BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 0877–0892
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp0877-0892

IMPLEMENTATION OF MAPPING-BASED MACHINE LEARNING ALGORITHM AS NON-STRUCTURAL DISASTER MITIGATION TO DETECT LANDSLIDE SUSCEPTIBILITY IN TAKARI DISTRICT

Sefri Imanuel Fallo[1*], Lidia Paskalia Nipu[2]

[1]*Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Universitas San Pedro*
[2]*Environmental Engineering Study Program, Faculty of Engineering and Planning, Universitas San Pedro*
Jln. Ir. Soekarno, Kupang, 85112, Indonesia

Corresponding author's e-mail: ***fallosefriimanuel@gmail.com***

ABSTRACT

Article History:

Received: 21st November 2023
Revised: 7th January 2024
Accepted: 10th March 2024
Published: 1st June 2024

Keywords:

Machine Learning;
Landslide Susceptibility;
Non-Structural Mitigation;
Mapping.

This research is primarily dedicated to providing a comprehensive exposition of the methodology applied in the deployment of a cartographic-based machine learning algorithm designed for the precise identification of areas susceptible to landslides within the geographical confines of the Takari District. This research delves into the application of mapping-based machine learning algorithms in the domain of non-structural disaster mitigation, with a specific emphasis on the detection of landslide susceptibility within the Takari District. A range of machine learning algorithms, including Support Vector Machine, Naive Bayes Classifier, Ordinal Logistic Regression, Random Forest, and Decision Tree, were harnessed to evaluate rainfall data within the context of landslide susceptibility. An evaluation of model performance, anchored in accuracy and Kappa metrics, unveiled that both the Ordinal Logistic Regression and Random Forest models exhibited noteworthy precision, reaching a commendable 74.36%. Nevertheless, a meticulous examination of Kappa values disclosed the ascendancy of the Random Forest model, which achieved a superior Kappa value of 0.5397. As portrayed in the visual representation provided, it becomes manifest that the Random Forest algorithm's prognostications yield 66 instances of cloudy atmospheric conditions, 48 occurrences of light precipitation, and 3 episodes of moderate rainfall. These predictions are influenced by several factors, including average temperature, humidity levels, wind speed, duration of sunlight, and wind direction at maximum speed. Consequently, this comprehensive analysis underscores the Random Forest algorithm as the most efficacious model for landslide susceptibility prediction. Furthermore, the study seamlessly integrated overlay maps, encompassing the Slope Inclination Map of the Takari District, Geological Map of the Takari District, and Soil Type Map of the Takari District, to contribute to the formulation of a definitive map delineating the susceptibility to landslides in the Takari District. Furthermore, further research could conduct spatial validation of the model predictions using additional datasets or remote sensing data to validate the accuracy of the landslide susceptibility map and ensure its applicability across different geographical regions.

How to cite this article:

S. I Fallo and L. P. Nipu., "IMPLEMENTATION OF MAPPING-BASED MACHINE LEARNING ALGORITHM AS NON-STRUCTURAL DISASTER MITIGATION TO DETECT LANDSLIDE SUSCEPTIBILITY IN TAKARI DISTRICT," *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 0877-0892, June, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Natural disasters are unpredictable events that can occur at any time and in any location, resulting in both tangible and intangible losses for society. Indonesia, in particular, is highly susceptible to various types of disasters, including earthquakes, tsunamis, floods, and landslides [1]. Among these, landslides are the most prevalent and widespread disaster, affecting nearly every region in Indonesia, particularly during periods of heavy rainfall. This is primarily due to the fact that approximately 45% of the country's land area consists of mountainous terrain that is highly vulnerable to landslides and erosion [2]. Typically occurring in hilly areas, especially during the rainy season, landslides can cause significant damage to properties, loss of life, and destruction of infrastructure [3]. The occurrence of landslides is often attributed to land instability, which is frequently exacerbated by human activities such as improper land use, including constructing settlements on slopes or excessive soil or sand extraction in lower areas [4].

These triggers for landslides are further intensified by intense and prolonged rainfall, which saturates the soil, reducing its ability to retain water and resulting in downward movement. Numerous factors, including rainfall, topography (slope and length of slope), vegetation, soil type, soil texture, soil moisture, and geology, contribute to the occurrence of landslides. Large-scale landslides can have catastrophic consequences, causing extensive damage to buildings, roads, bridges, and settlements [5]. Presently, landslides have become a widespread issue across Indonesia, leading to loss of life, property damage, and long-term impacts on land and livelihoods. Tehrani highlights several parameters that contribute to landslides, including rainfall, topography (slope and length of slope), vegetation, soil type, soil texture, soil moisture, and geology [6].

Takari District, located in Kupang Regency, East Nusa Tenggara Province (NTT), is situated in a mountainous area highly prone to landslides, particularly during periods of heavy rainfall [2]. Recently, a landslide disaster occurred in Takari Subdistrict on Friday, February 17, 2023, resulting in the complete disruption of transportation access connecting South Central Timor Regency, North Central Timor Regency, Belu Regency, Malaka Regency, and East Timor. Climate change and unpredictable high-intensity rainfall are believed to be the triggering factors for landslides in Takari Subdistrict. This disaster has had a significant impact on the tourism sector, particularly in terms of road accessibility and safety [7]. Tourists express concerns and apprehension about visiting destinations that are prone to such disasters [8]. Undoubtedly, Takari Subdistrict serves as a crucial route to reach tourist destinations in South Central Timor Regency, North Central Timor Regency, Belu Regency, and Malaka Regency. Therefore, early detection of landslide vulnerability through mapping-based machine learning algorithms becomes essential as a means of disaster mitigation.

This research paper focuses on elucidating the methodology employed to implement a mapping-based machine learning algorithm for the detection of landslide susceptibility in Takari District. The outcomes of the analysis, which delineate the areas identified as high-risk zones, will be presented. Additionally, the implications of these findings will be discussed, along with potential strategies for disaster mitigation and risk reduction specifically tailored for Takari District.

The primary objective of this study is to contribute to the field of disaster management by showcasing the efficacy of mapping-based machine learning algorithms in identifying landslide susceptibility [9]. The results obtained from this research will serve as a valuable resource for policymakers, researchers, and practitioners involved in disaster risk reduction and land-use planning. Equipped with this knowledge, stakeholders will be able to make informed decisions aimed at safeguarding lives and minimizing the adverse impacts of landslides in Takari District.

2. RESEARCH METHODS

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. 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 [7].

The analysis technique used in this research is Machine Learning. Machine Learning, a subfield of artificial intelligence, facilitates computer systems to acquire knowledge from data and generate predictions or decisions without explicit programming [10]. By recognizing patterns and trends within data, Machine Learning empowers computer systems to autonomously learn and adapt. Over the past few decades, Machine Learning has emerged as a prominent and influential technology. Its effectiveness and efficiency in solving problems have been demonstrated across various domains, including facial recognition, sentiment analysis, and anomaly detection [3]. Machine Learning distinguishes itself from traditional approaches that rely on explicit programming. Instead, it enables computer systems to learn from available data and create models capable of making predictions or decisions on unseen data. Within the realm of research, Machine Learning has become an invaluable tool. Leveraging techniques and algorithms specific to Machine Learning, researchers can analyze complex data, unveil latent patterns, and generate accurate predictions. Consequently, researchers can gain novel insights, address intricate research inquiries, and solve complex problems [11]. In other words, Machine Learning is a powerful technology that enables computer systems to learn from data and make predictions or decisions. Its capacity to recognize patterns and adapt autonomously has established its significance across various fields. Within research, Machine Learning facilitates the analysis of intricate data, the identification of concealed 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.

