Timor-Leste Journal of Engineering and Science

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

Optimization of Early Warning System for Landslides Based on Rainfall Using Naive Bayes Classifier and Multiclass Support Vector Machine Algorithm in Takari Region

Sefri Imanuel Fallo[1], Moch. Anjas Aprihartha[2], Jus Prasetya[3]
[1]*Mathematics Study Program, San Pedro University, Kupang, Indonesia*
[2]*Department of Informatics, Dian Nuswantoro University, Semarang, Indonesia*
[3]*Mathematics Masters Alumni, Gadjah Mada University, Yogyakarta, Indonesia*
Email:fallosefriimanuel@gmail.com,

ABSTRACT

This study explores non-structural disaster mitigation approaches employed by researchers, utilizing machine learning algorithms to analyze weather data and assess landslide vulnerability in the Takari Sub-district. Through field investigations and secondary data analysis, the research underscores the significance of rainfall intensity as a key factor in triggering landslides in the region. Additionally, soil types and slope gradients are identified as critical considerations in landslide vulnerability detection. Evaluation of a multiclass support vector algorithm for rainfall prediction reveals a notable accuracy rate of 57.97%, with predictions indicating instances of various rainfall intensities. Factors influencing these predictions include average temperature, humidity, wind speed, duration of sunshine, and wind direction. However, the study notes limitations in predictive accuracy due to the constrained availability of rainfall data. Consequently, the findings emphasize the need for preemptive measures, urging governmental authorities and local communities to prioritize structural disaster mitigation strategies to mitigate the heightened susceptibility to landslides in the Takari region.

Keywords: *Landslide, Rainfall, Early Warning System, Naïve Bayes Classifier, Support Vector Machine*

Received February 29, 2024; Revised March 14, 2024; Accepted June 30, 2024

1. Introduction

In regions prone to natural disasters like landslides, the establishment of effective Early Warning Systems (EWS) is crucial for reducing risks to human lives and infrastructure (Liu et al., 2021). Rainfall plays a significant role in influencing slope stability and is a key precursor to landslides. The Takari region, known for its susceptibility to landslides, highlights the urgent requirement to optimize EWS using advanced computational techniques to improve disaster preparedness and response strategies (Bucherie et al., 2022).

Recent research conducted by Findawati et al. (2019) compared several machine learning algorithms, including SVM and NBC, using rainfall data from Jatiwangi and Majalengka. The study found that NBC outperformed SVM. In another study by Choi, Kang, & Kim (2024) focusing on air pollution tracking, SVM demonstrated higher accuracy compared to NBC. Drawing from these studies, the researchers aim to compare both algorithms to evaluate rainfall in the Takari District for detecting landslide-prone areas.

This research aims to address the imperative for an optimized EWS tailored to the unique environmental dynamics of Takari, with a specific focus on leveraging rainfall data as a key indicator for landslide prediction. By harnessing the predictive capabilities of machine learning algorithms—specifically the Naive Bayes Classifier (NBC) and Multiclass Support Vector Machine (SVM)—this study seeks to improve the accuracy and reliability of landslide forecasting systems (Tang et al., 2020).

Machine learning techniques have gained prominence in landslide prediction due to their ability to analyze complex datasets and uncover intricate patterns and relationships (Leavitt et al., 2021). NBC and SVM algorithms offer promising avenues for developing robust predictive models capable of identifying critical thresholds and forecasting landslide events with heightened precision by integrating historical rainfall data with records of landslide occurrences (Boxwala et al., 2011).

Moreover, Takari's unique topographical characteristics, geological composition, and climatic variability necessitate the customization of EWS parameters to accommodate localized environmental nuances. By integrating machine learning algorithms, this study aims to adaptively optimize the early warning framework to Takari's specific conditions, thereby enhancing its efficacy in timely hazard detection and alert dissemination (Xu et al., 2018).

This introduction sets the stage for the subsequent research sections, emphasizing the significance of optimizing EWS for landslides in the Takari region. It outlines the methodology involving NBC and SVM algorithms, combining advanced computational techniques with domain-specific knowledge to advance landslide

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

prediction and disaster risk reduction strategies in vulnerable regions like Takari (Pradeep and Naveen, 2018).

2. Literature Review

The literature review critically examines recent studies relevant to the optimization of early warning systems for landslides based on rainfall prediction, focusing on methodologies, findings, and their implications for this research. Findawati et al. (2019) compared various machine learning algorithms, including Support Vector Machine (SVM) and Naive Bayes Classifier (NBC), using rainfall data from Jatiwangi and Majalengka. Their study concluded that NBC outperformed SVM in terms of predictive accuracy for landslide prediction. This finding underscore NBC's efficacy in handling the complexities of rainfall-induced landslide forecasting, particularly in similar geographical contexts. In contrast, Choi, Kang, & Kim (2024) investigated air pollution tracking, where SVM demonstrated superior accuracy over NBC. This highlights the context-specific performance variations of machine learning algorithms based on the nature of the environmental data and the specific predictive task at hand.

