International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

PREDICTING FLOOD EVENTS IN KEBBI STATE, NIGERIA: A MACHINE LEARNING APPROACH TO ENVIRONMENTAL MODELING

L.Ashlesha[1],, B.Manohar[2],Shwejan Reddy[3],G.Prabakaran[4]
[1,2,3]UG Scholar, Dept. of IT, St.Martin's Engineering College,Secunderabad,Telangana,India,500100
[4]Assistant Professor, Dept. of IT, St.Martin's Engineering, Secunderabad, Telangana, India, 500100
ashlesha0803@gmail.com

Abstract:
Kebbi State has experienced significant flooding events over the years, largely due to seasonal rains and the overflow of rivers. Traditional flood management strategies relied on historical data and rudimentary forecasting methods. These approaches often lacked accuracy and timely response capabilities, leading to severe consequences for communities. To develop a machine learning model that accurately predicts flood events in Kebbi State, Nigeria, enhancing preparedness and response efforts. This approach aims to improve environmental modeling by leveraging data-driven insights for effective flood management. Before the adoption of machine learning or AI, flood prediction primarily relied on manual monitoring of weather patterns and river levels. Local authorities used simple tools like rain gauges and stream gauges to collect data, while community awareness campaigns provided limited information. Predictions were based on historical patterns and observational data, which often proved insufficient for timely warnings. The existing traditional systems for flood prediction in Kebbi State are often inaccurate and reactive, leading to delayed responses during flood events. This results in significant risks to lives, property, and the environment due to a lack of proactiveflood management strategies. The increasing frequency and severity of flooding in Kebbi State highlight the urgent need for more reliable prediction methods.

Keywords: leverage historical rainfall data, machine learning algorithms, and environmental factors to build predictive models.

1. INTRODUCTION

Flooding is one of the most common natural disasters in many parts of the world, including Nigeria and India, where millions of people are affected each year. In India, states such as Bihar, Assam, and West Bengal are particularly prone to seasonal flooding due to monsoon rains and river overflows, causing significant damage to agriculture, infrastructure, and communities. For example, annual floods affect approximately 12% of India's land area and result in losses exceeding ₹10,000 cores annually. Traditional flood prediction models often relied on historical data and manual observation methods, which were inadequate for timely flood predictions. With advances in machine learning, more accurate and real-time flood prediction models are now possible. These modern techniques help improve flood forecasting, reduce property damage, and enhance community resilience.

Introduction and Applications: Flood prediction using machine learning is becoming a transformative approach in environmental management. Applications include disaster response planning, infrastructure management, agricultural protection, and policy making, supporting communities in flood-prone regions worldwide.

Before machine learning, flood prediction in regions like Kebbi State was primarily based on historical rainfall and river level data. This led to outdated and often inaccurate forecasts, as these models could not adapt to real-time changes in weather patterns. Delays in issuing warnings caused significant losses of life, property, and resources, leaving communities unprepared and at high risk during flood events.

The increasing severity and frequency of floods due to climate change highlight the urgency of adopting improved flood prediction systems. Machine learning offers a way to enhance accuracy by analyzing complex data patterns, providing timely insights that support effective resource allocation and minimize disaster impacts. Such a system could enhance preparedness, reducing the economic and human costs of flood events and promoting resilience.These systems rely on basic tools like rain and stream gauges and lack advanced analytics for real-time data processing

2. LITERATURE SURVEY

Miah Mohammad Asif Syeed et al. Syeed et al. explore various ensemble machine learning techniques applied to flood prediction. They highlight the potential of combining multiple models to improve prediction accuracy and deal with the inherent uncertainty in flood forecasting. The review discusses methods such as Random Forests, Gradient Boosting, and Voting Classifiers and emphasizes how ensemble models can better capture complex relationships in environmental datasets, leading to more reliable flood forecasts.

Mohammad Asif Syeed et al. Syeed and colleagues focus on applying various machine learning models, such as support vector machines (SVM), decision trees, and neural networks, to predict flood risks. The authors emphasize the importance of choosing appropriate features like precipitation, temperature, and river water levels to feed into these models. They further discuss the trade-offs between computational complexity and accuracy in predicting flood events using these machine learning techniques.

