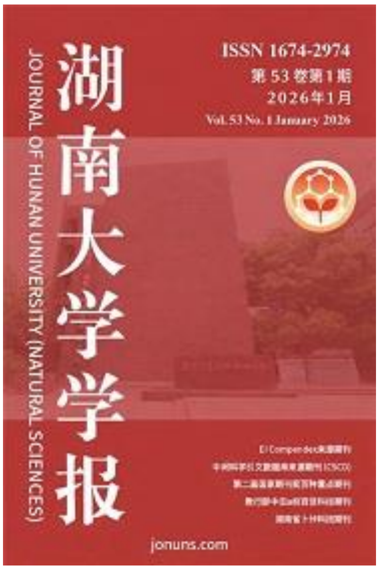
湖南大学学报

JOURNAL OF HUNAN UNIVERSITY (NATURAL SCIENCES)

ISSN 1674-2974

第 53 卷第 1 期

2026年1月

Vol. 53 No. 1 January 2026

jonuns.com

Journal of Hunan University
(Natural Sciences)

Vol. 53 No. 1
January 2026

Available online at
https://jonuns.com

Open Access Article https://doi.org/10.55463/issn.1674-2974.53.1.6

Geospatial Flood Risk Mapping Using a Probabilistic Naive Bayes Model

Devni Prima Sari [1*], Riry Sriningsih [1], Meira Parma Dewi [1], Azhari Syarief [2], Dina Agustina [3], Yuni Wardana[1]

[1]Department of Mathematics, Universitas Negeri Padang, Indonesia,

[2]Department of Geography, Universitas Negeri Padang, Indonesia,

[3]Data Analytics, Mathematical Modelling, and Forecasting Research Group, Universitas Negeri Padang, Indonesia,

* Corresponding author: devniprimasari@fmipa.unp.ac.id

Article History:

Received: December 29, 2025
Revised: January 26, 2026
Accepted: February 11, 2026
Published: February 27, 2026

Abstract: Flood events occur recurrently in Padang City, Indonesia, generating substantial social, economic, and environmental consequences. This study aims to develop a geospatial flood vulnerability map using a probabilistic Naïve Bayes model integrated with GIS-based spatial analysis in ArcGIS. The model incorporates six key conditioning factors: rainfall, slope, soil type, landform, geology, and land use.

The Naïve Bayes classifier achieved an overall accuracy of 97.69%, indicating high predictive capability and model reliability. The resulting vulnerability map categorizes the study area into three classes—low, moderate, and high vulnerability. High-vulnerability zones are predominantly concentrated in the western part of Padang City, primarily due to low-lying topography, upstream surface runoff accumulation, and tidal influences.

This study presents a statistically grounded and computationally efficient framework that integrates probabilistic machine learning with spatial analysis for urban-scale flood vulnerability assessment. Compared to conventional deterministic approaches, the proposed method offers improved adaptability, rapid processing, and strong predictive performance. The framework provides valuable decision-support tools for flood risk mitigation, urban planning, and sustainable land management and can be applied to other flood-prone regions with comparable environmental characteristics.

Keywords: Naïve Bayes classifier; flood vulnerability mapping; GIS-based spatial analysis; probabilistic modeling; urban flood risk; disaster mitigation.

基于概率朴素贝叶斯模型的地理空间洪水风险制图

摘要：洪水事件在印度尼西亚巴东市反复发生，造成显著的社会、经济和环境影响。本研究旨在结合基于 ArcGIS 的地理信息系统（GIS）空间分析方法，构建一种基于概率朴素贝叶斯模型的地理空间洪水脆弱性制图方法。模型综合考虑六个关键影响因子：降雨量、坡度、土壤类型、地貌、地质条件和土地利用类型。

朴素贝叶斯分类器的总体预测精度达到 97.69%，表明该模型具有较高的预测能力和良好的可靠性。研究结果将研究区域划分为低、中和高三类脆弱性等级。其中，高脆弱性区域主要集中在巴东市西部地区，这主要归因于低洼地形、上游地表径流汇集以及潮汐影响等因素。

本研究提出了一种在统计学基础上构建且计算效率较高的综合框架，将概率机器学习方法与空间分析技术相结合，用于城市尺度的洪水脆弱性评估。与传统确定性方法相比，该方法具有更强的适应性、更快的处理效率和更优的预测性能。该框架可为洪水风险缓解、城市规划与可持续土地管理提供有效的决策支持工具，并可推广应用于具有类似环境特征的其他洪水易发地区。

关键词：朴素贝叶斯分类器；洪水脆弱性制图；基于GIS的空间分析；概率建模；城市洪水风险；灾害缓解

1. Introduction

Floods are among the most common types of natural disasters in Indonesia, and their impacts often result in significant economic, social, and environmental losses. Padang City, one of the areas vulnerable to this disaster, faces serious challenges due to the region's topography, high rainfall, and increasingly rapid urbanization. In addition, poor drainage management and uncontrolled land use worsen the situation, increasing the risk of flooding in densely populated and low-lying areas. In this context, disaster mitigation is a top priority that must be designed strategically and data-driven to reduce the impact of flooding on the community.

