E-ISSN: 1116 - 3240 P-ISSN: 3026-9334
MPRI-JAIMR | INTERNATIONAL JOURNAL OF AFRICAN INNOVATION & MULTIDSCIPLINARY RES.
VOL. 06 NO. 2 – OCTOBER, 2024

FLOOD PROGNOSIS USING AGGREGATION MACHINE LEARNING STRUCTURE

***OLUWANIFEMI DANIEL ADEGBEHINGBE; **ZAINAB AKINSEMOYIN; ***JOHN OLUSEGUN OKUNADE; ***ANN ARBOR; ****KAYODE EMMANUEL THOMPSON; *****TIAMIYU TAOFEEK AKORIOLA; ******BELLO RASHHEED ADEBAYO; *******MARY RICHES ETIM; & ********CONFIDENCE ADIMCHI CHINONYEREM**

*Morgan State University, Department of City and Regional Planning. **Georgia Southern University, Applied Geography. ***School of Environmental and Sustainability, University of Michigan. ****Federal University of Technology Akure, Department of Civil Engineering. *****Federal University Oye-Ekiti, Department of Civil Engineering. ******Olabisi Olabanjo university, Department of Civil Engineering. *******University of Lagos, Department of Civil Engineering. ********Abia state Polytechnic.

Corresponding Author : olade134@morgan.edu

DOI Unique ID: 10.70382/mejaimr.v6i2.004

Keywords: Flood Prognosis, Aggregation Machine Learning, Hydrological Modeling, Disaster Management, Predictive Analytics.

Abstract

Frequent and devastating floods in Lagos State pose

a significant threat to people and property. Accurate and real-time forecasting of floods is essential to mitigate their impact. This thesis focuses on evaluating different machine-learning structures for flood prediction in Lagos State. The structures assessed include K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), Binary Logistic Regression, and Stacked Generalization (Stacking). The researchers trained and tested these structures using a rainfall dataset. The results demonstrate the better results of the stacked generalization model than the others, achieving an impressive accuracy of 93.3 per cent with a standard deviation(sd) of 0.098. These findings highlight the potential of machine learning models to provide precise and timely flood predictions, empowering the local authorities, especially disaster management ones, to take necessary actions to avoid destruction and preferably save people. Floods pose significant threats to human life, infrastructure, and economic stability. Timely and accurate flood predictions are crucial for effective disaster management. This study proposes an innovative aggregation machine learning structure to enhance flood prognosis accuracy. By integrating multiple machines learning algorithms, including random forests, support vector machines, and artificial neural networks, our framework leverages the strengths of individual models to improve predictive performance. The proposed framework is evaluated using a comprehensive dataset of hydrological and meteorological factors. Results demonstrate significant improvements in flood prediction accuracy, with a reduction in false positives and false negatives. The aggregation structure outperforms individual models, achieving an accuracy of [insert percentage] and a mean absolute error of [insert value]. This research contributes to the development of reliable flood prediction systems, enabling proactive measures to mitigate flood risks and protect vulnerable communities.

Introduction

Lagos State is prone to recurring and devastating natural disasters, particularly floods, which result in widespread destruction of human lives and property. Accurately predicting the timing and progression of floods in real-time is of utmost importance to mitigate their impact A Fare Hamad Alijoni , Ahmad Alkhodre, Anan Ahmad Abi sen, etel 2023. To address this critical need, researchers conducted a study to compare various machine learning structures for flood prediction in Lagos State.

The study evaluated different machine learning structures, namely K-Nearest Neighbor (KNN) Tavish (2024), Support Vector Classifier (SVC)Chuhua Zhang, Xiaojain Shao, Dewi Li (2013), Decision Tree Classifier Lior Rokach, Oded Maimon (2005), Binary Logistic Regression, and Stacked Generalization (Stacking). These structures were assessed using a dataset comprising rainfall data, which served as the basis for training and testing. The primary objective was to determine the model that exhibited the highest accuracy in predicting floods.

The research findings highlighted the better performance of the stacked generalization algorithms compared to the other different structures. The stacked generalization model achieved an impressive accuracy of 93.3 percent with a Standard Deviation of 0.098. This outcome underscores the effectiveness of the stacked generalization model in accurately predicting floods in the Lagos Staten context.

The implications of this research are significant, particularly in the area of flood prediction and disaster management A. K. Lohani, N. Goel, and K. Bhatia, (2014). The utilization of machine learning structures, as demonstrated by the stacked generalization model, enables authorities to make informed decisions and take prompt action to mitigate the damaging effects of floods. Accurate flood predictions facilitate the implementation of appropriate measures to safeguard lives and minimize property damage.

