [8]. While previous studies have used SVM for flood hazard mapping or K-Means for region clustering separately, there is a gap in the literature on combining these methods specifically for flood loss prediction. Furthermore, most existing research focuses on predicting flood events or spatial vulnerability, with little emphasis on quantitative estimation of economic losses.

To address this gap, this study proposes a novel two-stage machine learning approach that integrates K-Means Clustering as a feature engineering technique to improve the Support Vector Machine model's accuracy in predicting flood losses in Gorontalo Province. The first stage involves clustering historical flood events into groups based on their characteristics (e.g., rainfall, casualties, damaged infrastructure). The resulting cluster labels are then incorporated as new contextual features in the second stage, where the SVM model is trained to predict economic losses. This methodology allows the model to first understand the "type" of flood event before estimating its financial impact, leading to more nuanced and accurate predictions. The main innovation of this research lies in the clustering-based feature engineering framework in this two-stage ML pipeline, which is designed explicitly for flood loss estimation, unlike conventional single-stage modeling. This approach is expected to provide a more robust tool for disaster management stakeholders, facilitating faster, more targeted mitigation and recovery planning.

2. RESEARCH METHOD

2.1. Research Stages

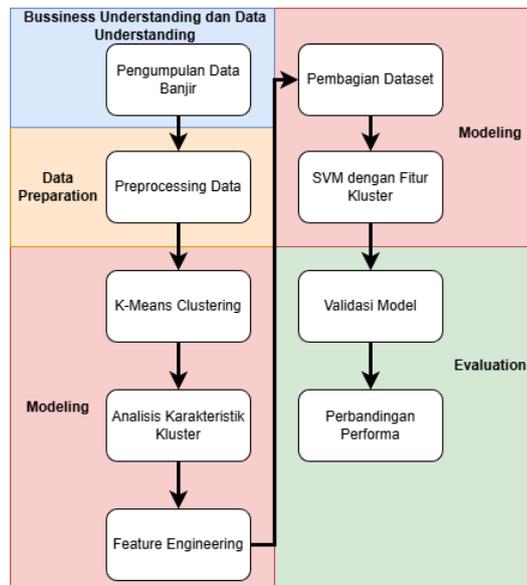


Figure 1. Research Stages

This study adopts the *Cross-Industry Standard Process for Data Mining* (CRISP-DM) methodology, modified to accommodate the *two-stage machine learning* approach [9]. This quantitative research design aims to develop a predictive model for flood loss estimation by integrating *unsupervised* and *supervised learning* techniques.

- a. Flood Data Collection: Data was obtained from four government agencies in Gorontalo Province: BPBD (flood event and impact data), BPS (demographic and socioeconomic data), BMKG (rainfall data), and the Social Services Agency (data on refugees and damage). The dataset covers 294 flood events from 2010 to 2024, including variables such as the number of victims, the number of flooded houses, infrastructure damage, rainfall, and potential economic losses.
- b. *Data Preprocessing*: preparing data for analysis. Data is cleaned by imputing missing values and normalizing each variable to a uniform scale using the *Standard Scaler*. Outliers are also identified and handled to avoid interference with the analysis results. For categorical variables such as the location of occurrence, one-hot encoding is used for transformation.
- c. K-Means Clustering (Stage 1): flood event data is grouped based on similar characteristics. First, the optimal number of clusters will be determined by looking at the *silhouette* value. After that, the K-Means algorithm is run using variables such as the number of victims, the number of flooded houses, and rainfall. As a result, each flood event is assigned a cluster label reflecting its similarity to other events.
- d. Cluster Characteristics Analysis: Descriptive statistics for each cluster will be examined to understand each cluster's profile.
- e. *Feature Engineering*: additional features will be prepared for the prediction model. The earlier cluster labels are used as new features. The most influential features are selected using Pearson correlation analysis. The result is a dataset enriched with additional information from the clustering process.
- f. Dataset Division: The data is split into two parts: training and testing. The division is carried out by considering the proportions of each cluster to ensure representativeness. As a baseline, 70% of the data is used to train the model, while the remaining 30% is used to test the model's performance. Cross-validation is also used to assess the model's reliability. The dataset will be split using several other ratios to find the ratio with the highest accuracy.
- g. A prediction model was developed using a Support Vector Machine with Cluster Features. The model will be trained on two datasets for comparison: the first dataset contains only basic features, while the second dataset adds cluster information as additional features. Model parameters are tuned to achieve optimal performance.
- h. Model Validation: The model is validated to measure its accuracy. The evaluation metrics used are RMSE () and R-squared to assess how closely the model's predictions match the actual values. Statistical tests are also used to ensure that differences in model performance are significant.
- i. Performance Comparison: The two developed models were compared. The comparison was based on evaluation metrics. The results of this comparison indicate whether adding cluster features improves prediction accuracy.
- j. Interpretation of research analysis results. At this stage, the author used AI tools to develop the presentation. The tools used included ChatGPT to help researchers find interesting synonyms and equivalents, and Grammarly to improve the structure of English sentences.

