International Journal of Innovative Research and Scientific Studies, 8(5) 2025, pages: 201-210

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Flood forecasting using machine learning methods in a visual programming environment

Gulzat Ziyatbekova[1,2,3], Dauren Darkenbayev[1,4*]

[1]*Al-Farabi Kazakh National University, Almaty, Kazakhstan.*
[2]*Almaty Technological University, Almaty, Kazakhstan.*
[3]*Institute of Information and Computational Technologies Almaty, Kazakhstan.*
[4]*Kazakh National Women's Teacher Training University, Almaty, Kazakhstan.*

Corresponding author: Dauren Darkenbayev (*Email: dauren.darkenbayev1@gmail.com*)

Abstract

Floods are among the most destructive natural phenomena, significantly impacting the environment, economy, infrastructure, and public safety. Effective forecasting of such emergencies is crucial, especially in the context of global climate change and the increasing population density in coastal and low-lying areas. This work aims to develop and analyze flood forecasting models using machine learning algorithms within the Orange visual software environment. The study employed five algorithms: Random Forest, Decision Tree, Gradient Boosting, AdaBoost, and K-Nearest Neighbors. A comparative analysis of these models was conducted using key classification metrics, including Accuracy, Precision, Recall, and AUC. Special emphasis was placed on visualizing the results and assessing the usability of models within the Orange environment. The findings can be valuable for educational purposes, such as teaching students to work with real environmental data, as well as for practical applications in early warning systems and flood monitoring.

Keywords: Algorithms, Analytics platform, Flood, Forecasting, Machine learning, Visual analysis.

DOI: 10.53894/ijirss.v8i5.8598
Funding: This work is supported by the Research Institute of Mathematics and Mechanics at Al-Farabi Kazakh National University, Kazakhstan (Grant number: AP19678157).
History: Received: 27 May 2025 / **Revised:** 1 July 2025 / **Accepted:** 3 July 2025 / **Published:** 17 July 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Floods are among the most destructive natural phenomena, significantly affecting infrastructure, agriculture, and public safety [1]. In recent decades, the frequency and intensity of such disasters have increased due to climate change and anthropogenic load. In this regard, the need for early warning and forecasting systems based on objective data is increasing [2].

Modern machine learning methods enable the analysis of large volumes of climatic, hydrological, and geographical data to predict the likelihood of floods. Among the available platforms for data analysis, the Orange environment deserves special attention, as it allows the implementation of the entire machine learning cycle from pre-processing to visualization of results using a visual interface [3-8].

The development of intelligent analysis methods aimed at solving applied problems in the field of ecology and civil defense is one of the urgent tasks of modern science and education. In particular, the creation of accessible and reliable flood forecasting models can significantly reduce damage from natural disasters and increase the readiness of emergency services to respond. The use of Orange provides a unique opportunity for students, teachers, and specialists to quickly master machine learning technologies without the need for programming. This paper integrates machine learning and visual analytics methods, which allow the use of intelligent analysis tools in educational projects and applied research. The proposed approach ensures not only forecasting accuracy but also ease of implementation, which distinguishes it from more complex software solutions that require in-depth training in programming [9-15].

The purpose of the work is to build and evaluate flood forecasting models based on machine learning methods using the Orange visual environment.

To achieve this goal, the following tasks were set:

- Prepare and load a dataset containing parameters affecting flood risk into the Orange environment;
- Build classification models using various algorithms (Random Forest, Decision Tree, Gradient Boosting, AdaBoost, K-Nearest Neighbors)
- Evaluate the accuracy and stability of models using cross-validation methods;
- Analyze the significance of features and interpret the results using Orange visual analysis tools.

In recent years, numerous studies have been published on the application of machine learning methods to natural forecasting problems. In particular, the work of Mosavi et al. [16] considers the use of neural networks, decision trees and ensemble methods for flood forecasting. Yaseen et al. [17] conducted an extensive comparative analysis of classification algorithms in flood modeling and showed the high efficiency of ensemble models. The article by Das and Gupta [18] emphasizes the importance of using factors such as precipitation, water levels in rivers, and soil moisture in building forecast models. Orange, as a visual data analysis tool, has gained recognition in the educational environment and scientific research due to its accessibility and integration with machine learning libraries [19].

A distinctive feature of this work is the comprehensive implementation of modeling stages within the Orange environment, without the need for coding, which makes the proposed approach particularly accessible to a broad audience of users, from students to practicing specialists. Unlike previously published works, where the focus is primarily on the use of software libraries and programming languages (e.g., Python, R), this study employs visual model building, significantly simplifying the process and enabling rapid hypothesis testing. Additionally, unlike most studies limited to a single model, this research conducts a comparative analysis of several popular machine learning algorithms (Random Forest, Decision Tree, Gradient Boosting, AdaBoost, K-Nearest Neighbors) on a single dataset, allowing for the identification of the most effective method in terms of accuracy and interpretability of results. It is anticipated that the findings will be valuable for educational institutions, specialists in civil protection and environmental monitoring, and will serve as a foundation for developing more accurate and reproducible models of early flood warning systems.

2. Materials and Methods

To implement the tasks, the Orange software tool was used as a visual data analysis environment with machine learning support, developed in Python. Orange provides an intuitive interface based on the concept of block model construction (workflow), which allows the user to form a data processing chain without the need for programming. This makes the platform especially relevant for applied research and educational purposes.