Machine Learning holds significant promise in facilitating the mitigation of non-structural disasters, encompassing both natural phenomena such as landslides, floods, droughts, and earthquakes, as well as anthropogenic calamities like forest fires, air pollution, and social conflicts [12]. In this particular context, Machine Learning can be effectively harnessed to discern patterns and trends within data pertaining to non-structural disasters, thereby enabling enhanced modeling, prediction, and decision-making processes. Moreover, Machine Learning can play a pivotal role in expediting and enhancing the accuracy of decision-making during emergency scenarios. By leveraging Machine Learning algorithms, computer systems possess the capability to analyze real-time data derived from diverse sources, including weather and fire sensors, as well as social media platforms, thereby furnishing pertinent and precise information to authorities for the purpose of implementing effective mitigation strategies [13]. In summation, Machine Learning harbors considerable potential in supporting the mitigation of non-structural disasters. Its aptitude for recognizing patterns and trends within data, coupled with its ability to generate accurate predictions and models, renders it an invaluable tool in non-structural disaster mitigation endeavors. Through appropriate implementation, Machine Learning stands poised to make a significant contribution towards the preservation of societal well-being and environmental integrity in the face of non-structural disasters.

2.1. Support Vector Machine

Support Vector Machines (SVM) is a machine learning methodology that is founded upon the principle of Structural Risk Minimization (SRM) [10]. Its fundamental aim is to discern an optimal hyperplane within the input space, which can effectively segregate different classes [14]. 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. The Support Vector Machine algorithm can be implemented through the following steps:

1. Partitioning the data into testing and training datasets.
2. Determining the cost value.
3. Establishing the values of standard deviation and alpha.
4. Applying the RBF kernel.

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

5. Establishing the predicted outcome using RBF Kernel

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

2.2. Naïve Bayes Classifier

The Naïve Bayes Classifier is a machine learning algorithm employed for classification purposes, leveraging probabilistic principles [15]. It operates on the basis of Bayes' theorem and makes the assumption 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. The Naïve Bayes Classifier algorithm can be executed through the following sequential procedures [16]:

1. Computation of the probabilities associated with each pre-defined class.
2. Calculation of the mean and standard deviation pertaining to each individual feature.
3. Identification of the training and testing datasets.
4. Prediction assessment by inputting the testing data into the Gaussian density function.
5. Subsequent evaluation of the likelihood.
6. Determination of the posterior probability.
7. Subsequent normalization of the probability values to ensure consistency and comparability.

$$
\begin{aligned}
NP\left(Y = 1 \mid X_i\right) &= \frac{P\left(Y = 1 \mid X_i\right)}{P\left(Y_1 \mid X_i\right) + P\left(Y_2 \mid X_i\right) + \cdots + P\left(Y_n \mid X_i\right)} \\
NP\left(Y = 2 \mid X_i\right) &= \frac{P\left(Y = 2 \mid X_i\right)}{P\left(Y_1 \mid X_i\right) + P\left(Y_2 \mid X_i\right) + \cdots + P\left(Y_n \mid X_i\right)} \\
&\vdots \\
NP\left(Y = n \mid X_i\right) &= \frac{P\left(Y = n \mid X_i\right)}{P\left(Y_1 \mid X_i\right) + P\left(Y_2 \mid X_i\right) + \cdots + P\left(Y_n \mid X_i\right)}
\end{aligned} \tag{3}
$$

8. Establishing the predicted outcome

$$
Y_{MAP} = \arg\max_{Y_j \in Y}\left(P\left(Y_j\right)\prod_{i=1}^{n} P\left(X_i \mid Y_j\right)\right) \tag{4}
$$

2.3. Ordinal Logistic Regression

Ordinal Logistic Regression is a statistical modeling approach utilized to examine and predict the relationship between an ordinal dependent variable and one or more independent variables [17]. In the context of ordinal logistic regression, the dependent variable encompasses multiple ordered categories, such as rating scales or Likert scales. The model estimates the cumulative probabilities associated with each category, relative to a reference category, while accounting for the ordinal nature of the outcome variable [18]. The Ordinal Logistic Regression algorithm can be executed through the following steps:

1. Examination of the Multicollinearity Assumption.
2. Determination of the logit model.

$$\begin{aligned} Logit\ \mathrm{P}(Y \le 1 \mid x) = g_1(x) = \log\left(\frac{\mathrm{P}(Y \le 1 \mid x)}{\mathrm{P}(Y > 1 \mid x)}\right) \\ = \alpha_1 - \sum_{k=1}^{p} \beta_k x_k \\ \vdots \\ Logit\ \mathrm{P}(Y \le n \mid x) = g_n(x) = \log\left(\frac{\mathrm{P}(Y \le n \mid x)}{\mathrm{P}(Y > n \mid x)}\right) \\ = \alpha_n - \sum_{k=1}^{p} \beta_k x_k \end{aligned} \tag{5}$$

3. Estimation of cumulative probabilities.

$$\begin{aligned} \mathrm{P}(Y \le 1 \mid x) = \frac{\exp\left(\alpha_1 - \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\alpha_1 - \sum_{k=1}^{p} \beta_k x_k\right)} \\ \vdots \\ \mathrm{P}(Y \le n \mid x) = \frac{\exp\left(\alpha_n - \sum_{k=1}^{p} \beta_k x_k\right)}{1 + \exp\left(\alpha_n - \sum_{k=1}^{p} \beta_k x_k\right)} \end{aligned} \tag{6}$$

4. Calculation of probability values.
5. Comparison of probability values to ascertain the predicted outcome.

2.4. Random Forest

Random Forest is a machine learning algorithm that combines multiple decision trees to make predictions. Each tree is built on a random subset of the training data and features, and the algorithm selects the best split points based on certain criteria [13]. The final prediction is made by aggregating the predictions from all the trees. Random Forest is known for its ability to handle high-dimensional data and reduce overfitting [19]. It is widely used in various fields, such as finance, healthcare, and image recognition. The Random Forest algorithm can be implemented through the following steps:

1. Selection of k samples from the dataset D, randomly and with replacement.
2. Utilizing the dataset D to construct the i-th decision tree.
3. In constructing the i-th decision tree, the CART methodology can be employed. The CART methodology employs information gain to determine each node in the tree. The calculation of information gain can be computed using the Equation.

$$Gain(A) = Info(D) - Info_A(D) \tag{7}$$

   The value of $Info(D)$ can be determined using the formula $Info_A(D)$, as presented below:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{8}$$

4. In order to predict the target class of the test data using the Random Forest algorithm, the test data is passed through the rules established by each individual tree. The predictions generated by each tree may

differ or coincide. Therefore, the final prediction is determined by selecting the class that is most frequently predicted among the trees.

