A Web Based System to Predict Floods in Malawi

By: Fatima Kanthema
Harrison Chikhadza
Mayamiko Kadzanja
Richard Luke
Precious Phiri
Francisco Magwira

ABSTRACT

Natural calamity disrupts our daily life and brings many sufferings in our life. Among the natural calamities, flood is one of the most catastrophic. Predicting flood helps us to take necessary precautions and save human lives. Several types of data (meteorological condition, topography, river characteristics, and human activities) are used to predict flood water level in an area. In this paper, we propose a web-service based flood prediction system by incorporating machine learning models to predict the extent of floods in Malawi. This will facilitate the monitoring of the various flood-intensifying factors, contributing in increasing the flood water level in an area. Eventually, the decision makers would able to take measures to control those factors and to reduce the intensity of flooding in an area.

Keywords

Web Based Application, Machine learning Algorithms, Data and Metrics, proposed system architecture

1. INTRODUCTION

Even in this twenty first century after so many technological innovations humans are helpless in the hand of natural disasters. There are different natural disasters like, floods, volcanic eruptions, earthquakes, and tsunamis. Flood is considered as the most catastrophic among the other natural disasters. Flood causes the highest number of fatalities and greater economic damage in comparison to other natural disasters.

As for example, d Malawi almost every year faces erratic rainfall and which leads to heavy precipitation causes flash as well as river line floods. For example, precipitation in January 2015 was four times higher than average and caused flooding in 15 of 28 districts: This natural disaster affected over 1,300,000 Malawians. In 2022, the UN reports that 46 people died, 18 people reported missing and 206 were injured from floods caused by cyclone ANA. This natural disaster affected 995,072 Malawians. As a result, Malawi lost economic gains through death of life, livestock and loss of property, thereby pushing more people into absolute poverty. Precipitation is one of the indicators used to determine likelihood of flooding.

2. DATA AND METRICS

The data consists of a set of locations and various features associated with each location, such as soil type, elevation, proximity to bodies of water and past

flood history from 2010 to 2022. The target label for each location is a binary value indicating whether or not a flood occurred within the next year.

The model will be evaluated on its ability to accurately predict the probability of a flood occurring at each location. The metric will be average absolute error between the predicted probabilities and the true labels. The model with the lowest average absolute error will be chosen as the best model.

The goal is to minimize the Root Mean Squared Error:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

RMSE is the standard deviation of the prediction errors (Residuals), indicating the data points far from the regression line. In addition, the RMSE gives the spread out these residuals are.

3. Machine Learning Algorithms

In machine learning algorithm, two variables like x and y to are used. Variable x in function (f) is mapped to an output variable (Y): Y = f(X). The various algorithms used for prediction are discussed below.

3.1 Random Forest Classifier

Random Forest classifiers fall under the broad umbrella of ensemble-based learning methods. They are simple to implement, fast in operation and have proven to be highly successful in various domains. The key principle underlying the random forest

approach comprises the construction of many "simple" decision trees in the training stage and the majority vote (mode) across them in the classification stage. Much like in the case of naïve Bayes– and k-nearest neighbor–based algorithms, random forests are famous in part due to their simplicity on the one hand and generally good performance on the other.

3.2 Ensembling

This ensemble learning technique aggregates the results of multiple de-correlated decision trees collected in a "forest" to output its classification result. In concept, it is very similar to a Random Forest Classifier and only differs from it in constructing the decision trees in the forest.

Creating a correlation matrix representing each model prediction and then the best combination will be using those who have low correlation.

Basically, different predictions means that each model is capturing some information that the other is not and thus, sharing the point of views will boost the final result.

The following python concepts are used to predict the rainfall.

- Sklearn (al. M. e., 2021)

The Sklearn is a library for python which feature algorithm like SVM, Random Forest etc for machine learning analysis. It is used to build models.

- NumPy

In python we used NumPy library for scientific computing. It is a core library which provides tools a and high performance for a given array objects.

- Pandas

Pandas library is an open source library that is used to make analysis of data and to use easily. It provides high performance and easy to use data structure.

- Matplotlib

Matplotlib is a python library used to create plots and graphs. It provides variety of bar charts, histogram and error charts.

- Seaborn

Seaborn library is based on data visualization like matplotlib. It's used to represent statistical plotting.