3. RESULTS AND DISCUSSION

3.1. Data Preprocessing Results

Two hundred ninety-four data entries were successfully collected, including the following variables: Date/Time of Event, Location of Event, Injuries/Illnesses, Victims, Evacuated, Severely Damaged Houses, Mildly Damaged Houses, Flooded Houses, Damaged Public Facilities, Damaged Bridges, Rainfall, and Potential Losses. The data preprocessing stage was successfully carried out on 294 flood event data. After the data cleaning process, which included handling missing values and outliers, 160 data sets were obtained that were ready for further analysis.

Table 1. Flood Data Summary

no	Date/Time of Event	Location of Event	Injuries/Illnesses	Victims	Evacuated	Severely Damaged Houses	Moderately Damaged Houses

1	January 19, 2017	Boalemo	0	0	3640	0	0
2	03/07/2024	Bone Bolango	0	8503	53	14	0
3	02/06/2017	Gorontalo	0	3877	200	1	0
4	July 24, 2020	Gorontalo City	0	32,511	4705	719	76
5	January 22, 2012	Pohuwato	0	3040	0	10	0

Slightly damaged houses	Flooded Houses	Damaged Public Facilities	Damaged Bridges	Rainfall (mm ³)	Potential Losses
12	728	0	0	304.85	Rp.298,844,544
26	1821	0	0	402	Rp2,056,441,938
3	720	0	0	293	Rp232,473,030
660	5248	0	0	366	IDR 89,482,320,714
5	0	0	0	142.43	Rp1,137,494,400

3.1.1 Data Normalization

Based on the figure, the three variables have been normalized to a scale of 0.0 to 1.0, centered at 0.0.

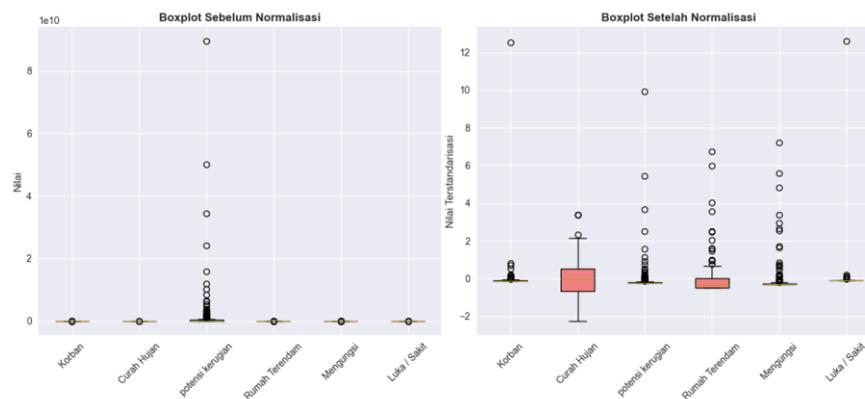


Figure 2. Comparison of Data Before and After Normalization

3.1.2 One-Hot Encoding

After *one-hot encoding*, the location variable was successfully converted from a categorical to a numerical variable, resulting in 17 additional columns.

