



Pengembangan Model Prediksi Kerugian Banjir di Provinsi Gorontalo Menggunakan Machine Learning Menggunakan Support Vector Regression

Mariyo W. Yani^{1*}, Rampi Yusuf², Hermila A³

^{1,2}Program Studi Sistem Informasi, Universitas Negeri Gorontalo, Indonesia

³Program Studi Pendidikan Teknologi Informasi, Universitas Negeri Gorontalo, Indonesia

E-mail: mariyowyani@gmail.com

Article Info

Article history:

Received February 05, 2026

Revised February 24, 2026

Accepted February 27, 2026

Keywords:

Flood Loss Prediction,
Machine Learning, K-Means
Clustering,
Support Vector Machine,
Disaster Mitigation.

ABSTRACT

Accurate prediction of economic losses due to flooding is a crucial aspect in supporting effective disaster mitigation planning in Gorontalo Province. This study proposes the use of Support Vector Regression (SVR). Historical flood data for the period 2010–2024 was obtained from relevant government agencies and used as the basis for analysis. In the first stage, the data was normalized using Standard Scaler and One-hot Encoding for categorical feature transformation. The second stage, namely the development of a loss prediction model using SVR. The model will be drilled using several scenarios to obtain the highest RMSE and R2 results. The test results show that the SVR model has an RMSE of Rp686,073,887 and an R² of 0.7091. These findings indicate that the SVR model is quite reliable in predicting flood losses using historical data. The resulting model has the potential to be an informative decision support tool for stakeholders in designing more targeted and data-driven flood mitigation and response strategies.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article Info

Article history:

Received February 05, 2026

Revised February 24, 2026

Accepted February 27, 2026

Keywords:

Prediksi Kerugian Banjir,
Support Vector Regression,
Bencana banjir,
Support Vector Machine,
Mitigasi Bencana.

ABSTRAK

Prediksi kerugian ekonomi akibat banjir secara akurat merupakan aspek krusial dalam mendukung perencanaan mitigasi bencana yang efektif di Provinsi Gorontalo. Penelitian ini mengusulkan Penggunaan Support Vector Regression (SVR). Data kejadian banjir historis periode 2010–2024 diperoleh dari instansi pemerintah terkait dan digunakan sebagai dasar analisis. Pada tahap pertama, Data dinormalisasi menggunakan Standard Scaler dan One-hot Encoding untuk tranformasi fitur kategorikal. selanjutnya pada tahap kedua, yaitu pembangunan model prediksi kerugian menggunakan SVR. Model akan dilatih menggunakan beberapa skenario untuk mendapat hasil RMSE dan R2 tertinggi. Hasil pengujian menunjukkan bahwa model SVR yang RMSE sebesar Rp686.073.887 dan R² sebesar 0,7091. Temuan ini mengindikasikan bahwa model SVR cukup andal dalam memprediksi kerugian akibat banjir menggunakan data-data historis. Model yang dihasilkan berpotensi menjadi alat pendukung keputusan yang informatif bagi para pemangku kepentingan dalam merancang strategi mitigasi dan respons banjir yang lebih terarah dan berbasis data.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