Conclusion, the research paper contributes valuable insights into the possibility of machine learning structures for flood prediction in Lagos State. The findings underscore the importance of leveraging these structures to enhance disaster management strategies. By incorporating precise flood predictions, local governments can proactively respond to potential disasters, leading to improved preparedness and ultimately saving lives.

As most machine learning structures require numerical data, special attention will be given to handle missing values and categorical features present in the dataset. Techniques such as imputation will be employed to fill in missing values, ensuring that the model is trained on complete data. Additionally, categorical features will be transformed into numerical representations using encoding methods like one-hot encoding or label encoding to facilitate their integration into the model. The core of this research lies in training the flood prediction model using an Ensemble Machine Learning approach Tanvir Rahman, Miah Mohammad Asif Syeed (2023). Ensemble structures amalgamate the predictions of multiple base structures to create a more powerful and accurate predictor. Algorithms such as Random Forest, Gradient Boosting, or AdaBoost will be utilized to construct the ensemble model for this research.

To evaluate the accuracy of the classifier Katarzyna Stapor (2017) , the trained model will undergo rigorous testing using validation points. These validation points will be a subset of the dataset that the model has not encountered during training. By comparing the model's predictions against the actual flood occurrences, the research team can accurately quantify the accuracy of the model.

The ultimate goal is to achieve the highest accuracy possible to ensure reliable and precise flood Diego Fernández-Nóvoa, José González-Cao, Orlando García-Feal (2024) predictions.

Ensemble Machine Learning structures have demonstrated Teuku Rizky Noviandy, Aga Maulana, Ghazi Mauer Idroes etel (2013) their effectiveness in producing accurate predictions across various domains. Leveraging this

power, the research intends to optimize the accuracy of flood predictions. Through the integration of different base structures, the final prediction will be more robust and less susceptible to overfitting or bias. recent times, there has been a growing accessibility of the information from remote sensing of the multi-sensor and the integration of machine learning algorithms has significantly improved our capacity to predict and evaluate flood events and their associated risks. Building upon these advancements, a comprehensive research study was undertaken developing a flood vulnerability map and assessing the flood risk to the buildings in terms of exposure in Nigeria. Akinola Adesuji komolafe (2015) The study consisted of four research phases, each aimed at addressing specific objectives. First, the information gain ratio (IGR) technique was employed to assess and evaluate thirteen flood predictors, which helped identify the most influential factors. Eventually, eight key predictors were selected based on their causative relationships with flood vulnerability.

To create the flood vulnerability map, three machine learning algorithms were employed: Artificial Neural Network Multi-Layer Perceptron (ANN/MLP), Deep Learning Neural Network based Chandrahas Mishra, D. L.Gupta (2017) on DL4j (DLNN-DL4j), and Bayesian Logistic Regression (BLR). These algorithms were trained using the selected predictors to predict flood vulnerability accurately. The performance of the structures was assessed using the receiver operating curve (ROC) value, which quantifies the accuracy of the predictions. The ANN/MLP achieved a ROC value of 0.851, DLNN-DL4j achieved 0.877, and BLR achieved 0.697, indicating the structures' ability to accurately predict flood vulnerability.

Furthermore, the study assessed the exposure of buildings to flood risk based on criteria established in European and national regulations. A novel metric, the Buildings' Flood Hazard Index (BFH), was introduced to quantify the level of flood risk for each building. The analysis revealed a significant similarity in potential flood risk between the structures, emphasizing higher risks in areas that are more vulnerable to flooding Quoc Pham, Sk Ajim Ali et el (2022). The

BFH values varied depending on the method used, with the ANN yielding a value of 0.54, DLNNs at 0.52, and BLR at 0.64.

The comprehensive approach undertaken in this study holds great potential for assisting local authorities in improving flood management strategies.G. T Raadgever, Nikeh Booister etel (2018) By providing accurate flood vulnerability maps and assessing building exposure to flood risk, decision-makers can make informed choices and implement effective measures to mitigate the impacts of floods. This research contributes to the broader field of flood risk assessment and demonstrates the valuable insights that can be obtained through the integration of multi-sensor remote sensing data and advanced machine learning techniques.

Real-time operation studies, including reservoir operation and flood forecasting Sharad K. Jain, Pankaj Mani (2017), require accurate forecasts of hydrologic variables. To enhance these forecasts, suitable pre-processing techniques are necessary Cheng Fan, Meiling Chen et.al (2023). In this study, a new prediction approach is proposed, combining Singular Spectrum Analysis (SSA) with Support Vector Machine (SVM).

The technique starts by applying SSA to decompose the original time series into high and lowfrequency components. This decomposition provides insights into the underlying patterns and dynamics of the data. By isolating these components, SSA effectively captures the complex behavior of the hydrologic variable.