Although significant improvements have been made in computing power and distributed hydrological modeling, integrating uncertainty into flood forecasting remains challenging and up-to-date. In recent years, various statistical methods have been developed and effectively applied in flood vulnerability mapping. Currently, Machine Learning (ML) or Artificial Intelligence methods, which are advanced [1] soft-computing approaches for natural disaster prediction and assessment, are widely used in flood studies. These methods are based on effective and objective mathematical algorithms for analysis and prediction [2]. Some popular ML methods used for flood vulnerability assessment include Artificial Neural Networks [3], Logistic Model Trees [4], Support Vector Machines, Logistic Regression [5], Adaptive Neuro-Fuzzy Inference Systems [6], and Neural-Fuzzy approaches [7].

The main challenge in flood vulnerability assessment and mapping is the lack of a universal model that can be applied in all areas with a consistent level of accuracy. Each area has unique characteristics that affect flood patterns, such as topography, hydrology, land use, and socio-economic factors. Therefore, an approach sensitive to local context is needed to produce accurate and relevant flood vulnerability mapping. Integration of spatial data with appropriate analysis techniques is key to identifying critical factors and developing models that can be adjusted to various regional conditions. Thus, there is a need for ongoing research to explore the possibility of model selection for accurate identification and mapping of flood-prone areas. With this aim, we will conduct research with the Naïve Bayes model. This model is applied in the city of Padang, which is the capital of West Sumatra Province, Indonesia.

Probabilistic approaches such as Naive Bayes have been widely used in environmental risk analysis to overcome these limitations. This approach has the advantage of handling complex, uncertain data using the principles of probability [8]. Naive Bayes, for example, provides a simple, fast method for generating predictions from historical data. This method assumes independence among variables, meaning that each factor is not assumed to influence the others [9], [10].

This study aims to analyze flood risk in Padang City using the Naive Bayes algorithm as a probabilistic approach. The main focus is to produce a classification of flood hazard levels based on key parameters, including soil type, geological type, landform type, slope class, land use type, and rainfall. By using the Naive Bayes algorithm, this study is expected to produce an accurate and relevant flood hazard-level zoning map to support more effective, data-driven flood risk mitigation planning.

In addition, this research is expected to address technical challenges in probabilistic-based flood risk mapping, such as integrating environmental and social data into the model and ensuring that modeling results can be translated into accurate mitigation measures. With a systematic approach, this research is expected to provide not only theoretical contributions in the development of risk mapping models but also practical benefits for disaster risk reduction efforts, especially in flood-prone areas in Indonesia.

2. Research Methods

This study applied a probabilistic Naive Bayes model integrated with ArcGIS-based spatial analysis to assess flood vulnerability in Padang City. The research workflow, shown in Figure 1, was designed to ensure methodological transparency and reproducibility.

The first stage is a literature review, which involves examining theories, methods, and prior research on flood vulnerability analysis and the application of the Naive Bayes algorithm. In this stage, key parameters such as rainfall, slope, soil type, landform, geology, and land use are identified as important factors in flood risk analysis. The second stage is data collection, which involves primary and secondary data. Primary data are collected through interviews with experts to identify the main risk factors for flooding. Meanwhile, secondary data are collected from thematic maps, including rainfall, slope, soil type, and land use. The third stage is data processing, during which all spatial data is converted to a format suitable for analysis in ArcGIS. Continuous data is discretized to facilitate probabilistic analysis [11].

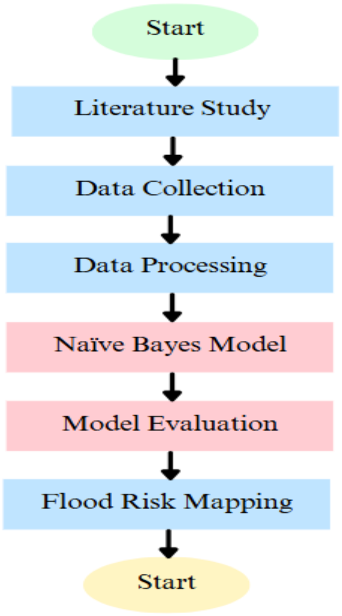

Figure 1. Essential connections

The fourth stage is the creation of a Naive Bayes model, which includes the calculation of initial probability (prior), likelihood, and posterior probability to determine the level of flood hazard. This model works with the principle of simple probability for risk classification. The fifth stage is model evaluation, in which the model's predictions are compared with historical data to assess model accuracy. This stage ensures the validity of the model’s results. The last stage is flood vulnerability mapping, which integrates the Naive Bayes results into a spatial map. This map provides low, medium, and high-risk zoning, assisting stakeholders in flood mitigation planning. With these stages, this study produces a comprehensive and applicable flood risk analysis to support risk mitigation in Padang.

3. Naive Bayes

Naive Bayes is a probability-based algorithm often used in data analysis and classification. This algorithm is based on the assumption of independence between variables, meaning that each variable is considered to have no influence on the others in determining the outcome [12]. Although this assumption is rarely fully met in real-world data, its simplicity and efficiency make Naive Bayes a popular method in various applications [13].