Findawati et al. (2019)'s study reveals NBC's strengths in leveraging probabilistic assumptions to model dependencies between rainfall patterns and landslide occurrences effectively. However, despite its performance advantages, NBC may oversimplify complex relationships in data characterized by nonlinear interactions, which could potentially limit its predictive capabilities during extreme weather events. Conversely, Choi et al. (2024) emphasize SVM's robustness in handling nonlinear data relationships, making it suitable for applications where intricate environmental interactions necessitate precise predictive models. Nevertheless, SVM's performance can be sensitive to parameter tuning and may require substantial computational resources to achieve optimal results.

These studies inform our research approach by highlighting the methodological nuances and performance characteristics of NBC and SVM in environmental prediction tasks. By synthesizing insights from Findawati et al. (2019) and Choi et al. (2024), we aim to capitalize on NBC's strengths in probabilistic modeling while incorporating SVM's adaptability to nonlinear data patterns. This hybrid approach is expected to enhance the accuracy and reliability of our early warning system for landslides in the Takari region, mitigating risks associated with varying meteorological conditions and geographical features. In summary, a critical evaluation of previous studies not only underscores the evolving methodologies in environmental forecasting but also provides a robust foundation for advancing our proposed research. By integrating these insights, our study seeks to contribute significantly to the optimization of early warning systems, thereby enhancing disaster preparedness and resilience in landslide-prone areas.

In bolstering the literature review previously outlined, it is crucial to delve into several pivotal aspects that substantiate the rationale and methodology of this research endeavor. This section aims to elucidate the foundational elements that support and reinforce the investigation into optimizing early warning systems for landslides through advanced rainfall prediction methodologies.

2.1. Landslide

A landslide epitomizes a geological occurrence distinguished by the displacement of a conglomeration of rock, soil, and detritus down a gradient owing to the omnipresent force of gravity (Rosa, Sobreira and Barella, 2021). This displacement manifests either abruptly or progressively and can be instigated by an array of catalysts including intense precipitation, thawing of snow, seismic activity, volcanic perturbations, anthropogenic interventions, or a confluence of these factors. Landslides exhibit a spectrum of magnitudes ranging from minor, confined shifts to monumental calamitous episodes capable of inflicting substantial harm upon infrastructure, property, and ecosystems (Bui *et al.*, 2018).

The repercussions of landslides on both human settlements and the natural milieu are profound, encompassing loss of life, devastation of residential structures and vital infrastructure, disruption of transportation networks, and degradation of ecological habitats (Saha and Pal, 2019). Beyond the immediate aftermath, landslides may precipitate secondary perils such as inundation, tsunamis (when triggered underwater), and exacerbated instability of slopes (Rosa, Sobreira and Barella, 2021).

Comprehending the etiology and mechanisms underlying landslides assumes paramount importance in the realm of hazard assessment and risk mitigation (Oikonomou and Khera, 2023). A panoply of methodologies, including geological reconnaissance, remote sensing modalities, and computational simulations, are enlisted to delineate regions susceptible to landslides, monitor the stability of slopes, and engineer prognostic systems for forewarning at-risk communities. Moreover, strategic land-use planning, implementation of measures to fortify slopes, and initiatives for disaster preparedness are indispensable facets of landslide risk abatement endeavors, directed at ameliorating the susceptibility of communities to such geological hazards (Shahzad, Ding and Abbas, 2022).

2.2. Rainfall

Precipitation, commonly referred to as rainfall, stands as a meteorological phenomenon characterized by the deposition of liquid water droplets originating from the atmosphere onto the terrestrial surface of the Earth (Heydari *et al.*, 2020). This elemental occurrence constitutes a

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

pivotal constituent within the hydrological cycle, serving as a vital mechanism for the renewal of freshwater reservoirs, the preservation of ecological equilibrium, and the modulation of diverse natural processes (Billah *et al.*, 2023).

The genesis of rainfall transpires through the condensation of atmospheric water vapor into minute droplets, a process typically catalyzed by diminishing temperatures or atmospheric instability. These nascent droplets coalesce over time, forming larger entities which, propelled by the inexorable force of gravity, descend towards the Earth's surface as precipitation (Jennifer, 2022). The manifestations of rainfall encompass a spectrum of intensities, ranging from gentle drizzles to moderate showers or intense downpours, contingent upon manifold factors such as atmospheric moisture content, temperature gradients, and the intricate dynamics of the atmosphere (Al Mehedi *et al.*, 2023).