2.1. Research Data

The flood_prediction_dataset.csv dataset was used as the initial material, containing synthetically generated and realistically modeled data on meteorological and hydrological parameters. The set includes the following parameters [20-26]:

- Rainfall - amount of precipitation (mm);
- River_Level - water level in the river (m);
- Soil_Moisture - soil moisture (%);
- Temperature - air temperature (°C);
- Humidity - relative air humidity (%);
- Wind_Speed - wind speed (m/s);
- Flood - target binary variable (0 - no flood, 1 - flood).

The data were pre-processed: missing values were excluded, there were no categorical variables, and no feature scaling was required.

	Name	Type	Role	Values
1	Rainfall	N numeric	feature	
2	Soil Moisture	N numeric	feature	
3	River Level	N numeric	feature	
4	Temperature	N numeric	feature	
5	Season	C categorical	feature	Autumn, Spring, Summer, Winter
6	Flood	C categorical	feature	No, Yes

Figure 1.
Structure of the initial dataset for flood modeling in the Orange environment.

Figure 1 shows the structure of the original dataset used to build flood forecasting models. The dataset includes both numerical features (precipitation, soil moisture, river level, air temperature) and categorical variables (season, presence/absence of flood). This structure allows for the efficient use of classification algorithms in the Orange environment to identify patterns preceding the occurrence of floods.

2.2. Software implementation

During the work, a data analysis scheme was implemented in Orange, including the following stages:

- Loading data using the File widget;
- Primary analysis of the data structure using Data Table, Box Plot, Distributions;
- Splitting data into training and test samples using the cross-validation method (10-fold);
- Building and training machine learning models:
  1. Gradient Boosting;
  2. Decision Tree;
  3. Random Forest;
  4. AdaBoost;
  5. k-Nearest Neighbors.
- Evaluating models by key quality metrics (Accuracy, Precision, Recall, F1-score, AUC) using the Test & Score widget;
- Visualizing results using ROC Analysis.

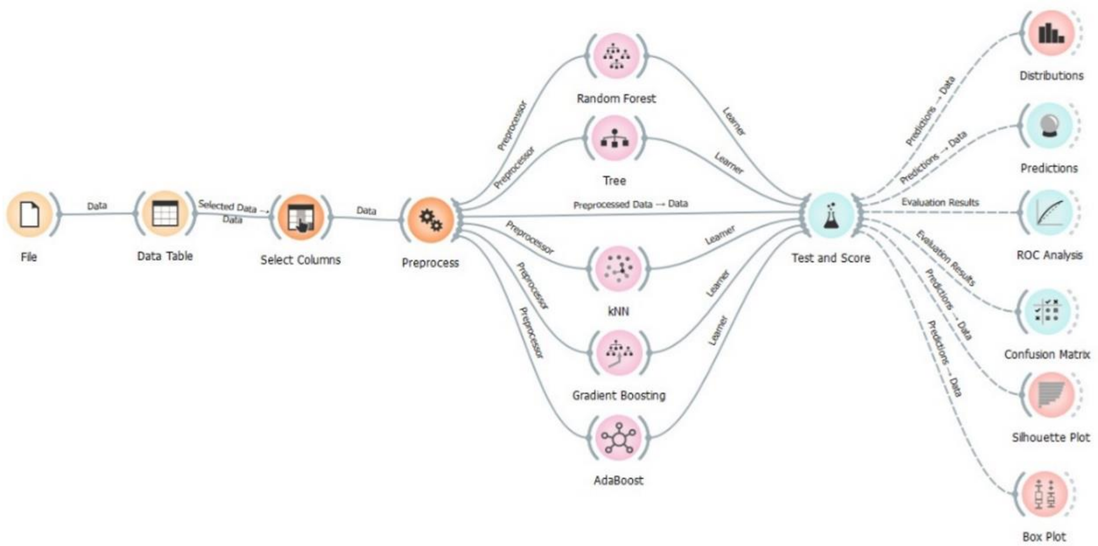

Figure 2.
A framework for building and evaluating machine learning models in the Orange environment.

Figure 2 illustrates the complete modeling process in the Orange visual environment: from loading and preprocessing data to building models (Random Forest, Decision Tree, kNN, Gradient Boosting, AdaBoost) and their subsequent evaluation using quality metrics (ROC analysis, confusion matrix, distribution plots, etc.).

2.3. Algorithms and Mathematical Foundations of Forecasting

To build the forecasting models, both basic and ensemble classification algorithms were selected. This approach allows for a comparative analysis of the models' performance and the identification of the most stable solutions [5, 27-29].

Random Forest. This is an ensemble of M decision trees trained on random subsamples and feature subsets:

$$\bar{y} = majority\_vote(h_1(X), h_2(X), \ldots, h_M(X)) \quad (1)$$

$h_i(X)$ - classification of the i -th tree;

The output is the most frequent answer among all trees (voting). Random forest reduces variance and increases resistance to overfitting compared to a single tree.

Decision Tree. The decision tree model recursively partitions the feature space, selecting attributes that minimize entropy or maximize information gain:

$$IG(S,A) = H(S) - \sum_{\upsilon \in Values(A)} \frac{|S_0|}{S} H(S_0) \quad (2)$$

where, $H(S)$ - entropy of a set of objects S; A - attribute (for example, precipitation level);

IG - Information gain.

k-Nearest Neighbors. Model-free method: classification is based on the classes of the k closest points in the training set:

$$\bar{y} = \arg\max_{c \in \{0,1\}} \sum_{i=1}^{k} \mathrm{II}(y_i = c) \quad (3)$$

where, II is the indicator function; y_i is the class label for the i -th nearest neighbor (in Euclidean metric):

$$d(X, X_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2} \quad (4)$$

Gradient Boosting. Gradient Boosting is an ensemble method in which several weak models (usually decision trees) are trained sequentially, with each subsequent model attempting to correct the mistakes of the previous ones.