The main advantages of Naive Bayes are its ease of implementation and high computational speed, making it well-suited to large datasets. In addition, this algorithm can handle both categorical and continuous variables with a simple conversion process [14]. Naive Bayes is often used for simple classification tasks such as spam detection [15], [16], sentiment analysis [17], [18], or early risk prediction on less complex datasets [19], [20]. In the context of flood vulnerability analysis, Naive Bayes can be an efficient tool to provide fast results on simple datasets.

Table 1. Variable description

Variables	Class	Class Score	Probability
Soil Type (X_1)	Red Yellow Pedsolik Comp	1	0.198150
	Alluvial	2	0.245109
	Regosol	3	0.016009
	Organosol	4	0.000711
	Latosol	5	0.466382
	Andosol	6	0.073639
Geological Types(X_2)	Painan Formation	1	0.141942
	Alluvium	2	0.265386
	Intrusive Rocks	3	0.036286
	Volcanic Rocks	4	0.456777
	Limestone	5	0.016720
	Metamorphic Rocks	6	0.023479
	Alluvial Fan	7	0.059409

Variables	**Class**	**Class Score**	**Probability**
Landform Types(X_3)	Volk Mountains Complex	1	0.662398
	Sand Dunes	2	0.015653
	Coastal Alluvial Data	3	0.013518
	Back Swamp	4	0.060121
	Flood Plain	5	0.009961
	Natural Dike	6	0.036642
	Volcanic Hills	7	0.009961
	Chalk Hills	8	0.014586
	Alluvial Fan	9	0.026681
	Fluvio Vulcan Fan	10	0.088936
	Human Change	11	0.011384
	River Burn	12	0.005336
	Inter-Shore Depression	13	0.008538
	Gisik Strait	14	0.009605
	Piroska Flow Terrace	15	0.026681
Slope Class(X_4)	<15	1	0.294913
	15 – 40	2	0.191747
	> 40	3	0.513340
Type of Land Use(X_5)	Bush / Alang Alang	1	0.058698
	Forest	2	0.644255
	Mangrove forest	3	0.001067
	Irrigated Rice Fields	4	0.090004
	Plantation / Garden	5	0.035575
	Housing area	6	0.094984
	Steam Power Plant	7	0.000711
	Empty / Bare Land	8	0.008894
	Harbor	9	0.000711
	River	10	0.006048
	Fields / Fields	11	0.016720
	Industry	12	0.005336
	Trade and Services	13	0.012451
	Mining	14	0.003202
	Public Service Facilities	15	0.007115
	Sports Facilities	16	0.001423
	Office	17	0.002846
	Farm	18	0.000711
	Sand / Land Dunes	19	0.000356
	Rainfed Rice Fields	20	0.004269
	Defense and Security	21	0.003202
	Public Cemetery	22	0.000356
	Swamp Forest	23	0.001067
Rainfall (X_6)	2500 – 3000	1	0.001070
	3000 – 3500	2	0.027390
	3500 – 4000	3	0.331550
	4000 – 4500	4	0.146210
	4500 – 5000	5	0.212020
	> 5000	6	0.281750
Flood Hazard Class(C)	Low	1	0.678410
	Currently	2	0.270720
	Tall	3	0.050870

The principle behind Naive Bayes is Bayes' theorem, which is also known as Bayes' Rule [21]. Bayes' theorem is used to calculate conditional probability, which is the probability of an event occurring given information about past events. Overall, Naive Bayes is an ideal algorithm for initial analysis or situations where the data is relatively simple and does not have many variable interactions. However, for more complex data with significant relationships between variables, approaches such as Bayesian Belief Networks [22] are more appropriate, as they can account for dependencies between variables. Thus, Naive Bayes can serve as a quick and efficient initial method before moving on to more complex models for deeper analysis.

In real-world applications, there are multiple predictor variables, and in classification problems, there are multiple output classes. The classes can be represented as, $\mathrm{C}_1, \mathrm{C}_2, \dots, \mathrm{C}_\mathrm{n}$ and the predictor variables can be represented as vectors, $x_1, x_2, \dots, x_n$ [23]. The goal of the Naive Bayes algorithm is to measure the conditional probability of an event with a feature vector $x_1, x_2, \dots, x_n$ belonging to a certain class C_i,

$$P(C_i|x_1,x_2,\dots,x_n)=\frac{P(x_1,x_2,\dots,x_n|C_i).P(C_i)}{P(x_1,x_2,\dots,x_n)} \text{ for } 1<i<k \qquad (1)$$

By calculating the above equation, we get:

$$P(x_1,x_2,\dots,x_n|C_i).P(C_i)=P(x_1,x_2,\dots,x_n,C_i)$$

$$\begin{aligned} P(x_1,x_2,\dots,x_n,C_i) &= P(x_1|x_2,\dots,x_n,C_i).\,P(x_2,\dots,x_n,C_i) \\ &= P(x_1|x_2,\dots,x_n,C_i).P(x_2|x_3,\dots,x_n,C_i).P(x_3,\dots,x_n,C_i) \\ &\ \vdots \\ &= P(x_1|x_2,\dots,x_n,C_i).P(x_2|x_3,\dots,x_n,C_i)\dots P(x_{n-1}|x_n,C_i).P(x_n|C_i).P(C_i) \end{aligned}$$

However, the conditional probabilities, which are $P\left(x_j|x_{j+1},\dots,x_n,C_i\right)$ summed to be $P\left(x_j|C_i\right)$ because each predictor variable is independent in Naive Bayes.