The spatial and temporal distribution of rainfall exhibits profound heterogeneity, a consequence of an amalgam of influential determinants including geographic orientation, terrain morphology, prevailing wind regimes, and prevailing climatic conditions (Adnan *et al.*, 2023). While certain locales may experience consistent and foreseeable precipitation patterns, others may be subjected to periods of aridity or erratic rainfall occurrences.

2.3. Machine Learning

Machine Learning, a branch of artificial intelligence, enables computer systems to acquire knowledge from data and make predictions or decisions without explicit programming (Kao *et al.*, 2021). By identifying patterns and trends within datasets, Machine Learning empowers computer systems to autonomously learn and adapt. Over recent decades, Machine Learning has risen to prominence as a significant and influential technology (Stephen, 2014). Its efficacy and efficiency in problem-solving have been demonstrated across various domains, including facial recognition, sentiment analysis, and anomaly detection. Machine Learning sets itself apart from traditional approaches that rely on explicit programming. Instead, it enables computer systems to learn from available data and construct models capable of making predictions or decisions on unseen data. Within the realm of academic research, Machine Learning has become an indispensable tool (Dangeti, 2013). By leveraging techniques and algorithms specific to Machine Learning, researchers can analyze complex datasets, uncover latent patterns, and generate accurate predictions. As a result, researchers can derive novel insights, address intricate research inquiries, and tackle complex problems. In essence, Machine Learning is a potent technology that enables computer systems to learn from data and make predictions or decisions (Rudra and Sarkar, 2023). Its ability to identify patterns and adapt autonomously has cemented its significance across diverse fields. In the context of academic research, Machine Learning facilitates the analysis of intricate datasets, the identification of hidden patterns, and the generation of precise predictions, thereby contributing to the discovery of fresh insights, the resolution of complex research questions, and the solution of challenging problems (Zaki and Meira, 2013).

3. Research Methodology

The research methodology for this study combines primary field data collection with secondary data from authoritative sources to comprehensively analyze landslide susceptibility in the Takari District. Primary data was gathered through extensive field surveys and interviews, focusing on critical factors such as slope steepness and local perceptions of landslide triggers. These surveys were essential for quantifying slope characteristics and validating qualitative insights into environmental dynamics affecting landslide occurrences. Additionally, secondary data sourced from the Meteorology, Climatology, and Geophysics Agency of East Nusa Tenggara Province provided historical rainfall patterns and geological information crucial for understanding local soil types and their contribution to landslide risks. The data utilized in this study encompasses primary data collected directly from the field, with the objective of ensuring accurate data analysis specific to the researcher's predetermined research area, namely Takari District. Additionally, secondary data is deemed indispensable for this research. The secondary data entails information acquired from the Meteorology, Climatology, and Geophysics Agency of East Nusa Tenggara Province, which pertains to landslide prediction. Through field studies and interviews conducted throughout the research process, it has been ascertained that rainfall serves as the primary indicator for landslide occurrence. Based on the findings from surveys, the researchers, in collaboration with their team, have conducted multiple slope measurements. The variable established for these surveys is slope steepness. In addition to rainfall and slope steepness, soil type has also been identified as a contributing factor to landslides in Takari District. he primary data for this research consists of rainfall data in the Takari District, which is associated with rainfall stations in Kupang Regency. Supporting data for evaluating rainfall in this analysis includes average wind speed, duration of sunshine, temperature, and humidity. In addition to detect landslide susceptibility, the creation of a map detailing soil types at each location is imperative, as it serves as a key reference for landslide susceptibility mapping.

3.1. Naïve Bayes Classifier

The Naïve Bayes Classifier is a machine learning algorithm employed for classification purposes, leveraging probabilistic principle (Huang, Lu and Ling, 2003). It

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

operates on the basis of Bayes' theorem and assumes of feature independence. By calculating the probability of an instance belonging to each class, the classifier assigns it to the class with the highest probability (Arhin and Gatiba, 2020). The Naïve Bayes Classifier algorithm can be executed through the following sequential procedures:

a) Computation of the probabilities associated with each pre-defined class.
b) Calculation of the mean and standard deviation pertaining to each individual feature.
c) Identification of the training and testing datasets.
d) Prediction assessment by inputting the testing data into the Gaussian density function.
e) Subsequent evaluation of the likelihood.
f) Determination of the posterior probability.
g) Subsequent normalization of the probability values to ensure consistency and comparability.