3.2. Clustering Analysis (Stage 1)

3.2.1 Determining the Optimal Number of Clusters

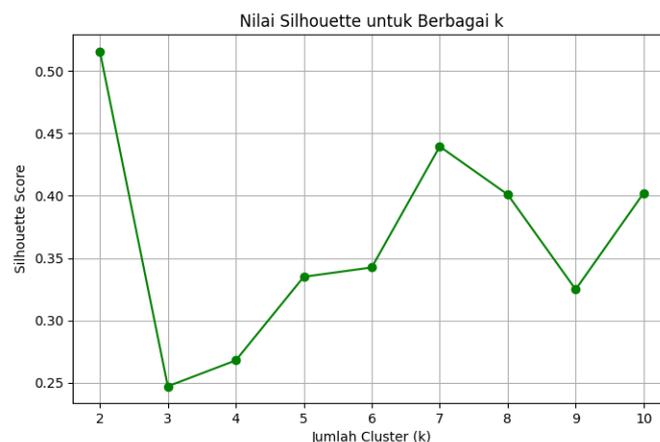


Figure 3. *Silhouette Score* for the optimal number of clusters k

As seen in the figure, the number of clusters with the highest silhouette value is k=2. A high *silhouette* value indicates that the data within a cluster are more similar to each other than those across clusters.

3.2.2 Characteristics of Each Cluster

From the results of determining the number of clusters using the silhouette score, the flood clusters will be divided into two, namely cluster zero and cluster 1, with cluster 0 labeled as "Moderate Flood" and cluster 1 labeled as "Extreme Flood," which have the following characteristics:

1. Cluster 0 – Moderate Flood: average rainfall < 300 mm³, average losses < 50 billion, average number of submerged houses < 4000 houses, and average number of victims < 6000 people.
2. Cluster 1 – Extreme Flood: average rainfall > 300 mm³, average losses > 60 billion, average number of flooded houses > 4000 houses, and average number of victims > 16,000 people.

3.3 Feature Engineering

3.3.1 Pearson Correlation Analysis



Figure 4. Heatmap of Pearson Correlation Analysis

The graph above shows the correlation between the variables. As seen in the image, the level of correlation ranges from white to red. The redder a variable is, the higher its correlation with other variables. After conducting a Pearson correlation test, it was found that the variables of severely damaged houses, slightly damaged houses, moderately damaged houses, damaged public facilities, and damaged bridges are correlated with the variable of potential losses.

3.3.2 Dataset Division

The model was created with options for selecting the ratio of training data and testing data, enabling more in-depth analysis.