**Corresponding Author:**Mariyo W. Yani^{1*}

Program Studi Sistem Informasi, Universitas Negeri Gorontalo, Indonesia

E-mail: mariyowyani@gmail.com**PENDAHULUAN**

Banjir merupakan salah satu bencana hidrometeorologi yang memiliki frekuensi kejadian tinggi di wilayah tropis, termasuk Indonesia. Tingginya intensitas dan dampak yang ditimbulkan menjadikan pengelolaan risiko banjir sebagai isu strategis dalam pembangunan daerah. Di Provinsi Gorontalo, kejadian banjir menunjukkan tingkat kerentanan yang signifikan. Berdasarkan (Data Informasi Bencana Indonesia, 2024), selama periode 2010–2024 tercatat 294 peristiwa banjir yang berdampak pada 1.226.356 jiwa di seluruh kabupaten/kota. Kondisi tersebut dipengaruhi oleh curah hujan tahunan yang relatif tinggi, yakni sekitar 190 mm/tahun (Badan Pusat Statistik Provinsi Gorontalo, 2021), laju deforestasi mencapai 33.492 hektare pada periode 2017–2021 (Forest Watch Indonesia, 2024), serta perubahan dinamika pola aliran sungai (Laban et al., 2025). Alih fungsi lahan dari hutan menjadi kawasan terbangun dan pertanian berimplikasi pada peningkatan limpasan permukaan dan berkurangnya daya resap tanah (Oo & Humphries, 2026), sehingga memperbesar potensi dan intensitas banjir.

Selain menimbulkan kerusakan fisik pada infrastruktur dan permukiman, banjir juga berdampak pada aspek ekonomi secara luas. Estimasi nasional menunjukkan bahwa kerugian akibat banjir mencapai sekitar Rp30 triliun per tahun. Apabila dihitung secara proporsional berdasarkan jumlah penduduk, Provinsi Gorontalo diperkirakan mengalami potensi kerugian sebesar Rp400–600 miliar per tahun. Besaran kerugian ini sangat dipengaruhi oleh durasi proses pemulihan pascabencana. Mekanisme perencanaan rehabilitasi dan rekonstruksi melalui Kajian Kebutuhan Pasca-Bencana (JITUPASNA), yang menjadi dasar penyusunan Rencana Rehabilitasi dan Rekonstruksi (R3P), dilaksanakan setelah fase tanggap darurat berakhir (BNPB, 2025). Waktu pelaksanaannya bervariasi, berkisar antara 6 minggu hingga 3 tahun, sedangkan di Kota Gorontalo rata-rata membutuhkan satu minggu untuk setiap kecamatan. Keterlambatan dalam proses pemulihan berpotensi meningkatkan kerugian tidak langsung, termasuk terganggunya rantai pasok dan berhentinya aktivitas ekonomi masyarakat (Wei et al., 2023).

Fenomena tersebut menunjukkan adanya kebutuhan mendesak terhadap sistem prediktif yang mampu memberikan estimasi risiko dan kerugian secara lebih cepat dan akurat. Pendekatan konvensional yang bersifat reaktif belum sepenuhnya mampu menjawab tantangan kompleksitas bencana. Dalam konteks ini, perkembangan Machine Learning menawarkan alternatif solusi berbasis data. Algoritma Support Vector Machine (SVM) telah terbukti memiliki kemampuan prediktif yang baik dengan rata-rata akurasi sebesar 85%, bahkan mencapai 95% pada penggunaan kernel linier (Khan et al., 2024). Sementara itu, metode K-Means efektif dalam mengelompokkan wilayah rawan bencana ke dalam lima klaster dengan nilai silhouette coefficient sebesar 0.7483035806217097 (Sudrajat et al., 2025).

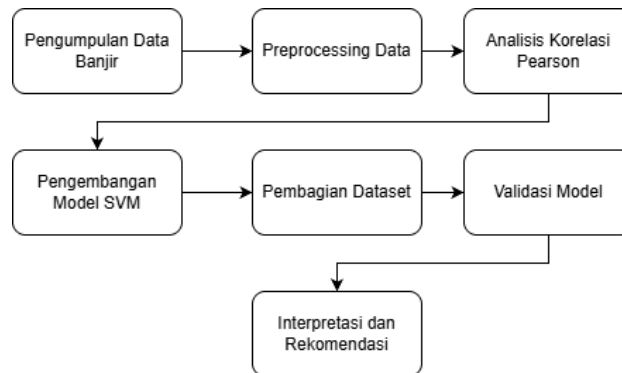
Berdasarkan latar belakang tersebut, penelitian ini mengembangkan model prediksi machine learning menggunakan Support Vector Regression dalam memprediksi kerugian banjir di Provinsi Gorontalo. Model dibangun menggunakan data rekaman bencana yang diperoleh dari BPBD, BPS, BMKG, dan Dinas Sosial Provinsi Gorontalo. Penelitian ini bertujuan menghasilkan model prediktif yang dapat mendukung pengambilan keputusan dalam mitigasi dan penanggulangan bencana secara lebih efektif, sekaligus berkontribusi pada percepatan proses pemulihan pascabencana melalui estimasi kerugian yang berbasis data dan terukur.