2.5. Decision Tree

A decision tree is a predictive model that employs a hierarchical structure resembling a tree to make informed decisions or predictions based on input features [20]. It is a supervised learning algorithm that recursively divides the data based on the values of different features, resulting in a tree-like structure with nodes representing decision points and edges signifying potential outcomes. At each node, a decision is made by evaluating a specific feature value, leading to subsequent nodes or leaf nodes, which represent the ultimate prediction or decision [21]. The Decision Tree algorithm can be implemented through the following steps:

1. The initial step involves identifying the parent node for the decision tree.
2. In the context of Decision Trees, entropy serves as a metric for assessing the degree of randomness within the information being analyzed. Higher entropy values indicate a greater level of uncertainty and complexity in drawing meaningful conclusions from the available information.

$$H(x) = -\sum_{i=1}^{m} \mathrm{P}(\mathrm{x}_i)\log_2 \mathrm{P}(\mathrm{x}_i) \tag{9}$$

3. Information gain refers to the measure of the information acquired pertaining to a random variable or signal through the observation of another random variable. It quantifies the disparity between the entropy of the parent node and the weighted average entropy of the child nodes. This metric serves as an indicator of the extent to which the introduction of additional variables contributes to the overall understanding and predictive power of the model.

$$IG(S,A) = H(S) - \sum_{i=0}^{n} P(x)H(x) \tag{10}$$

4. Gini impurity is a mathematical measure that quantifies the frequency at which a randomly selected element from a given set would be inaccurately classified if it were randomly labeled based on the distribution of labels within a specific subset [20]. This metric provides insights into the level of impurity or disorder within the dataset, with lower values indicating a higher degree of purity and more accurate labeling.

$$Gini(E) = 1 - \sum_{j=1}^{c} p_j^2 \tag{11}$$

5. Choose one that has a higher Gini gain. Gini gain is higher for outlook. So we can choose it as our root node.

2.6. Non-Structural Disaster Mitigation

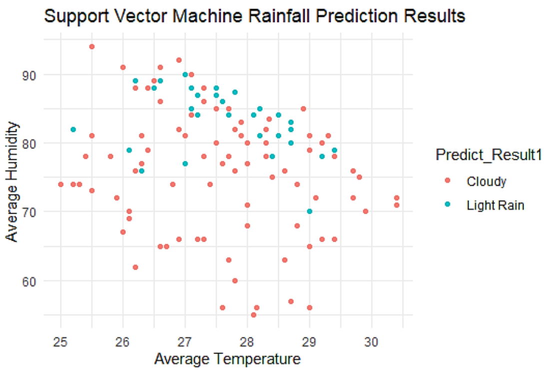
Non-structural disaster mitigation pertains to the application of tactics and approaches that do not entail the physical modification or construction of infrastructures as a means to mitigate the consequences of both natural and man-made disasters [22]. In contrast to relying exclusively on solutions rooted in physical structures, non-structural methodologies concentrate on alternative facets, including policy formulation, planning initiatives, educational endeavors, and community involvement. At this stage, mitigation is carried out through mapping based on the observations made in the field [23]. The mapping process involves assessing slope classes, soil types, geology, and evaluating the predictions of rainfall data obtained from testing the best-fit model using machine learning algorithms. Subsequently, an overlay is performed on the generated map, which yields the final landslide vulnerability map.

3. RESULTS AND DISCUSSION

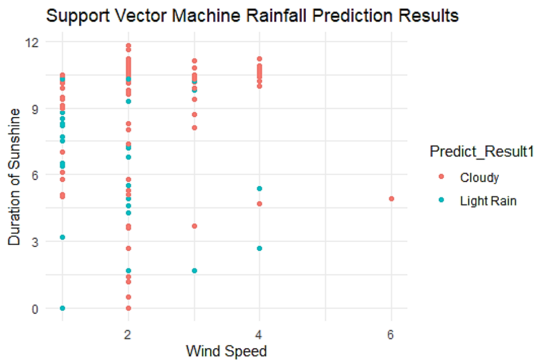
This segment represents the culmination of our research, where we unveil the results of our endeavors in utilizing mapping-based machine learning algorithms as a non-structural approach to disaster mitigation, specifically focused on the identification of landslide susceptibility within the Takari District. 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 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

3.1 Support Vector Machine

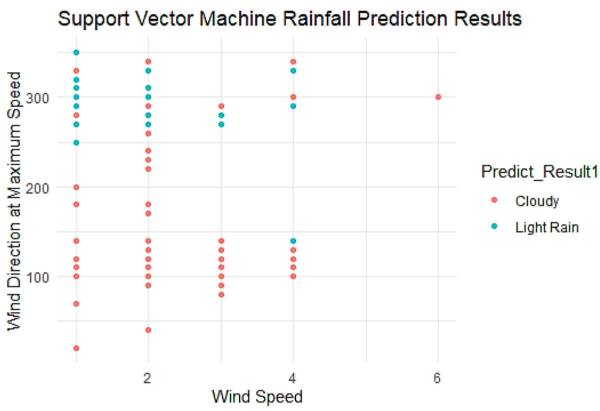
The multiclass SVM algorithm employs a radial kernel, which is known for its high accuracy [24]. The output results of this algorithm are presented in the figure below.



(a)



(b)



(c)

Figure 1. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the Support Vector Machine (SVM) Algorithm

Based on the illustration above, it can be elucidated that the SVM prediction yields 87 instances of overcast conditions and 30 instances of light rain. Several factors influencing these occurrences include average temperature, average humidity, wind speed, duration of sunshine, and wind direction at maximum speed. The comprehensive statistical outcomes using the SVM algorithm are elucidated in the table below.

Table 1. The Comprehensive Statistical Summary Derived from The SVM Algorithm

Overall statistics	Value
Accuracy	0.6581
95% CI	(0.5647, 0.7433)
No Information Rate	0.7436
P-Value [Acc > NIR]	0.9849
Kappa	0.2764

3.2 Naïve Bayes Classifier

Bayes' theorem underpins one of the most straightforward supervised classification methods [15]. 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.