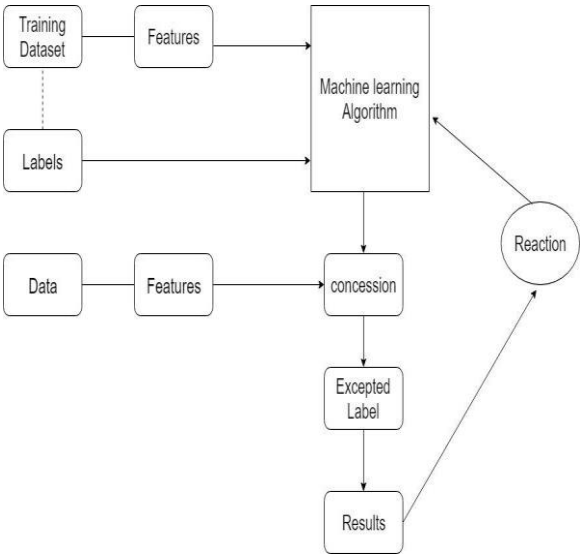

Fig.1: Proposed flood prediction model architecture

4. Proposed system architecture

Labeled data set are used for training. The features are extracted from the training data set and the set of features are given as input to the machine learning algorithm.

Figure 1 shows the proposed system architecture. Labeled data set are used for training. The features are extracted from the training data set and the set of features are given as input to the machine learning algorithm.

4.1 Report

A report is a specific form of describing and identifying and examining issues arise in case of an event. Such an event may be described as high rainfall and change in weather pattern which may increase the chance of flood occurrence.

4.2 Analysis

A report in ML defines the accuracy of the data. The data are examined and evaluated by breaking its variables to uncover their dependencies. These help to understand better about the data.

4.3 Monitor

The data are constantly measured, and the performance is monitored to provide the accuracy of the output from the given dataset.

4.4 Prediction

The obtained historical dataset is trained, and an algorithm is applied to obtain an output. The forecast method is used for the prediction analysis.

4.5 Simulate

The simulation in machine learning helps us to forecast the changes that have never happened before and to obtain scenarios outside the historical bounds.

Fig.2: Prediction Methodology

5. Conclusion and Results

In this paper, the simple machine learning algorithm to predict the accuracy of the flood occurrence is implemented. The desired algorithm shows the results of occurrence of flood in the upcoming year. When compared with the other algorithms, the decision tree algorithm gives more accurate results and provide high performance accuracy and easy to understand. The decision tree also generate model for nonlinear dataset. This nonlinear can be applied to find the accuracy of linear or logistic dataset. As the compared results shows that the decision tree gives more accuracy compared to other simple machine learning algorithm. As the gathered dataset can provide huge volume of variables it can't be implemented in a simple machine learning algorithm. For a huge amount of data set it can be implemented in neural network which will provide more accuracy and output of the provided dataset. As the neural network uses fuzzy state machine act it can produce multiple results with different probabilities. It can provide historical dataset with more mutable and adaptable form.

REFERRENCE

[1] M. Panahi et al., "Flood spatial prediction modeling using a
hybrid of meta-optimisation and support vector regression
modeling," Catena, vol. 199, no. January 2020, p. 105114, 2021.

[2] R. Tabbussum and A. Q. Dar, "Performance evaluation of Artificial intelligence paradigms—artificial neural networks,
fuzzy logic, and adaptive neuro-fuzzy inference system for flood
prediction," Environ. Sci. Pollut. Res., vol. 28, no. 20, pp. 25265–
25282.

[3] C. N. Egbinola, H. D. Olaniran, and A. C. Amanambu, "Flood
management in cities of developing countries: the example of
Ibadan, Nigeria," J. Flood Risk Manag., vol. 10, no. 4, pp. 546–
554, 2017.

[4] S. A. A. El-Magd, B. Pradhan, and A. Alamri, "Machine learning algorithm for flash flood prediction mapping in Wadi El-Laqeita
and surroundings, Central Eastern Desert, Egypt," Arab. J.
Geosci., vol. 14, no. 4, 2021.

[5] T. W. Gallien, B. F. Sanders, and R. E. Flick, "Urban coastal flood prediction: Integrating wave overtopping, flood defenses
and drainage," Coast. Eng., vol. 91, pp. 18–28, 2014.