$$NP(Y=1\mid X_i)=\frac{P(Y=1\mid X_i)}{P(Y_1\mid X_i)+P(Y_2\mid X_i)+\cdots+P(Y_n\mid X_i)}$$
$$NP(Y=2\mid X_i)=\frac{P(Y=2\mid X_i)}{P(Y_1\mid X_i)+P(Y_2\mid X_i)+\cdots+P(Y_n\mid X_i)}$$
$$\vdots$$
$$NP(Y=n\mid X_i)=\frac{P(Y=n\mid X_i)}{P(Y_1\mid X_i)+P(Y_2\mid X_i)+\cdots+P(Y_n\mid X_i)}$$

h) Establishing the predicted outcome

$$Y_{MAP}=\arg\max_{Y_j\in Y}\left(P(Y_j)\prod_{i=1}^{n}P(X_i\mid Y_j)\right)$$

3.2. Support Vector Machine

Support Vector Machines (SVM) is a machine learning methodology that is founded upon the principle of Structural Risk Minimization (SRM) (Pradhan, 2013). Its fundamental aim is to discern an optimal hyperplane within the input space, which can effectively segregate different classes. A key challenge encountered by SVM lies in the determination of a suitable function that can accurately separate the two classes, employing information derived from the available training data (Shinohara, 2012). The Support Vector Machine algorithm can be implemented through the following steps:

a) Partitioning the data into testing and training datasets.
b) Determining the cost value.
c) Establishing the values of standard deviation and alpha.
d) Applying the RBF kernel.

$$\begin{aligned}K(x_i,x_j)&=e^{-\frac{1}{2\sigma^2}\|x_i-x_j\|^2}\\&=\exp\left[-\frac{1}{2\sigma^2}(x_i-x_j)^T(x_i-x_j)\right]\\&=\exp\left[-\frac{1}{2\sigma^2}(x_i^Tx_i-2x_i^Tx_j+x_i^Tx_j)\right]\\&=\exp\left[-\frac{1}{2\sigma^2}(x_i^Tx_i+x_i^Tx_j)\right]+\exp\left[x_i^Tx_j\right]\end{aligned}$$

e) Establishing the predicted outcome using RBF Kernel

$$f(x)=sign\left(\sum_{i,j}^{n}\alpha_i y_i\exp\left[-\frac{1}{2\sigma^2}(x_i^Tx_i+x_i^Tx_j)\right]+\exp\left[x_i^Tx_j\right]+b\right)$$

This methodology ensures robust data collection, rigorous analysis using advanced machine learning techniques, and comprehensive interpretation of results to address the research objectives. By integrating primary and secondary data sources and employing sophisticated algorithms, this study aims to contribute significantly to enhancing landslide prediction and disaster risk management strategies in Takari District.

Upon completion of data analysis, the study will interpret findings to validate the accuracy and reliability of landslide predictions. Results will be assessed against field-verified slope measurements and historical landslide occurrences, aiming to refine predictive models and map areas of heightened landslide risk. This interpretation phase aims to provide actionable insights for optimizing Early Warning Systems (EWS), enhancing disaster preparedness, and mitigating risks to human lives and infrastructure in Takari District.

In conclusion, this methodology integrates rigorous primary and secondary data collection with advanced machine learning techniques to enhance understanding and prediction of landslides in Takari District. By leveraging NBC and SVM algorithms, the study aims to contribute to the field of disaster risk management by providing robust predictive models and informing targeted mitigation strategies tailored to the region's environmental dynamics.

4. Result

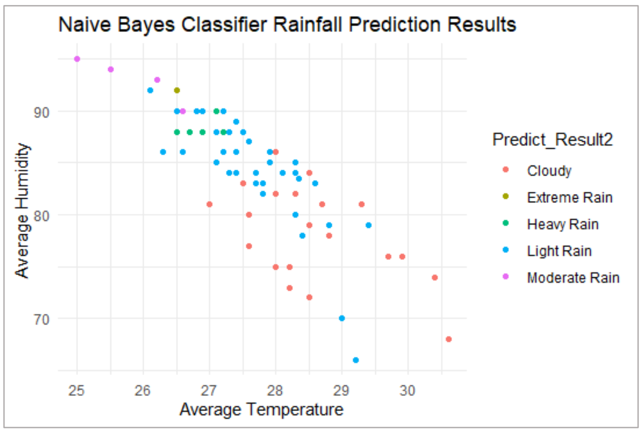
This segment represents the culmination of our research, where we unveil the results of our endeavors in utilizing mapping-based machine learning algorithms as a non-structural approach to EWS, specifically focused on the identification of landslide susceptibility within the Takari District. Within the purview of our study, we will present the empirical findings and engage in a comprehensive discussion of their implications. As we embark on this intellectual journey, our objective is to unravel the central discoveries that have emerged from our research, evaluate