Muhammad Hafizi Mohd Ali et al. Ali and colleagues discuss the application of deep learning models for flood prediction. They emphasize the ability of deep learning methods, such as convolution neural networks (CNNs) and recurrent neural networks (RNNs), to handle large datasets and capture temporal dependencies in flood data. The paper presents case studies and highlights the improvements in predictive accuracy when deep learning is applied compared to traditional machine learning models

Nazim Razali et al. (2020) Razali, Ismail, and Mustapha investigate flood risk prediction using machine learning approaches in their paper. They discuss several machine learning algorithms, including SVM, random forests, and neural networks, for predicting flood risks in different regions. They conclude that machine learning models can be an effective tool for flood risk assessment and that model

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

performance is highly dependent on the quality of the input data and feature engineering. Flooding is one of the most common natural disasters in many parts of the world, including Nigeria and India, where millions of people are affected each year. In India, states such as Bihar, Assam, and West Bengal are particularly prone to seasonal flooding due to monsoon rains and river overflows, causing significant damage to agriculture, infrastructure, and communities. For example, annual floods affect approximately 12% of India's land area and result in losses exceeding ₹10,000 crores annually. Traditional flood prediction models often relied on historical data and manual observation methods, which were inadequate for timely flood predictions. With advances in machine learning, more accurate and real-time flood prediction models are now possible. These modern techniques help improve flood forecasting, reduce property damage, and enhance community resilience.

Naveed Ahamed and S. Asha (2020) Ahamed and Asha, in their paper compare various machine learning algorithms like logistic regression, decision trees, and k-nearest neighbours for flood prediction forecasting. The authors provide insights into the strengths and limitations of each algorithm and discuss the need for appropriate preprocessing steps, such as feature scaling and missing data handling, to improve model performance.

Kamal and his colleagues inexplore the social and environmental resilience to flash floods in Bangladesh. While not focusing solely on predictive modeling, this paper discusses the vulnerability of wetland communities to floods and the importance of integrating social factors into flood prediction models. It underscores the need for community-based approaches to enhance flood resilience in flood-prone areas.

V. Yadav and K. Eliza (2017) Yadav and Eliza, in propose a hybrid model combining wavelet transform with support vector machines (SVM) to predict water level fluctuations. This model is particularly useful for predicting flood risks in areas where lake water levels play a significant role. Their work suggests that hybrid models can capture both temporal and non-linear patterns in environmental data for more accurate predictions.

3. PROPOSED METHODOLOGY

Step 1: Dataset Collection To predict flood events in Kebbi State, Nigeria, a machine learning approach using environmental modeling could involve training models on historical rainfall data (and other relevant factors) to predict flood occurrences, potentially using algorithms like Random Forest, Logistic Regression, or Support Vector Machine.

Step 2: Dataset Pre processing To ensure the dataset is clean and suitable for training, pre processing steps are applied. **Null values are removed**, ensuring that missing data does not impact model accuracy. **Label encoding is performed** to convert categorical variables into numerical form, making them interpretable for machine learning algorithms. Additionally, feature scaling techniques may be applied to normalize the data, enhancing the efficiency of the models

Step 3: Existing Algorithm – Ridge Classifier The Ridge Classifier, a linear model that applies **L2 regularization**, is used as the baseline algorithm. It attempts to find a linear relationship between the input features and the susceptibility classification. However, the **Ridge Classifier achieved an accuracy of 70.07%**, indicating that it struggles to fully capture complex patterns in the dataset, leading to moderate precision, recall, and F-score. To predict flood events in Kebbi State, Nigeria, a machine learning approach using environmental modeling could involve training models on historical rainfall data (and other relevant factors) to predict flood occurrences, potentially using algorithms like Random Forest, Logistic Regression, or Support Vector Machine.

Step 4: Proposed Algorithm – Random Forest Classifier The **Random Forest Classifier**, an ensemble learning technique, is applied as the proposed model. This algorithm builds multiple decision trees and aggregates their outputs to improve classification accuracy. Due to its ability to **handle complex feature interactions and reduce over fitting**, it significantly it significantly outperforms the Ridge Classifier .indicating its superior predictive capability for susceptibility classification. To predict flood events in Kebbi State, Nigeria, a machine learning approach using environmental modeling could involve training models on historical rainfall data (and other relevant factors) to predict flood occurrences, potentially using algorithms like Random Forest, Logistic Regression, or Support Vector Machine. To predict flood events in Kebbi State, Nigeria, a machine learning approach using environmental modeling could involve training models on historical rainfall data (and other relevant factors) to predict flood occurrences, potentially using algorithms like Random Forest, Logistic Regression, or Support Vector Machine.