3.4 SVM Model Development (Stage 2)

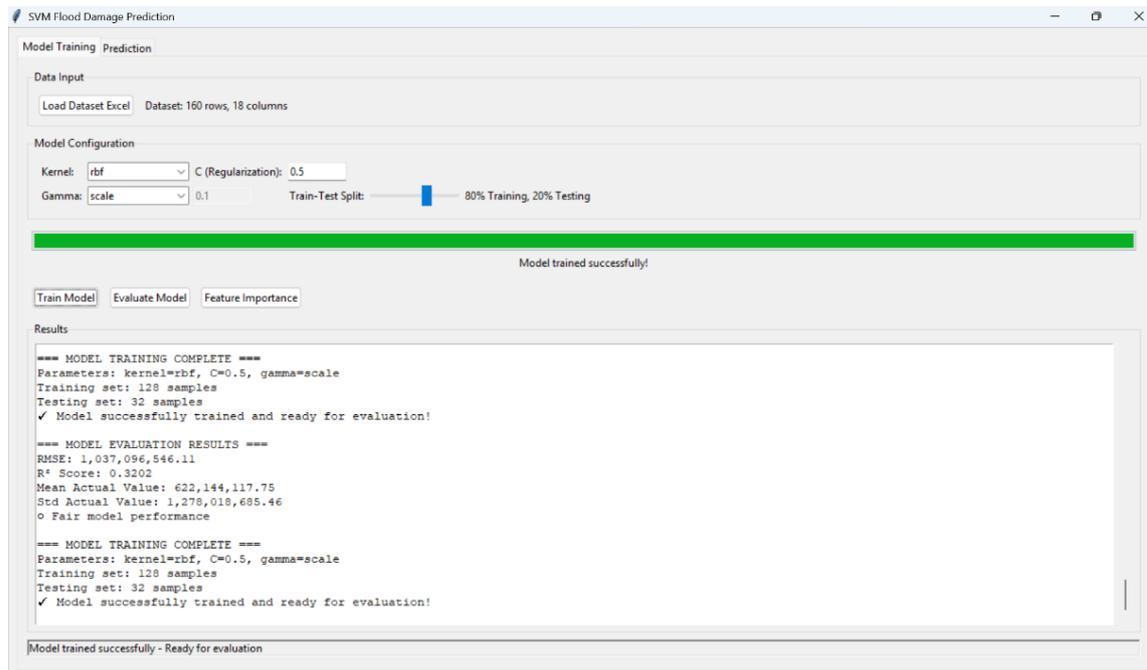


Figure 5. Application *Interface*

The model was created in Python using the Visual Studio Code editor. The model was designed with an interface to help users more easily predict losses. There are two pages, namely *Model Training* and *Prediction*.

On the *Model Training* and *Prediction* pages, there are several menus, namely:

- Data Input* is used to input the dataset to be processed. The input dataset is in Excel format (.xlsx).
- Model Configuration* used to tune the C and gamma parameters, as well as the ratio of *the dataset* to the *training data*.
- The train model* button is used to train data from the dataset that has been entered in the '*Data Input*' menu.
- The Evaluate Model* button is used to test validity using RMSE and R^2 .
- The Feature Importance* button displays a graph of the correlation between variables.
- Result* column, used to display the Model Evaluation results.
- The Prediction* page is used to predict flood damage values based on the input dataset.

3.5 Model Validation

Table 2. Model Performance Validation Table

Evaluation Metrics	Model 1 (Basic Features)	Model 2 (Basic Features + Clustering)
RSME	Rp 686,073,887	IDR 656,780,799
R^2	0.7091	0.7334

The model was tested using the same configuration. The configuration was selected based on the highest model performance, namely Kernel = Sigmoid, $C = 0.3$, Gamma = 0.1, and a dataset split ratio of 80:20. The results from the testing showed that Model 1 produced an RSME = Rp 686,073,887 and $R^2 = 0.7091$, while Model 2 produced an RSME = Rp 656,780,799 and $R^2 = 0.7334$.

3.6 Model Validation

Table 3. Model Performance Comparison Table

Evaluation Metrics	Model 1 (Basic Features)	Model 2 (Basic Features + Clustering)	Improvement
RSME	IDR 686,073,887	IDR 656,780,799	4.27
R ²	0.7091	0.7334	3.43

Based on the table above, performance improves between model 1, which uses only a dataset containing basic variables, and model 2, which is trained on a dataset that has been clustered beforehand using k-means.

3.7 Discussion of Finding

The findings of this study indicate that the two-stage approach namely clustering (K-Means) to map “flood regimes” and then making them additional features in the SVM model provides a significant improvement in performance. RMSE decreased by approximately 4.27% and R² increased by ±3.43% compared to models without cluster features. Conceptually, these results are consistent with the literature, which confirms that separating latent heterogeneity through clusters often makes the predictor-target relationship more “local” and easier for supervised models to learn, thereby increasing accuracy. Cluster-based two-stage studies in applied computing and logistics, for example, show that a good clustering stage can strengthen the performance of the subsequent prediction stage because the model works on a more homogeneous distribution [10], [11]. In the context of flood disasters, clustering has also proven useful for distinguishing areas/events with different causes and severity levels. Recent research utilizes K-Means and DBSCAN to cluster flood areas or episodes prior to further modeling, and reports improved risk/vulnerability mapping because different spatial-environmental structures are successfully captured by clusters. This is in line with the interpretation in this study that the two main clusters (“moderate flooding” vs. “extreme flooding”) are not only triggered by rainfall intensity, but also by exposure factors (number of houses/people affected), which in turn strongly influence the magnitude of losses [12], [13].

The relevance of exposure as a major driver of losses is also consistent with recent research over the past five years at the household and community levels, which shows that indicators such as the number of affected homes, building type, and density/area of exposure are important determinants in shaping the vulnerability curve and the magnitude of financial damage. Thus, the decision to include “exposure” features (flooded homes, victims, refugees) and add cluster labels as contextual features to SVM in this study is in line with current best practices [14], [15], [16].