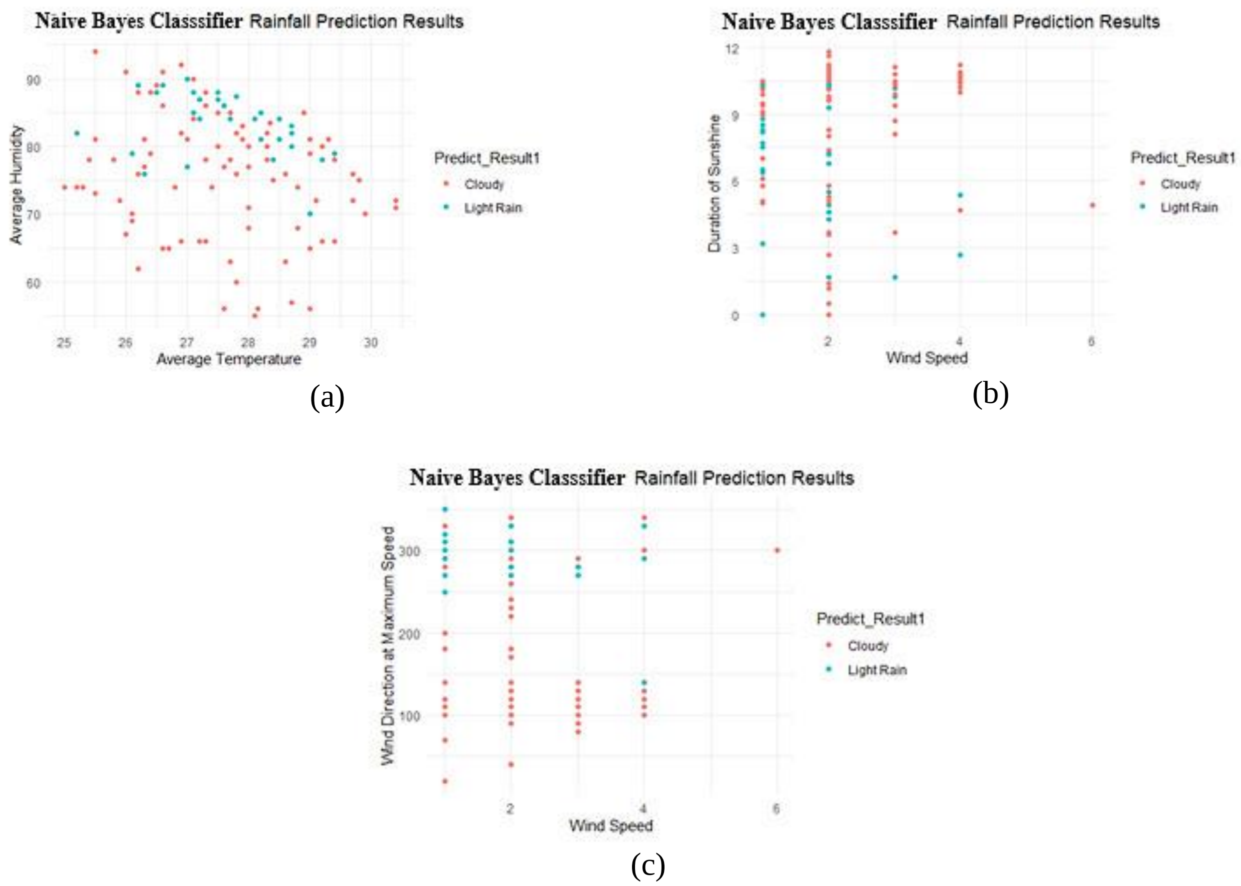


Figure 2. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the Naïve Bayes Classifier Algorithm

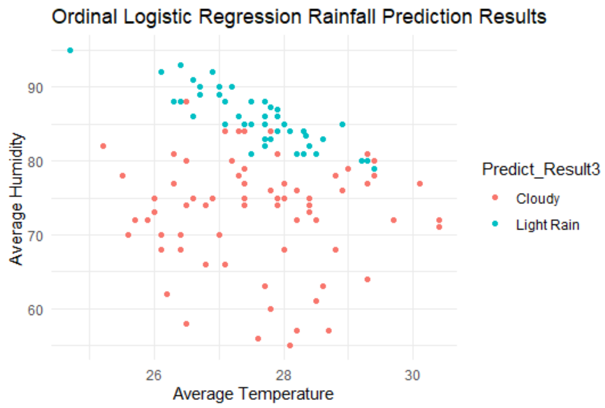
Drawing insights from the provided illustration, it is discernible that the Naïve Bayes Classifier's prognostications manifest as follows: 60 instances of overcast conditions, 7 occurrences of substantial rainfall, 43 episodes of light precipitation, 1 instance of exceptionally heavy downpour, and 5 cases of moderate rainfall. These prognostications are intricately influenced by various determinants, including the mean temperature, average humidity levels, wind velocity, duration of solar irradiation, and the prevailing wind direction at maximum speeds. A comprehensive exposition of the statistical outcomes engendered through the utilization of the Naïve Bayes Classifier algorithm is duly delineated in the ensuing table.

Table 2. The Comprehensive Statistical Summary Derived from Naïve Bayes Classifier Algorithm

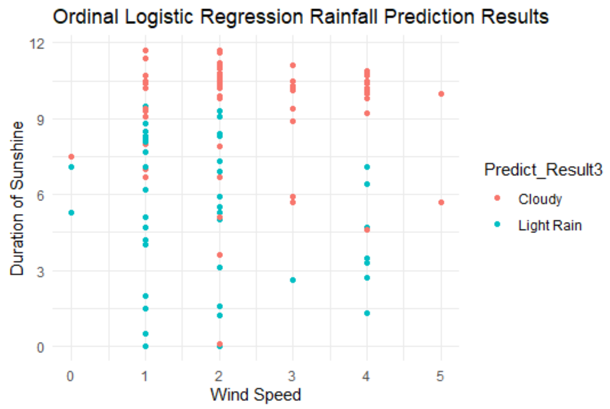
Overall statistics	Value
Accuracy	0.6466
95% CI	(0.5524, 0.7331)
No Information Rate	0.5172
P-Value [Acc > NIR]	0.003328
Kappa	0.3815

3.3 Ordinal Logistic Regression

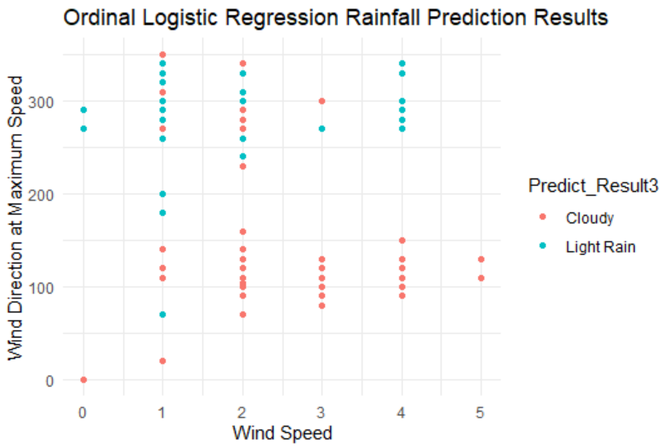
Ordinal logistic regression is a type of regression analysis used when the dependent variable is ordinal, meaning it has ordered categorical values. The result present below:



(a)



(b)



(c)

Figure 3. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the Ordinal Logistic Regression Algorithm

Based on the illustration above, it can be elucidated that the Ordinal Logistic Regression prediction results in 68 instances of overcast conditions, 48 instances of light rain, and 1 instance of moderate rain. Several factors influencing these occurrences include average temperature, average humidity, wind speed, duration of sunshine, and wind direction at maximum speed. The comprehensive statistical outcomes using the Ordinal Logistic Regression algorithm are detailed in the table below.