Let's have a training set:

$$D = \{(x_i, y_i)\}_{i=1}^{n}, x_i \in R^m, y_i \in R \quad (5)$$

The model is constructed as a sum of weak models:

$$F(x) = \sum_{t=1}^{T} \gamma_t h_t(x) \quad (6)$$

where $h_t(x)$ - weak model at i -th iteration (usually a decision tree), γ_t model weight, selected taking into account the minimization of the loss function. At each step, t the algorithm minimizes the loss function $L(y, F(x))$, for example:

$$F_t(x) = F_{t-1}(x) - \eta \nabla F_{t-1} L(y, F_{t-1}(x)) \quad (7)$$

Where is the learning rate, and the gradient is taken over the prediction values? For classification problems, such as binary flood prediction (risk/no risk), the logistic loss function is used:

$$L(y, F(x)) = \log\left(1 + e^{-yF(x)}\right) \quad (8)$$

AdaBoost. AdaBoost (Adaptive Boosting) works by iteratively building an ensemble of weak classifiers, where each subsequent one focuses on the mistakes of the previous ones.

Let the binary problem be: $y_i \in \{-1, +1\}$, and the initial weights of all objects are equal to:

$$\omega_i^{(1)} = \frac{1}{n}, i = 1, \ldots, n \quad (9)$$

At each iteration $t = 1, \ldots, T$:

1. A weak classifier $h_t(x)$ is trained, minimizing the error:

$$\varepsilon_t = \sum_{i=1}^{n} \omega_i^{(t)} \cdot \mathrm{II}(h_t(x_i) \neq y_i) \quad (10)$$

2. The confidence coefficient of the classifier is calculated:

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right) \quad (11)$$

3. The weights of objects are updated:

$$\omega_i^{(t+1)} = \omega_i^{(t)} \cdot \exp(-\alpha_t y_i h_t(x_i)), i = 1, \ldots, n \quad (12)$$

Then the weights are normalized:

$$\omega_i^{(t+1)} := \frac{\omega_i^{(t+1)}}{\sum_j^n \omega_j^{(t+1)}} \tag{13}$$

The final prediction of the model:

$$H(x) = sign\left(\sum_{i=1}^{T} \alpha_t h_t(x)\right) \tag{14}$$

3. Results

Gradient Boosting and AdaBoost were the most effective methods for solving the problem of flood prediction based on selected features. This demonstrates the high ability of ensemble methods to accurately classify under conditions of a limited and possibly unbalanced data set. The Decision Tree model showed the worst result and may be less suitable for this task.

To evaluate the effectiveness of the models, cross-validation with a split into 10 folds (Stratified Cross Validation) was used. The following algorithms were tested:

1. Random Forest;
2. Decision Tree;
3. k-Nearest Neighbors (kNN);
4. Gradient Boosting;
5. AdaBoost.

Table 1.
Comparative table of results.

Model	AUC	CA	F1-score	Precision	Recall	LogLoss
Random Forest	0.990	0.985	0.985	0.985	0.985	0.190
Decision Tree	0.919	0.985	0.985	0.985	0.985	0.541
k-NN (k=5)	0.990	0.985	0.985	0.985	0.985	0.400
Gradient Boosting	1.000	1.000	1.000	1.000	1.000	0.000
AdaBoost	1.000	1.000	1.000	1.000	1.000	0.000

Gradient Boosting and AdaBoost performed best across all metrics, including AUC = 1.000, CA = 1.000, LogLoss = 0.000, indicating perfect classification on this sample. Random Forest and kNN also performed well (AUC = 0.990, CA = 0.985), but with a slight increase in log loss. Decision Tree performed worse than other models in AUC (0.919) and showed the highest log loss (LogLoss = 0.541), indicating lower reliability.