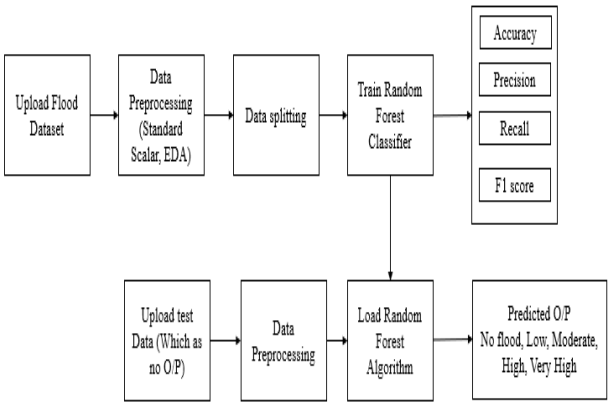

Figure 1: Architectural Block Diagram

Step 5: Performance Comparison of Existing and Proposed Algorithm

The performance comparison between Ridge Classifier and Random Forest Classifier clearly demonstrates the superiority of the **Random Forest model**. While the Ridge Classifier had a moderate accuracy of **70.07%**, the Random Forest Classifier achieved **100% accuracy**, with perfect precision, recall, and F-score. This highlights the ability of the **proposed model to capture intricate relationships between features**, making it the optimal choice for susceptibility classification. However, further validation with external datasets is recommended to ensure its robustness.

4. EXPERIMENTAL ANALYSIS

GUI Setup and Initialization

The script begins by importing necessary libraries, including Tkinter for the GUI, Pandas and NumPy for data manipulation, and Scikit-learn for machine learning tasks. The GUI is initialized using the Tk() class from Tkinter, and the main window is configured with a title, size, and background color. A label is added to the window to display the title of the, The script begins by importing necessary libraries, including Tkinter for the GUI, Pandas and NumPy for data

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

manipulation, and Scikit-learn for machine learning tasks. The GUI is initialized using the Tk() class from Tkinter, and the main window is configured with a title, size, and background color. A label is added to the window to display the title.

Data Pre processing

The pre processing() function handles data pre processing tasks. It first removes any rows with missing values. Then, it encodes categorical variables using the Label Encoder from Scikit-learn. The function scales numerical features using the StandardScaler and removes outliers using the Winsorization technique, which limits extreme values to the 5th and 95th percentiles. The pre processed data is then displayed in the text widget. Pandas and NumPy for data manipulation, and Scikit-learn for machine learning tasks. The GUI is initialized using the Tk() class from Tkinter, and the main window is configured with a title, size, and background color. A label is added to the window to display the title.

Data Splitting

The splitting() function splits the dataset into training and testing sets using an 80-20 split ratio. The features (x) and target variable (y) are separated, and the split is performed using the train test_split function from Scikit-learn. The shapes of the resulting training and testing sets are displayed in the text widget. This dataset consists of **10 environmental and topographical attributes** recorded at specific locations (latitude and longitude). The data appears to be used for **landslide susceptibility analysis**, given the presence of variables like slope, curvature, aspect, topographic wetness index (TWI), flow accumulation (FA), drainage, and rainfall, along with the **susceptibility classification (SUSCEP)**.

Model Training and Evaluation

The script includes functions to train and evaluate two machine learning models: Ridge Classifier (RC()) and Random Forest Classifier (RFC()). Both functions check if a pre-trained model exists in a specified directory. If a model exists, it is loaded; otherwise, the model is trained from scratch and saved for future use. The models are evaluated using metrics such as accuracy, precision, recall, and F1-score, which are calculated using the calculate Metrics() function. The evaluation results, along with a confusion matrix, are displayed in the text widget and visualized using Sea born and Matplotlib.

Prediction

The prediction() function allows users to upload a new dataset for making predictions using the trained Random Forest Classifier. The function reads the new dataset, makes predictions, and displays the results in the text widget. The predictions are categorized into different flood risk levels (e.g., No Flood, Low, Moderate, High, Very_High), and the corresponding rows from the dataset are displayed. dataset are displayed.