Table 3. The Comprehensive Statistical Summary Derived from Ordinal Logistic Regression Algorithm

Overall statistics	Value
Accuracy	0.7436
95% CI	(0.6546, 0.8198)
No Information Rate	0.5812
P-Value [Acc > NIR]	0.0001874
Kappa	0.5208

3.4 Random Forest

Random Forest is a powerful and versatile machine learning algorithm that belongs to the ensemble learning category. The result presented below:

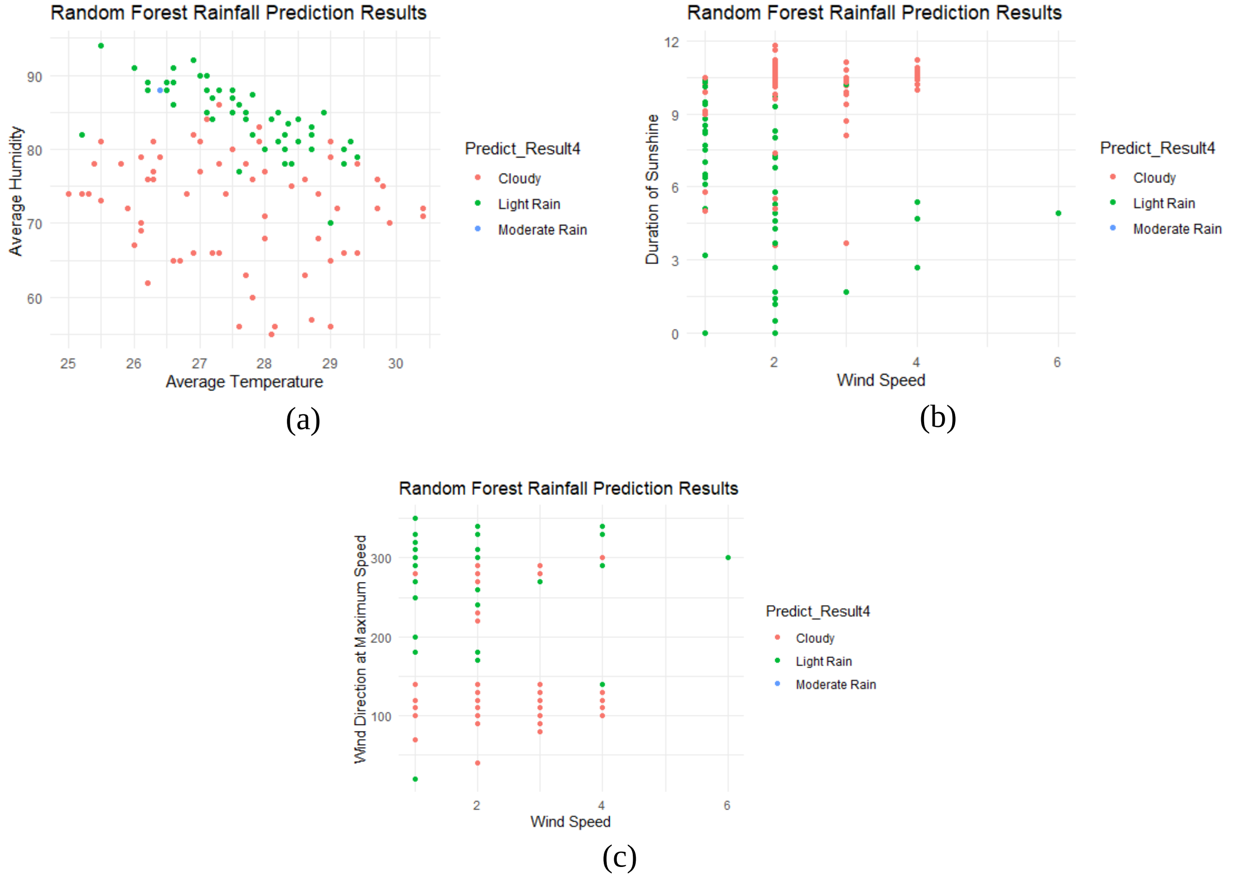


Figure 4. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the Random Forest Algorithm

As depicted in the image above, it can be elucidated that the Random Forest prediction results in 66 cases of overcast weather, 48 instances of light rain, and 3 occurrences of moderate rainfall. These predictions are influenced by several factors, including average temperature, humidity levels, wind speed, duration of sunlight, and wind direction at maximum speed. A detailed breakdown of the statistical outcomes generated by the Random Forest algorithm can be found in the table presented below.

Table 4. The Comprehensive Statistical Summary Derived from Random Forest Algorithm

Overall statistics	Value
Accuracy	0.7436
95% CI	(0.6546, 0.8198)
No Information Rate	0.5641
P-Value [Acc > NIR]	4.398e-05
Kappa	0.5397

3.5 Decision Tree

A decision tree is a popular and intuitive machine learning model used for both classification and regression tasks. It's a tree-like structure where each internal node represents a "decision" based on a feature, each branch represents the outcome of that decision, and each leaf node represents the final prediction or classification [20]

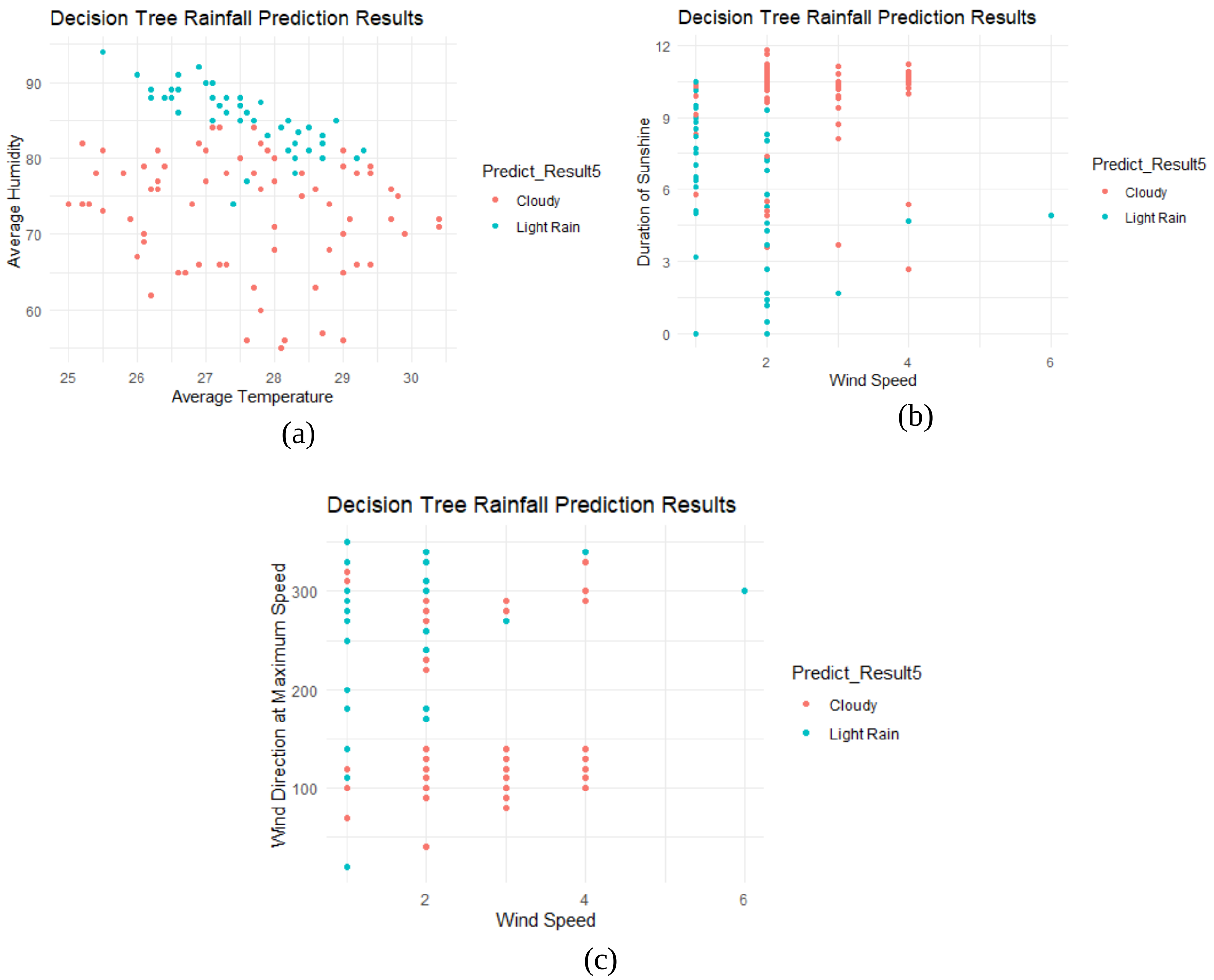


Figure 5. (a)(b)(c) Graphical Visualization Depicting the Anticipated Precipitation Predictions Achieved Through the Utilization of the Decision Tree Algorithm

According to the figure depicted above, it can be clarified that the Decision Tree prediction results in 71 cases of cloudy and 46 occurrences of light rain. These predictions are influenced by several factors, such as average temperature, average humidity, wind speed, duration of sunlight, and wind direction at maximum speed. The overall statistical results produced by the Decision Tree algorithm are detailed in the table presented below.