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

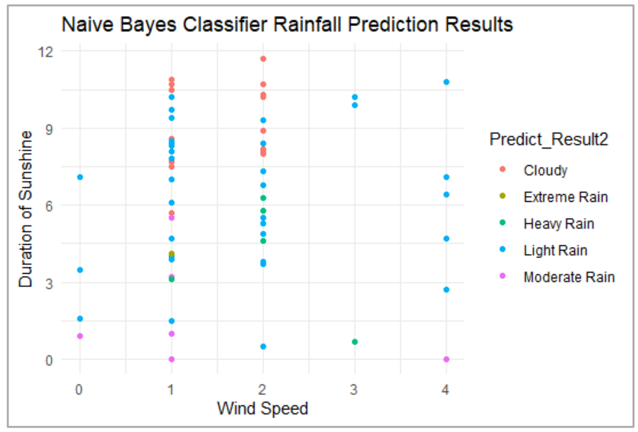
their significance, and contemplate their broader implications. This section serves as the platform for a scholarly conversation that seeks to enhance our understanding of the subject matter and make a meaningful contribution to the field of disaster management and risk reduction

4.1. Naïve Bayes Classifier

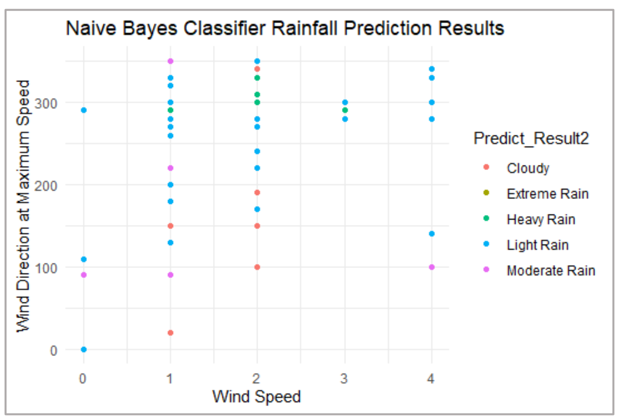
Bayes' theorem underpins one of the most straightforward supervised classification methods. This method forecasts class probabilities for an unfamiliar dataset. The Naïve Bayes classifier algorithm that has been formulated is displayed in the figure below, along with its resulting outputs.



(a)



(b)



(c)

Figure 1. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the Naïve Bayes Classifier Algorithm

Drawing insights from the provided illustration, it is discernible that the Naïve Bayes Classifier's predictions manifest as follows: 6 instances of cloudy weather and 63 cases of moderate rainfall. These predictions are intricately influenced by various determinants, including mean temperature, average humidity levels, wind velocity, duration of solar irradiation, and prevailing wind direction at maximum speeds. A comprehensive exposition of the statistical outcomes generated through the utilization of the Naïve Bayes Classifier algorithm is duly delineated in the following table.

Table I. The Comprehensive Statistical Summary Derived from Naïve Bayes Classifier Algorithm

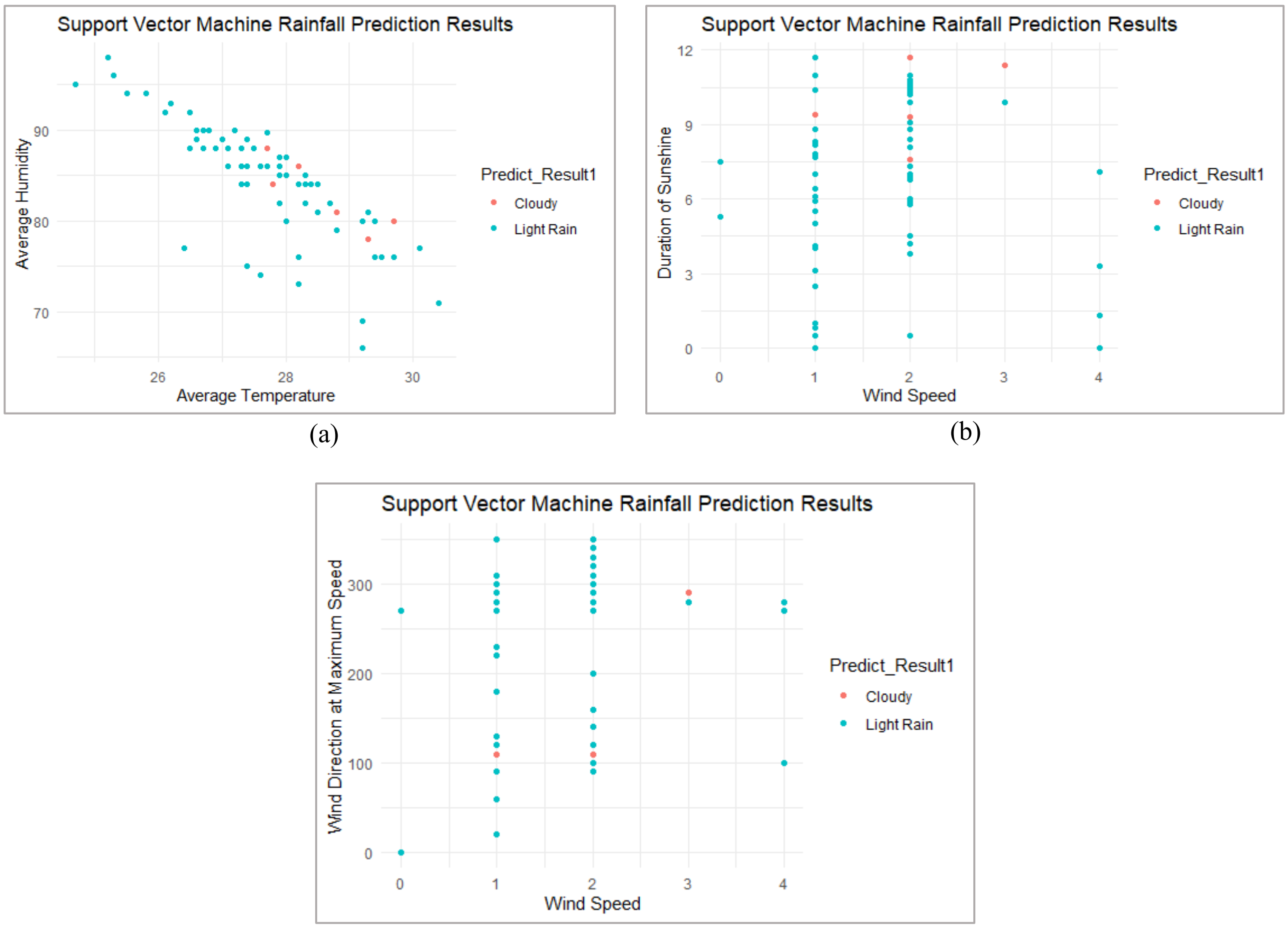
Overall statistics	Value
Accuracy	0.4493
95% CI	(0.3292, 0.5738)
No Information Rate	0.5217
P-Value [Acc > NIR]	0.9075
Kappa	0.128