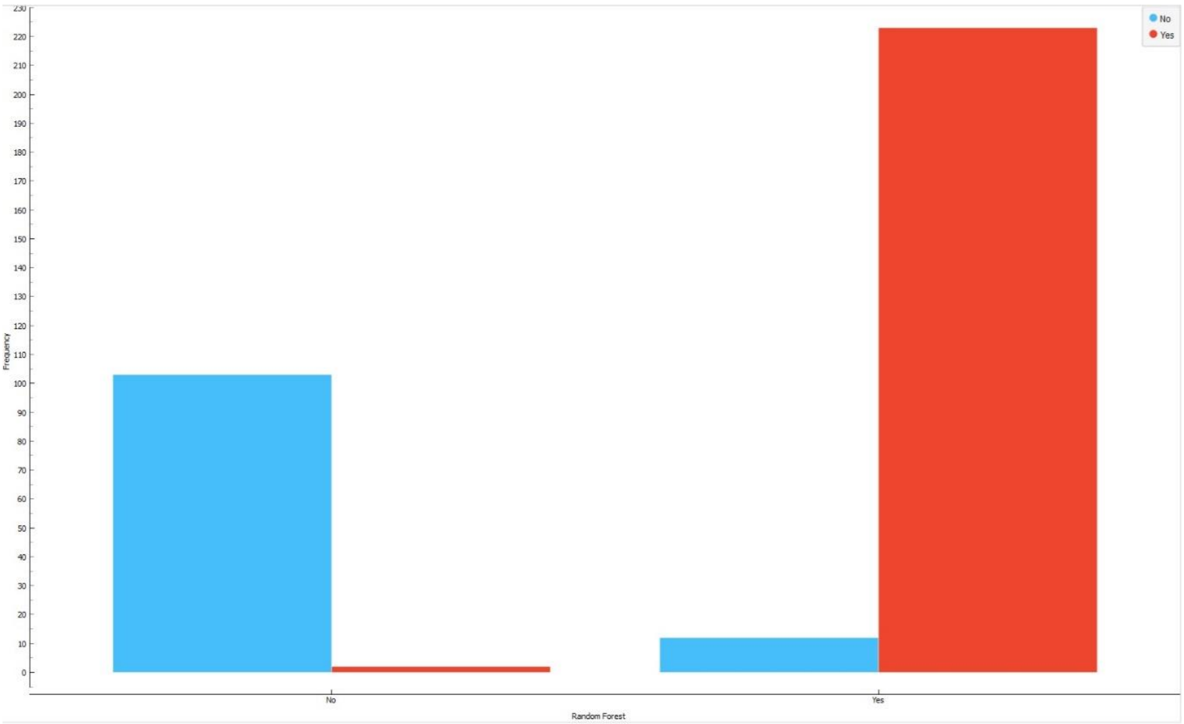

Figure 3.
Distribution of the target variable "Flood" in the dataset used to train the Random Forest model.

To evaluate the efficiency of the Random Forest model in the binary classification task of the Flood event, the distribution of the model predictions was analyzed in relation to the true class labels. Figure 3 shows the visualization of

the classification results, where the abscissa axis displays the values predicted by the model (No - no flood, Yes - flood), and the ordinate axis indicates the frequency of observations.
The color scheme of the graph indicates the true classes:

- Blue color corresponds to cases where the flood did not actually occur;
- Red color indicates cases where the flood did occur.

The distribution analysis indicates that when predicting "Yes" (the presence of a flood), the majority of observations correspond to actual flood cases, demonstrating a high sensitivity (recall) of the model concerning the positive class. Similarly, when predicting "No," the model accurately identifies cases without floods, confirming a high level of specificity. The number of false positives (FP) and false negatives (FN) is minimal, suggesting a low error rate and supporting the reliability of the Random Forest model for classifying meteorological and hydrological phenomena. Overall, the visual and quantitative results affirm the high accuracy and robustness of the Random Forest model in flood prediction based on the available dataset.

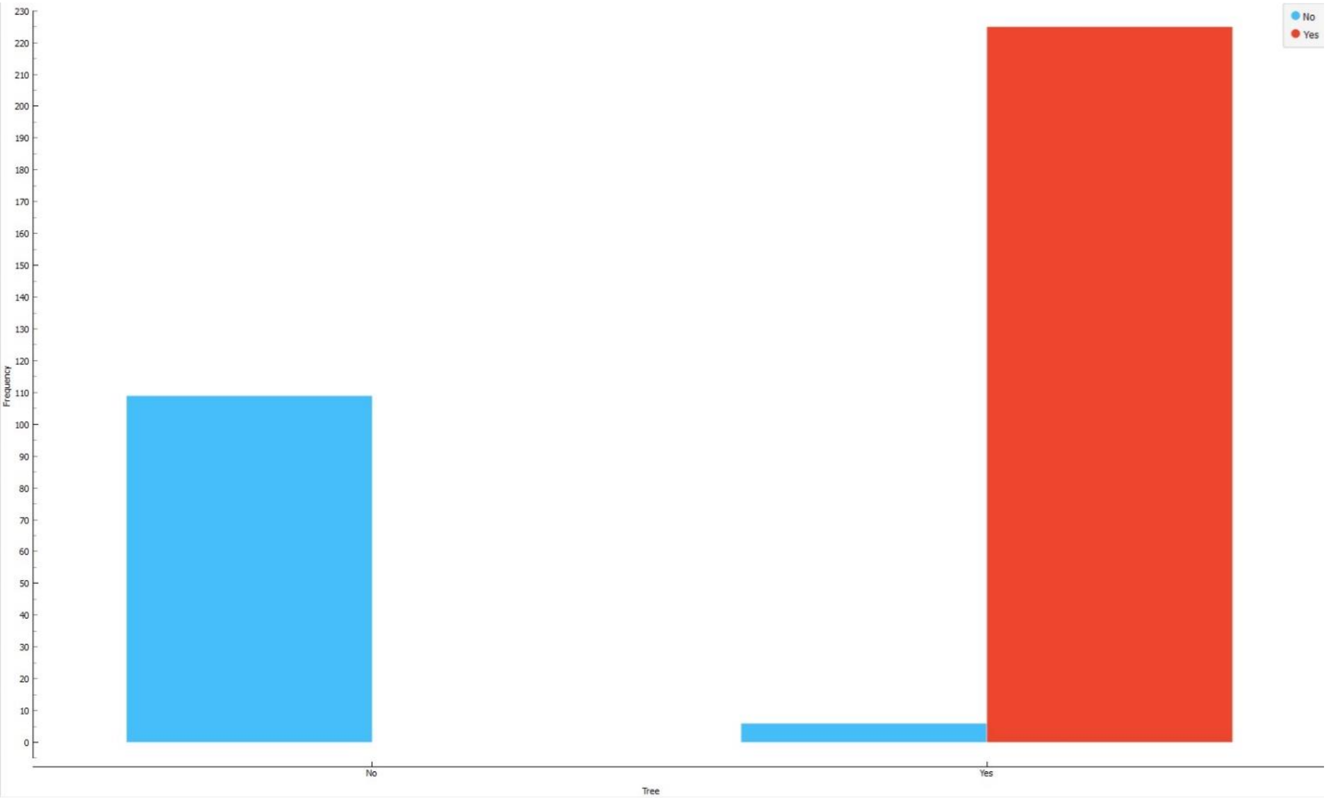

Figure 4.
Distribution of decision tree model predictions across event classes for the Flood event.