Dataset Description

This dataset consists of **10 environmental and topographical attributes** recorded at specific locations (latitude and longitude). The data appears to be used for **landslide susceptibility analysis**, given the presence of variables like slope, curvature, aspect, topographic wetness index (TWI), flow accumulation (FA), drainage, and rainfall, along with the **susceptibility classification (SUSCEP)**.

Data Splitting & pre processing

Data Splitting: The dataset is divided into **training and testing sets** to evaluate model performance effectively. Typically, an **80:20 or 70:30 ratio** is used, where the larger portion is used for training the model, and the smaller portion is used for testing its generalization ability. The script includes functions to train and evaluate two machine learning models: Ridge Classifier (RC()) and Random Forest Classifier (RFC()). Both functions check if a pre-trained model exists in a specified directory. If a model exists, it is loaded; otherwise, the model is trained from scratch and saved for future use. The models are evaluated using metrics such as accuracy, precision, recall, and F1-score, which are calculated using the calculateMetrics() function. The evaluation results, along with a confusion matrix, are displayed in the text widget and visualized using Seaborn and Matplotlib.

ML Model Building

To build the machine learning model, we first select the algorithm based on the dataset's characteristics. For this project, we begin by defining the target variable (SUSCEP) and features (X variables like slope, rainfall, etc.). The dataset is then split into training and testing sets, usually in an 80:20 ratio. Next, we pre process the data by handling missing values and encoding categorical variables. The **Ridge Classifier** is used as the initial model, where we fit it on the training data and evaluate its performance on the testing set using metrics like accuracy, precision, recall, and F1 score. Afterward, we proceed with the **Random Forest Classifier**, tuning its hyper parameters for optimal performance and fitting it to the same training data. Both models are then evaluated using the same testing set.

