International Journal of Computer Sciences and Mathematics Engineering Vol. 2 No. 2 (2023)
Published by: CV. HAWARI

International Journal of Computer Sciences and Mathematics Engineering
Journal homepage: www.ijecom.org

Analysis of Artificial Intelligence Machine Learning Technology for Mapping and Predicting Flood Locations in Pahlawan Batu Bara Village

Zulham Sitorus[1], Eko Hariyanto[2], Fahmi Kurniawan[3]
[1,3] Computer System Studies Program, Faculty of Science and Technology, University Pembangunan Panca Budi, Medan, Indonesia

Article Info

Article history:

Received Nov 25, 2023
Revised Nov 27, 2023
Accepted Jan 31, 2023

Keywords:

Artificial Intelligence Technology, Machine Learning, Flooding

ABSTRACT

This research proposes the application of artificial intelligence technology, especially machine learning, to improve flood predictions. Flooding is a serious threat that can cause major losses to society and the environment. In an effort to overcome this problem, machine learning methods are used to analyze historical data related to weather, rainfall, topography, drainage systems and other factors that influence the occurrence of floods. Machine learning algorithms such as neural networks, decision trees, and other models to predict the potential for flooding in an area. Data collected from weather sensors, satellite maps and other data sources is used to train the model so that it is able to identify patterns that lead to flooding conditions. The research results show that the machine learning approach is able to increase the accuracy of flood predictions with a better level of reliability compared to traditional methods. The implementation of artificial intelligence technology in flood prediction has great potential to provide early warning to the public and authorities, thereby reducing the negative impacts caused by flood disasters. It is hoped that this research can become the basis for developing a more effective early warning system in dealing with the threat of flooding in the future.

Corresponding Author:

Zulham Sitorus,
Computer System Studies Program, Faculty of Science and Technology, University Pembangunan Panca Budi, Medan, Indonesia
Email: zulhamsitorus@dosen.pancabudi.ac.id

1. INTRODUCTION

Artificial intelligence (AI) and machine learning (ML) technologies have an increasingly important role in efforts to predict and reduce the impact of natural disasters, including floods. Floods are one of the natural disasters that often occur and can cause extensive damage to infrastructure and the environment and harm human life. The use of artificial intelligence and machine learning technology in flood prediction is relevant because of its ability to process and analyze data quickly and efficiently. Various factors such as weather conditions, rainfall patterns, topography, soil type, drainage, land use,

ISSN 2962-4274

and river systems can be studied by machine learning algorithms to identify patterns and relationships between variables related to the potential for flooding.
Data collected from various sources such as weather sensors, satellites, river water level measurements, and field observations can be processed by machine learning models. Through a training process using historical data, ML models can learn to recognize patterns that indicate situations that have the potential to cause flooding. Apart from predictions, this technology can also be used to develop early warning systems. By utilizing machine learning algorithms that continuously learn from real-time data, early warning systems can provide more accurate and timely information to authorities and the public regarding the potential for flooding. The application of artificial intelligence and machine learning technology in flood prediction has great potential to improve preparedness in facing this natural disaster. With more accurate information and better predictions, it is hoped that losses caused by flooding can be reduced and more effective mitigation measures can be taken.
Judging from the geographical and geological characteristics of the Indonesian region, it is one of the areas prone to flood disasters. About 30% of the 500 rivers in Indonesia cross densely populated areas. More than 220 million people, some of whom are poor and live in flood-prone areas. In general, flood disasters occur in the western part of Indonesia which receives higher rainfall than the eastern part. Based on morphological conditions, the cause of flooding is because the relief of Indonesia's landscape is very varied and there are many rivers flowing through it. Areas prone to flooding are exacerbated by deforestation or changes in land use that do not pay attention to water catchment areas. Changes in land use which then result in flood disasters can be proven, among others, in urban areas along the coast, especially those flowed by rivers. Uncontrolled deforestation also causes a high and uncontrolled increase in water flow (run off) in residential areas, giving rise to flash floods and environmental damage in river basin areas.
The increased flow rate in seawater and rainfall that falls as a result of not/less seeping into the ground, so that rainfall becomes surface flow. Surface flows that move on the ground surface erode the soil and carry it to the river body, therefore the river flow not only increases the discharge but also adds material resulting from erosion. The big focus in this research is machine learning which can apply how to periodically detect complex patterns and make intelligent decisions based on data available in previous years. Machine learning can study patterns from historical data contained in databases from previous years to map floods from rainfall and high tides from the sea. This can be done to find out how much function the flood mapping results have in the Pahlawan Village area.
Demographics and economics has increased the high level of vulnerability to disaster events. Deforestation, erosion of hilly areas, land burning, and environmental destruction are real examples that occur in the daily lives of people in various regions. As a result, people in that environment experience suffering and hardship or people who are far from that environment. In this way, disaster risk reduction education becomes a very important vehicle for creating a culture of preparedness and alertness in facing the threat of disaster, as well as an embodiment of education for sustainable development.