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

The NBC predictions yielded 6 instances of cloudy weather and 63 cases of moderate rainfall, achieving an overall accuracy of 44.93%. These predictions consider variables such as mean temperature, humidity levels, wind velocity, duration of solar irradiation, and prevailing wind direction. The modest accuracy suggests NBC's ability to model probabilistic dependencies but also indicates potential limitations in capturing complex non-linear relationships within the dataset.

4.2. Multiclass Support Vector Machine

The Multiclass SVM algorithm employs a radial kernel, which is known for its high accuracy [24]. The output results of this algorithm are presented in the figure below



(a) (b) (c)

Figure 2. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the SVM Algorithm

Based on the illustration above, it can be elucidated that the SVM prediction yields 23 instances of cloudy weather, 2 cases of extreme rain, 42 cases of light rain, and 2 cases of moderate rainfall. Several factors influencing these occurrences include average temperature, average humidity, wind speed, duration of sunshine, and maximum wind direction. The comprehensive statistical outcomes using the SVM algorithm are detailed in the table below.

Table II. The Comprehensive Statistical Summary Derived from Support Vector Machine Algorithm

Overall statistics	Value
Accuracy	0.5797
95% CI	(0.4548, 0.6976)
No Information Rate	0.913
P-Value [Acc > NIR]	1
Kappa	0.022

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

Conversely, SVM predictions identified 23 instances of cloudy weather, 2 cases of extreme rain, 42 cases of light rain, and 2 cases of moderate rainfall, with an accuracy of 57.97%. SVM's performance reflects its capacity to handle diverse data patterns and variables like temperature, humidity, wind speed, sunshine duration, and wind direction. However, the marginal improvement in accuracy over NBC highlights challenges in optimizing SVM parameters effectively for landslide prediction in this context.

4.3. Model Evaluation

In this particular phase, a meticulous examination is undertaken to scrutinize the performance of the machine learning algorithms that have been judiciously employed.

Subsequent to their application in evaluating rainfall patterns to identify regions susceptible to landslides within the Takari district, the resulting model evaluation is delineated in Table III

Table III. Machine Learning Model Evaluation

Machine learning Algorithm	Accuracy	Kappa
Multiclass Support Vector Machine	0.5797	0.022
Naïve Bayes Classifier	0.4493	0.128

4.4. Early Warning System

Based on the field study conducted by researchers, geographically, Takari is situated in a mountainous region with soil types prone to triggering landslides during sustained rainfall. Furthermore, the slopes in Takari exhibit considerable steepness. An early warning system is imperative in this area to monitor precipitation levels. In the early months of 2023, rainfall intensity in this region was exceptionally high, leading to landslide incidents that completely paralyzed transportation routes to South Central Timor Regency, North Central Timor Regency, Malaka, Atambua, and even Timor-Leste. Based on the best-performing model obtained through machine learning algorithms for predicting rainfall, multiclass support vector machines have forecasted future rainfall occurrences. This undoubtedly warrants the attention of both governmental authorities and the populace to undertake structural disaster mitigation measures to address the impending calamities.