The equation becomes:

$$P(x_1, x_2, \ldots, x_n, C_i)$$
$$= P(x_1|C_i).P(x_2|C_i)\ldots P(x_{n-1}|C_i).P(x_n|C_i).P(C_i)$$
$$P(x_1, x_2, \ldots, x_n, C_i) = \left[\prod_{j=1}^{n} P(x_j|C_i)\right].P(C_i)$$

for $1 < i < k$

4. Results and Discussion

4.1. Description of Flood Risk Determining Variables

The data used in this study are from interviews, the Indonesian Agency for Meteorological, Climatological and Geophysics (Badan Meteorologi, Klimatologi, dan Geofisika, or simply BMKG), and previous studies related to flood hazards. Details of each variable are shown in Table 1. The variables in Table 1 represent values for Padang City divided into 2811 grids, with 80% of the data designated as training data and 20% as testing data.

4.2. Classification with the Naïve Bayes Method

The Naïve Bayes classification process in this study uses 2249 test data randomly selected from 2811. In the Naive Bayes algorithm, the basic assumption is that all variables that affect classification are independent of each other. This means that other variables do not influence each variable's contribution to the final probability, so the interaction or relationship between variables is not taken into account, as shown in Figure 2.

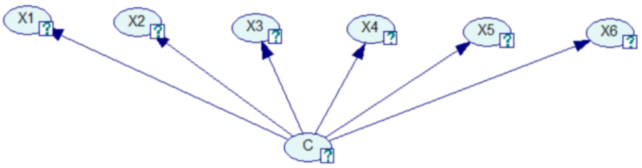

Figure 2. Naive Bayes structure

In this study, the variables of soil type (X_1), geology type (X_2), landform type (X_3), slope (X_4), land use type (X_5), and rainfall (X_6) have separate effects on the results of flood risk classification (C).

The prior probabilities for each class are shown in Table 2.

Table 2. Prior Probability on Each Class

Flood Hazard Class(C)	Score	Amount	Probability
Low	1	1518	0.67497
Currently	2	616	0.2739
Tall	3	115	0.05113
Total		2249	1

Table 2 provides an overview of the distribution of flood hazard levels based on risk categories. The data shows that low flood hazards dominate, while high flood hazards are relatively rare. This information can be used to prioritize flood mitigation efforts in areas with medium and high risks.

We need to determine $P(X|C_i)$ for each of x_j at X and C_i. The probability calculation $X$with the condition of using C_i training data using variables X_1up to X_6. All the results of this calculation are shown in Table 3-Table 8.

Table 3. (Calculation of probability $P(x_1|C_i)$

	2) C	1	2	3
X1	**1**	0.270341	0.056270	0.008264
	2	0.029528	0.649518	0.842975
	3	0.000656	0.046624	0.099174
	4	0.000656	0.001608	0.024793
	5	0.593176	0.236334	0.016529
	6	0.105643	0.009646	0.008264
		1	1	1

Table 4. Calculation of probability $P(x_2|C_i)$

	C	1	2	3
X2	**1**	0.194754	0.028892	0.008197
	2	0.024918	0.728732	0.950820
	3	0.043934	0.030498	0.008197
	4	0.655082	0.046549	0.008197
	5	0.021639	0.008026	0.008197
	6	0.028852	0.020867	0.008197
	7	0.030820	0.136437	0.008197
		1	1	1

Table 5. Calculation of probability$P(x_3|C_i)$

	C	1	2	3
X3	**1**	0.94129159	0.06656101	0.00769231
	2	0.00260926	0.05863708	0.00769231
	3	0.00130463	0.02377179	0.14615385
	4	0.00130463	0.12044374	0.46923077
	5	0.00065232	0.01743265	0.10000000
	6	0.00130463	0.12836767	0.00769231
	⋮	⋮	⋮	⋮
	14	0.00065232	0.03645008	0.00769231
	15	0.00391389	0.08874802	0.00769231
		1	1	1

Table 6. Calculation of probability $P(x_4|C_i)$

	C	1	2	3
X4	**1**	0.01051940	0.86914378	0.98305085
	2	0.24523340	0.12116317	0.00847458
	3	0.74424721	0.00969305	0.00847458
		1	1	1

Table 7. Calculation of probability $P(x_5|C_i)$

	C	1	2	3
X5	**1**	0.03439325	0.13302034	0.00724638
	2	0.91499027	0.04381847	0.00724638
	3	0.00194679	0.00312989	0.00724638
	4	0.00064893	0.27386541	0.17391304
	5	0.02271252	0.08137715	0.00724638
	6	0.00064893	0.21596244	0.59420290