As can be seen from the graph, the decision tree model correctly classifies observations in most cases. The "No" column (prediction: there will be no flood) is dominated by blue, indicating high accuracy in determining the absence of floods. However, a small share of red in the same column indicates false negative errors cases when a flood occurred, but the model did not predict it. At the same time, the "Yes" column (prediction: there will be a flood) is almost entirely colored red, indicating high sensitivity (recall) of the model; it effectively identifies cases in which a flood is actually observed. A low number of blue segments in this column indicates rare false positives. Thus, the decision tree model demonstrates balanced performance, with good interpretability and fairly high accuracy in the tasks of monitoring extreme natural phenomena.

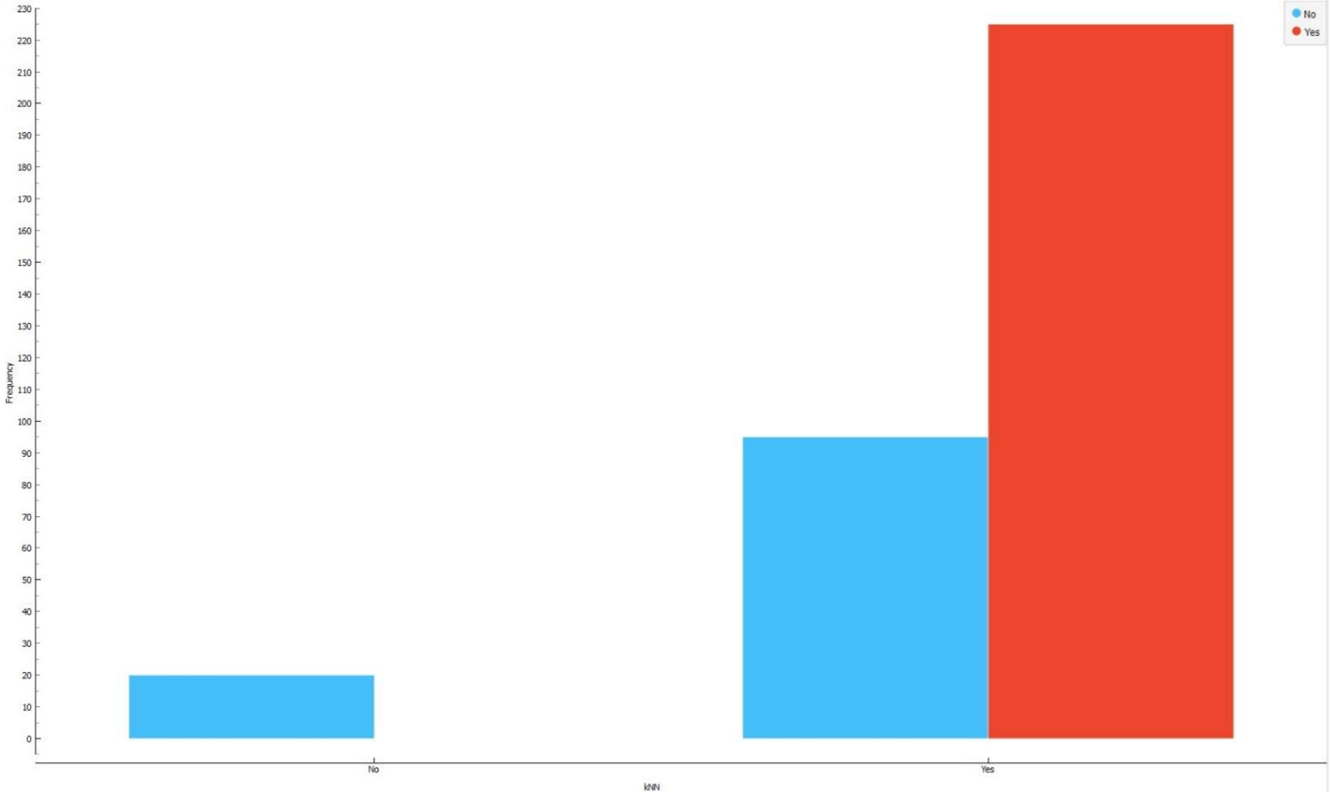

Figure 5.
Distribution of k-nearest neighbors (kNN) model predictions across event classes for the Flood event.