METODE

Penelitian ini dirancang sebagai studi kuantitatif dengan pendekatan analitik prediktif. Proses penelitian dilaksanakan selama empat bulan (Agustus–November) dengan pengumpulan data dari instansi pemerintah yang terkait dengan penanggulangan bencana banjir di Provinsi Gorontalo. Kerangka kerja analisis mengadaptasi model CRISP-DM yang dimodifikasi.

. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Pengembangan Model Machine Learning

Sumber: Rancangan Peneliti, 2025

Populasi penelitian mencakup seluruh data kejadian banjir yang tercatat di Provinsi Gorontalo, dengan unit analisis berupa rekaman peristiwa banjir. Data primer diperoleh dari BPBD, BPS, BMKG, dan Dinas Sosial Provinsi Gorontalo dalam bentuk dokumen resmi dan arsip historis, sedangkan data sekunder bersumber dari INARISK, DIBI, BPS, dan BMKG. Pengumpulan data dilakukan melalui studi dokumentasi dan telaah pustaka terhadap data yang tersedia.

Tahap awal dilakukan *data preprocessing* melalui imputasi nilai hilang, normalisasi menggunakan *Standard Scaler*, penanganan *outlier*, serta transformasi variabel kategorikal dengan *one-hot encoding*.

Seleksi variabel dilakukan menggunakan uji korelasi Pearson. Dataset dibagi menjadi data pelatihan dan pengujian dengan rasio awal 70:30 serta diuji pada beberapa skenario pembagian, disertai penerapan *cross-validation* untuk meningkatkan keandalan model. Pemodelan dilakukan menggunakan Support Vector Machine (SVR). Parameter model dioptimalkan melalui *hyperparameter tuning*. Evaluasi kinerja dilakukan menggunakan RMSE dan *R-squared*, serta diuji signifikansinya secara statistik.

HASIL PENELITIAN

Sebanyak 294 entri data kejadian banjir berhasil dikumpulkan yang mencakup variabel tanggal/waktu kejadian, lokasi, jumlah korban luka/sakit, korban terdampak, jumlah pengungsi, rumah rusak berat, rusak sedang, rusak ringan, rumah terendam, fasilitas umum rusak, jembatan rusak, curah hujan, serta potensi kerugian. Setelah dilakukan tahapan pembersihan data (*data cleaning*) dan penanganan *missing values*, diperoleh 160 data yang layak digunakan untuk analisis lanjutan.

Tahap praproses dilanjutkan dengan normalisasi menggunakan metode *StandardScaler*, yaitu teknik standarisasi (normalisasi Z-score) yang mentransformasi setiap fitur sehingga memiliki rata-rata (μ) nol dan deviasi standar (σ) satu (Pinheiro et al., 2025). Transformasi ini bertujuan untuk mengatasi perbedaan skala antarvariabel, mengingat variabel seperti korban dan potensi kerugian memiliki rentang nilai yang jauh lebih besar dibandingkan variabel lain. Visualisasi melalui boxplot menunjukkan bahwa distribusi data