2. RESEARCH METHOD

The research methods carried out will later be able to help the people of Pahlawan Village find out flood-prone spots in the areas that will be used by the community and also to improve the quality of human resources. The designed system can also provide broader knowledge to develop machine learning concepts in science.
This research begins by collecting information from the data that you want to highlight in this title and is taken from several official websites, such as BMKG, BPS, and BPBD Batu Bara district. in this research is flood disaster data in the Batu Bara district. Apart from that, the use of machine learning techniques can be used to increase the resilience and preparedness of the data obtained against flood disasters.

ISSN 2962-4274

In conducting this research, data was collected using a sampling and investigation process, by collecting several documents such as data samples from the BMKG and BPBD as well as information from the community. Sustainability in this research also carries out an observation system and literature studies as support for fulfilling the data needed to conduct research.

Machine learning refers to a method that allows computers to have the ability to learn and perform work automatically. The machine learning process is carried out through certain algorithms so that work orders to the computer can be done automatically (Hirani, 2018). Machine learning is carried out through 2 phases, namely the training phase and the application phase. The training phase is the modeling process of the algorithm used which will be studied by the system through training data, while the application phase is the modeling process which the system has learned through the training phase will be used to produce a certain decision, using testing data. Machine learning can be done in two ways, namely supervised learning and unsupervised learning. Unsupervised learning is the processing of sample data carried out without requiring the final result to have a form that conforms to a certain form, by using several data samples at once. The application of unsupervised learning can be found in the visualization process or data exploration. Supervised learning is the processing of data samples x which will be processed in such a way, to produce output that matches the final result y. Supervised learning can be applied to the classification process (Hirani, 2018).

According to An-hiu Lap-Pui hau and Liang Wang (2018:617), machine learning was developed by professionals from various fields such as mathematics, neurotechnology, biotechnology, computer science, electricity, and electronic engineering. They use techniques and human learning abilities to program computer systems so that machines will imitate human behavioral intelligence and use human reasoning as a model. Machines are given the ability to learn, cognitive processing skills, further improving performance on certain tasks from big data. Machines are also taught deep learning to learn from datasets and perform functions that they are not specifically programmed to do As machines are taught to imitate human behavior in exponential speed, capacity, and memory, machine learning outperforms conventional statistical approaches, achieving human parity in speech recognition and becoming a Digital Assistant for humans, for example Siri and Alexa.

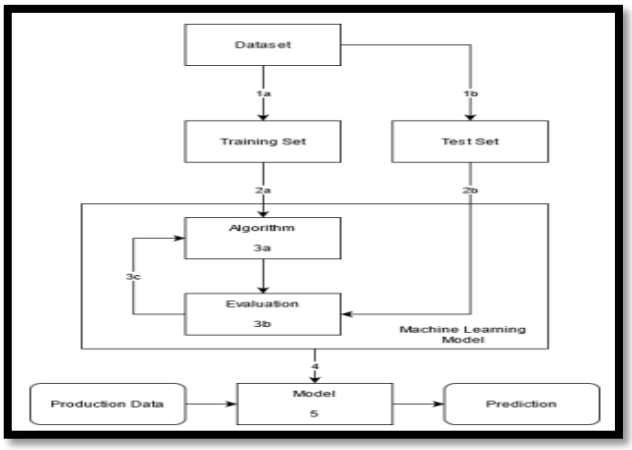

Figure 1. Machine Learning Process

Machine Learning begins by processing existing data, aiming to obtain interconnected information and to determine which attributes are, then creating a model that can be used to explain the condition of the existing system, until finally a decision is made based on the model. In general, the way Machine Learning works is by processing a series of data called a data set, which comes from a system by determining the values of the system, determining which are attributes and which are responses, and then creating a model based on these values, so that when there is new data, the expected value will be by the expectations of the model obtained.