As can be seen in Figure 5, the kNN model predicts most observations as "Yes" indicating the presence of a flood. This is evident from the majority of cases in the "Yes" column, of which a significant portion actually corresponds to real floods (red segments). However, this category also contains a notable proportion of false positive predictions (blue segments), suggesting that non-flood cases are misclassified as floods. The "No" category predominantly contains blue segments, indicating that the model can recognize some safe situations, but such predictions are relatively infrequent. These results suggest that the model has high recall for the positive class ("Yes"), but its precision is compromised, leading to a large number of false positives. This behavior is typical for models where the nearest neighbors are determined by a metric that is not always robust to multicollinearity and large-scale differences between features. Therefore, although the kNN model can detect floods, it tends to overgeneralize the positive class and requires further tuning (such as selecting the optimal k value and data normalization) to reduce false positives.

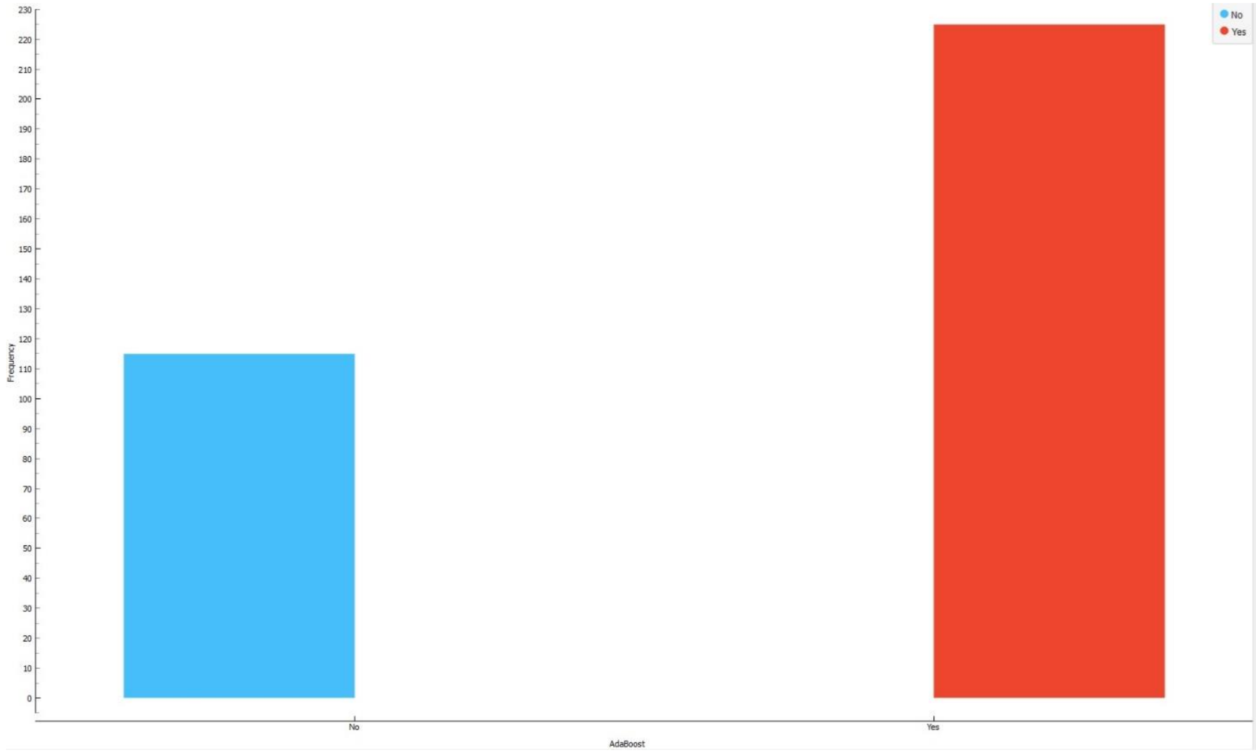

Figure 6.
Comparative Analysis of Flood Classifications: AdaBoost vs Gradient Boosting.

To evaluate the efficiency of the Gradient Boosting and AdaBoost models, we plotted the distribution of predictions by classes: "Yes" (predicted the presence of a flood) and "No" (predicted its absence). As can be seen from the plots, both models demonstrate an almost identical distribution of labels, in which the predominant number of objects are classified as "Yes". This suggests that both models produce the same predictions when processing the same input data. This distribution of predictions indicates high sensitivity of the models to features associated with the risk of flooding. Such behavior is especially important in problems where it is necessary to minimize the number of false negative situations in which a flood occurred but was not predicted. This may also indicate the presence of stable patterns in the data or a possible imbalance of classes in the training set. Quantitative metrics confirm visual observations: accuracy, recall, and F1-measure for the Gradient Boosting and AdaBoost models were the same, indicating comparable efficiency of these methods. Despite the differences in the training principles, adaptive weight adjustment in AdaBoost and loss function optimization in Gradient Boosting, both models demonstrated equally high generalization ability and successfully identified key features affecting flood risk. Thus, it can be concluded that both algorithms demonstrated equally high predictive ability on the considered dataset. This confirms the stability and robustness of ensemble methods in natural hazard classification problems. The choice between them in practical applications may be based not on the accuracy of predictions, but on other characteristics, such as training time, resistance to overfitting, computational complexity, and interpretability of the model. The combined visual and statistical analysis allows us to state that both models are effective tools for flood monitoring and early warning, and can be recommended for use in emergency forecasting systems.

