Implmenting Machine Learning Techniques To Forecast Floods In Bangladesh Based On Historical Data

Sadiq Uddin Bhuiyan S.M. Toufique Arman Zaman Ahmed Lateef Jubaer Bin Islam

Supervisor: Dewan Ziaul Karim

Abstract

Flooding is a complex phenomenon that, due to its nonlinear and dynamic character, is difficult to anticipate. As a result, the prediction of floods has emerged as a critical area of study in the field of hydrology. Numerous researchers have handled this topic in various ways, spanning from physical models to image processing, however the time steps and precision are insufficient for all applications. This report looks at machine learning approaches for forecasting weather conditions and criteria and assessing the associated margins of uncertainty. The evaluated outputs enable more accurate and precise flood prediction for a variety of applications, including transportation systems.

Problem Statement

All over the world, floods have devastating impacts on life and property. The accurate prediction of floods, the communities that may be affected and the cost in damages they may cause can seriously help emergency services be allocated more efficiently. **What models are most useful for these predictions, and how accurate are they?**

Research Objectives

- Collecting data on weather patterns in Bangladesh as a whole, as well as district specific weather data
- Using machine learning models to predict the likelihood of floods given certain weather conditions
- Comparing different machine learning models in terms of their accuracy

Workflow of Methodology

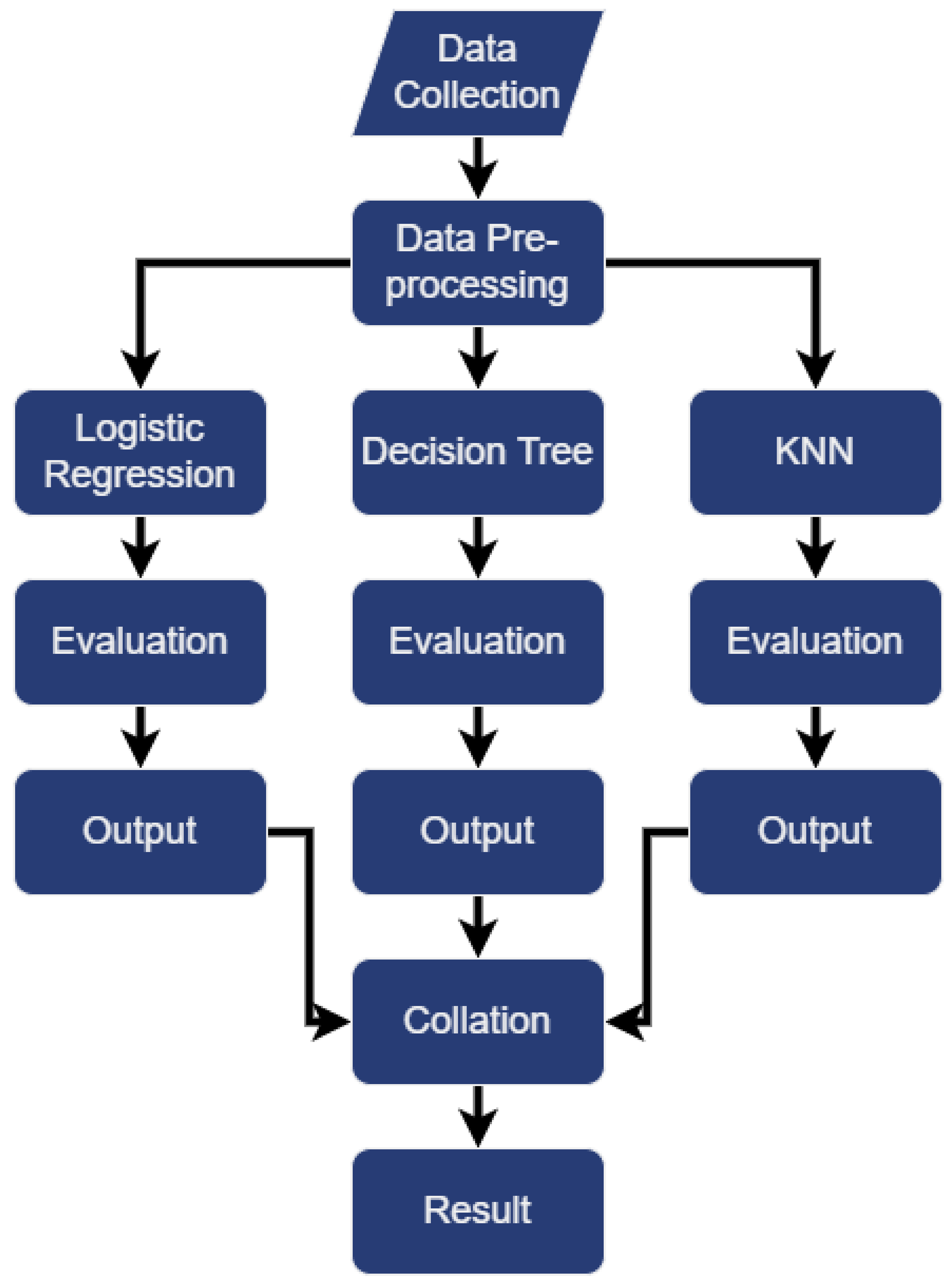


Figure 1. Illustration of Workflow

Dataset Preprocessing

The dataset consisted of weather data from multiple stations since 1949 and totalled 20,000 data points. The labels included relevant weather data such as rainfall, temperature data and cloud cover During preprocessing, irrelevant labels like station name were removed, units were standardized and duplicate entries were removed. Following that, missing values were filled using imputation. Finally, the dataset is separated into test and train sets, with 75% being the train set and 25% being the test set.

Experimental Results

Models	Accuracy	Precision	F1-Score	Error
Decision Tree	96.57%	97.94%	97.84%	3.4%
KNN	96.75%	98.09%	97.95%	3.3%
Logistic Regression	94.61%	97.60%	96.64%	5.4%

Table 1. Performance Comparison in Terms of Accuracy, Precision and F1-Score

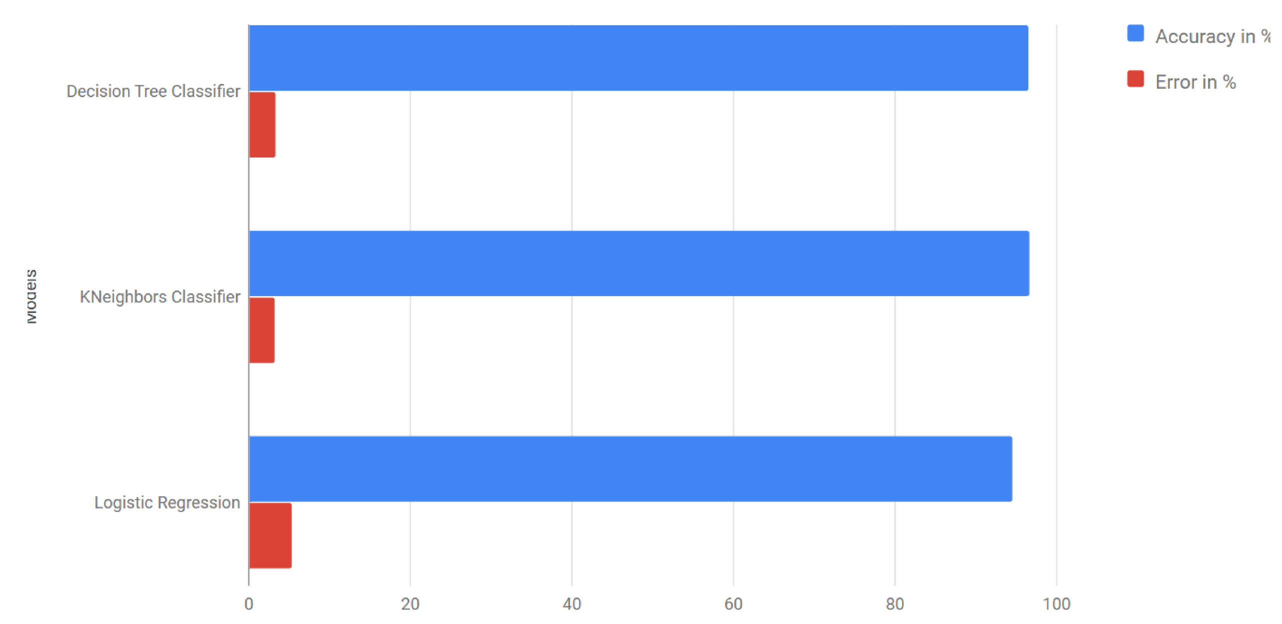
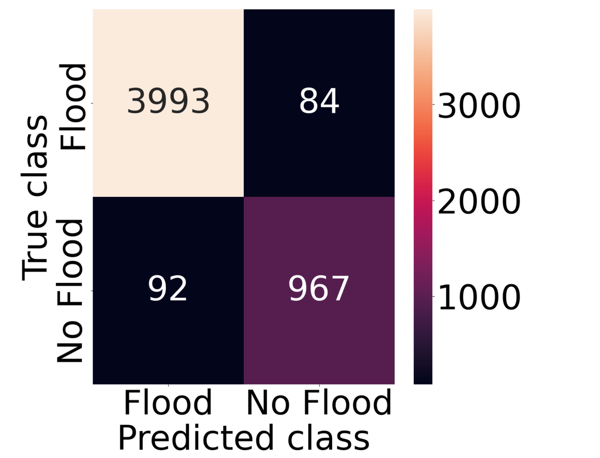
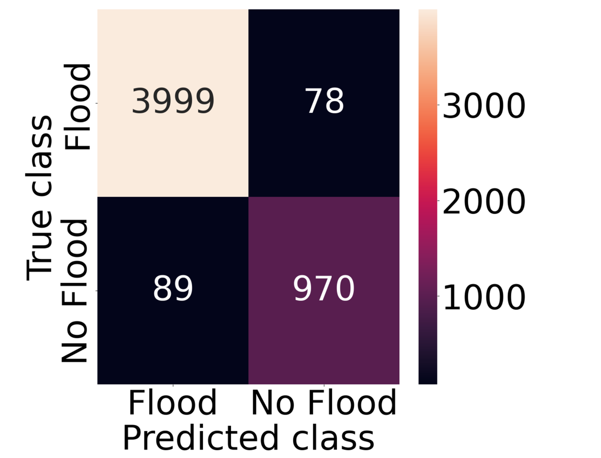