	⋮	⋮	⋮	⋮
	22	0.00064893	0.00312989	0.00724638
	23	0.00064893	0.00625978	0.00724638
		1	1	1

Table 8. Calculation of probability $P(x_6|C_i)$

	C	1	2	3
X6	**1**	0.00065617	0.00482315	0.00826446
	2	0.00328084	0.05787781	0.16528926
	3	0.18700787	0.63344051	0.71900826
	4	0.16272966	0.11254019	0.08264463
	5	0.26115486	0.12379421	0.01652893
	6	0.38517060	0.06752412	0.00826446
		1	1	1

In this section, we will calculate the conditional probability of an event with a vector of variables $x_1, x_2, \dots, x_6$ included in a certain class C_i using Equation (1).

$$P(C_i|x_1, x_2, \dots, x_6) = \frac{P(x_1, x_2, \dots, x_6|C_i).P(C_i)}{P(x_1, x_2, \dots, x_6)}$$

$$\text{for } 1 < i < 3$$

Because $X_1, X_2, \dots, X_6$ it is independent, then

$$P(x_1, x_2, \dots, x_6|C_i) = P(x_1|C_i).P(x_2|C_i) \dots P(x_6|C_i)$$

So that

$$P(C_i|x_1, x_2, \dots, x_6) = \frac{P(x_1|C_i).P(x_2|C_i) \dots P(x_6|C_i)P(x_1, x_2, \dots, x_6|C_i).P(C_i)}{P(x_1, x_2, \dots, x_6)}$$

$$\text{for } 1 < i < 3$$

The calculation $P(C_i|x_1, x_2, \dots, x_6)$ was carried out for training data consisting of 562 grids for each flood hazard level.

Suppose a calculation $P(C_1|x_1, x_2, \dots, x_6)$ is performed for the value $X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1$.

$$P(C_1|X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1) = \frac{P(X_1 = 2|C_1).P(X_2 = 4|C_1).P(X_3 = 5|C_1).P(X_4 = 1|C_1).\ P(X_5 = 13|C_1).P(X_6 = 1|C_1).P(C_1)}{P(X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1)}$$

$$P(C_1|X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1) = \frac{3.81475E - 14}{3.29438E - 08} = 1.15796\text{E} - 06$$

In the same way, obtained $P(C_2|X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1) = 0.747440813$ and $P(C_3|X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1) = 0.252558029$. So, the flood hazard level for $X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1$ is $\max[P(C_i|X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1)]$ which is $P(C_2|X_1 = 2, X_2 = 4, X_3 = 5, X_4 = 1, X_5 = 13, X_6 = 1) = 0.747440813$. The flood hazard level in this example is medium (2). Later, the value will be obtained for all test data.

4.3. Evaluation of the Naïve Bayes Model

A confusion matrix is a tool often used to evaluate the performance of classification models, including Naive Bayes. This matrix depicts the number of correct and incorrect predictions for each class on the testing data, thus providing deep insight into the model's performance, as shown in Table 9.

Table 9. Naive Bayes Confusion Matrix

		Actual Value		
		1	**2**	**3**
Predicted value	**1**	389	4	0
	2	0	133	1
	3	0	8	27

To evaluate model performance, accuracy metrics are used to describe the model's ability to perform classification.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total of Data}}$$

$$\text{Accuracy} = \frac{389 + 133 + 27}{562} = 97.69\%$$

The evaluation results show that the Naive Bayes model has a very good performance in classifying flood hazards, with an accuracy value reaching 97.69%. This model effectively distinguishes between low, medium, and high flood hazard classes with a high level of accuracy. This indicates that the simple probabilistic approach of Naive Bayes can be an efficient solution for predicting flood risk when the data are relatively simple and structured.

4.4. Flood Hazard Level Mapping

Flood vulnerability mapping in ArcGIS is an essential step toward understanding flood risk spatially and in depth. This process aims to produce a flood hazard-level zoning map that can serve as a basis for disaster mitigation planning.

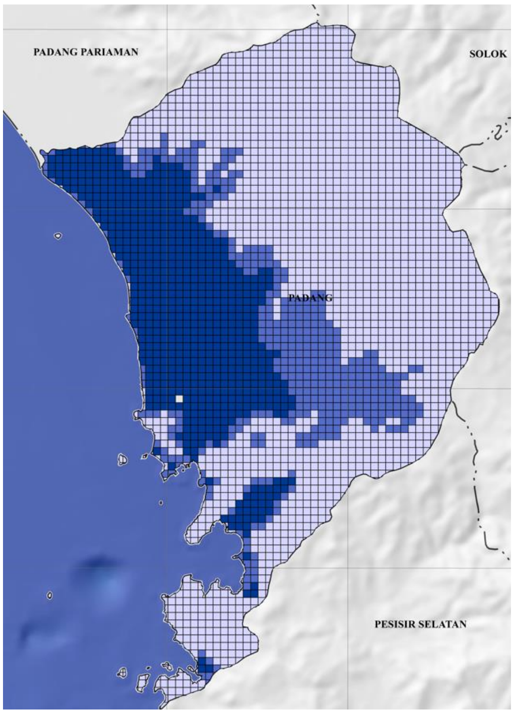

Figure 3. Flood Hazard Level Map

Each conditioning parameter was systematically analyzed and integrated to classify areas into low, moderate, and high flood hazard levels using probabilistic inference based on the Naïve Bayes model. The Naïve Bayes classifier was selected due to its superior predictive accuracy compared to the Bayesian Belief Network model, as well as its computational efficiency and robustness in handling multi-parameter spatial datasets.