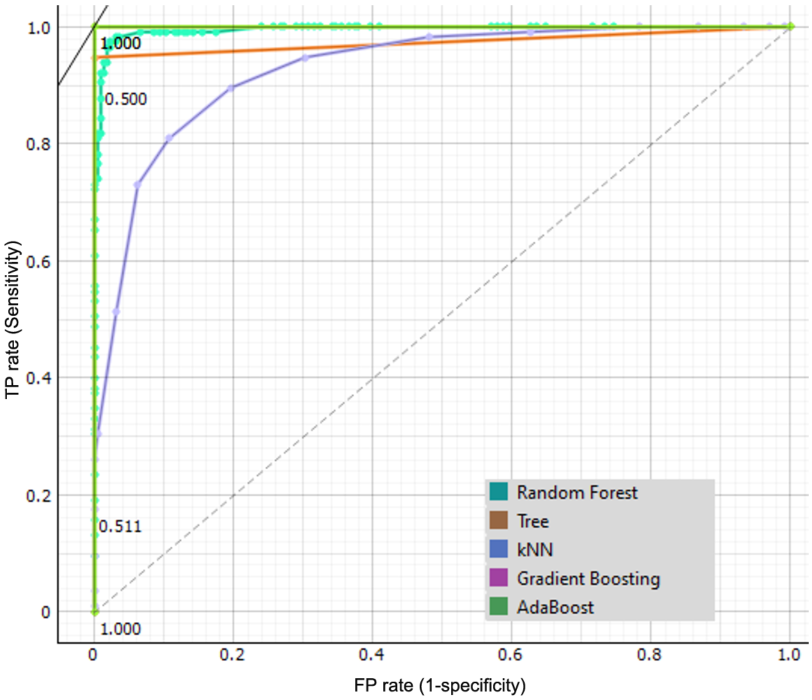

Figure 7.
ROC curves for comparing classification algorithms.

To assess the quality of the classifiers, a ROC analysis was performed (Figure 7). The text discusses a comparison of several classification algorithms, including Random Forest, Decision Tree, k-Nearest Neighbors (kNN), Gradient Boosting, and AdaBoost. ROC curves illustrate the relationship between true positive rate (TPR, sensitivity) and false positive rate (FPR, 1 − specificity) as the classification threshold varies. The Random Forest method demonstrated the highest performance, with its ROC curve approaching the upper left corner, indicating high sensitivity and specificity. Even at low FPR levels, the model achieved TPR values close to 1, suggesting minimal type I errors and excellent class separation ability. The area under the curve (AUC) approaching 1 further confirms the algorithm's high efficiency. Gradient Boosting and AdaBoost models showed results comparable to Random Forest, highlighting the effectiveness of ensemble methods for this problem. The Decision Tree classifier exhibited moderate performance, while the kNN algorithm showed less accurate predictions, particularly at low FPR values. Additionally, equal error costs for false positives and false negatives (FP Cost = FN Cost = 500) were set, with a positive class prior probability of 34%, providing a realistic evaluation in the context of imbalanced classes. In conclusion, the Random Forest algorithm is preferred for this classification task, especially when accurate positive class detection is critical, such as in flood detection.

4. Conclusion

The article implements and analyzes a flood forecasting system using machine learning methods in the Orange visual environment. The conducted studies have shown that modern classification algorithms can effectively solve early warning problems of natural emergencies, while providing high values of key metrics such as Accuracy, Precision, Recall, AUC, and LogLoss. The modeling results indicated that the best performers across all evaluation indicators are the ensemble methods Gradient Boosting and AdaBoost, which demonstrated ideal values of AUC (1.000), classification accuracy (CA = 1.000), and zero LogLoss, indicating a high ability to accurately recognize flood cases. The Random Forest and k-nearest neighbors (kNN) methods also showed stable and high results, only trailing the leaders in some logarithmic loss metrics. The Decision Tree model, despite its simplicity and interpretability, demonstrated the poorest performance, especially for the LogLoss metric, which limits its use in conditions requiring high reliability. Particular attention in the study was paid to visual data analysis and the usability of the Orange environment. The ability to build models without programming, visually configure data processing pipelines, and visually present results makes Orange an effective tool for both educational purposes and prototyping applied intelligent systems in ecology, civil defense, and environmental monitoring.

The presented article demonstrates the practical value of integrating visual analytical platforms and machine learning methods for solving predictive analysis problems. The results obtained can be used in the development of early flood warning systems, in educational projects, and in scientific research aimed at increasing resilience to natural disasters.

References

[1] M. Kalimoldayev, G. Ziyatbekova, T. Mazakov, S. Jomartova, A. Mazakova, and I. Suleimen, "Development of river flow modeling methodology," in *2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, 15-17 May 2024 2024, pp. 20-23

[2] G. Balakayeva, G. Kalmenova, D. Darkenbayev, and C. Phillips, "Development of an application for the thermal processing of oil slime in the industrial oil and gas sector," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska,* vol. 13, no. 2, 2023.

[3] Z. Wang, C. Lai, X. Chen, B. Yang, S. Zhao, and X. Bai, "Flood hazard risk assessment model based on random forest," *Journal of Hydrology,* vol. 527, pp. 1130-1141, 2015/08/01/ 2015. https://doi.org/10.1016/j.jhydrol.2015.06.008

[4] N. Razali, S. Ismail, and A. Mustapha, "Machine learning approach for flood risks prediction," *IAES International Journal of Artificial Intelligence,* vol. 9, no. 1, pp. 73–80, 2020. https://doi.org/10.11591/ijai.v9.i1.pp73-80

[5] Y. Wu, Z. Q. Zhang, X.Hu, W., and S. Si, "Prediction of flood sensitivity based on logistic regression, eXtreme Gradient Boosting, and random forest modeling methods," *Water Science and Technology,* vol. 89, no. 10, pp. 2605–2624, 2024.https://doi.org/10.2166/wst.2024.146

[6] S. Nevo *et al.*, "Flood forecasting with machine learning models in an operational framework," *Hydrology and Earth System Sciences,* vol. 26, no. 15, pp. 4013-4032, 2022.

