Water Scarcity Prediction System using Machine Learning approach and Resource Estimation

Shreyansh
Dept of CSE,Sathyabama Institute of Science and Technology, Chennai, India
shreyanshdaeas44@gmail.com

Vikash Chaurasia
Dept of CSE,Sathyabama Institute of Science and Technology,Chennai,India.
vikas9598@gmail.com

Dr.T.Shashikala,
Dept of CSE,Sathyabama Institute of Science and Technology,Chennai,India.
dean.computing@sathyabama.ac.in

Abstract-

Water Scarcity is one of the most risky and unusual issues looked by the world. Water shortage can mean shortage in accessibility because of physical deficiency, or shortage in access because of the disappointment of foundations to guarantee a customary stockpile or because of an absence of satisfactory framework.

Water shortage as of now influences each and every nation of the world. Water has consistently been one of the rare assets where and as we probably am aware just nearly 3% of the water on the planet is drinkable. The developing populace has likewise gotten one of the key components influencing the speed of depletion of water assets. Numerous locales over the world, particularly the bone-dry districts have even begun giving indications of the weariness of a large portion of their water assets. The paper has coordinated the forecast framework with the anticipated provincial water information alongside the number of inhabitants around there to foresee the assessed measure of water required.

Keywords: Rainwater, Required water, Prediction, Scarcity.

I. INTRODUCTION

Water shortage is one of the difficult issues in the cutting edge world. It influences all mainland's and around 3 billion individuals around the world in any event a month out of consistently. In excess of 15 percent of the absolute populace around the globe needs access to clean drinking water. The water emergency is expanding radically and has gotten one of the significant issues over the globe. A report proposes that the United States of America alone squanders 495 billion litter of new water every week. As just however one-hundredth of surface water is crisp and reasonable for human utilization, it turns out to be vital that we spare water all together that our people in the future endure. The unchecked use of water and extraordinary climate have intensified things and in a matter of seconds, there will be a water deficiency, anomalies in market interest, groundwater shrinkage, among different difficulties.

From time to time the updates on shortage of water in a specific region over the world has been a typical news bringing about an enormous number of passing's because of absence of water inside the human body causing different infections, for example, drying out, and so forth.

Regardless of having the information on how much harm a dry spell can do to a nation or the world overall. The administration can do little to facilitate the agony of the individuals languishing. The losses of life in a dry season are disturbing and the instances of dry spells in pretty much every nation has gotten practically normal and are expanding each year. Taking the case of the 2016 dry season in India, one of the most noticeably awful dry seasons in history of the country influenced around 330 million individuals. Shortage doesn't just influence people straightforwardly yet in addition purposes numerous relative issues which can cause a chain of issues on the planet.

A dry spell in a specific zone extraordinarily influences the vegetation around there and for the most part pulverizes it. This causes a great deal of issues for the poor segment of the general public fundamentally contained labourers and furthermore influences the working effectiveness of the individuals. These everything all together lead

Water shortage is one of the significant issues in the cutting edge world. It as of now influences each landmass and around 2.8 billion individuals around the globe in any event one month out of consistently. More than 1.2 billion individuals need access to clean drinking water. The water emergency has gotten one of the significant worries over the globe. A report recommends that the only us squanders 7 billion gallons of drinking water every day. As just short of what one percent of earth's surface water is reasonable for human utilization, it becomes urgent that we spare water with the goal that our people in the future endure. The unchecked utilization of water and extraordinary climate conditions have compounded the circumstance and right away there will be a new water deficiency, abnormalities in organic market, groundwater shrinkage, among different difficulties.

From time to time the updates on shortage of water in a specific zone over the world has been a typical news bringing about an enormous number of passing because of absence of water to ascend in the costs of the day by day articles and along these lines the dry season in a solitary area of a nation influences the entire country. These evil impacts now and then spread to rather an enormous territory and result in a worldwide emergency.

The harm it dispenses on the individuals and the correct working of the general public can be minimalized with an expectation framework that could really tell the assessed measure of assets that will be expected to handle such a circumstance.