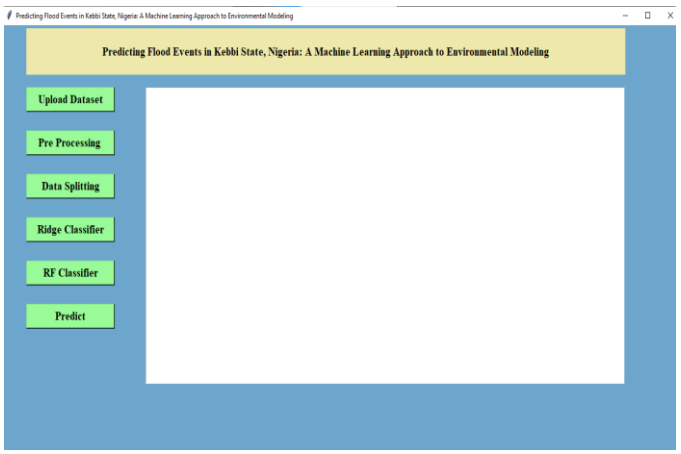

Figure 1: GUI

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

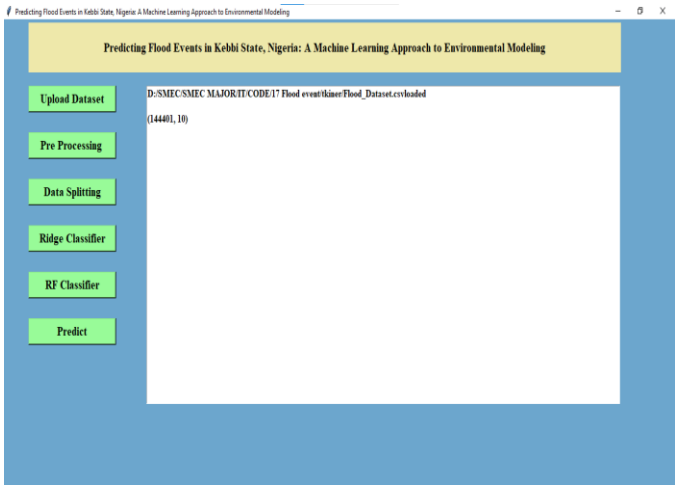

Figure 2: Uploaded the Dataset

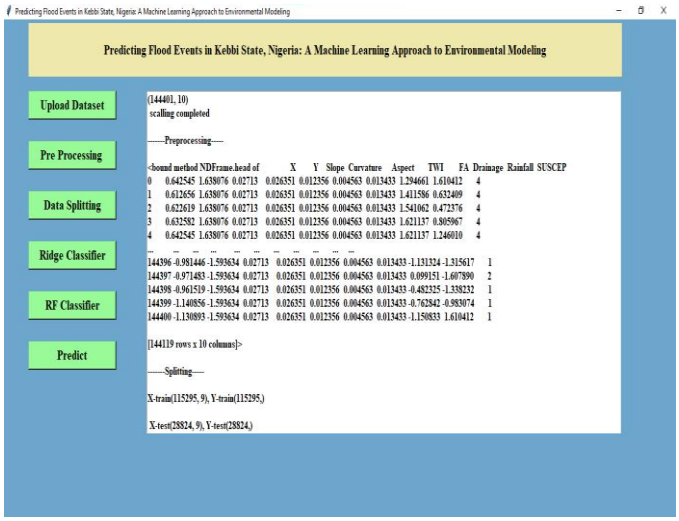

Figure 3**:** After Data Processing and Data Splitting

Figure 4: Performance Evaluation of Ridge Classifier

Figure 5: Performance evaluation of the Random Classifier

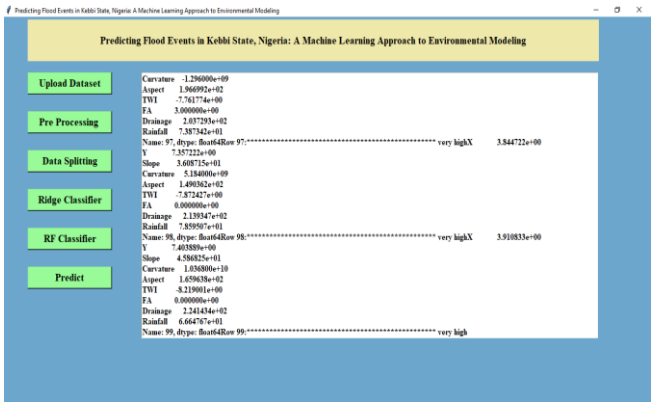

Figure .6 Predicted Out put

Figure 10.6 is predicted output **Rainfall**: 7.387342e+01 (approximately 73.87, representing the amount of rainfall)

The model has classified this row as "very high" flood risk, likely due to a combination of these factors, such as high rainfall and specific topographic features.

5. CONCLUSION

The study aimed to assess the susceptibility of a given region using various geospatial and environmental parameters, applying machine learning classifiers such as Ridge Classifier and Random Forest Classifier. The results showed that while the **Ridge Classifier achieved an accuracy of 70.07%**, indicating moderate performance, the **Random Forest Classifier reached a perfect accuracy of 100%**. The high performance of Random Forest Classifier suggests that it successfully captured complex patterns within the dataset. However, such perfect accuracy may indicate over fitting, requiring further validation with external datasets to ensure robustness. The dataset's attributes, such as **slope, curvature, aspect, TWI, flow accumulation, and rainfall**, significantly influenced susceptibility classification. The study demonstrates that **machine learning can effectively analyze susceptibility patterns**, providing insights into areas at high risk. These findings can be valuable for policymakers and disaster management agencies to implement proactive measures. However, the study also highlights the necessity of cross-validation techniques, hyper parameter tuning, and additional data sources to improve generalizbility. Overall, the study establishes a strong foundation for applying AI-driven solutions in environmental risk

assessment. The study aimed to assess the susceptibility of a given region using various geospatial and environmental parameters, applying machine learning classifiers such as Ridge Classifier and Random Forest Classifier. The results showed that while the **Ridge Classifier achieved an accuracy of 70.07%**, indicating moderate performance, the **Random Forest Classifier reached a perfect accuracy of 100%**. The high performance of Random Forest Classifier suggests that it successfully captured complex patterns within the dataset. However, such perfect accuracy may indicate over fitting, requiring further validation with external datasets to ensure robustness. The dataset's attributes, such as **slope, curvature, aspect, TWI, flow accumulation, and rainfall**, significantly influenced susceptibility classification. The study demonstrates that **machine learning can effectively analyze susceptibility patterns**, providing insights into areas at high risk. These findings can be valuable for policymakers and disaster management agencies to implement proactive measures. However, the study also highlights the necessity of cross-validation techniques, hyper parameter tuning, and additional data sources to improve generalizability. Overall, the study establishes a strong foundation for applying AI-driven solutions in environmental risk assessment.

