Flood Prediction using Logistic Regression for Kerala State

Saiesh Naik
Department of Computer
Engineering Vidyavardhini's College of
Engineering And Technology
(Mumbai University) Vasai, India

Anang Verma
Department of Computer Engineering
Vidyavardhini's College of
Engineering And Technology
(Mumbai University)   Vasai, India

Srushti Ashok Patil
Department of Computer Engineering
Vidyavardhini's College of
Engineering And Technology
(Mumbai University) Vasai, India

Prof. Anil Hingmire
Department of Computer Engineering
Vidyavardhini's College of Engineering and Technology
Vasai, India

Abstract—**Flood is a major disaster in India, which causes a huge damage to the living world. The model built here is an approach towards the determination of flood which can be used further for various implementations. It uses a Machine learning approach-Supervised Learning known as Logistic Regression. Logistic Regression is a type of Binary Classification which gives the output in the form of 0s and 1s, which makes it easier to pre-determine whether the event (flood) is about to occur or not. The main aim is to study previous data of States and give a good-fitting approach to be used in future. In Future Machine Learning will replace all the human efforts and abilities in order to make things easier. This is one of those applications of ML that will prove beneficial to the Living Life in-order   to   sustain the upcoming event (Flood).**

Keywords;- Machine Learning, Logistic Regression, Linear Regression, Rainfall data, Flood.

I. INTRODUCTION

It is seen that now-a-days Machine Learning is playing a huge role in every field, it also includes research about various events depending on the previous event related data. It is the ability to learn based on experiences which is only possible when we have some original, precise, complete data. Thus, In-order to apply ML we need to assemble all the required data. The data then needs to be pre-processed so that it could be carried forward for further operations or functioning. It is necessary to maintain knowledge about the residing or available data to use it in the best way possible. We can use or try different Algorithms to get the accuracy based on the existing data. Flood prediction is an important consideration, due to changing climatic conditions. We have used Logistic Regression to come-up with the best outcome. We have considered KERALA STATE for the maximum use of the built- system. Floods are the most damaging natural disaster in this world. On the occasion of heavy flood, it can destroy a whole community. It is crucial to develop a flood prediction system as a mechanism to predict and reduce the flood risk. It proves necessary for alerting resident to take early action such as evacuate quickly to a safer and higher   place.   Aim

is to specify the contribution of ML in different models. The dataset for the amount of rainfall in various states in India is provided on data.gov.in. We have provided dataset consisting of rainfall details of Kerala of previous 115 years, it clearly defines the annual as well as the monthly rainfall data which proves this system more accurate, and it confirms its reliability, efficiency and confident dependence as well.This model gives us a well defined idea of how the Logistic Regression Algorithm works well with a precise data. This algorithm solves half of the case because of its binary classified nature. The goal of this particular system is to contribute to development of ML as well and to improve the conditions of Living Life in case of the calamity (Flood).

II. OBJECTIVES

A EXISTINGALGORITHM

Flood Early Warning System

As the name suggests, Flood Early Warning System (FLEWS) is as system by which flood induced hazards can be minimized and prevented. There are different organizations which are working on flood forecasting and early warning at national, continental as well as global scale.

In a flood prediction system the most significant info is constant hydro-meteorological perceptions which are given by climate radar satellites and auto hydro meteorological station systems (Billa et al, 2006; Budhakooncharoen, 2004). This data's which are real time data can be used in various ways to evaluate flood risks and issues of flood warning. Apart from real life data, probabilistic weather forecasts (Numerical Weather Prediction-NWP) are also playing an important role in providing input for hydrological models to generate warnings scenarios(Burger et al 2009;Thielen et al 2010).Other than having conjectures of the most significant information (precipitation) a model should be chosen that describes and mimics catchment responses for flood early warning.
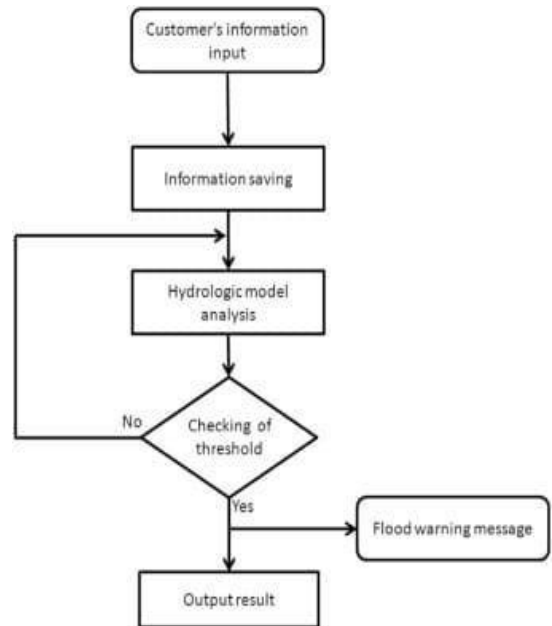
B IMPLEMENTEDMETHODOLOGY