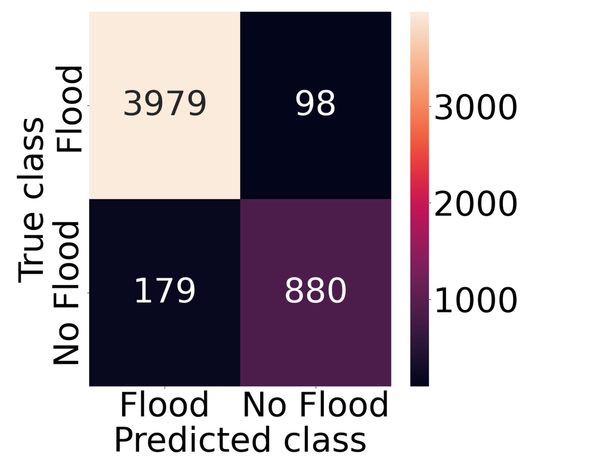
Figure 2. Bar Graph on the Accuracy of the Models



Confusion matrix of Decision Tree



Confusion matrix of KNN



Confusion matrix of Logistics Regression

Experimental Analysis

The Decision Tree model yielded 96.57% accuracy, 97.94% precision, 97.84% F1-Score, and 3.4% error. All results were better than Logistic Regression but worse than KNNs. KNN gave us 96.75 percent testing accuracy. The assessment precision and F1-Score are also the highest accurate, with 98.09% and 97.95%. Its 3.3% inaccuracy makes it more trustworthy. The Logistics Regression model gave us 94.61% minimum testing precision. The highest error rate of 5.4% makes it unpredictable compared to the Decision Tree and KNN models. Precision and F1-Score were 97.60% and 96.64%, respectively, which is somewhat less accurate than the models mentioned previously.

Related Work

- N Gauhar, S Das K Moury (2021), Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm
- Y Zhou, Z Cui, K Lin, S Sheng, H Chen, S Guo C Xu (2022), Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques

Conclusion

The study has effectively demonstrated the significance of data-driven approaches in mitigating the devastation caused by floods in the region. Through the use of machine learning algorithms and statistical models, the thesis demonstrates that flood events can be predicted with a reasonable degree of accuracy. By incorporating historical data into these models, researchers can identify patterns and trends that contribute to the occurrence of floods and create early warning systems to alert vulnerable communities. This proactive approach can substantially improve disaster preparedness and response efforts, ultimately saving lives and minimizing property damage. However, there are numerous areas that require additional research and future work.

References

[1] Noushin Gauhar, Sunanda Das, and Khadiza Sarwar Moury. Prediction of flood in bangladesh using k-nearest neighbors algorithm. In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pages 357–361, 2021.

[2] Yanlai Zhou, Zhen Cui, Kangling Lin, Sheng Sheng, Hua Chen, Shenglian Guo, and Chong-Yu Xu. Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. *Journal of Hydrology*, 604:127255, 2022.

Department of Computer Science and Engineering