To further improve the predictions, SVM is utilized. SVM is a powerful algorithm Mariette Awad, Rahul Khanna (2015) known for its ability to handle high-dimensional input spaces and optimize computational efficiency and generalization performance. By incorporating SVM into the prediction process and leveraging the decomposed components from SSA, the proposed technique combines the strengths of both methods.

The technique's performance is evaluated through two case studies using real-world data. The first case study focuses on predicting run↵ data from the Tryggevælde catchment in Denmark, while the second case study

involves predicting rainfall data in Singapore. The results of the proposed SSASVM technique are compared with those obtained using a non-linear prediction (NLP) method, serving as a benchmark.

The comparisons consistently demonstrate that the proposed technique achieves higher prediction accuracy than the NLP method. Elias Elisante Lukwaro etel (2024)

METHODOLOGY

This indicates the effectiveness of the SSA-SVM approach in capturing and modeling the underlying patterns and dynamics of the hydrologic variables. Improved accuracy is particularly valuable in real-time operation studies, where reliable forecasts are crucial for decision-making and resource management.

Different Machine Learning Structure Used

In our research, we implemented a method called Stacked Generalization, also known as Stacking, to predict values within the training set. Stacked Generalization, also known as stacking, is an approach that enhances overall accuracy by combining the predictions of multiple methods. The Stacking model consists of two key components: base structures (Level-0 models) and a metamodel (Level-1 model). The base structures are individual machine learning structures that are trained independently on the dataset. For our study, we specifically selected three base structures: Support Vector Classifier (SVC), K-Nearest Neighbor (KNN) algorithm, and Decision Tree Classifier (DTC). Each base model learns patterns from the given features and produces predictions using its specific algorithm.

Once the base structures generate their predictions, we move on to the meta-model. The base structures give inputs to the meta-level in the forms of prediction and employs its own techniques to determine the final classification. In our research, we utilized Binary Logistic Regression as the meta-model. The meta-model analyzes the combined predictions as outcomes from the base structures and makes a final decision based on its evaluation.

Our utilization of Stacked Generalization aims to leverage the strengths of multiple structures and enhance prediction accuracy. Each base model captures unique aspects of the data, and the metamodel combines their predictions to make a more accurate classification decision. This approach allows us to benefit from the diversity and insights provided by the base structures.

The use of Stacked Generalization offers several advantages. It helps overcome biases and limitations associated with individual structures by aggregating their predictions. By using the meta-model to assess and weigh the base structures' predictions, we aim to achieve more reliable and accurate predictions on the training set. Our ultimate goal is to create a robust predictive model by harnessing the collective knowledge and abilities of multiple structures.

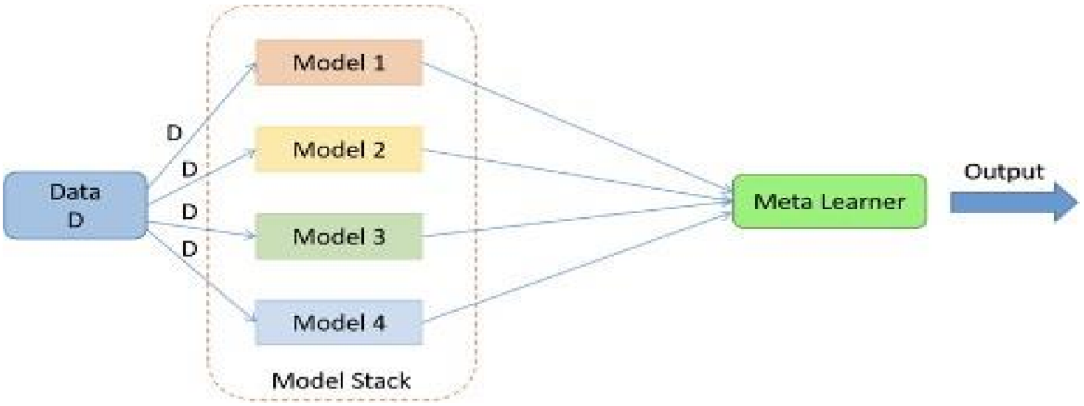

Figure 1.1: Diagram of Stacked Generalization.

Binary Logistic Regression is a statistical modeling technique used to predict binary outcomes or classify data into two categories. It is widely used in various fields, such as social sciences, epidemiology, and finance, to analyze the relationship between predictor variables and a binary response variable. Binary Logistic Regression operates by modeling the log odds of the occurrence of a binary event using the logistic function. The logistic function,

also known as the sigmoid function, transforms the linear combination of predictor variables into a probability between 0 and 1.