The next finding on moderate performance improvement also indicates that there are still large errors at the “tail” of the distribution, which is common in heavy-tailed disaster loss data. Statistical and loss risk management literature suggests log transformation or tail-sensitive model families (e.g., approaches that pay attention to tail characterization) to stabilize variance and reduce the dominance of extreme events on square-based metrics such as RMSE. The practical implication is that, in further work, the evaluation can be augmented with more robust metrics (MAPE/sMAPE or quantile/pinball loss) and/or by modeling the target on a log scale so that the residuals are more balanced across clusters [17], [18], [19].

From a methodological perspective, the selection of SVM as the primary model is also in line with recent flood studies that found SVM to be competitive for flood status/risk classification/estimation, especially when patterns are non-linear and decision boundaries are complex. However, to make the claim of improvement due to the addition of cluster labels more convincing, best practices in the literature suggest verification through statistical testing of cross-fold paired errors (e.g., paired t-test/Wilcoxon) and rigorous verification of potential data leakage, namely ensuring that the cluster algorithm is only trained on the training data in each fold before being used to label the test data. These steps are commonly reported in modern two-stage studies so that performance improvements are not merely descriptive [20].

Substantively, the policy consequence of separating regimes through clusters is differentiated responses: in the “moderate flooding” cluster, the focus of intervention can be on protecting household assets and basic facilities; while in the “extreme flooding” cluster, the priority shifts to critical infrastructure, mass evacuation, and continuity of public services. Recent literature on ML-based flood risk quantification and the integration of socio-economic indicators also encourages the use of similar pipelines to support adaptive and contextual decision-making. By expanding hydro-spatial features (elevation/DEM, distance to rivers, inundation area) and testing alternative models (RF, XGBoost/LightGBM) while maintaining transparency through model explanation techniques, further studies have the potential to improve both performance and the clarity of policy recommendations [21].

Although the two-stage pipeline applied in this study provides improved predictive performance and offers a richer approach to understanding the heterogeneous structure of flood losses, these findings still have a number of important implications and limitations. In terms of implications, the results show that combining unsupervised and supervised techniques allows policymakers to identify two different risk regimes and design more targeted mitigation strategies based on the severity of events. This approach also opens up opportunities for the development of more adaptive decision support systems, especially when combined with spatial and real-time data such as satellite rainfall or up-to-date demographic information. However, this study is limited by the relatively small sample size after data cleaning, which may reduce the representativeness of temporal and spatial patterns. In addition, the heavy-tailed distribution of losses, the potential for land use changes over the data period, and the risk of distribution shift in future flood events may affect model generalization. Another limitation is the use of a single clustering method and a single kernel in SVM, so that exploration of alternative methods (e.g., Gaussian Mixture Models, hierarchical clustering, or RBF kernel) could enrich the understanding of underlying patterns and potential performance improvements. Thus, interpretation of the findings should consider these methodological limitations and encourage further research to expand the scope of features, increase sample size, and evaluate the pipeline in the context of more dynamic data.

4. CONCLUSION

This study proposes a two-stage machine learning approach by integrating K-Means Clustering as a feature engineering process to improve the accuracy of the Support Vector Machine model in predicting economic losses due to flooding in Gorontalo Province. Through the analysis of 294 historical flood events and a comprehensive preprocessing process, the developed two-stage model was able to identify two main clusters of flood events, namely clusters with moderate flood characteristics and extreme flood clusters. This cluster information proved to provide significant contextual value to the prediction model, as it captured differences in rainfall intensity patterns, number of victims, number of houses affected, and damage levels that were not optimally represented by basic features.

The evaluation results show that the addition of cluster features improves model performance, with a 4.27% decrease in RMSE and a 3.43% increase in R^2 compared to models without cluster features. These findings prove that the integration of unsupervised learning in the supervised learning process can enrich the data structure and improve the model's ability to recognize complex and heterogeneous flood damage patterns. This achievement also fills a gap in the literature, as most previous studies only utilized SVM or K-Means separately and focused more on flood-prone classification rather than quantitative estimation of economic losses. In practical terms, this approach can serve as the basis for developing a decision support system for BPBD and related agencies to accelerate the initial estimation of post-disaster damage, which currently takes up to 90 days. This model has the potential to help the government allocate resources more quickly and accurately, as well as support data-based mitigation planning in areas with different risk characteristics. In addition, the identification of “moderate flood” and “extreme flood” clusters provides a new basis for differentiating response strategies and risk management policies.