The resulting output is a probabilistic flood hazard map that effectively identifies priority areas for mitigation and risk-reduction interventions. This map serves as a valuable decision-support tool for policymakers and stakeholders, as it not only visualizes spatial variations in flood hazard levels but also highlights the dominant contributing factors influencing flood risk. Through this integrated probabilistic and geospatial approach, the study contributes to advancing data-driven flood mitigation strategies and enhancing the effectiveness of spatial planning in Padang City and other regions with similar characteristics.

As illustrated in Figure 3, high flood hazard levels are predominantly concentrated in the western part of Padang City. This pattern is primarily attributed to the region’s low-lying topography and its proximity to the coastline. The western area functions as a natural accumulation zone, where surface runoff from higher elevations converges. Moreover, its coastal location exposes it to additional tidal flooding risks. When intense rainfall coincides with high tidal conditions, the likelihood and severity of flooding in this area increase substantially.

5. Conclusions

This study demonstrates the scientific value of integrating a probabilistic Naïve Bayes classifier with GIS-based spatial analysis for urban-scale flood vulnerability mapping. The proposed framework provides a statistically robust and computationally efficient alternative to conventional deterministic approaches and more complex Bayesian models. Despite its relative methodological simplicity, the model achieved a high predictive accuracy of 97.69% while preserving spatial interpretability and analytical transparency.

The resulting flood vulnerability map offers practical and actionable insights for identifying priority zones for mitigation and risk-reduction strategies in Padang City. By combining probabilistic modeling with geospatial analysis, the study contributes to strengthening evidence-based urban planning and disaster risk management practices.

Future research should consider incorporating temporal rainfall variability, climate change projections, and hybrid machine learning approaches to further improve predictive performance, model generalizability, and spatial resolution. Expanding the framework to multi-temporal and multi-hazard assessments would also enhance its applicability in broader environmental risk analysis contexts.

Declarations

Author Contributions

Conceptualization, D.P.S.; methodology, R.S.; software, Y.W.; validation, A.S., and D.P.S.; formal analysis, D.P.S.; investigation, A.S..; resources, M.P.D.; data curation, A.S..; writing—original draft preparation, D.P.S.; writing—review and editing, R.S.; visualization, D.A.; supervision, M.P.D.; project administration, M.P.D.; funding acquisition, D.P.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

Data available in a publicly accessible repository: The data presented in this study are openly available in [repository name e.g., FigShare] at [doi], reference number [reference number].

Funding

Funding information is not available.

Acknowledgements

We gratefully acknowledge the support of the Institute for Research and Community Service (LPPM), Universitas Negeri Padang, under Contract Number: 1655/UN35.15/LT/2024.

Institutional Review Board Statement

Every participant in the study gave their informed consent.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. All ethical standards have been fully observed, including those related to plagiarism, informed consent, research misconduct, data fabrication and/or falsification, duplicate publication and/or submission, and redundant publication.

References

[1] BUI D.T., TSANGARATOS P., NGO P.T.T., PHAM T.D., and PHAM B.T. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. Science of The Total Environment. 2019 Jun 10;668:1038–1054.

[2] DOU J., YUNUS A.P., BUI D.T., SAHANA M., CHEN C.W., ZHU Z., WANG W., and PHAM B.T. Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. Remote Sensing. 2019 Mar 15;11(6):638.

[3] FALAH F., RAHMATI O., ROSTAMI M., AHMADISHARAF E., DALIAKOPOULOS I.N., and POURGHASEMI H.R. Artificial Neural Networks for Flood Susceptibility Mapping in Data-Scarce Urban Areas. Spatial Modeling in GIS and R for Earth and Environmental Sciences. 2019;323–336.

[4] KHOSRAVI K., PHAM B.T., CHAPI K., SHIRZADI A., SHAHABI H., REVHAUG I., PRAKASH I., and BUI D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Science of The Total Environment. 2018 Jun 15;627:744–755.

[5] NANDI A., MANDAL A., WILSON M., and SMITH D. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. Environmental Earth Sciences. 2016 Mar 1;75(6):1–16.

[6] AHMADLOU M., KARIMI M., ALIZADEH S., SHIRZADI A., PARVINNEJHAD D., SHAHABI H., and PANAHI M. Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). 2018;34(11):1252–1272.