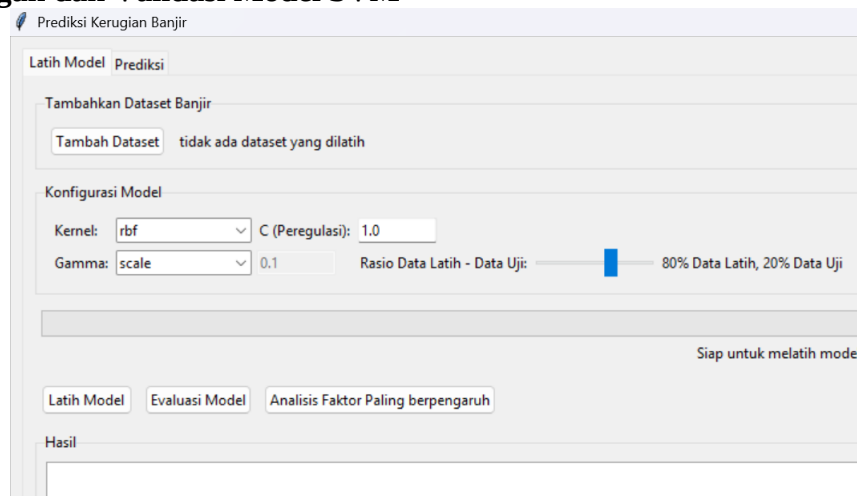
sebelum normalisasi memiliki rentang yang sangat lebar, terutama pada variabel korban, curah hujan, dan potensi kerugian. Setelah normalisasi, seluruh variabel terpusat di sekitar nol dengan sebaran yang lebih seragam. Secara metodologis, penggunaan boxplot untuk mengevaluasi distribusi data didasarkan pada konsep *five-number summary* yang mencakup nilai minimum, kuartil pertama, median, kuartil ketiga, dan maksimum (BABURA et al., 2025). Dengan demikian, tahapan ini memastikan bahwa proses pembelajaran mesin tidak didominasi oleh variabel dengan skala besar.

Selanjutnya dilakukan *one-hot encoding* (OHE) untuk mentransformasikan variabel kategorikal “Tempat Kejadian” menjadi representasi numerik. Proses ini menghasilkan penambahan jumlah variabel dari 12 menjadi 17 kolom. OHE mengubah setiap kategori unik menjadi variabel biner independen sehingga menghilangkan asumsi ordinal semu yang dapat mengganggu algoritma berbasis jarak, termasuk Support Vector Machine (SVM). Penerapan OHE krusial dalam meningkatkan akurasi dan reliabilitas model karena setiap kategori diperlakukan sebagai entitas independen (Abdulsalam & Güney, 2025).

Analisis Korelasi Pearson

Analisis korelasi Pearson dilakukan untuk mengidentifikasi faktor yang memengaruhi potensi kerugian. Korelasi Pearson mengukur kekuatan hubungan linier antarvariabel dengan rentang nilai -1 hingga +1 (Dufera et al., 2023). Hasil analisis menunjukkan bahwa variabel rumah rusak berat, rusak ringan, rusak sedang, jumlah pengungsi, fasilitas umum rusak, dan jembatan rusak memiliki korelasi positif terhadap potensi kerugian. Temuan ini menegaskan bahwa tingkat kerusakan infrastruktur dan besarnya paparan populasi merupakan determinan utama dalam pembentukan nilai kerugian.

Pengembangan dan Validasi Model SVM



Gambar 2. Pengembangan dan Validasi Model SVM

Model dikembangkan menggunakan kernel Sigmoid dengan parameter gamma = 0,1 dan C = 0,3. Evaluasi dilakukan menggunakan Root Mean Square Error (RMSE) dan koefisien determinasi (R^2). RMSE mengukur rata-rata kesalahan prediksi dalam satuan yang sama dengan variabel target, sehingga semakin kecil nilainya semakin baik (Chicco et al., 2021). Sementara itu, R^2 menunjukkan proporsi variasi target yang mampu dijelaskan model (Gao, 2024).

Model SVM menghasilkan RMSE sebesar Rp686.073.887 dan R^2 sebesar 0,7091. Secara konseptual, akurasi tinggi mengindikasikan bahwa model berhasil mempelajari pola data dan mampu melakukan generalisasi pada data uji (Chicco et al., 2021).