Table 5. The Comprehensive Statistical Summary Derived from Decision Tree Algorithm

Overall statistics	Value
Accuracy	0.7009
95% CI	(0.6093, 0.782)
No Information Rate	0.6068
P-Value [Acc > NIR]	0.02207
Kappa	0.429

3.6 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 appraising rainfall patterns for the purpose of identifying regions susceptible to landslides within the Takari district, the resulting model evaluation is delineated herewith:

Table 6. Model Evaluation

Machine Learning Algorithm	Accuracy	Kappa
Multiclass Support Vector Machine	0.6581	0.2764
Naïve Bayes Classifier	0.6466	0.3815
Ordinal Logistic Regression	0.7436	0.5208
Random Forest	0.7436	0.5397
Decision Tree	0.7009	0.4290

In light of the model evaluation outcomes, predicated upon the examination of accuracy and Kappa metrics, it becomes apparent that both the Ordinal Logistic Regression and Random Forest models manifest the most robust accuracy, standing at 74.36%. Nonetheless, a juxtaposition of the Kappa values across these algorithmic models divulges that the Random Forest model attains the zenith Kappa value, amounting to 0.5397. Consequently, it may be inferred that the preeminent model, ascertained through a comprehensive scrutiny of accuracy and Kappa metrics, is the Random Forest algorithm.

3.7 Mapping the Geographic Condition in Takari District

Upon completion of the slope gradient map, which was generated based on the categorization of slope classes, the subsequent task involves identifying the susceptible landslide areas through precise measurements of cliff formations during on-site surveys. A number of locations exhibiting vulnerability to landslides were discovered within Noelmina Village, Oesusu Village, and Takari Sub-district.

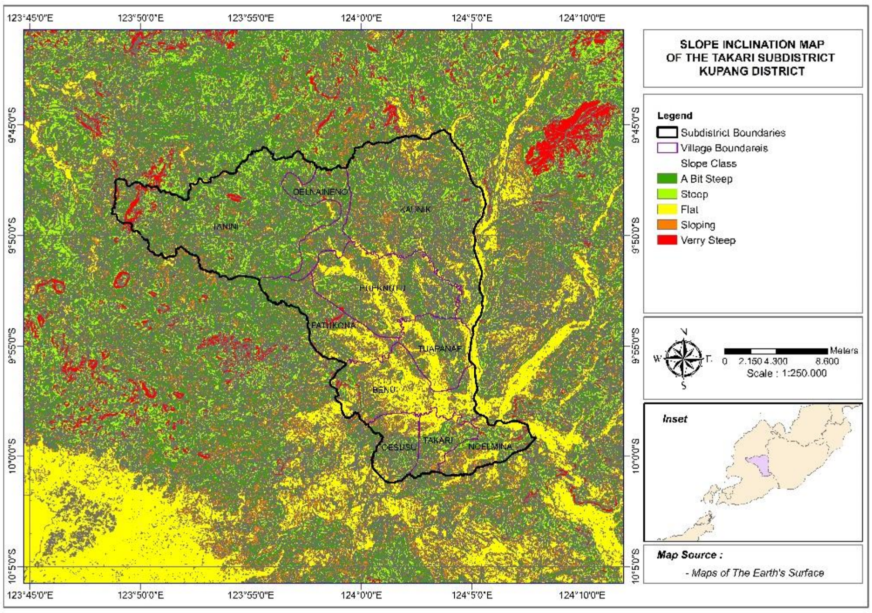


Figure 6. Slope Inclination Map of the Takari District

It is worth noting that certain sites remain classified as level terrains, thereby mitigating any associated hazards. The spatial distribution of these aforementioned locations is visually depicted in the cartographic representation presented below.

Furthermore, the inclusion of a geological map is imperative in the detection of landslide-prone areas [24]. Geological mapping serves the purpose of identifying rock formations, geological structures, and hydrogeological conditions prevalent within a given region. The data derived from such mapping exercises can be effectively employed in landslide vulnerability modeling, facilitating the identification of susceptible regions through a comprehensive analysis of their geological attributes.

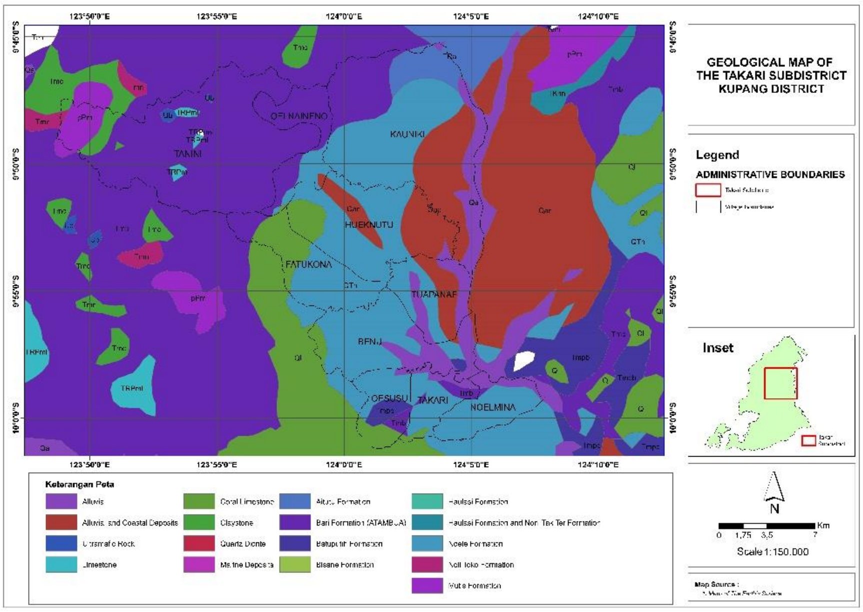


Figure 7. **Geological Map of the Takari District**

By utilizing machine learning algorithms and geological data as attributes, the identification of landslide vulnerability can be executed with heightened efficacy and precision. This approach enables the utilization of a more comprehensive range of geological information in order to predict the degree of landslide susceptibility and facilitate informed decision-making regarding disaster mitigation strategies **[6]**.