[7] HONG H., PANAHI M., SHIRZADI A., MA T., LIU J., ZHU A.X., CHEN W., KOUGIAS I., and KAZAKIS N. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. Science of The Total Environment. 2018 Apr 15;621:1124–1141.

[8] NAIEM S., KHEDR A.E., IDREES A.M., and MARIE M.I. Enhancing the Efficiency of Gaussian Naïve Bayes Machine Learning Classifier in the Detection of DDOS in Cloud Computing. IEEE Access. 2023;11:124597–124608.

[9] PERETZ O., KOREN M., and KOREN O. Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. Engineering Applications of Artificial Intelligence. 2024 Oct;136:108972.

[10] AMIN A., ADNAN A., and ANWAR S. An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes. Applied Soft Computing. 2023 Apr;137:110103.

[11] SARI D.P., ROSHA M., and ROSADI D. Exploring the Discretization of Peak Ground Acceleration Variables: A Comparative Analysis of K-means and K-medoids. Philippine Journal of Science. 2024;153(5):1661–1668.

[12] SCHONLAU M. The Naive Bayes Classifier. In: Springer; 2023. p. 143–160.

[13] CHEN H., HU S., HUA R., and ZHAO X. Improved Naive Bayes classification algorithm for traffic risk management. EURASIP Journal on Advances in Signal Processing. 2021 Dec 22;2021(1):30.

[14] KUMAR R., KRISHNA GOSWAMI B., MOTIRAM MHATRE S., and AGRAWAL S. Naive Bayes in Focus: A Thorough Examination of its Algorithmic Foundations and Use Cases. International Journal of Innovative Science and Research Technology (IJISRT). 2024 Jun 7;2078–2081.

[15] MYTHILI J., DEEBESHKUMAR B., ESHWARAMOORTHY T., and AJAY J.N. Enhancing Email Spam Detection with Temporal Naive Bayes Classifier. In: 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT). IEEE; 2024. p. 1–6.

[16] K A., and HALDER S. Detection of Multilingual Spam SMS Using Naïve Bayes Classifier. In: 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA). IEEE; 2023. p. 89–94.

[17] GAUR P., VASHISTHA S., and JHA P. Twitter Sentiment Analysis Using Naive Bayes-Based Machine Learning Technique. 2023;367–376.

[18] LASE Y.Y., LUBIS A.R., ELYZA F., and SYAFLI S.A. Mental Health Sentiment Analysis on Social Media TikTok with the Naïve Bayes Algorithm. In: 2023 6th International Conference of Computer and Informatics Engineering (IC2IE). IEEE; 2023. p. 186–191.

[19] KANDEL R., and BAROUD H. A data-driven risk assessment of Arctic maritime incidents: Using machine learning to predict incident types and identify risk factors. Reliability Engineering & System Safety. 2024 Mar;243:109779.

[20] WANG B., CHEN Y., and LI Z. A novel Bayesian Pay-As-You-Drive insurance model with risk prediction and causal mapping. Decision Analytics Journal. 2024 Dec;13:100522.

[21] PAJILA P.J.B., SHEENA B.G., GAYATHRI A., ASWINI J., NALINI M., and R S.S. A Comprehensive Survey on Naive Bayes Algorithm: Advantages, Limitations and Applications. In: 2023 4th International Conference on Smart Electronics and Communication (ICOSEC). IEEE; 2023. p. 1228–1234.

[22] SARI D.P., ROSADI D., EFFENDI A.R., DANARDONO, and ROSHA M. Integrating Peak Ground Acceleration as a Damage Factor in Risk-Based Premium Rate Assessment using K-medoids Bayesian Networks. Journal of Applied Science and Engineering. 2024;28(6):1361–1369.

[23] CHEBIL W., WEDYAN M., ALAZAB M., ALTURKI R., and ELSHAWEESH O. Improving Semantic Information Retrieval Using Multinomial Naive Bayes Classifier and Bayesian Networks. Information. 2023 May 3;14(5):272.

参考文:

[1] BUI D.T., TSANGARATOS P., NGO P.T.T., PHAM T.D., 和 PHAM B.T. 基于优化模糊规则特征选择技术与树模型集成方法的山洪易发性建模. Science of The Total Environment. 2019年6月10日;668:1038–1054.

[2] DOU J., YUNUS A.P., BUI D.T., SAHANA M., CHEN C.W., ZHU Z., WANG W., 和 PHAM B.T. 基于LiDAR DEM的地震与降雨诱发滑坡易发性GIS多统计模型与数据挖掘方法评估. Remote Sensing. 2019年3月15日;11(6):638.

[3] FALAH F., RAHMATI O., ROSTAMI M., AHMADISHARAF E., DALIAKOPOULOS I.N., 和 POURGHASEMI H.R. 数据匮乏城市地区洪水易发性制图的人工神经网络方法. Spatial Modeling in GIS and R for Earth and Environmental Sciences. 2019;323–336.

[4] KHOSRAVI K., PHAM B.T., CHAPI K., SHIRZADI A., SHAHABI H., REVHAUG I., PRAKASH I., 和 BUI D.T. 伊朗北部Haraz流域山洪易发性建模的决策树算法比较研究. Science of The Total Environment. 2018年6月15日;627:744–755.