II. LITERATURE SURVEY

Amir AghaKouchak.[2] proposed hybrid statistical dynamical framework for meteorological drought prediction an Application to the south-western United States. The results The results of the statistical precipitation model are also generated for the same forecast periods. This step involves the conditional forecast of precipitation anomaly given the global teleconnection indices

Ranvish C Deo and Mehmat Shini.[3] proposed a prediction model of future drought is an effective mitigation tool for assessing adverse consequences of drought events on vital water resources, agriculture, ecosystems and hydrology. Data-driven models and predictions are using machine learning algorithms are promising tenets for these purposes as they require very less developmental time, minimal inputs and are relatively less complex than the dynamic or physical model.

Jiyoung rehe and Jungo Yun al.[4] proposed different reduction of drought impacts could also be achieved through sustainable drought management and proactive measures against drought disaster. Accurate and timely provision of drought information is important. During this study, drought forecasting models to supply high-resolution drought information supported drought indicators for ungauged areas were developed.

Haikumg park, Kiyun Ping and Dong Kung Lee. [5] Proposed a system named severe drought area prediction (SDAP) model, to estimate soil moisture index (SMI) maps

III. METHODOLOGY

The Proposed device works on identifying the sources by operating on a prediction system based on the Naive Bayesian Classifier to are expecting the drought location of a rustic with the aid of processing the rainfall information in that place.
The version proposed here has been considered as a classification trouble. [6] Classification is that the problem of identifying which of a group of categories (sub-populations) a replacement commentary belongs, on the thought of a schooling set of facts containing observations (or instances) whose class club is recognised . The classification works at the prediction of the drought regions by producing one of the two consequences i.e. “Yes” and “No” to the matter statement - if the place is going to own a drought or not.

Naive Bayes methods [7] are a set of supervised learning algorithms primarily based on making use of Bayes’ theorem with the “naive” assumption of conditional independence between each pair of functions given the value of the class variable. Bayes’ theorem states the subsequent relationship, given magnificence variable y and dependent feature vector x1 through xn,:

Using naive conditional independence assumption that:

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y),$$

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) P(x_1, \dots x_n \mid y)}{P(x_1, \dots, x_n)}$$

for all i, this relation is simplified to the following formulae:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \ldots, x_n)}$$

Since P(x1,…,xn) is consistent given the input, we can use classification rule which is cited below:

$$P(y \mid x_1, \ldots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

and we are able to use Maximum A Posteriori (MAP) estimation to estimate P(y) and P(xi|y); the previous is then the relative frequency of sophistication Y inside the training set. The special naive Bayes classifiers differ especially by way of the assumptions they make concerning the distribution of P(xi|y) .In spite of their apparently over-simplified assumptions, naive Bayes classifiers have worked pretty nicely in lots of real-global situations, famously report class and spam filtering. They require a little quantity of coaching information to estimate the specified parameters.

Naive Bayes rookies and classifiers are often extraordinarily fast as compared to more state-of-the-art methods. The decoupling of the category conditional characteristic distributions means every distribution are often independently envisioned as a one-dimensional distribution. This, in turn, facilitates to alleviate problems stemming from the curse of dimensionality.

On the turn side, even though naive Bayes is called a decent classifier, it's far recognized to be an awful estimator, so the chance outputs from predict_proba are not to be taken too seriously.

The set of rules produces the areas wherein the possibilities of drought are there and then fetches the populace of that area. [8]Spain's National Statistics Institute said that average family water consumption in Spain became 137 litres in keeping with character according to day in 2012. The populace is then used to offer an estimation using the formulae:

$$res = \sum (P) * est$$

, where res = resources needed,

P = Population of that area,

est = estimated amount of water used per person in a day

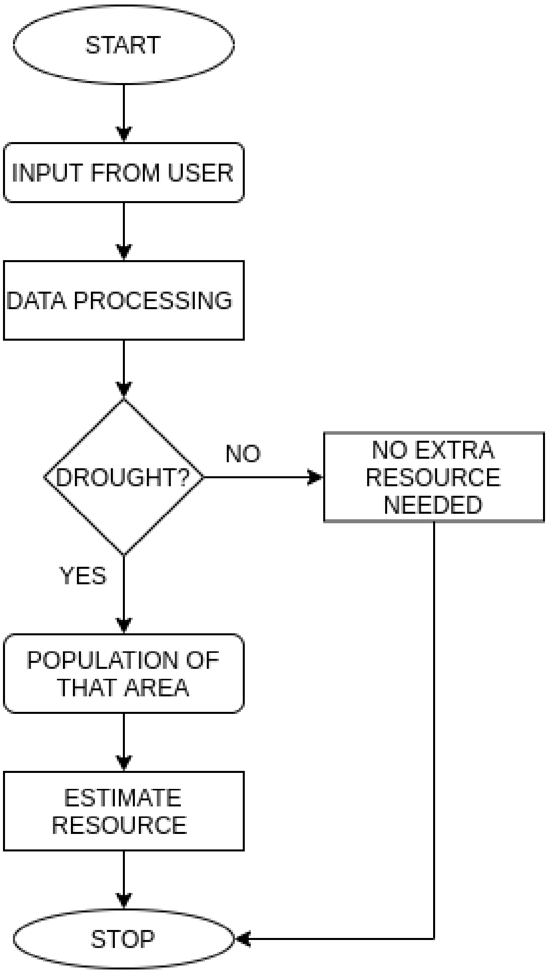


Fig 1. System Design Flow

IV. PERFORMANCE ANALYSIS

To evaluate the performance of the proposed method, numerous different datasets are used. The proposed algorithm is evaluated based on different parameters such as true positive (TP), false positive (FP), True Negative (TN) and false negative (FN). TP is described as when a system predicts a drought when the drought is observed in that year. FP is described as when a system predicts a drought but no drought is observed in that particular area in that year. TN is when no drought is predicted and no drought is observed in that region in the year of prediction. FN is when no drought is predicted but the area seems to have the drought.

The measure of prediction is defined as the combination of two criteria are defined as follows

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

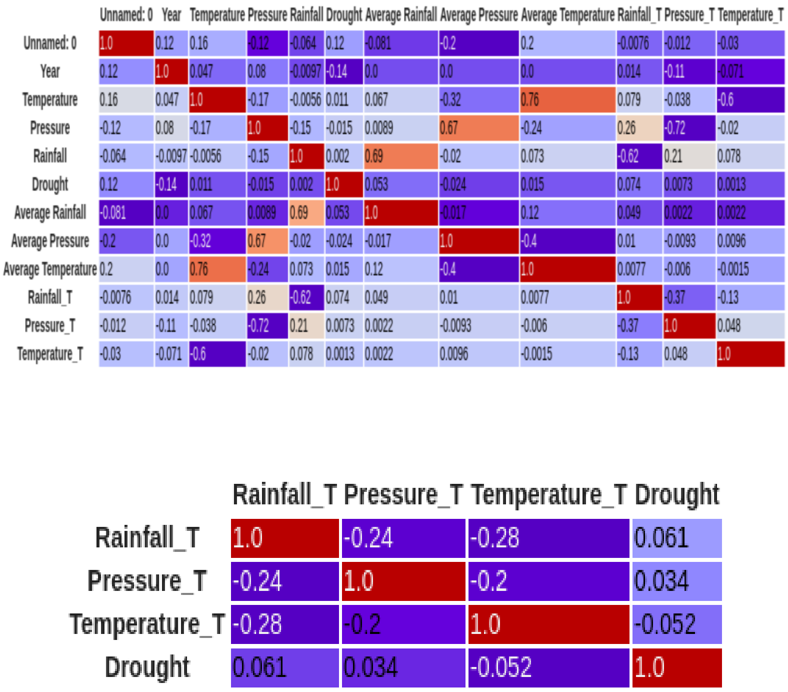
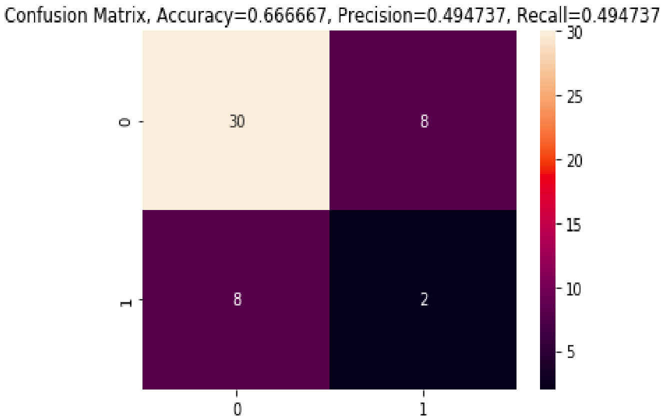
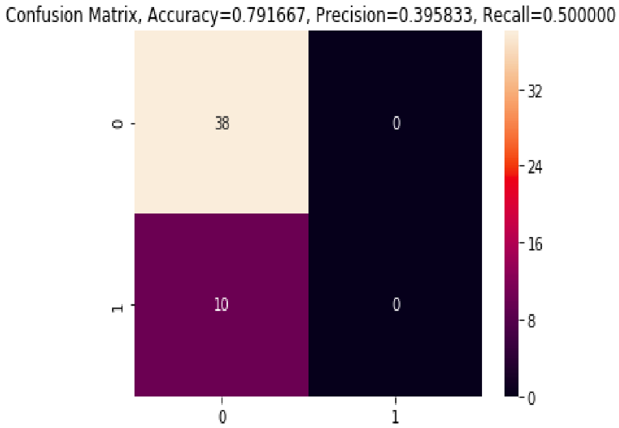