The composition of soil plays a pivotal role in determining landslide susceptibility. Various soil types, such as clay, loam, and fine-grained soil, exhibit a predisposition to instability and are prone to landslide occurrences during periods of intense precipitation or seismic activity. Conversely, soil types characterized by greater stability, such as sandy or coarse-grained soil, display a heightened resistance to landslides. Soils that possess low density, high moisture content, and high water retention capacity tend to exhibit a higher degree of susceptibility to landslides. Moreover, soil texture, clay content, and groundwater saturation also exert an influence on soil stability and the potential for landslides. Presented below is a mapping of soil types within the Takari Sub-district.

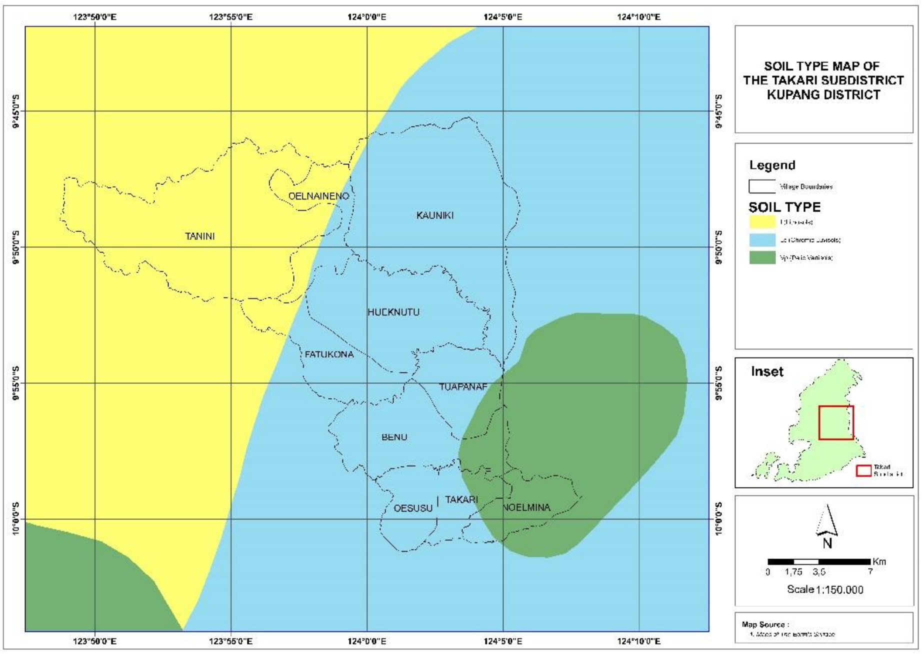


Figure 8. **Soil Type Map of the Takari District**

Comprehending the various soil types and their influence on landslide susceptibility is of paramount importance in landslide disaster mitigation. By incorporating the consideration of soil types in the analysis and mapping of landslide vulnerability, appropriate mitigation measures can be implemented to safeguard areas prone to landslides.

3.8 Non-Structural Disaster Mitigation

Following a thorough geospatial analysis of slope inclination, geological composition, and soil types, the creation of a landslide susceptibility map for the Takari district was undertaken. This endeavor centered on the integration of data from the aforementioned three thematic maps within the ArcGIS platform. The resultant cartographic representation categorizes the susceptibility to landslides into four distinct gradations, namely: low, moderate, high, and very high. The produced landslide susceptibility map is visually presented in the subsequent diagram.

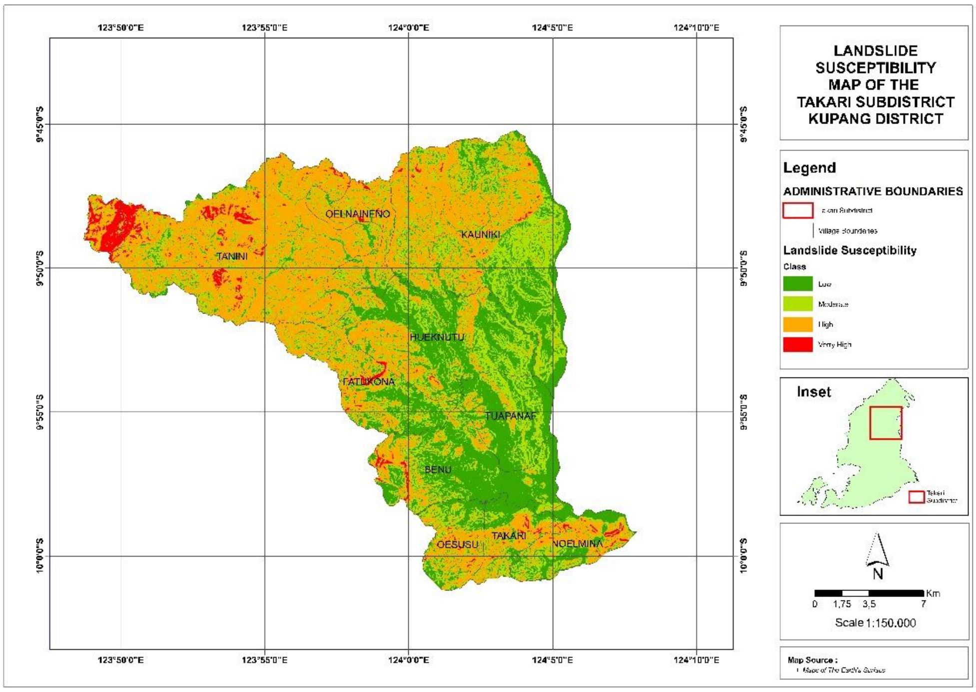


Figure 8. Landslide Susceptibility Map of the Takari District

Based on the final mapping outcomes, areas characterized by high landslide susceptibility are represented in orange, while the regions with the highest susceptibility are depicted in red. According to the susceptibility map, the areas falling within the category of high and very high landslide susceptibility encompass Noelmina, Takari, Oesusu, Fatukona, Tanini, and Oelnaino. Vigilance is particularly crucial in these areas during the rainy season.

Drawing from the model evaluation results of machine learning algorithms, the Random Forest algorithm emerges as the most robust model. The predictive output for rainfall data testing generated by this algorithm anticipates 48 instances of light rain and 3 occurrences of moderate rainfall. It is imperative to underscore the significance of this prediction; continuous and heavy rainfall poses a substantial hazard to regions already identified as having a high and very high susceptibility to landslides.

4. CONCLUSIONS

In light of the evaluation of model performance, which hinged on the analysis of accuracy and Kappa metrics, it becomes evident that both the Ordinal Logistic Regression and Random Forest models exhibit the highest level of accuracy, standing at 74.36%. However, a comparison of Kappa values between these algorithmic models reveals that the Random Forest model achieves the highest Kappa value, reaching a peak value of 0.5397. Consequently, it can be inferred that, through a comprehensive assessment of accuracy and Kappa metrics, the Random Forest algorithm emerges as the most superior model.

Following an extensive geospatial analysis encompassing slope inclination, geological composition, and soil types, the development of a landslide susceptibility map for the Takari district was embarked upon. This undertaking revolved around the integration of data from the aforementioned three thematic maps within the ArcGIS platform. The resulting cartographic representation classifies landslide susceptibility into four distinct categories: low, moderate, high, and very high.

Based on the ultimate mapping results, areas displaying heightened susceptibility to landslides are denoted in orange, while those with the utmost susceptibility are rendered in red. The regions falling within the high and very high landslide susceptibility categories comprise Noelmina, Takari, Oesusu, Fatukona, Tanini, and Oelnaino. Vigilance in these regions is of paramount importance, particularly during the rainy season.