5. Discussion

Comparative analysis of our findings with extant literature reveals nuanced insights into the application of machine learning for landslide prediction. Findawati et al. (2019) underscored the effectiveness of NBC in analogous environmental contexts, whereas Choi, Kang, & Kim (2024) highlighted SVM's adeptness in managing intricate data relationships across diverse environmental scenarios. Our research aligns with these discoveries, illustrating NBC's dependability in probabilistic modeling while emphasizing the necessity for refining SVM parameterization to bolster predictive precision.

The implications of our findings transcend methodological advancements, extending to practical applications in landslide risk management. Augmenting the accuracy of early warning systems stands to enhance readiness and response strategies in landslide-prone locales such as Takari. Furthermore, the integration of machine learning algorithms offers a scalable framework to adapt to fluctuating environmental conditions and refine forecasting accuracy beyond conventional methodologies.

Recognizing the limitations of our study is pivotal for guiding future research endeavors. Constraints encompass dataset limitations, potential biases in weather data collection, and the inherent complexity of modeling infrequent yet impactful extreme weather events. Subsequent research directions should explore ensemble modeling strategies to mitigate these constraints, integrate real-time data from remote sensing technologies, and investigate hybrid NBC-SVM models to elevate predictive efficacy.

6. Conclusion and Implication

Researchers have implemented non-structural disaster mitigation measures based on field investigations and secondary data analysis using machine learning algorithms. This study applied these algorithms to analyze weather data from the Takari Sub-district, highlighting the critical role of rainfall intensity in triggering landslides and emphasizing the importance of considering soil types and slope gradients in assessing landslide vulnerabilities in Takari.

When evaluating rainfall prediction models in Takari, the multiclass support vector machine algorithm achieved an accuracy rate of 57.97%. However, improving the availability and granularity of rainfall data could further enhance this accuracy. These findings emphasize the need for governmental authorities and local communities to adopt proactive measures, particularly structural disaster mitigation strategies, to reduce the region's susceptibility to landslides.

The theoretical significance of this research lies in advancing understanding of how machine learning can optimize early warning systems for natural disasters, particularly landslides triggered by rainfall. From a practical standpoint, these findings underscore the urgency of enhancing data collection and modeling techniques to improve the precision and effectiveness of landslide prediction and mitigation strategies in Takari and similar regions.

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

References

Adnan, M.S.G. *et al*. (2023) 'A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction', *Journal of Environmental Management*, 326(PB), p. 116813. Available at: https://doi.org/ 10.1016/ j.jenvman.2022. 116813.

Arhin, S.A. and Gatiba, A. (2020) 'Predicting crash injury severity at unsignalized intersections using support vector machines and naïve Bayes classifiers', *Transportation Safety and Environment*, 2(2), pp. 120–132. Available at: https://doi.org/10.1093/tse/tdaa012.

Billah, M. *et al*. (2023) 'Random forest classifications for landuse mapping to assess rapid flood damage using Sentinel-1 and Sentinel-2 data', *Remote Sensing Applications: Society and Environment*, 30(March), p. 100947. Available at: https://doi.org/10.1016 /j.rsase.2023.100947.

Boxwala, A.A. *et al*. (2011) 'Using statistical and machine learning to help institutions detect suspicious access to electronic health records', *Journal of the American Medical Informatics Association*, 18(4), pp. 498–505. Available at: https://doi.org/10.1136/amiajnl-2011-000217.

Bucherie, A. *et al*. (2022) 'A comparison of social vulnerability indices specific to flooding in Ecuador: principal component analysis (PCA) and expert knowledge', *International Journal of Disaster Risk Reduction*, 73(February), p. 102897. Available at: https://doi.org/10.1016/j.ijdrr.2022.102897.

Bui, D.T. *et al*. (2018) 'Landslide detection and susceptibility mapping by AIRSAR data using support vector machine and index of entropy models in Cameron Highlands, Malaysia', *Remote Sensing*, 10(10). Available at: https://doi.org/10.3390/rs10101527.

Dangeti, P. (2013) *Statistics for Machine Learning Build*, *Journal of Chemical Information and Modeling*.

Heydari, M. *et al*. (2020) 'Application of holt-winters time series models for predicting climatic parameters (Case study: Robat Garah-Bil station, Iran)', *Polish Journal of Environmental Studies*, 29(1), pp. 617–627. Available at: https://doi.org/10.15244/pjoes/100496.

Huang, J., Lu, J. and Ling, C.X. (2003) 'Comparing naive bayes, decision trees, and SVM with AUC and accuracy', *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 553–556. Available at: https://doi.org/10.1109/icdm.2003.1250975.

Jennifer, J.J. (2022) 'Feature elimination and comparison of machine learning algorithms in landslide susceptibility mapping', *Environmental Earth Sciences*, 81(20), pp. 1–23. Available at: https://doi.org/10.1007/s12665-022-10620-5.