[7] C. C. Ekwueme, C. J. Eze, and B. C. Okoro, "Machine learning based prediction of urban flood susceptibility from selected rivers in a tropical catchment area.," *Civil Engineering Journal,* vol. 7, no. 6, pp. 1107–1118. , 2021.

[8] Y. Gao, "Research on the application of machine learning technology in hydrological flood prediction," *Journal of Computer, Signal, and System Research,* vol. 2, no. 2, pp. 28–34, 2025.https://doi.org/10.71222/se6cyv71

[9] K. Qian, A. Mohamed, and C. Claudel, "Physics informed data driven model for flood prediction: Application of deep learning in prediction of urban flood development," *arXiv preprint arXiv:1908.10312,* 2019. https://doi.org/10.48550/arXiv.1908.10312

[10] C. Deng and W. Wang, "A two-stage partitioning monthly model and assessment of its performance on runoff modeling," *Journal of Hydrology,* vol. 592, p. 125829, 2021.https://doi.org/10.1016/j.jhydrol.2020.125829

[11] P. A. Mendoza, J. McPhee, and X. Vargas, "Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment," *Water Resources Research,* vol. 48, no. 9, 2012. https://doi.org/10.1029/2011WR011089

[12] E. F. Wood *et al.*, "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water," *Water Resources Research,* vol. 47, no. 5, 2011.

[13] C. Paniconi and M. Putti, "Physically based modeling in catchment hydrology at 50: Survey and outlook," *Water Resources Research,* vol. 51, no. 9, pp. 7090-7129, 2015. https://doi.org/10.1002/2015WR017780

[14] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting," *J Math Stat,* vol. 8, no. 3, pp. 330-338, 2012. https://doi.org/10.3844/jmssp.2012.330.338

[15] Y.-M. Chiang, L.-C. Chang, M.-J. Tsai, Y.-F. Wang, and F.-J. Chang, "Dynamic neural networks for real-time water level predictions of sewerage systems-covering gauged and ungauged sites," *Hydrology and Earth System Sciences,* vol. 14, no. 7, pp. 1309-1319, 2010.

[16] A. Mosavi, P. Ozturk, and K.-W. Chau, "Flood prediction using machine learning models: Literature review," *Water,* vol. 10, no. 11, p. 1536, 2018. https://doi.org/10.3390/w10111536

[17] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *Journal of Hydrology,* vol. 569, pp. 387-408, 2019.

[18] S. Das and A. Gupta, "A novel flood probability index based on radar rainfall and soil moisture estimates for a small vegetated watershed in Southeast Brazil," *Atmosphere,* vol. 16, no. 6, p. 633, 2025.

[19] D. Darkenbayev, A. Altybay, Z. Darkenbayeva, and N. Mekebayev, "Intelligent data analysis on an analytical platform," *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska,* vol. 14, no. 1, pp. 119-122, 2024. https://doi.org/10.35784/iapgos.5423

[20] D. T. Bui *et al.*, "A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area," *Science of The Total Environment,* vol. 701, p. 134413, 2020.

[21] Z. Guo, J. P. Leitao, N. E. Simões, and V. Moosavi, "Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks," *Journal of Flood Risk Management,* vol. 14, no. 1, p. e12684, 2021. https://doi.org/10.1111/jfr3.12684

[22] Y. Abou Rjeily, O. Abbas, M. Sadek, I. Shahrour, and F. Hage Chehade, "Flood forecasting within urban drainage systems using NARX neural network," *Water Science and Technology,* vol. 76, no. 9, pp. 2401-2412, 2017. https://doi.org/10.2166/wst.2017.409

[23] Y. Tang *et al.*, "flood forecasting based on machine learning pattern recognition and dynamic migration of parameters," *Journal of Hydrology: Regional Studies,* vol. 47, p. 101406, 2023/06/01/ 2023. https://doi.org/10.1016/j.ejrh.2023.101406

[24] F.-J. Chang, K. Hsu, and L.-C. Chang, *Flood forecasting using machine learning methods*. MDPI. https://doi.org/10.3390/books978-3-03897-549-6, 2019.

[25] H. I. Kim, Y. S. Lee, and B. Kim, "Real-time flood prediction applying random forest regression model in urban areas," *Journal of Korea Water Resources Association,* vol. 54, no. 12, pp. 1119-1130, 2021.

[26] K. Khosravi *et al.*, "A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran," *Environmental Monitoring and Assessment,* vol. 190, no. 2, pp. 1–18., 2018.

[27] M. M. A. Syeed, M. Farzana, I. Namir, I. Ishrar, M. H. Nushra, and T. Rahman, "Flood prediction using machine learning models," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022: IEEE, pp. 1-6.

[28] H. E. Aydin and M. C. Iban, "Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with SHapley Additive exPlanations," *Natural Hazards,* vol. 116, no. 3, pp. 2957-2991, 2023.

[29] A. Altybay, A. Nakiskhozhayeva, and D. Darkenbayev, "Numerical simulation and parallel computing of the acoustic wave equation," in *AIP Conference Proceedings (Vol. 3085, No. 1). AIP Publishing*, 2024.