Fig. 2. Processed rainwater data for the model

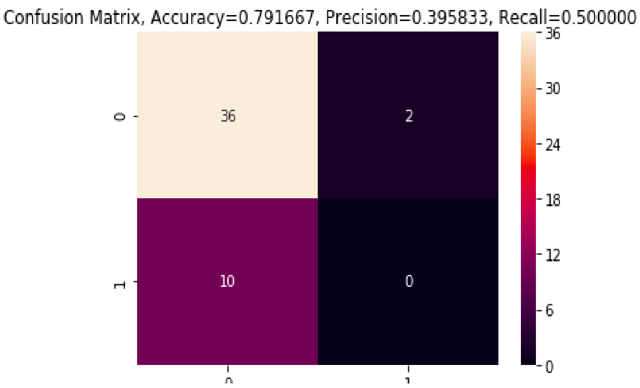
The data can be analysed on the basis of the confusion matrix obtained by using different classification algorithms and then the most efficient among them is chosen to be the prediction model for the project. On implementation of the confusion matrix we get three values on which the model is decided i.e. accuracy, precision and recall.



Gaussian NB



Naive Bayes



Random Forest Classifier

V. FUTURE VERSIONS

The present version of the prediction model can be improved by using proper and highly reliable data and also implementing the other constraints into consideration like the amount of water resources supplied each year in the form of water tankers and supplied water, current and reliable ground data on regional basis which is updated on a regular basis.

The accuracy can be increased exponentially if all such constraints are considered in the proposed model.

VI. CONCLUSION

The main feature of this project is the ability to predict the estimated amount of water resource needed in case the region suffers a drought in the upcoming year.

The ability of the prediction model to predict the drought in a region based on the rainfall data based on the random forest classification model with a 0.791667 % accuracy makes the model highly reliable and can be implemented with proper set of data for better results.

VII. ACKNOWLEDGEMENT

We are very thankful to our guide Dr. T. Sasikala M.E., Ph.D., Dean, School of Computing, for her valuable guidance, suggestion, and encouragement throughout this work. We convey thanks to Dr. S. Vigneshwari M.E., Ph.D., and Dr. L. Lakshmanan M.E., Ph.D., Heads of Department, Computer Science and Engineering, for providing full support during the reviews. I also thank all teaching and non-teaching staff of the department for their support.

VIII. REFERENCES:

[1] https://www.sciencedaily.com/terms/water_scarcity.htm

[2] Jasechko, S.; Perrone, D.; Befus, K. M.; Bayani Cardenas, M.; Ferguson, G.; Gleeson, T.; Luijendijk, E.; McDonnell, J. J.; Taylor, R. G.; Wada, Y.; et al. Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination. _Nat. Geosci._2017, _10_, 425.

[3] Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning, 2nd ed.; Springer: New York, NY, USA, 2017; pp. 43–100.

[4] Janža, M. Modelling heterogeneity of Ljubljana polje aquifer using Markov chain and geostatistics. Geologija 2009, 52/2, 233–240.

[5] Nayak, P.; Sudheer, K.; Rangan, D.; Ramasastri, K. Short-term flood forecasting with a neurofuzzy model.
Water Resour. Res. 2005,41.

[6] https://en.wikipedia.org/wiki/Statistical_classification

[7] H. Zhang (2004). The optimality of Naive Bayes. Proc. FLAIRS.

[8] https://blog.ferrovial.com/en/2015/03/how-many-liters-of-water-does-a-person-need-per-day/