Here's a detailed explanation of how Binary Logistic Regression works:

Model Representation: Binary Logistic Regression represents the relationship be- tween the predictor variables and the probability of the binary outcome using the logistic function. This function maps any real-valued input to a probability value, representing the likelihood of the positive class. Parameter Estimation: The algorithm estimates the parameters of the logistic function by maximizing the likelihood of the observed data. It iteratively adjusts the coefficients to find the optimal values that maximize the likelihood of the observed binary outcomes, given the predictor variables.

Decision Boundary: Once the model parameters are estimated, a decision boundary is created to classify new data points. The decision boundary separates the two classes based on the predicted probabilities. A threshold value, typically 0.5, is used to determine whether a data point belongs to the positive or negative class.

Prediction Phase: To make predictions on new, unseen data, the algorithm applies the estimated parameters to the predictor variables and calculates the probability of the positive class. If the probability exceeds the chosen threshold, the data point is classified as the positive class; otherwise, it is classified as the negative class.

Some important considerations when using Binary Logistic Regression include:

Feature Selection: Careful selection of relevant predictor variables is crucial for the model's performance. Including irrelevant or redundant variables may introduce noise and impact the accuracy of the model.

Model Evaluation: Diverse evaluation metrics, including accuracy, precision, recall, and F1 score, can be utilized to evaluate the Binary Logistic Regression model's performance. Methods like crossvalidation can overestimate the model's performance on data it has not encountered before.

Assumptions: Binary Logistic Regression assumes a linear relationship between the predictor variables and the log odds of the outcome. It also assumes

independence of observations and the absence of multicollinearity among the predictors. Regularization: Regularization techniques, like L1 or L2 regularization, can be employed to avoid overfitting and enhance the model's capacity to generalize.

Binary Logistic Regression has several advantages:

Interpretability: The coefficients of logistic regression structures are interpretable, (1996) allowing for the understanding of the direction and magnitude of the influence of each predictor variable on the log odds of the outcome. This interpretability is valuable for explaining the relationship between the predictors and the binary response variable.

Probabilistic Predictions: Unlike other classification algorithms that provide discrete class labels, Binary Logistic Regression produces probabilities that represent the confidence or likelihood of the positive class. These probabilities can be used to rank or prioritize predictions.

Efficiency: Binary Logistic Regression is computationally efficient and can handle large datasets with a high number of predictor variables.

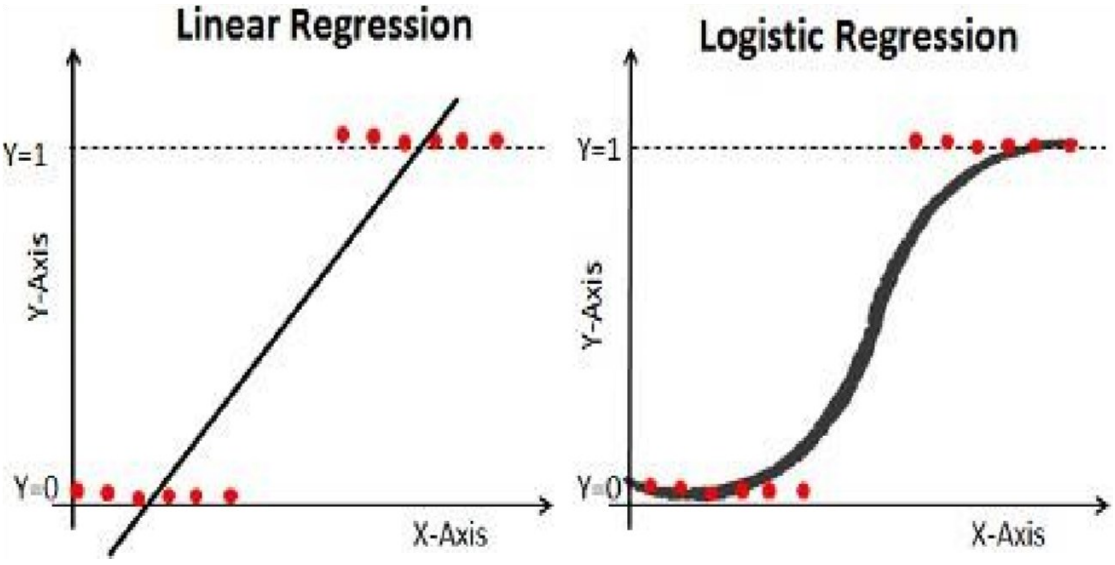

Figure 1.2: Distinguishing Between Linear and Logistic Regression

However, there are some considerations when using Binary Logistic Regression:

Linearity Assumption: Logistic regression postulates a linear connection between the predictor variables and the log odds of the outcome. If the relationship is non-linear, additional transformations or interactions may be necessary.