ISSN 2962-4274

Random forest is a bagging method, namely a method that generates several trees from sample data where the creation of one tree during training does not depend on the previous tree, then decisions are taken based on the majority of votes (Wibowo, Saikhu, & Soelaiman, 2016). The two concepts that are the basis of random forests are building an ensemble of trees via bagging with replacement and random feature selection for each tree built. First, every sample taken from the dataset for training trees can be used again for training other trees. Second, the features used during training for each tree are a subset of the features contained in the dataset (Wibowo et al., 2016). Ensemble-based classification will have maximum performance if the basic learners have low correlation. An ensemble must build a weak basic learner, because a strong learner will most likely have a high correlation and usually also cause overfit, whereas a random forest minimizes correlation and maintains classification strength by randomizing the training process, namely by selecting several features at random. of all the features that exist in each training tree, then use them using the selected features to get optimal tree branching. In contrast to the tree training process for ordinary decision trees, the tree training process that is part of a random forest does not use a pruning process, but branching will continue until the leaf limit size is reached (Wibowo et al., 2016). Random forest has two main parameters, namely: m, the number of trees to be used, and k, which is the maximum number of features considered during the branching process. The more the m value, the better the classification results, while the recommended k value is the square root or logarithm of the total number of features (Wibowo et al., 2016).

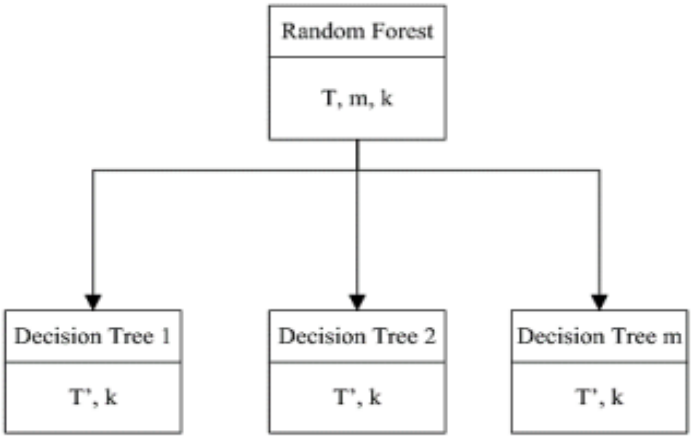

Figure 2. Rendom Forest Illustration

3. RESULTS AND DISCUSSIONS

Machine Learning (ML) involves the use of various techniques and algorithms to enable computers to learn from data and make predictions or decisions without being explicitly programmed. Several concepts and formulas are often used in machine learning, depending on the type of model and the purpose of the analysis. Random Forest is a machine learning algorithm based on ensemble learning, which combines several decision trees to make predictions. This collection of decision trees is generated randomly and independently and then combines the prediction results from each tree to determine the final prediction.

3.1. Research Data

This research was conducted in several areas prone to flooding in Pahlawan Village, Batu Bara Regency, North Sumatra in 2022 at a time of high-intensity rain which could cause flooding, covering 12 hamlets.

ISSN 2962-4274

Table 1. List of Pahlawan Village Hamlets Classified as Flood Prone Zones 2022

No.	Hamlet Name
1.	Hamlet Adi Daya Hamlet
2.	Amanah Hamlet
3.	Bogak Hamlet
4.	Pabrik Hamlet
5.	Sejahtera Hamlet
6.	Sejarah Hamlet
7.	Bandar Hamlet
8.	Bunga Jumpa Hamlet
9.	LobaiAbbas Hamlet
10.	Nelayan Hamlet
11.	Nilam Hamlet
12.	Wan Ahad Hamlet

3.2. Results and Data Processing

The rainfall map is obtained from the results of interpolation of monthly rainfall data which is then scored for each classification used. Rainfall scoring values can be seen in the following table.