However, this study has several limitations, particularly in terms of the reduced amount of data after the cleaning process, the potential for distribution shifts due to changes in land cover and regional development, and the heavy-tailed nature of the loss distribution, which can affect the stability of predictions in extreme events. In addition, this study only uses one clustering method (K-Means) and one main kernel in SVM. Therefore, further research is recommended to explore alternative clustering methods such as Gaussian Mixture Models and hierarchical clustering, as well as to test other predictive models such as Random Forest, Gradient Boosting, and XGBoost. The addition of hydrological and spatial variables such as elevation, distance to rivers, and flood area is also recommended to improve the accuracy of future models.

ACKNOWLEDGMENTS

Thank you to everyone involved in this research who has completed the report and article. Then to Gorontalo State University, which has fully supported this research.

REFERENCES

- [1] B. P. S. P. Gorontalo, “Indikator Iklim Gorontalo, 2021,” 2021. [Online]. Available: <https://gorontalo.bps.go.id/id/statistics-table/2/OTgjMg==/indikator-iklim-gorontalo.html>
- [2] M. Ridwan and J. Sarjito, “Studi Kajian Dampak Perubahan Tutupan Lahan terhadap Kejadian Banjir di Daerah Aliran Sungai,” *ENVIRO: Journal of Tropical Environmental Research*, vol. 26, no. 1, p. 38, 2024, doi: 10.20961/enviro.v26i1.93145.