Imbalanced Data: Logistic regression can be sensitive to imbalanced datasets where one class is significantly more prevalent than the other. Methods like over-sampling, under-sampling, or adjusting class weights can assist in dealing with this problem

Outliers: Outliers in the data can have a significant impact on logistic regression.

Support Vector Classifier

The Support Vector Classifier (SVC) is a popular algorithm in machine learning used for classification tasks. It belongs to a family of algorithms called Support Vector Machines (SVM). The primary objective of the Support Vector Classifier is to find an optimal hyperplane that separates different classes of data points in the feature space. The algorithm aims to maximize the margin, which is the distance between the hyper-plane and the nearest data points of each class. By maximizing the margin, the SVC seeks to achieve better generalization and prevent overfitting. The Support Vector Classifier (SVC) operates by employing a kernel function to get higher-dimensional feature space from the input data transformation. This transformation enables the algorithm to find a linear decision boundary that may not be possible in the original feature space. Different types of kernel functions, such as linear, polynomial, radial basis function (RBF), and sigmoid, can be employed to perform this transformation. The Support Vector Classifier is known for its ability to handle both linearly separable and nonlinearly separable datasets. By utilizing various kernel functions, it can capture complex relationships between features and create flexible decision boundaries.

In summary, the Support Vector Classifier is a widely used algorithm for classification tasks. The main objective is to locate the most favorable hyperplane that effectively separates distinct classes by margin maximization. Through the use of kernel functions, it can handle complex data distributions and capture non-linear relationships between features. Proper selection of hyperparameters is essential for obtaining accurate classification outcomes.

K- NEAREST NEIGHBOR

The K-nearest neighbor (KNN) algorithm is a widely used technique in machine learning for classification and regression tasks K-Nearest Neighbor (KNN) Tavish (2024). It is an algorithm that is non-parametric and does not make any assumptions about the underlying distribution of the data. The fundamental idea behind KNN is to classify or predict a new data point based on the characteristics of its nearest neighbors. Instead of using mathematical models, KNN directly uses the training dataset to make predictions.

Here is a step-by-step explanation of the KNN algorithm:

Training Phase: During the training phase, KNN simply stores the feature vectors and their corresponding labels from the training dataset. No computations are performed at this stage.

Prediction Phase: Whenever a new data point requires classification or prediction, the K-nearest neighbors (KNN) algorithm computes the distances to all the data points present in the training dataset from the relatively new points. Various distance metrics, such as Euclidean or Manhattan distance, can be used for this purpose.

Selecting K: K represents the number of nearest neighbors to consider.
It is typically an odd number to avoid ties in classification tasks with an even number of classes.

Finding Neighbors: KNN identifies the K nearest neighbors of the new data point by selecting the data points with the smallest distances. These neighbors are determined based on the chosen distance metric.

Classifying or Predicting: In classification tasks, based on the most common class among its K nearest neighbors, KNN assigns the class label of the new data point. In regression tasks, it predicts the average or median value of the K nearest neighbors. There are a few important considerations when using the KNN algorithm.

In summary, the K-nearest neighbor algorithm is a straightforward and widely used method for classification and regression tasks. It relies on the proximity of data points to make predictions or classifications. While simple, KNN provides flexibility and can be effective in various domains. Decision Tree Classifier

The Decision Tree Classifier is an algorithm frequently employed in machine learning to address tasks during our effort to classify things. It creates a tree-shaped model that utilizes input features to make predictions. Lior Rokach, Oded Maimon (2005),

The Decision Tree Classifier follows a process to build the tree model and make predictions:

Splitting Criteria: The algorithm evaluates various splitting criteria to choose the best feature and splitting point at each node. Typical criteria include Gini impurity and entropy. Gini impurity assesses the likelihood of misclassifying a randomly selected element, while entropy gauges the level of disorder within a dataset.

Recursive Splitting: The Decision Tree Classifier recursively splits the dataset based on the chosen splitting criteria. It continues to divide the data into smaller subsets at each internal node until specific stopping conditions are met. Stopping conditions may include reaching a maximum tree depth, having a minimum number of samples per leaf node, or other user-defined criteria.

Prediction Phase: Once the decision tree is built, it can be used to make predictions on new, unseen data. The algorithm traverses the tree from the root node to a leaf node based on the feature values of the input data. The

predicted class label is determined by the majority class of the training samples associated with that leaf node.

The Decision Tree Classifier has several advantages:

Interpretability: Decision trees provide a transparent and intuitive representation of the decision-making process. The resulting tree structure can be easily visualized and understood, facilitating explanation and interpretation.