Table 2. Rainfall Scoring

No.	Class	Harkat	Weight	Score
1.	>2500mm (really wet)	5	5	25
2.	2001-2500 (wet)	4		20
3.	1501-2000 (medium/humid)	3		15
4.	1001-1500mm (dry)	2		10
5.	<1000mm (very dry)	1		5

The soil type map was obtained by digitizing the soil type map of North Sumatra Province, then a clip was carried out in the research area and then a score value was given according to the reference used. Soil-type scoring values can be seen in the following table.

Table 3. Soil Type Scoring

No.	Type of Soil	Class	Harkat	Wight	Score
1.	Aluvial, Planosol, Hidromorf Kelabu, Laterik, Air Tanah	Very Smooth	5	4	20
2.	Latosol	Smooth	4		16
3.	Tanah Hutan Coklat, Tanah Mediteranian	Currently	3		12
4.	Andosol, Laterik, Grumosol, podsol, Podsolic	Rough	2		8
5.	Regosol, Litosol, Organosol, Renzina	Very Rough	1		4

ISSN 2962-4274

The land use map was obtained from Landsat Image 8 for 2020-2023 which was carried out by Supervised Classification with the following stages:

a. Image cropping is carried out to narrow down the area that corresponds to the research area study so that it can facilitate the image classification process.
b. Geometric correction is carried out on images by identifying Ground Control Points (GCP) or tie points that are easy to determine such as river forks or road intersections. The GCP accuracy value is shown by the Root Mean Square Error (RMSerror) value. RMS-error states the error value of the geometric correction process. Good accuracy is indicated by a very small RMS-error value close to zero. "Root Mean Square Error (RMSE) is the magnitude of the error rate in prediction results, where the smaller (closer to 0) the RMSE value is, the more accurate the prediction results will be." Calculation of RMS-error using the following equation:

$$RMS_{error} = \sqrt{(X' - X_{orig}) + (Y' - Y_{orig})}$$

Information:
X' and Y' = Coordinates of the output image
Xorig and Yorig = Original image coordinates (input)

c. Radiometric correction is carried out in two stages, namely radiometric calibration and atmospheric correction. Radiometric calibration is used to convert digital number values into reflectance, while atmospheric correction is used to remove atmospheric bias in the image.
Image classification aims to group and segment homogeneous features using supervised image classification. Image classification in the research area will be divided into several classes, namely:
o Open land, rice fields
o Agriculture, settlements, dry land
o Shrub
o Plantations
o Forest
d. After the classification is carried out, accuracy is measured by carrying out Ground Truth, namely taking points in the field/research location using GPS by providing attribute data at these points according to the actual conditions in the field. So the result of the classification is a land cover map, namely a map that provides information on areas according to their designation. So changes in the physical form of land cover can be seen every ten years from 2015 to 2020. Then after that, scoring is carried out. Land use scoring values can be seen in Table

Table 4. Land Use Scoring

No.	Class	Harkat	Wight	Score
1.	Pemukiman, Badan Air	5		25
2.	Sawah, Tambak	4		20
3.	Ladang, Tegalan, Kebun	3	5	15
4.	Semak Belukar	2		10
5.	Hutan	1		5

e. The height map is obtained from DEM (Digital Elevation Model) data processing. Altitude scoring values are presented in Table.

ISSN 2962-4274

Table 5. Land Use Scoring

No.	Class	Harkat	Wight	Score
1.	0 m- 50 m	5		25
2.	51 – 100 m	4		20
3.	101 – 150 m	3	5	15
4.	151 – 250 m	2		10
5.	>250 m	1		5

3.3. Tabulation Graph of the Prediction Process

This section explains the results of each table which is the main part in mapping and predicting with a level of accuracy which is processed using the concept of artificial intelligence machine learning algorithm by inputting the amount of data to be processed at each flood location.