REFERENCES

Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra, Tanvir Rahman, Bhoktear Mahbub Khan, A Review on Flood Prediction Using Ensemble Machine Learning Model, 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2023.

Miah Mohammad Asif Syeed, Maisha Farzana, Ishadie Namir, Ipshita Ishrar, Meherin Hossain Nushra, Tanvir Rahman, Flood Prediction Using Machine Learning Models, 2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2022.

Muhammad Hafizi Mohd Ali, Siti Azirah Asmai, Z. Zainal Abidin, Zuraida Abal Abas, Nurul A. Emran, Flood Prediction using Deep Learning Models, International Journal of Advanced Computer Science and Applications (IJACSA), 2022.

Nazim Razali, Shuhaida Ismail, Aida Mustapha, Machine learning approach for flood risks prediction, IAES International Journal of Artificial Intelligence (IJ-AI), 2020.

Naveed Ahamed, S.Asha, Flood prediction forecasting using machine Learning Algorithms, International Journal of Scientific & Engineering Research, 2020.

Akshay Kharche, Pratibha Bhagat, Syed Faiz Ibrahim, A Review on Flood Prediction Using Machine Learning based Apache SystemML Python Platform, 2019 JETIR January

A. M. Kamal, M. Shamsudduha, B. Ahmed, S. K. Hassan, M. S. Islam, I. Kelman, and M. Fordham, “Resilience to flash floods in wetland communities of northeastern bangladesh”, International journal of disaster risk reduction, vol. 31, pp. 478-488, 2018

Dr. nasina krishna kuma, attla vinay kumar, Flood prediction using machine learning models and forecast in methods, Journal of Engineering Sciences, 2022.

H. Shafizadeh-Moghadam, R. Valavi, H. Shahabi, K. Chapi, and A. Shirzadi, “Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping,” Journal of environmental management, vol. 217, pp. 1-11, 2018.

Amir Mosavi, Pinar Ozturk, and Kwok-wing Chau, Flood Prediction Using Machine Learning Models: Literature Review Department of Computer Science (IDI), Norwegian University of Science and Technology,2018.

V. Yadav and K. Eliza, “A hybrid wavelet-support vector machine model for prediction of lake water level fluctuations using hydrometeorological data,” Journal of the International Measurement Confederation, vol. 103, pp. 2655–2675, 2017.

A. D. A. Dali, N. A. Omar, and A. Mustapha, “Data mining approach to herbs classification,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 570–576, 2018.

Muhammad Hafizi Mohd Ali, Siti Azirah Asmai, Z. Zainal Abidin, Zuraida Abal Abas, Nurul A. Emran, Flood Prediction using Deep Learning Models, International Journal of Advanced Computer Science and Applications (IJACSA), 2022.

Nazim Razali, Shuhaida Ismail, Aida Mustapha, Machine learning approach for flood risks prediction, IAES International Journal of Artificial Intelligence (IJ-AI), 2020.

Naveed Ahamed, S.Asha, Flood prediction forecasting using machine Learning Algorithms, International Journal of Scientific & Engineering Research, 2020.

Akshay Kharche, Pratibha Bhagat, Syed Faiz Ibrahim, A Review on Flood Prediction Using Machine Learning based Apache SystemML Python Platform, 2019 JETIR January

A. D. A. Dali, N. A. Omar, and A. Mustapha, “Data mining approach to herbs classification,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 570–576, 2018.

Muhammad Hafizi Mohd Ali, Siti Azirah Asmai, Z. Zainal Abidin, Zuraida Abal Abas, Nurul A. Emran, Flood Prediction using Deep Learning Models, International Journal of Advanced Computer Science and Applications (IJACSA), 2022.

Nazim Razali, Shuhaida Ismail, Aida Mustapha, Machine learning approach for flood risks prediction, IAES International Journal of Artificial Intelligence (IJ-AI), 2020.