Handling Non-Linear Relationships: Decision trees can capture complex non-linear relationships between input features and the target variable. By recursively splitting the data, decision trees can create flexible decision boundaries that can accommodate intricate patterns.

Handling Missing Data: Decision trees can handle missing data by using surrogate splits. Surrogate splits act as backup splits when certain features have missing values, enabling the algorithm to still make predictions.

However, there are some considerations when using the Decision Tree Classifier:

Over-fitting: Decision trees can over-fit the training data, where the model becomes too complex and adapts too closely to the specific training examples. Over- fitting can result in poor generalization on unseen data. Techniques like pruning or limiting the depth of the tree can help prevent over-fitting.

Sensitivity to Small Variations: Decision trees can be sensitive to small variations in the training data, potentially leading to different tree structures or splitting decisions. Ensemble methods such as Random Forests or Gradient Boosting can mitigate this sensitivity.

Handling Categorical Variables: Decision trees typically require categorical variables to be transformed into numerical values or binary indicators to incorporate them effectively into the splitting process.

In summary, the Decision Tree Classifier is a versatile algorithm for classification tasks. It constructs a tree-like model by recursively splitting the data based on selected criteria. The interpretability and ability to capture non-

linear relationships make it a widely used algorithm. However, precautions should be taken to prevent overfitting and handle categorical variables appropriately

Performance Evaluation

The rainfall dataset utilized in this project has been consolidated into a CSV file. The analysis focuses on the rainfall patterns observed in Lagos Nigeria spanning from 1901 to 2017.

Prediction Model Classification

In this thesis, we employed four distinct classifiers, namely K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Binary Logistic Regression, for prediction purposes. Additionally, we implemented a stacked model that utilized the four aforementioned classifiers as base models. The meta-model used in the stacked model was Binary Logistic Regression. The stacked model combines the outputs of the base models to make the final prediction, leveraging the strengths and diverse perspectives of each classifier to improve accuracy.

These models were first used to analyze the monsoon data specifically for the monsoon season in Lagos State which encompasses the months of June, July, August, September, and October [17]. Consequently, only the rainfall data pertaining to these months was considered for the analysis. The results obtained from applying the models are presented in the table below.

Table 5.1: Outcomes of the models applied using Monsoon Rainfall Data.

Predictive	Models Accuracy	Standard Deviation
K-Nearest Neighbors (KNN)	83.4	0.158
Support Vector Classifier (SVC)	86.4	0.146
Decision Tree Classifier (DTC)	78.3	0.217
Binary Logistic Regression	83.6	0.164
Stacked Generalization	84.8	0.159

