**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 our previous works, we proposed a belief rule-based flood prediction system in a desktop environment. 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, Predictive Analytics, Flood Forecasting

# INTRODUCTION

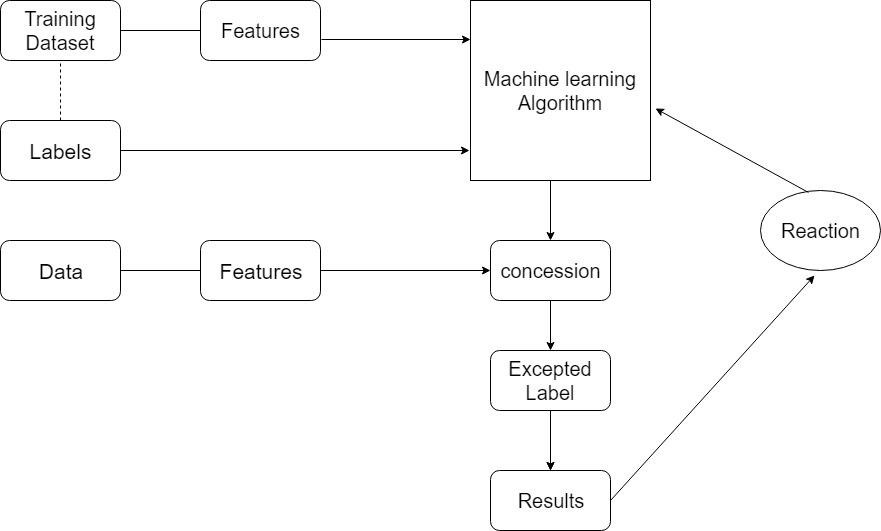
Even in this twenty first century after so many technological innovations humans are helpless in the hand of natural disasters. There are different natural disasters like, floods, volcanic eruptions, earthquakes, and tsunamis. Flood is considered as the most catastrophic among the other natural disasters [1]. Flood causes the highest number of fatalities and greater economic damage in comparison to other natural disasters.

As for example, a devastating flood occurred here in Malawi in

2015 and again in 2019 caused by overflowing of water.

# DATA AND METRICS

|  |
| --- |
| 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. Logistic Regression When the nature of the dependent variable is binary logistic regression is used. It is used to solve binary classification problems. This algorithm is best compared to linear regression algorithm because predicting the value of binary variable linear regression was not suitable. It will predict values only in particular range like outside 0 and 1.  In logistic regression linear relationship isn’t required. There is no multi collinearity in this algorithm, Linear Regression Linear regression algorithm is the most frequently used model for predictive analysis. Therefore, in regression simple regression to analysis the relationship between the dependent(y) variables and the independent variables to predict the accuracy outcomes and represent it in a statistical graphical representation. Decision Tree The final predicting model is decision tree. Generally, the decision tree is a predicting tool which split the data continuously according to the given certain data parameters. It is a type of supervised learning where a non-parametric method is approached for regression and classification problems. The model first targets the variables to predict value of the variables from the given data by analyzing the decision rules. By this way the accuracy and the output are determined by this decision model  The following python concepts are used to predict the rainfall.  • Sklearn  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. |

Fig.1: Proposed flood prediction model architecture

• 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.

# Proposed system architechture

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.

## 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.

## 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.

## Monitor

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

## 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.

## 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

# 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

(E. Toth\*; BAXTER E. VIEUX)