Drawing insights from the model evaluation outcomes of machine learning algorithms, the Random Forest model emerges as the most resilient model. Its predictions for rainfall data testing foretell 48 instances of light rain and 3 occurrences of moderate rainfall. It is imperative to underscore the significance of this prediction: continuous and heavy rainfall poses a substantial hazard to regions already identified as having a high and very high susceptibility to landslides.

REFERENCES

[1] M. Heydari, H. B. Ghadim, M. Rashidi, and M. Noori, “Application of holt-winters time series models for predicting climatic parameters (Case study: Robat Garah-Bil station, Iran),” *Polish J. Environ. Stud.*, vol. 29, no. 1, pp. 617–627, 2020, doi: 10.15244/pjoes/100496.

[2] N. W. Himalayas and J. Rohn, “Using PCA in evaluating event-controlling attributes of landsliding in the 2005 Kashmir earthquake region ,” *Nat. Hazards*, vol. 81, no. 3, pp. 1999–2017, 2017, doi: 10.1007/s11069-016-2172-9.

[3] I. Riadi, R. Umar, and F. D. Aini, “Analisis Perbandingan Detection Traffic Anomaly Dengan Metode Naive Bayes Dan Support Vector Machine (Svm),” *Ilk. J. Ilm.*, vol. 11, no. 1, pp. 17–24, 2019, doi: 10.33096/ilkom.v11i1.361.17-24.

[4] E. Yesilnacar and T. Topal, “Landslide susceptibility mapping : A comparison of logistic regression and neural networks methods in a medium scale study , Hendek region (Turkey),” vol. 79, pp. 251–266, 2005, doi: 10.1016/j.enggeo.2005.02.002.

[5] C. Highway, “Landslide Susceptibility Assessment Using Integrated Deep Learning Algorithm along the China-Nepal Highway,” 2018, doi: 10.3390/s18124436.

[6] M. Ho Park, M. Ju, S. Jeong, and J. Young Kim, “Incorporating interaction terms in multivariate linear regression for post-event flood waste estimation,” *Waste Manag.*, vol. 124, pp. 377–384, 2021, doi: 10.1016/j.wasman.2021.02.004.

[7] F. S. Tehrani, M. Calvello, Z. Liu, L. Zhang, and S. Lacasse, *Machine learning and landslide studies: recent advances and applications*, vol. 114, no. 2. Springer Netherlands, 2022. doi: 10.1007/s11069-022-05423-7.

[8] Z. Guo, Y. Shi, F. Huang, X. Fan, and J. Huang, “Geoscience Frontiers Landslide susceptibility zonation method based on C5 . 0 decision tree and K-means cluster algorithms to improve the efficiency of risk management,” *Geosci. Front.*, vol. 12, no. 6, p. 101249, 2021, doi: 10.1016/j.gsf.2021.101249.

[9] M. S. G. Adnan *et al.*, “A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction,” *J. Environ. Manage.*, vol. 326, no. PB, p. 116813, 2023, doi: 10.1016/j.jenvman.2022.116813.

[10] J. Huang, J. Lu, and C. X. Ling, “Comparing naive bayes, decision trees, and SVM with AUC and accuracy,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 553–556, 2003, doi: 10.1109/icdm.2003.1250975.

[11] S. V. Burger, *Introduction to Machine Learning with Python with R*. 2018. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/B9780123965028000139

[12] M. Ouyang, Y. Ito, and T. Tokunaga, “Quantifying the inundation impacts of earthquake-induced surface elevation change by hydrological and hydraulic modeling,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–13, 2021, doi: 10.1038/s41598-021-83309-7.

[13] H. Harsa *et al.*, “Machine learning and artificial intelligence models development in rainfall-induced landslide prediction,” *IAES Int. J. Artif. Intell.*, vol. 12, no. 1, pp. 262–270, 2023, doi: 10.11591/ijai.v12.i1.pp262-270.

[14] F. Huang, J. Zhang, C. Zhou, Y. Wang, J. Huang, and L. Zhu, “Faming Huang I Jing Zhang I Chuangbing Zhou I Yuhao Wang I Jinsong Huang I Li Zhu A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide sus- ceptibility prediction,” no. August 2019, pp. 217–229, 2020, doi: 10.1007/s10346-019-01274-9.

[15] S. V. Leavitt, R. S. Lee, P. Sebastiani, C. Robert Horsburgh, H. E. Jenkins, and L. F. White, “Estimating the relative probability of direct transmission between infectious disease patients,” *Int. J. Epidemiol.*, vol. 49, no. 3, pp. 764–775, 2021, doi: 10.1093/IJE/DYAA031.

[16] X. Tang, J. Li, M. Liu, W. Liu, and H. Hong, “Flood susceptibility assessment based on a novel random Naïve Bayes method: A comparison between different factor discretization methods,” *Catena*, vol. 190, no. March, p. 104536, 2020, doi: 10.1016/j.catena.2020.104536.

[17] A. A. Boxwala, J. Kim, J. M. Grillo, and L. Ohno-Machado, “Using statistical and machine learning to help institutions detect suspicious access to electronic health records,” *J. Am. Med. Informatics Assoc.*, vol. 18, no. 4, pp. 498–505, 2011, doi: 10.1136/amiajnl-2011-000217.

[18] L. Liu and X. yi You, “Water quality assessment and contribution rates of main pollution sources in Baiyangdian Lake, northern China,” *Environ. Impact Assess. Rev.*, vol. 98, no. November 2022, p. 106965, 2023, doi: 10.1016/j.eiar.2022.106965.

[19] M. H. Hasan, A. Ahmed, K. M. Nafee, and M. A. Hossen, “Use of machine learning algorithms to assess flood susceptibility

in the coastal area of Bangladesh," *Ocean Coast. Manag.*, vol. 236, no. November 2022, p. 106503, 2023, doi: 10.1016/j.ocecoaman.2023.106503.

[20] R. F. de M. M. A. Ponti, *Machine learning A Practical Approach on the Statistical Learning Theory*, vol. 45, no. 13. Springer International Publishing AG, 2018. [Online]. Available: https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA

[21] E. K. Sahin, "Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest," *SN Appl. Sci.*, vol. 2, no. 7, pp. 1–17, 2020, doi: 10.1007/s42452-020-3060-1.

[22] N. Shahzad, X. Ding, and S. Abbas, "A Comparative Assessment of Machine Learning Models for Landslide Susceptibility Mapping in the Rugged Terrain of Northern Pakistan," *Appl. Sci.*, vol. 12, no. 5, 2022, doi: 10.3390/app12052280.

[23] J. Liu *et al.*, "Assessment of Flood Susceptibility Using Support Vector Machine in the Belt and Road Region," *Nat. Hazards Earth Syst. Sci. Discuss.*, no. May, pp. 1–37, 2021.

[24] E. K. Sahin, "Implementation of free and open-source semi-automatic feature engineering tool in landslide susceptibility mapping using the machine-learning algorithms RF, SVM, and XGBoost," *Stoch. Environ. Res. Risk Assess.*, vol. 37, no. 3, pp. 1067–1092, 2023, doi: 10.1007/s00477-022-02330-y.