Kao, I.F. *et al*. (2021) 'Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts', *Journal of Hydrology*, 598(October 2020), p. 126371. Available at: https://doi.org/10.1016/j .jhydrol.2021.126371.

Leavitt, S. V. *et al*. (2021) 'Estimating the relative probability of direct transmission between infectious disease patients', *International Journal of Epidemiology*, 49(3), pp. 764–775. Available at: https://doi.org/10.1093/ IJE/DYAA031.

Findawati, Y., Astutik, I. I., Fitroni, A. S., Indrawati, I., & Yuniasih, N. (2019). Comparative analysis of Naïve Bayes, K Nearest Neighbor and C. 45 method in weather forecast. Journal of Physics: Conference Series, 1402(6), 066046. https://doi.org/10.1088/1742-6596/1402/6/066046

Liu, J. *et al*. (2021) 'Assessment of Flood Susceptibility Using Support Vector Machine in the Belt and Road Region', *Natural Hazards and Earth System Sciences Discussions*, (May), pp. 1–37.

Choi, Y., Kang, B., & Kim, D. (2024). Utilizing Machine Learning-based Classification Models for Tracking Air Pollution Sources: A Case Study in Korea. Aerosol and Air Quality Research, 24, 230222.

Al Mehedi, M.A. *et al*. (2023) 'Predicting the performance of green stormwater infrastructure using multivariate long short-term memory (LSTM) neural network', *Journal of Hydrology*, 625(PA), p. 130076. Available at: https://doi.org/10.1016/j.jhydrol.2023.130076.

Oikonomou, E.K. and Khera, R. (2023) 'Machine learning in precision diabetes care and cardiovascular risk prediction', *Cardiovascular Diabetology*, 22(1), pp. 1–16. Available at: https://doi.org/10.1186/s12933-023-01985-3.

Pradeep, K.R. and Naveen, N.C. (2018) 'Lung Cancer Survivability Prediction based on Performance Using Classification Techniques of Support Vector Machines, C4.5 and Naive Bayes Algorithms for Healthcare Analytics', *Procedia Computer Science*, 132, pp. 412–420. Available at: https://doi.org/10.1016/j.procs.2018.05.162.

Pradhan, B. (2013) 'Computers & Geosciences A

Vol.4, Issue.2, pp.19-27, 2024
Available online at: http://tljes.org/index.php/tljes/data

comparative study on the predictive ability of the decision tree , support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS', *Computers and Geosciences*, 51, pp. 350–365. Available at: https://doi.org/10.1016/j.cageo.2012.08.023.

Rosa, M.L., Sobreira, F.G. and Barella, C.F. (2021) 'Landslide susceptibility mapping using the statistical method of information value: A study case in Ribeirão dos Macacos basin, Minas Gerais, Brazil', *Anais da Academia Brasileira de Ciencias*, 93(1), pp. 1–19. Available at: https://doi.org/10.1590/0001-3765202120180897.

Rudra, R.R. and Sarkar, S.K. (2023) 'Artificial neural network for flood susceptibility mapping in Bangladesh', *Heliyon*, 9(6), p. e16459. Available at: https://doi.org/10. 1016/j.heliyon.2023.e16459.

Saha, T.K. and Pal, S. (2019) 'Exploring physical wetland vulnerability of Atreyee river basin in India and Bangladesh using logistic regression and fuzzy logic approaches', *Ecological Indicators*, 98(November 2018), pp. 251–265. Available at: https://doi.org/10.1016/j.ecolind.2018.11.009.

Shahzad, N., Ding, X. and Abbas, S. (2022) 'A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan', *Applied Sciences (Switzerland)*, 12(5). Available at: https://doi.org/10.3390/app12052280.

Shinohara, E.J. (2012) 'Assessment of SVM classification process for landslides identification Assessment of SVM classification process for landslides identification', (July 2014).

Stephen, M. (2014) *Machine Learning An Algorithmic Perspective Second Edition*. Available at: https://b-ok.cc/book/2543746/ef80cb.

Tang, X. *et al*. (2020) 'Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods', *Catena*, 190(March), p. 104536. Available at: https://doi.org/10.1016/j.catena.2020.104536.

Xu, H. *et al*. (2018) 'Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China', *Journal of Hydrology*, 563(March), pp. 975–986. Available at: https://doi.org/10.1016/j.jhydrol.2018.06.060.

Zaki, M.J. and Meira, M.J. (2013) *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Available at: https://books.google.com.tr/books?id=Gh9GAwAAQBAJ&lpg=PR9&dq=Data Mining and Analysis: FoundationsandAlgorithms&hl=t r&pg=PR9#v=onepage &q=Data Mining and Analysis: Foundations and Algorithms&f=false.