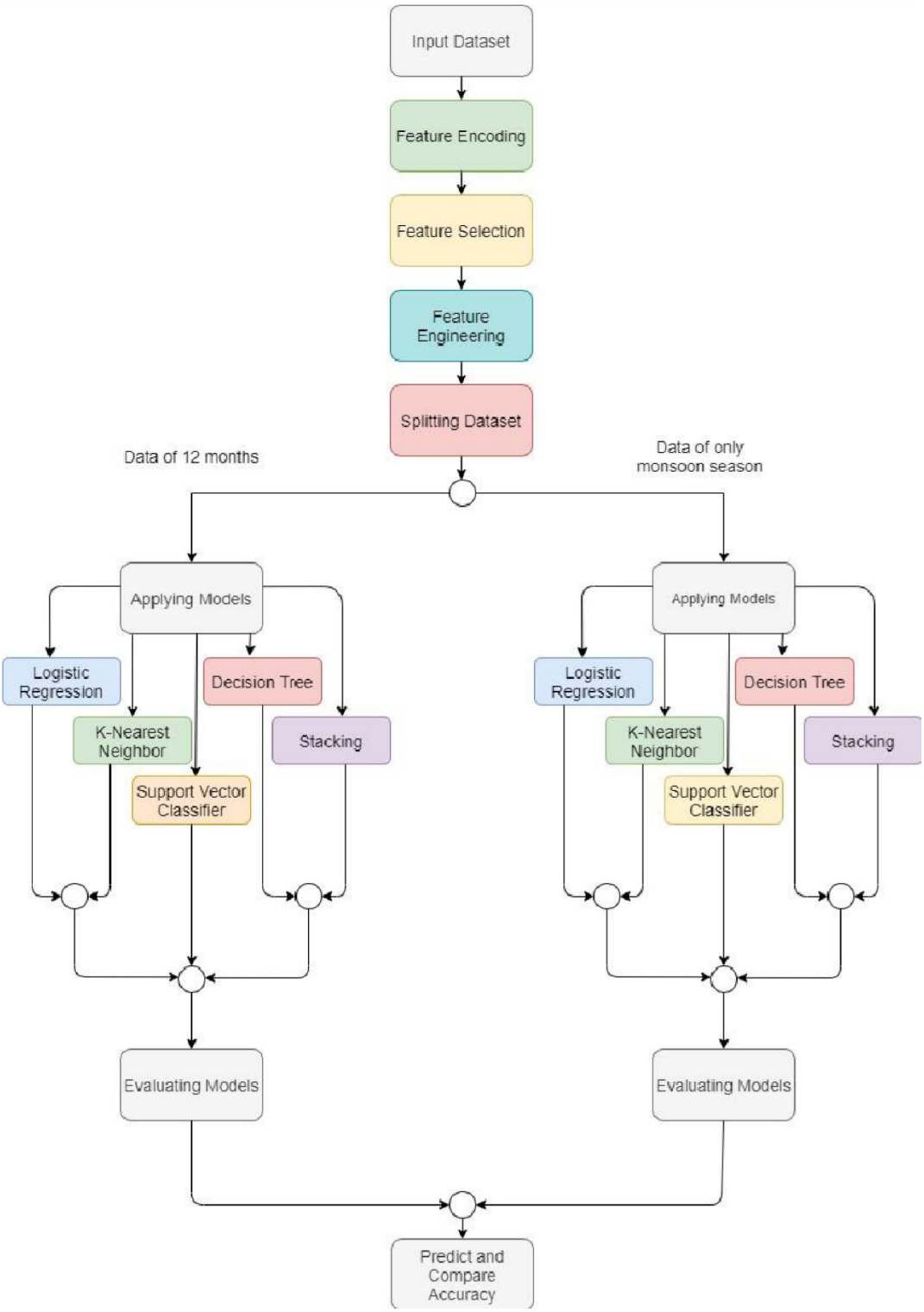

Figure 1.3: Our Proposed System is Illustrated

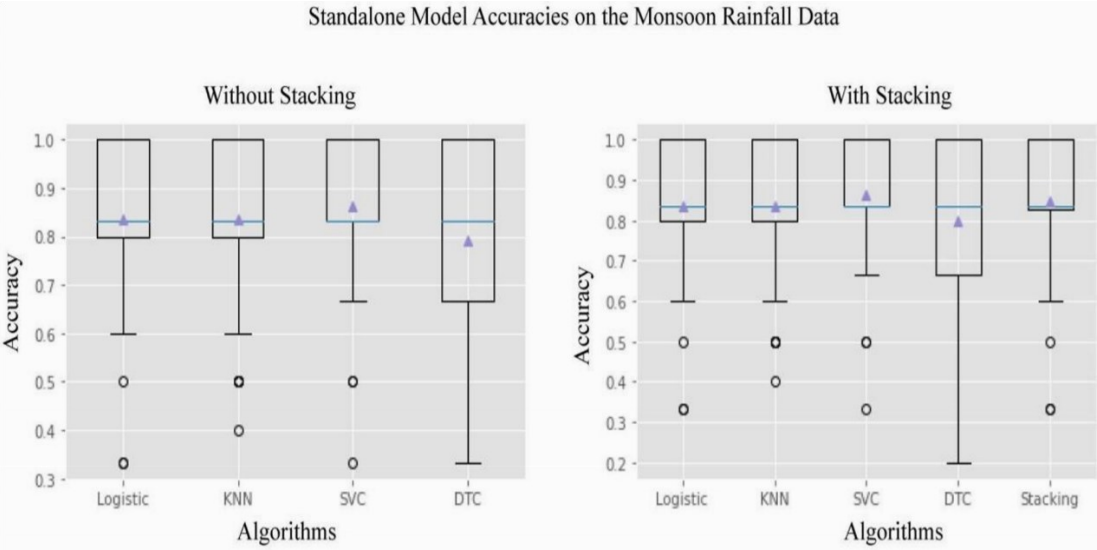

Figure 1.4: Whisker box plot for Standalone Model Accuracy on the Monsoon Rainfall Data Below is the table presenting the results obtained by utilizing the models on monsoon data from every month of the year. The objective is to improve accuracy by employing the same set of models. Shown in Table 1.

Table 1.2: Outcomes of the models applied using twelve months of Rainfall Data

Predictive	Models Accuracy	Standard Deviation
K-Nearest Neighbors (KNN)	74.6	0.172
Support Vector Classifier (SVC)	90.6	0.111
Decision Tree Classifier (DTC)	77.2	0.772
Binary Logistic Regression	93.0	0.103
Stacked Generalization	93.3	0.098

Upon examining the data presented in both tables, it is evident that utilizing rainfall data from all 12 months of the year yielded better accuracy for all models, except the Decision Tree Classifier (DTC), in comparison with using only the rainy season months. This suggests that predicting floods relies not only on rainfall during the monsoon season but also on precipitation throughout the entire year, alongside other contributing elements.

