Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, July : 2023

PREDICTION FORECASTING USING MACHINE LEARNING ALGORITHM WITH PAST AND REAL RECORDS OF FLOOD

VIJAYA RAMINENI, Associate Professor, Dept. of AI & IT, DVR & Dr.HS MIC college of Technology, A.P., India.

DURGA REDDY ESTHALAMURI, MCA Final year, Department of Computer Applications, DVR & Dr. HS MIC College of Technology, A.P., India.

Abstract:- Impact on daily life, flooding is one of the most pressing issues that Malaysia has been dealing with recently. Floods are a type of natural geohazard that typically occur because of consistently heavy rain. This natural phenomenon causes massive damage to the country's property and Gross Domestic Product (GDP). Floods are very harmful for nature, which are very complex to model. The flood prediction model will give risk reduction & it minimizes the future loss of human life. On 18 May 2016 a south Indian state Kerala was affected by flood. Machine learning is a method which provides intelligence to predict the result in future. The performance comparison of ML models is based on the speed, time and accuracy of the result. There exist a lot of machine algorithms which generate models with more accuracy. For flood prediction classification algorithms like decision tree and linear regression are used in this research.

INTRODUCTION

Floods are one of the worst affecting natural phenomena which causes heavy damage to property, infrastructure and most importantly human life. To prevent such disasters Machine learning model is created to predict the floods that can occur in the future. It's hard to create a predictive model because of its complexity. In this system the rainfall data is fed into four different Machine Learning models prior to this process, the data is cleaned and preprocessed, the dataset for training is split into Training set and Test set in the ratio of 7:3. Then the accuracy of each model is compared and the confusion matrix parameters are taken to evaluate and analyze. At the end the best model is chosen by comparing the accuracy. Machine learning provides capabilities to learn from past data. Also based on past data it generates models for future prediction. This technique will be very useful for flood

prediction. Earlier we need to give instruction to system for generating output and result. But now with the help of machine learning technique it generates models and gives result itself. Most of the work related to machine learning for flood either predict and helps to make precaution measure and suggest future flood. Kerala, one of the Southern state of India, experienced once in a century flood. There was high damage to the life and property. This motivated us to do study on this pattern of rainfall in the state of Kerala.

Even though there are numerous data-driven models for flood prediction, most studies focus on prediction using a single flood variable. The creation of various data-driven models may require unfeasible computing resources when estimating multiple flood variables. Furthermore, the trends of several flood variables can only be revealed by analysing long-term historical observations, which conventional data-driven models do not adequately support.

LITERATURE SURVEY

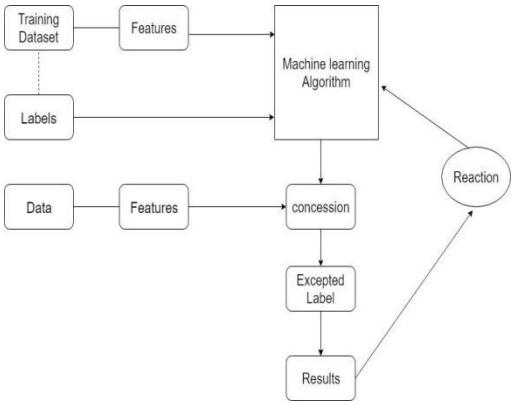
According to Ashizawa et al. [1], the entire GDP of Japan impacted by flood damage is at least 1% of the overall GDP of the nation. Tiggeloven et al. [2] stated that the top 15 countries, such as India, Bangladesh, China, and others, are vulnerable to flood occurrence at the present day and could be worst if no action is taken. Indeed, floods can cause a massive amount of money to repair the damage. Hence, flood occurrence can affect every country, including Malaysia. Shaari et al., [3] stated that from 2006 to 2010, there was nearly 1 million USD damage caused by floods which affected the nation's economic growth. There are many classifications of floods namely coastal floods, flash floods, ponding (or pluvial flooding), and river (or fluvial) floods [4]. Floods often occur, especially in Southeast Asia, including our country, Malaysia. The general types of flooding in Malaysia include riverbank overflow, flash floods, high tides [5], and monsoon floods [6]. Romali et al, [7] stated that Malaysian financial losses are estimated at nearly MYR 915 million annually on an average due to floods. According to Zerara [8], time series is a statistical method that can be applied in a broad range of longitudinal research designs. Typically, this time series design involves a single subject that is measured repeatedly at regular intervals over a large number of observations. Time series forecasting aims to predict an outcome based on the collection of historical data that can be used to build a quantitative model that explains the variables under consideration [9].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, July : 2023

