## ABSTRACT

Water Scarcity is one of the most risky and unusual issues looked at 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.

# TABLE OF CONTENTS

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

## 1.1 DROUGHT PREDICTION SYSTEM

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 become one of the significant issues over the globe. A report proposes that the United States of America alone squanders 495 billion litres 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 the 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 to 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 that fundamentally contains labourers and furthermore influences the working effectiveness of the individuals.

Water shortage is one of the significant issues in the 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 the 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. This problem needs to be assessed much more carefully and precisely. The correct estimation of drought can greatly influence the loss that used to occur in case of unprecedented drought to all kinds of beings from humans and animals to the seasonal crops. Drought in an area can not only affect the present condition in that area but also greatly affects the future prosperity and it takes a lot to recover through such an epidemic.

## 1.2 NEED OF THE PREDICTION SYSTEM

Drought is among the most disastrous natural hazards and occurs in virtually all geographical areas. Severe drought events in recent decades, including 2010–

2011 East Africa drought, 2011 Texas drought, 2012 U.S. Central Great Plains drought, and 2012–2015 California drought, have caused huge losses to agriculture, society, and ecosystems with profound impacts on crop production and water supply. Extensive impacts of drought in past decades at regional and global scales call for improved capability to cope with drought. Drought prediction plays a key role in drought early warning to mitigate its impacts. Drought is a complicated phenomenon and is among the least understood natural hazards due to its multiple causing mechanisms or contributing factors operating at different temporal and spatial scales. Drought mostly originates from precipitation deficit, while in certain cases it may result from the anomaly of other variables, such as temperature or evapotranspiration. Specifically, high temperature may lead to increased evaporation and reduced soil moisture, causing drought in agricultural sectors. Moreover, drought may not be a purely natural hazard, for human activities such as land use changes and reservoir operation may alter hydrologic processes and affect drought development. Overall, the development and evolution of drought result from complicated interactions among meteorological anomalies, land surface processes, and human activities. Drought may occur with multiple processes driving its onset, persistence or recovery, happening at a wide range of time scales (sub seasonal/weekly, seasonal, multiyear, or decadal) and across different spatial scales (local, regional, continental, and global). Drought is commonly characterized at the seasonal time scale. Recently, it has been highlighted that drought may occur at the sub seasonal scale. For example, the 2012 central U.S. drought with rapid onset in May is generally referred to as flash drought (typically occurs for a few days or weeks), which results from concurrent soil moisture deficit and anomalously high temperatures (and increased evaporation).

Drought prediction generally refers to the prediction of drought severity (e.g., values of a specific drought indicator). In certain cases, drought prediction also refers to other properties, such as drought duration and frequency, or phases, such as drought onset, persistence, and recovery. In this study, we mainly focus on the prediction of drought severity at the seasonal time scale, which centres around the current drought prediction efforts and is particularly related to the operational early warning to mitigate drought impacts.

## 1.3 ALGORITHM USED

### *1.3.1 CLASSIFICATION ALGORITHM*

Classification is a technique to categorize our data into a desired and distinct number of classes where we can assign labels to each class.

Applications of Classification are: speech recognition, handwriting recognition, biometric identification, document classification etc.

Classifiers can be:

Binary classifiers: Classification with only 2 distinct classes or with 2 possible outcomes.

example: Male and Female

example: classification of spam email and non-spam email

example: classification of author of book

example: positive and negative sentiment

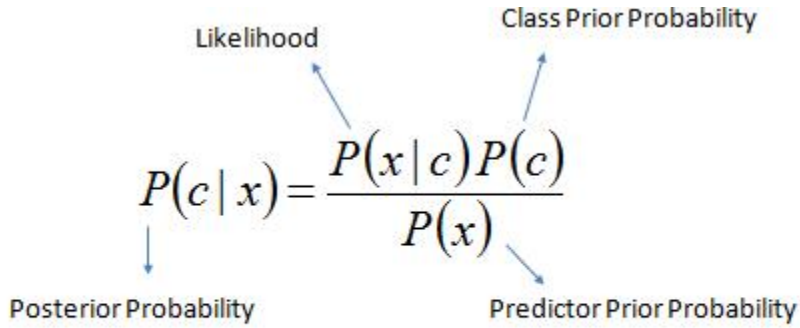
example: classification of mood/feelings in songs/music

example: classification of types of crops

#### *1.3.1.1. NAIVE BAYES CLASSIFIER*

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.




$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests.

Advantages:

- This algorithm requires a small amount of training data to estimate the necessary parameters.
- Naive Bayes classifiers are extremely fast compared to more sophisticated methods.

Disadvantages:

- Naive Bayes is known to be a bad estimator.

### *1.3.1.2 DECISION TREE CLASSIFIER*

Decision tree learning is a method commonly used in data mining.[1] The goal is to create a model that predicts the value of a target variable based on several input variables.

A decision tree is a simple representation for classifying examples. For this section, assume that all of the input features have finite discrete domains, and there is a single target feature called the "classification". Each element of the domain of the classification is called a class. A decision tree or a classification tree is a tree in which each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with an input feature are labeled with each of the possible values of the target or output feature or the arc leads to a subordinate decision node on a different input feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes, signifying that the data set has been classified by the tree into either a specific class, or into a particular probability distribution (which, if the decision tree is well-constructed, is skewed towards certain subsets of classes).

A tree is built by splitting the source set, constituting the root node of the tree, into subsets - which constitute the successor children. The splitting is based on a set of splitting rules based on classification features. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target variable, or when splitting no longer adds value to the predictions. This process of top-down induction of decision trees (TDIDT) is an example of a greedy algorithm, and it is by far the most common strategy for learning decision trees from data[citation needed.

In data mining, decision trees can be described also as the combination of mathematical and computational techniques to aid the description, categorization and generalization of a given set of data.

Data comes in records of the form:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, \ldots, x_k, Y)$$

The dependent variable, **Y** , is the target variable that we are trying to understand, classify or generalize. The vector **x** is composed of the features, x1, x2, x3, etc., that are used for that task.

Advantages:

- Simple to understand and interpret.
- Able to handle both numerical and categorical data.
- Requires little data preparation.
- Possible to validate a model using statistical tests.
- Performs well with large datasets.
- Mirrors human decision making more closely than other approaches.

Disadvantages:

- Trees can be very non-robust. A small change in the training data can result in a large change in the tree and consequently the final predictions.
- Decision-tree learners can create over-complex trees that do not generalize well from the training data. (This is known as overfitting.) Mechanisms such as pruning are necessary to avoid this problem (with the exception of some algorithms such as the Conditional Inference approach, that does not require pruning).
- For data including categorical variables with different numbers of levels, information gain in decision trees is biased in favor of attributes with more levels. However, the issue of biased predictor selection is avoided by the Conditional Inference approach, a two-stage approach, or adaptive leave-one-out feature selection.

#### *1.3.1.3 RANDOM FOREST CLASSIFIER*

Decision trees are a popular method for various machine learning tasks. Tree learning "comes closest to meeting the requirements for serving as an off-the-shelf procedure for data mining", say Hastie, "because it is invariant under scaling and various other transformations of feature values, is robust to inclusion of irrelevant features, and produces inspectable models. However, they are seldom accurate".