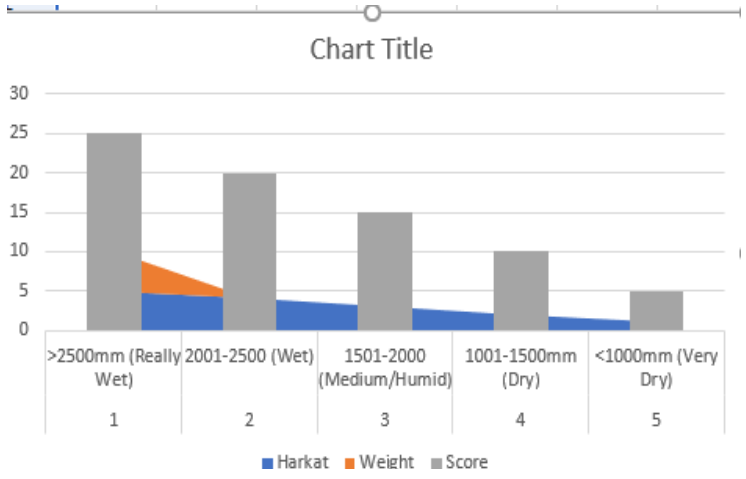

Figure 3. Rainfall Scoring

Figure 4. Soil Type Scoring

ISSN 2962-4274

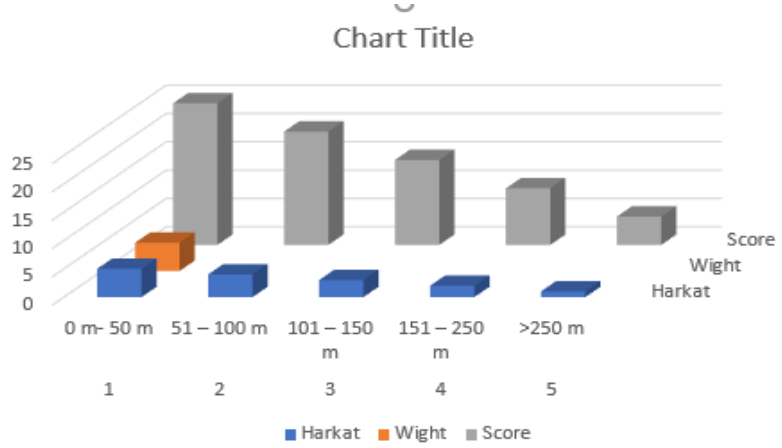

Figure 5. Soil Type Scoring

4. CONCLUSION

From the research results that have been carried out, it can be concluded that in mapping and predicting flood-prone areas in Pahlawan Village, Tanjung Tiram District, Batubara Regency, the data processing digitalization process carried out has succeeded in grouping the amount of data on flood-prone areas in 12 hamlets. in the Hero Village Comparison is carried out by comparing data on similar natural disasters with actual events, namely 2020 - 2023. Artificial Intelligence Machine Learning Algorithm is a process of processing historical data on previous events and is used as a basis for building knowledge that has been processed. using Random Forest.

REFERENCES

[1] *Anglin*, dkk. (2011).Analysis Effect Of Mother Tongue Using In Mathematics Learning Toward Conseptual Understanding Ability In Elementary Students. Dipresentasikan Dalam International Conference (ICREAM 5). Bandung.

[2] Ahmad, Abu. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning.

[3] Ahmad, Adeel. (2017, November 13). An overview of activation functions used in neural networks. Retrieved May 3, 2020, from adl1995.github.io website: https://adl1995.github.io/an-overview-of-activation-functions-used-inneural-networks.html

[4] Ali, Z. (2019, January 7). A simple Word2vec tutorial. Retrieved May 1, 2020, from Medium website: https://medium.com/@zafaralibagh6/a-simpleword2vec-tutorial-61e64e38a6a1

[5] Arni, D. (2018a, October 25). Apa Itu Text Mining ? Retrieved April 24, 2020, from Garuda Cyber Indonesia website: https://garudacyber.co.id/artikel/1254- apa-itu-text-mining

[6] Britz, D. (2016, July 8). Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs. Retrieved from WildML website: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part1-introduction-to-rnns/.

[7] Brownlee, J. (2017, October 3). How to Use Word Embedding Layers for Deep Learning with Keras. Retrieved from Machine Learning Mastery website: https://machinelearningmastery.com/use-word-embedding-layers-deeplearning-keras/.

[8] Nugroho, K. S. (2020, February 8). Dasar Text Preprocessing dengan Python. Retrieved from Medium website: https://medium.com/@ksnugroho/dasartext-preprocessing-dengan-python-a4fa52608ffe