The DATA concerned with factors that affect flood

will be provided. The provided data will be analyzed and used for the training. The data will be analyzed using basic approaches of Machine Learning-Linear/Logistic Regression. After the training the model will be tested. Based on the given input the model will predict if flood will occur or not. Real time data of rainfall from the month of March to May (for present year in Kerala state)Real time data of average rainfall in the first

10 days of June(for present year in Kerala state)Real time data of average increase in rainfall from the months of May of June(for present year in Kerala state). Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two potential qualities, for example, pass/fail which is represented by an indicator variable, where the two values are labeled "0" and "1". In the logistic model, the log-odds (the logarithm of the odds) for the value labeled "1" is a linear combination of one or more independent variables ("predictors"); the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The corresponding probability of the value labeled "1" can vary between 0 (certainly the value "0") and 1 (certainly the value "1"), hence the labeling; the function that converts log- odds to probability is the logistic function, hence the name. The unit of measurement for the log- odds scale is called a logit, from logistic unit, hence the alternative names. Analogous models with a different sigmoid function instead of the logistic function can also be used, such as the probit model. The defining characteristic of the logistic model is that increasing one of the independent variables multiplicatively scales the chances of the given result at a consistent rate, with every free factor having its own parameter; for a binary dependent variable this generalizes the odds ratio.
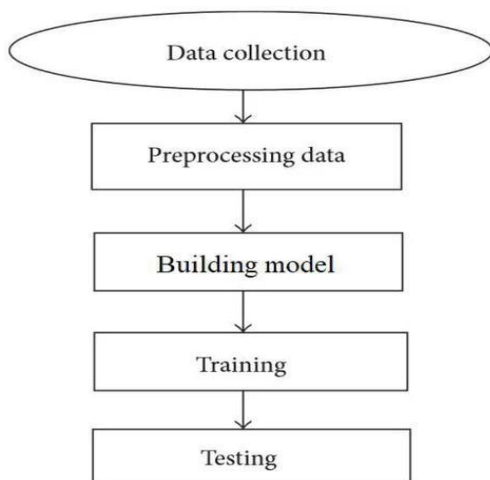
The binary logistic regression model has augmentations to multiple degrees of the needy variable: clear cut

yields with multiple qualities are demonstrated by multinomial strategic relapse, and if the different classes are requested, by ordinal strategic relapse, for instance the corresponding chances ordinal calculated model. The model itself essentially models likelihood of yield as far as information, and doesn't perform measurable order (it's anything but a classifier), however it very well may be utilized to make a classifier, for example by picking a cutoff esteem and grouping contributions with likelihood more noteworthy than the cutoff as one class, underneath the cutoff as the other; this is a typical method to make a binary classifier.

Flood column shows the classification of occurrence of the flood if the amount is greater than threshold value.

III. RESULT

Since, 75% of predicted flood was occurred, this model ends up being very valuable for additional explores and execution of avoidance or early flood cautioning framework. Among the natural disasters, floods are the most dangerous, making gigantic harm human life, framework, farming and the financial framework. Governments, accordingly, are feeling the squeeze to create dependable and precise maps of flood hazard territories and further arrangement for practical flood

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NTASU - 2020 Conference Proceedings

chance administration concentrating on avoidance, insurance and readiness. Flood forecast models are critical for risk appraisal and extraordinary occasions the board. Hearty and exact forecast exceptionally add to water plan of action the executives techniques, approach recommendations and investigation, and further departure demonstrating protest enrollment, suspect expectation and further wrongdoing examination.