- [3] F. W. Indonesia, "Kerusakan Hutan Alam di Gorontalo Disebut Semakin Parah Dipicu Pembangunan HTE untuk Biomassa," 2024. [Online]. Available: <https://fwi.or.id/kerusakan-hutan-alam-di-gorontalo-disebut-semakin-parah/>
- [4] BNPB, "Penyusunan dokumen ini difasilitasi oleh," 2022.
- [5] G. Graha *et al.*, "BNPB BADAN NASIONAL PENANGGULANGAN BENCANA I (Sembilan) Dokumen oB Januari)21."
- [6] Y. Tae, F. Nurdiansyah, and F. Marisa, "Prediksi Daerah Rawan Banjir di Kabupaten Malaka menggunakan Algoritma Support Vector Machine," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 6, pp. 11360–11366, 2024, doi: <https://doi.org/10.36040/jati.v8i6.11454>.
- [7] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informasi*, vol. 12, 2022, doi: [10.34010/jati.v12i1](https://doi.org/10.34010/jati.v12i1).
- [8] Á. López-Oriona, Y. Sun, and J. A. Vilar, "Improving the prediction accuracy of statistical models: A new hierarchical clustering approach," *Stat Comput*, vol. 35, no. 6, 2025, doi: [10.1007/s11222-025-10683-x](https://doi.org/10.1007/s11222-025-10683-x).
- [9] P. Chapman *et al.*, *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler, 2000. [Online]. Available: <https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf>
- [10] M. Li, X. Bi, L. Wang, and X. Han, "A method of two-stage clustering learning based on improved DBSCAN and density peak algorithm," *Comput Commun*, vol. 167, pp. 75–84, Feb. 2021, doi: [10.1016/J.COMCOM.2020.12.019](https://doi.org/10.1016/J.COMCOM.2020.12.019).
- [11] J. Wu, X. Liu, Y. Li, L. Yang, W. Yuan, and Y. Ba, "A Two-Stage Model with an Improved Clustering Algorithm for a Distribution Center Location Problem under Uncertainty," *Mathematics 2022, Vol. 10, Page 2519*, vol. 10, no. 14, p. 2519, Jul. 2022, doi: [10.3390/MATH10142519](https://doi.org/10.3390/MATH10142519).
- [12] J. Li, A. Zheng, W. Guo, N. Bandyopadhyay, Y. Zhang, and Q. Wang, "Urban flood risk assessment based on DBSCAN and K-means clustering algorithm," *Geomatics, Natural Hazards and Risk*, vol. 14, no. 1, p. 2250527, Dec. 2023, doi: [10.1080/19475705.2023.2250527](https://doi.org/10.1080/19475705.2023.2250527);WGROU:STRING:PUBLICATION.
- [13] H. Wang, S. Xu, H. Xu, Z. Wu, T. Wang, and C. Ma, "Rapid prediction of urban flood based on disaster-breeding environment clustering and Bayesian optimized deep learning model in the coastal city," *Sustain Cities Soc*, vol. 99, p. 104898, Dec. 2023, doi: [10.1016/J.SCS.2023.104898](https://doi.org/10.1016/J.SCS.2023.104898).
- [14] D. Rodríguez Castro *et al.*, "Key Drivers of Flash Flood Damage to Private Households," in *Journal of Flood Risk Management*, John Wiley and Sons Inc, Sep. 2025. doi: [10.1111/jfr3.70088](https://doi.org/10.1111/jfr3.70088).
- [15] T. Endendijk, W. J. W. Botzen, H. de Moel, J. C. J. H. Aerts, K. Slager, and M. Kok, "Flood Vulnerability Models and Household Flood Damage Mitigation Measures: An Econometric Analysis of Survey Data," *Water Resour Res*, vol. 59, no. 8, p. e2022WR034192, Aug. 2023, doi: [10.1029/2022WR034192](https://doi.org/10.1029/2022WR034192);REQUESTEDJOURNAL:JOURNAL:19447973;WGROU:STRING:PUBLICATION.
- [16] M. A. Belenguer-Plomer, I. Mendes, M. Lazzarini, O. Barrilero, P. Saameño, and S. Albani, "Estimating Flood-Affected Houses as an SDG Indicator to Enhance the Flood Resilience of Sahel Communities Using Geospatial Data," *Remote Sensing 2025, Vol. 17, Page 2087*, vol. 17, no. 12, p. 2087, Jun. 2025, doi: [10.3390/RS17122087](https://doi.org/10.3390/RS17122087).
- [17] T. Yan, Y. Lu, and H. Jeong, "Dependence Modelling for Heavy-Tailed Multi-Peril Insurance Losses," *Risks 2024, Vol. 12, Page 97*, vol. 12, no. 6, p. 97, Jun. 2024, doi: [10.3390/RISKS12060097](https://doi.org/10.3390/RISKS12060097).
- [18] N. Gupta and S. R. Chavan, "A comprehensive decision support system for the characterization of probability distribution tails for daily precipitation," *J Hydrol (Amst)*, vol. 626, p. 130282, Nov. 2023, doi: [10.1016/J.JHYDROL.2023.130282](https://doi.org/10.1016/J.JHYDROL.2023.130282).
- [19] Z. Li, J. Beirlant, and S. Meng, "Generalizing the log-Moyal distribution and regression models for heavy tailed loss data," *ASTIN Bulletin*, vol. 51, no. 1, pp. 57–99, Dec. 2019, doi: [10.1017/asb.2020.35](https://doi.org/10.1017/asb.2020.35).
- [20] J. Akbar and M. A. Setyo Yudono, "Water Level Classification for Detect Flood Disaster Status Using KNN and SVM," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 13, no. 3, pp. 298–302, Nov. 2024, doi: [10.32736/sisfokom.v13i3.2166](https://doi.org/10.32736/sisfokom.v13i3.2166).
- [21] Y. Liu *et al.*, "Flood Risk Assessment Combining Machine Learning with Multi-criteria Decision Analysis in Jiangxi Province, China," *International Journal of Disaster Risk Science 2025 16:5*, vol. 16, no. 5, pp. 858–869, Oct. 2025, doi: [10.1007/S13753-025-00669-8](https://doi.org/10.1007/S13753-025-00669-8).

BIOGRAPHY OF AUTHORS

Short Version of the Title -- only one line long (First Author et al)

First author's Photo (3x4cm)	Xxxx (9 pt)
Second author's photo (3x4cm)	Xxxx (9 pt)
Third author's photo (3x4cm)	Xxxx (9 pt)