PEMBAHASAN

Relevansi indikator paparan seperti jumlah rumah terendam dan korban terdampak sebagai determinan utama kerugian juga konsisten dengan studi terkini yang menekankan pentingnya variabel sosial-ekonomi dan kepadatan paparan dalam pembentukan kurva kerentanan (Asl et al., 2025).

Meskipun terjadi peningkatan performa, nilai RMSE yang masih relatif besar menunjukkan adanya karakteristik *heavy-tailed distribution* pada data kerugian bencana. Literatur manajemen risiko menyarankan transformasi logaritmik atau penggunaan metrik robust seperti MAPE atau *quantile loss* untuk meningkatkan stabilitas estimasi (Nourali & Shortridge, 2026). Selain itu, praktik terbaik metodologis menyarankan validasi lintas lipatan serta pengujian statistik kesalahan berpasangan guna memastikan peningkatan performa bukan akibat kebocoran data (*data leakage*) (Rainio et al., 2024).

Integrasi fitur hidrospasial tambahan serta eksplorasi model alternatif seperti Random Forest atau XGBoost/LightGBM berpotensi meningkatkan performa sekaligus memperkaya rekomendasi kebijakan (Suaif & Sylvianti Rahayu, 2025).

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa penggunaan model SVM efektif untuk prediksi kerugian banjir, sekaligus memberikan pemahaman struktural mengenai tipologi kejadian banjir di Provinsi Gorontalo.

KESIMPULAN

Penelitian ini bertujuan mengembangkan model prediksi kerugian akibat banjir algoritma SVR machine learning. Berdasarkan hasil analisis, model SVR mampu memprediksi kerugian akibat banjir menggunakan data-data historis banjir dan variabel terkait banjir. Penelitian menunjukkan kemampuan model dalam mengenali hubungan non-linier antara variabel penyebab dan dampak ekonomi yang ditimbulkan. Dengan demikian, penelitian ini memberikan kontribusi metodologis berupa penguatan pendekatan dua tahap sebagai strategi feature engineering yang efektif dalam pemodelan risiko bencana. Secara praktis, model yang dikembangkan berpotensi mendukung sistem mitigasi bencana berbasis data melalui estimasi kerugian yang lebih kontekstual dan adaptif terhadap karakteristik kejadian. Meskipun demikian, penelitian ini memiliki keterbatasan pada ruang lingkup wilayah yang terbatas pada Provinsi Gorontalo serta penggunaan variabel yang masih didominasi indikator dampak langsung. Oleh karena itu, pengembangan model lanjutan disarankan untuk menguji algoritma alternatif seperti Random Forest, Gradient Boosting, atau XGBoost guna membandingkan stabilitas dan performa prediksi. Selain itu, penambahan variabel hidrologi dan spasial—seperti elevasi, jarak terhadap aliran sungai, serta kepadatan penduduk—dapat memperkaya representasi risiko sehingga meningkatkan sensitivitas model terhadap faktor kerentanan wilayah. Validasi temporal dan spasial juga perlu dilakukan dengan menguji model pada periode waktu dan wilayah berbeda guna memastikan kemampuan generalisasi dan robustitas prediksi. Secara operasional, integrasi model ke dalam sistem peringatan dini berbasis web-GIS direkomendasikan agar dapat dimanfaatkan oleh pemerintah daerah dan pemangku kepentingan dalam perencanaan mitigasi, pengalokasian sumber daya, serta perumusan kebijakan penanggulangan bencana yang lebih responsif dan berbasis bukti.