For many years, time series forecasting has been an important research domain in meteorology [10], biology [11], and econometrics [12]. Generally, time series can have four characteristics: trends, seasonality, cycles, and noise [13]. Time series forecasting algorithms perform well with data that includes a time dimension and one or more properties [14].

PROPOSED SYSTEM

The objective of Flood Prediction using Machine Learning is to design a model to predict the flood using the rainfall data. The prediction of different models is taken and compared within each other to find the best model that has high accuracy. The flood can be predicted in different states of India in different months. The confusion matrix of different models in Machine learning is considered to evaluate the accuracy and precision of the system. 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. Figure 2 shows the steps involved in the prediction model.



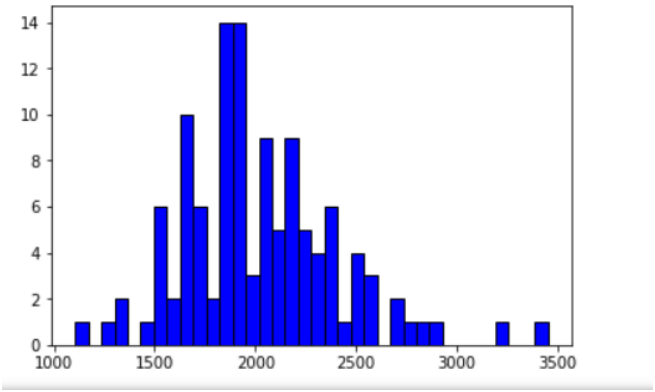
RELATED WORK

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

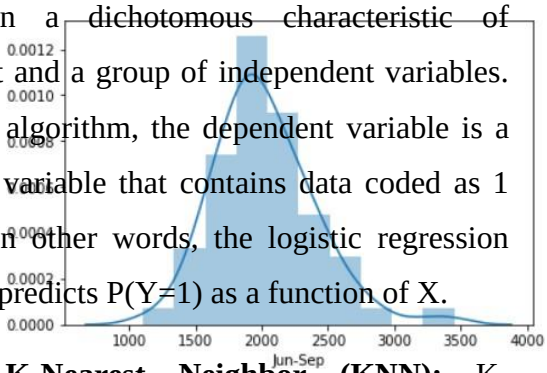
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 7, July : 2023

implemented in neural network which will



ALGORITHMS

Logistic Regression: Logistic Regression may be a machine learning algorithm that predicts the probability of a categorical variable. It is a statistical way of analyzing a group of knowledge that comprises quite one experimental variable that determines the result. The outcome is then measured with a dichotomous variable. The goal of this algorithm is to seek out the simplest model to explain the connection between a dichotomous characteristic of interest and a group of independent variables. In this algorithm, the dependent variable is a binary variable that contains data coded as 1 or 0. In other words, the logistic regression model predicts P(Y=1) as a function of X.



K-Nearest Neighbor (KNN): K-Nearest Neighbor is one among the supervised

```
x=pd.read_csv("kerala.csv")
y=pd.read_csv("kerala.csv")

y1=list(x["YEAR"])
x1=list(x["Jun-Sep"])
z1=list(x["JUN"])
w1=list(x["MAY"])
x2=list(x["Mar-May"])
x3=list(x["Jan-Feb"])
plt.plot(y1, x1,'*',color="Red")
plt.plot(y1, z1,'*',color="Green")
plt.plot(y1, w1,'*',color="Blue")
plt.show()
```

machine learning algorithms that stores all instances like training data points in an n-dimensional space. For real-valued data, the algorithm returns the mean of k nearest neighbors, and in case of receiving unknown discrete data, it analyses the closest k number of instances that is saved and returns the most common class as the result of the prediction. In the distance-weighted nearest neighbor algorithm, the contribution of each of the k neighbors is weighed according to their distance, giving higher weight to the closest neighbors. The K-Nearest Neighbor algorithm is a classification algorithm and is robust to noisy data as it averages the k-nearest neighbors. The algorithm first takes a bunch of labeled points and analyses them to find out the way to label the opposite points. Hence, to label a new point, it looks at the closest labeled points to that new point and has those neighbors vote, so whichever label most of the neighbors have is the label for the new point.