[5] NANDI A., MANDAL A., WILSON M., 和 SMITH D. 基于主成分分析与逻辑回归的牙买加洪水灾害制图. Environmental Earth Sciences. 2016年3月1日;75(6):1–16.

[6] AHMADLOU M., KARIMI M., ALIZADEH S., SHIRZADI A., PARVINNEJHAD D., SHAHABI H., 和 PANAHI M. 基于ANFIS与BBO及BAT算法集成的洪水易发性评估. 2018;34(11):1252–1272.

[7] HONG H., PANAHI M., SHIRZADI A., MA T., LIU J., ZHU A.X., CHEN W., KOUGIAS I., 和 KAZAKIS N. 结合自适应神经模糊推理系统与遗传算法和差分进化算法的洪水易发性评估. Science of The Total Environment. 2018年4月15日;621:1124–1141.

[8] NAIEM S., KHEDR A.E., IDREES A.M., 和 MARIE M.I. 提升高斯朴素贝叶斯分类器在云计算DDoS检测中的效率. IEEE Access. 2023;11:124597–124608.

[9] PERETZ O., KOREN M., 和 KOREN O. 朴素贝叶斯分类器——一种提升召回率与精确率的集成方法. Engineering Applications of Artificial Intelligence. 2024年10月;136:108972.

[10] AMIN A., ADNAN A., 和 ANWAR S. 基于进化计算与朴素贝叶斯的电信行业客户流失预测自适应学习方法. Applied Soft Computing. 2023年4月;137:110103.

[11] SARI D.P., ROSHA M., 和 ROSADI D. 峰值地面加速度变量离散化研究：K-means与K-medoids比较分析. Philippine Journal of Science. 2024;153(5):1661–1668.

[12] SCHONLAU M. 朴素贝叶斯分类器. Springer; 2023. p. 143–160.

[13] CHEN H., HU S., HUA R., 和 ZHAO X. 改进的朴素贝叶斯分类算法在交通风险管理中的应用. EURASIP Journal on Advances in Signal Processing. 2021;2021(1):30.

[14] KUMAR R., KRISHNA GOSWAMI B., MOTIRAM MHATRE S., 和 AGRAWAL S. 聚焦朴素贝叶斯：算法基础与应用场景的全面研究. International Journal of Innovative Science and Research Technology (IJISRT). 2024年6月7日;2078–2081.

[15] MYTHILI J., DEEBESHKUMAR B., ESHWARAMOORTHY T., 和 AJAY J.N. 基于时间特征的朴素贝叶斯分类器在电子邮件垃圾检测中的应用. In: 2024年通信、计算与物联网国际会议（IC3IoT）. IEEE; 2024. p. 1–6.

[16] K A., 和 HALDER S. 基于朴素贝叶斯分类器的多语言垃圾短信检测. In: 2023 IEEE第5届控制论、认知与机器学习应用国际会议（ICCCMLA）. IEEE; 2023. p. 89–94.

[17] GAUR P., VASHISTHA S., 和 JHA P. 基于朴素贝叶斯机器学习技术的Twitter情感分析. 2023;367–376.

[18] LASE Y.Y., LUBIS A.R., ELYZA F., 和 SYAFLI S.A. 基于朴素贝叶斯算法的TikTok社交媒体心理健康情感分析. In: 2023第6届计算机与信息工程国际会议（IC2IE）. IEEE; 2023. p. 186–191.

[19] KANDEL R., 和 BAROUD H. 基于数据驱动的北极海事事故风险评估：利用机器学习预测事故类型并识别风险因素. Reliability Engineering & System Safety. 2024年3月;243:109779.

[20] WANG B., CHEN Y., 和 LI Z. 一种新型贝叶斯按里程付费保险模型：风险预测与因果映射. Decision Analytics Journal. 2024年12月;13:100522.

[21] PAJILA P.J.B., SHEENA B.G., GAYATHRI A., ASWINI J., NALINI M., 和 R S.S. 朴素贝叶斯算法综述：优势、局限性与应用. In: 2023第4届智能电子与通信国际会议（ICOSEC）. IEEE; 2023. p. 1228–1234.

[22] SARI D.P., ROSADI D., EFFENDI A.R., DANARDONO, 和 ROSHA M. 基于K-medoids贝叶斯网络的风险保费评估中峰值地面加速度作为损害因子的整合研究. Journal of Applied Science and Engineering. 2024;28(6):1361–1369.

[23] CHEBIL W., WEDYAN M., ALAZAB M., ALTURKI R., 和 ELSHAWEESH O. 基于多项式朴素贝叶斯分类器与贝叶斯网络的语义信息检索改进研究. Information. 2023年5月3日;14(5):272.

Manuscript Information
Word count: 5,719 words (excluding references).

Peer-Review Record
Fast-track status: Not fast-tracked.

First-round reviews received: 3 reports.
Revision cycles completed: 3 rounds.
Final version submitted: February 11, 2026

Disclaimer / Publisher’s Note