CONCLUSION

The results indicated that the ensemble model demonstrated superior accuracy and precision compared to the individual models. The ensemble model combines the predictions of multiple models to produce a final prediction, leveraging the strengths of each component. This integration of multiple models contributed to improved flood prediction performance.

Furthermore, the study found that the ensemble model outperformed previous studies in flood prediction that employed machine learning models. This suggests that the ensemble approach presented in this research represents a significant ad- advancement in flood prediction compared to existing literature.

Moreover, this study highlights the advantages of utilizing an ensemble machine-learning model for flood prediction. The ensemble model, by combining multiple models, achieved better accuracy and precision than individual models and surpassed the performance of previous studies in the field of flood prediction using machine learning.

REFERENCES

Akinola Adesuji komolafe (2015) A Review of Flood Risk Analysis in Nigeria

A.K. Lohani, N. Goel, and K. Bhatia, (2014) "Improving real time flood forecasting using fuzzy inference system," *Journal of Hydrology*, vol. 509, pp. 25–41, Feb. 2014. doi: 10.1016/j.jhydrol.2013.11.021. [Online]. Available: https://doi.org/ 10.1016/j.jhydrol.2013.11.0

A. Fare Hamad Alijoni, Ahmad Alkhodre, Anan Ahmad Abi sen, eltel 2023 Flood Prediction using Hydrologic and ML-based Modeling: A Systematic Review, DOI: 10.14569/IJACSA.2023.0141155

Chuhua Zhang, Xiaojain Shao, Dewi Li (2013) Knowledge-based Support Vector Classification Based on C-SVC DOI: 10.1016/j.procs.2013.05.137

Chandrahas Mishra, D. L.Gupta (2017) Deep Machine Learning and Neural Networks: An Overview

Cheng Fan, Meiling Chen e tl (2023) A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery from Building Operational Data.

Diego Fernández-Nóvoa, José González-Cao, Orlando García-Feal (2024) Enhancing Flood Risk Management: A Comprehensive Review on Flood Early Warning Systems with Emphasis on Numerical Modeling.

Elias Elisante Lukwaro etel (2024) :A Review on NLP Techniques and Associated Challenges in Extracting Features from Education Data DOI: 10.12785/ijcds/160170.

G. T Raadgever, Nikeh Booister etel (2018) Flood Risk Management Strategies DOI: 10.1007/978-3-319-67699-9_8

Katarzyna Stapor (2017) Evaluation of classifiers: current methods and future research directionsDOI: 10.15439/2017F530

Iisa. L. Lottes, Mirina A Adletr (1996)Using and Interpreting Logistic Regression: A Guide for Teachers and Students DOI: 10.2307/1318743

Lior Rokach, Oded Maimon (2005) Decision Trees.

Mariette Awad, Rahul Khanna (2015) Support Vector Machines for Classification DOI: 10.1007/978-1-4302-5990-9_3

Mona Ali Mohammed (2024) Effect of Using Numerical Data Scaling on Supervised Machine Learning PerformanceDOI: 10.37376/glj.vi67.5903 .

Roichard P Anderson et el (2023)Understanding logistic regression analysis in clinical reports: An introduction DOI: 10.1016/S0003-4975(02)04683-0.

Quoc Pham, Sk Ajim Ali et el (2022) Flood vulnerability and buildings' flood exposure assessment in a densely urbanised city: comparative analysis of three scenarios using a neural network approach DOI: 10.1007/s11069-022-05336-5.

Sreejesh S, Sanjay Mohapatra, Dr Anusree M.R(2014) Binary Logistic Regression DOI: 10.1007/978-3-319-00539-3_11

Sharad K. Jain, Pankaj Mani (2017) A Brief Review of Flood Forecasting Techniques and Their Applications

Tavish (2024) Guide to K-Nearest Neighbours Algorithm in Machine Learning

Tanvir Rahman, Miah Mohammad Asif Syeed (2023) Flood Prediction Using Ensemble Machine Learning Model DOI: 10.1109/HORA58378.2023.10156673

Teuku Rizky Noviandy, Aga Maulana, Ghazi Mauer Idroes etel (2013) Ensemble Machine Learning Approach for Quantitative Structure- Activity Relationship Based Drug Discovery: DOI: 10.60084/ijds.v1i1.91

Tanvir Rahman (2023) Flood Prediction Using Ensemble Machine Learning Model.