Floods are the most harming cataclysmic event right now. On the event of overwhelming flood, it can crush an entire network. It is pivotal to build up a flood expectation framework as an instrument to anticipate and lessen the flood chance. It demonstrates fundamental for making occupant aware of make early actions, for example, evacuate quickly to a safer and higher place.

FLOOD YEAR	PREDICTED MONTH	OBSERVED MONTH
2013	JUL-SEP	AUG
2007	JUL-SEP	JUL-AUG
1994	JUL-SEP	AUG
1991	JUL-SEP	-
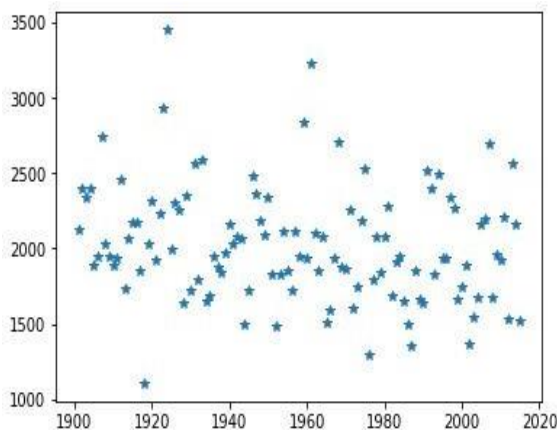
Large sample sizes are required for logistic regression to give adequate numbers in the two classifications of the reaction variable. The more illustrative factors, the bigger the example size required. Logistic relapse gives a helpful way to displaying the reliance of a parallel reaction variable on at least one informative factors, where the last can be either clear cut or ceaseless. The attack of the subsequent model can be evaluated utilizing various techniques.

IV. CONCLUSION

Presently the flood prediction systems are far from satisfactory. There is a dare need to improve the present systems in terms of data collection and representation. Hence there is a need of this prediction model as available meteorological models are not accurate in rainfalls predictions. This prediction system provides 86.08% of the accuracy. Therefore, this model proves to be quite useful for further research and implementation of prevention or early flood warning systems.

REFERENCES

[1] Danso-Amoako, E.; Scholz, M.; Kalimeris, N.; Yang, Q.; Shao, J. Predicting dam failure risk for sustainable flood retention basins: A generic case study for the wider greater manchester area. Computers, Environment and Urban Systems 2012, 36, 423-433.

[2] Bruen, M.; Yang, J. Functional networks in real-time flood forecasting—A novel application. Adv. Water Resour. 2005, 28, 899–909.

[3] Kim, B.; Sanders, B.F.; Famiglietti, J.S.; Guinot, V. Urban flood modeling with porous shallow-water equations: A case study of model errors in the presence of anisotropic porosity. J. Hydrol. 2015, 523, 680–692.

[4] Liang, X.; Lettenmaier, D.P.; Wood, E.F.; Burges, S.J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. Atmos. 1994, 99, 14415–14428.

[5] Costabile, P.; Macchione, F.; Natale, L.; Petaccia, G. Flood mapping using lidar dem. Limitations of the 1-D modeling highlighted by the 2-D approach. Nat. Hazards 2015, 77, 181–204.

[6] Haddad, K.; Rahman, A. Regional flood frequency analysis in eastern australia: Bayesian GLS regression-based methods within fixed region and ROI framework–quantile regression vs. Parameter regression technique. J. Hydrol. 2012, 430, 142–161.

[7] Mosavi, A.; Rabczuk, T.; Varkonyi-Koczy, A.R. Reviewing the novel machine learning tools for materials design. In Recent Advances in Technology Research and Education; Springer: Cham, Switzerland, 2017; pp. 50–58.

[8] Ortiz-García, E.; Salcedo-Sanz, S.; Casanova-Mateo, C. Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. Atmos. Res. 2014, 139,128–136.

[9] Dineva, A.; Várkonyi-Kóczy, A.R.; Tar, J.K. Fuzzy expert system for automatic wavelet shrinkage procedure selection for noise suppression. In Proceedings of the 2014 IEEE 18th International Conference on Intelligent Engineering Systems (INES), Tihany, Hungary, 3–5 July 2014; pp. 163–168.