DAFTAR PUSTAKA

Abdulsalam, H. A., & Güney, H. (2025). Investigating the Impact of One-Hot Encoding on Classification Performance and Time Complexity for Efficient Network Intrusion Detection. *Hittite Journal of Science and Engineering*, 12(4), 197–207. <https://doi.org/10.17350/HJSE19030000366>



- Asl, S. R., Rahman, A., Tate, E., Lehman, W., & Wing, O. (2025). Social vulnerability correlates of flood risk to crops and buildings. *Natural Hazards*, 121(7), 8137–8158. <https://doi.org/10.1007/s11069-025-07137-y>
- BABURA, B. I., Bulama, L. M., & Umar, D. A. (2025). Box-K-Means symbolic data analysis: A k-means algorithm for boxplot variables with application to global climate Variability. *Journal of Statistical Sciences and Computational Intelligence*, 1(3), 157–165. <https://doi.org/10.64497/jssci.12>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Dufera, A. G., Liu, T., & Xu, J. (2023). Regression models of Pearson correlation coefficient. *Statistical Theory and Related Fields*, 7(2), 97–106. <https://doi.org/10.1080/24754269.2023.2164970>
- Gao, J. (2024). R-Squared (R^2) – How much variation is explained? *Research Methods in Medicine & Health Sciences*, 5(4), 104–109. <https://doi.org/10.1177/26320843231186398>
- Khan, M., Hooda, B. K., Gaur, A., Singh, V., Jindal, Y., Tanwar, H., Sharma, S., Sheoran, S., Vishwakarma, D. K., Khalid, M., Albakri, G. S., Alreshidi, M. A., Choi, J. R., & Yadav, K. K. (2024). Ensemble and optimization algorithm in support vector machines for classification of wheat genotypes. *Scientific Reports*, 14(1), 22728. <https://doi.org/10.1038/s41598-024-72056-0>
- Laban, S., Hazairin, H., Nugraha, F., & Fajeriana, N. (2025). Pemodelan Hidrograf untuk Menilai Pengaruh Perubahan Penggunaan Lahan Terhadap Karakteristik Aliran Sungai Di Sub-DAS Tanralili, Sulawesi Selatan. *Jurnal Penelitian Pengelolaan Daerah Aliran Sungai*, 1. <https://doi.org/10.59465/jppdas.2025.9.1.1-22>
- Nourali, Z., & Shortridge, J. E. (2026). Predicting the impact of extreme weather on agricultural losses in the Delmarva peninsula using multi-step machine learning and financial crop loss data. *Stochastic Environmental Research and Risk Assessment*, 40(1), 13. <https://doi.org/10.1007/s00477-025-03157-z>
- Oo, T. Z., & Humphries, U. W. (2026). Hydrological modeling of flood impacts under land use and land cover change: A systematic review of tools, trends, and challenges. *MethodsX*, 16, 103724. <https://doi.org/10.1016/j.mex.2025.103724>
- Pinheiro, J. M. H., Oliveira, S. V. B. de, Silva, T. H. S., Saraiva, P. A. R., Souza, E. F. de, Godoy, R. V., Ambrosio, L. A., & Becker, M. (2025). The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks. *IEEE Access*, 13, 199903–199931. <https://doi.org/10.1109/ACCESS.2025.3635541>
- Rainio, O., Teuvo, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), 6086. <https://doi.org/10.1038/s41598-024-56706-x>
- Suaif, A., & Sylvianti Rahayu, E. (2025). Analisis Faktor Dan Pola Kejadian Banjir Di Bandar Lampung Menggunakan Arima, Random Forest, Dan Xgboost. *Jurnal Teknologi Komputer Dan Informatika*, 3(2). <https://doi.org/10.59820/tekomin.v3i2.339>
- Sudrajat, R., Hadiana, A. I., & Melina, M. (2025). Evaluasi Kualitas Kluster Wilayah Rawan Bencana Menggunakan K-Means dengan Silhouette dan Elbow Method. *Jurnal Algoritma*, 22(2), 127–139. <https://doi.org/10.33364/algoritma/v.22-2.2379>
- Wei, S., Zhou, Q., Luo, Z., She, Y., Wang, Q., Chen, J., Qu, S., & Wei, Y. (2023). Economic impacts of multiple natural disasters and agricultural adaptation measures on supply chains in China. *Journal of Cleaner Production*, 418, 138095. <https://doi.org/10.1016/j.jclepro.2023.138095>