Support Vector Machines: SVM uses a classifier that categorizes the info set by setting an optimal hyperplane between data. This classifier is chosen as it is incredibly versatile in the number of different kernel functions that can be applied, and this model can yield a high predictability rate. Support Vector Machine is one among the foremost popular and widely used clustering algorithms. It belongs to a gaggle of generalized linear classifiers and is taken into account as an extension of the perceptron. It was developed in the 1990s and continues to be the desired method for a high-performance algorithm with a little tuning.

CONCLUSION

In this paper, we proposed the machine learning system ignited by data cleaning and processing, replacing or removing the null values, model building and evaluation. At the end the flood prediction model has given different accuracy results from four different models. From the above results and analysis, the best algorithm for flood prediction is Logistic Regression with (99%).

Future work: Currently, these models produce a good accuracy for one day ahead but for future work these models need to be tuned to produce a good accuracy for multi-days ahead.

BIBILIOGRAPHY

[1]. BAXTER E. VIEUX, " Estimation of Rainfall forFlood Prediction from WSR-88D Reflectivity:A Case Study".

[2]. Hsu, K., Gupta, H.V., Sorooshian, S., 1995,"Artificial neural network modeling of

the rainfall–runoff process. Water Resource" Res. 31
(10), 2517–2530.
[3]. Zealand, C.M., Burn, D.H., Simonovic, S.P., J.Hydrol 1999, " Short term streamflow forecasting
using artificial neural networks.. 214, 32–48.
[4] K. A. Oladapo, S. A. Idowu, Y. . Adekunle, and F. . Ayankoya, "Categorization of Conditioning Variables for Pluvial Flood Risk Assessment," Int. J. Sci. Eng. Res., vol. 11, no. 8, pp. 355–368, 2020. [5] S. F. Zakaria, R. M. Zin, I. Mohamad, S. Balubaid, S. H. Mydin, and E. M. R. Mdr, "The development of flood map in Malaysia," AIP Conf. Proc., vol. 1903, no. November, 2017, doi: 10.1063/1.5011632. [6] F. Nurashikin Sungip et al., "The Impact of Monsoon Flood Phenomenon on Tourism Sector in Kelantan, Malaysia: A Review," Int. J. Eng. Technol., vol. 7, no. 4.34, p. 37, 2018, doi: 10.14419/ijet.v7i4.34.23577. [7] N. S. Romali, Z. Yusop, M. Sulaiman, and Z. Ismail, "Flood risk assessment: A review of flood damage estimation model for Malaysia," J. Teknol., vol. 80, no. 3, pp. 145–153, 2018, doi: 10.11113/jt.v80.11189. [8] M. Zerara, "Machine Learning and Statistical Methods for Time Series Forecasting : a Case Machine Learning and Statistical Methods for Time Series Forecasting : a Case Study with Water Demand," no. March, 2021, doi: 10.36227/techrxiv.15019698.v1. [9] A. Pant and R. S. Rajput, "Time Series Analysis of Gold Price Using R," no. January, 2018.

AUTHOR PROFILES:

Mrs Vijaya Ramineni completed her Master of Computer Applications (MCA), M.Tech in (CSE) from Acharya Nagarjuna University. She has published more than10 papers in indexing Journals, currently working as an Associate Professor in thedepartment of AI & IT at DVR & Dr HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District. Her areas of interest are Data Mining, Cloud Computing and Machine Learning.

Mr. Durga Reddy Esthalamuri, as MCA student in the Department of Computer Applications at DVR & Dr HS MIC College of Technology (Autonomous), Kanchikacherla, NTR District. He has completed BSC in Parvathaneni Brahmayya Siddhartha College of Arts & Science from Krishna University. His areas of interests are Networks, Machine Learning and Cloud Computing.