

# **RAINFALL IN INDIA FOR AGRICULTURE**

TEAM LEADER : ASWIN N S  
TEAM MEMBER 1: DIVAKAR U  
TEAM MEMBER 2: KAILASH N  
TEAM MEMBER 3: KRISHNA R

## **LITERATURE SURVEY**

### **1. RAINFALL IN INDIA FOR AGRICULTURE: A MACHINE**

#### **LEARNING APPROACH**

Weather plays an important role in agriculture production. Thus there is no aspect of crop culture that is immune to impact of weather. Weather factor contribute to optimal crop growth, development and yield. For rainfall variability needs to be expressed in terms of percentage so that minimum assured rainfall amounts at a certain level of probability. For optimal productivity at a given location crops must be such that their weather requirements match the temporal match of relevant weather elements. A detailed knowledge of rainfall regime at a place is an important prerequisite for agriculture planning and management. Soil fertility refers to the inherent capacity of soil to supply nutrients in adequate amount and in suitable proportion for crop growth and crop yield.

Significance of analysing Temperature, Rainfall and Soil:

1. It has a profound influence on crop growth, development and yields. Weather aberrations can cause physical damage to crops.
2. Well distributed rainfall is an important factor determining yield.
3. A defined range of maximum and minimum temperatures from the boundaries of observable growth.
4. The impact of climate change are most evident in crop productivity because this parameter

represents the component of greatest concern to producers, as well as consumers.

5. Warming temperatures associated with climate change will affect plant growth and development along with crop yield. Increase of temperature may cause yield declines between 2.5% and 10% throughout 21st century.

6. Global warming affects agriculture in a number of ways, including through changes in average temperature, rainfall.

Advantages of using Matlab in analysing data:

1. Datatypes and preprocessing capabilities designed for engineering and scientific data.
2. Thousand of prebuilt functions for statistical analysis, machine learning and signal processing.
3. Accelerated performance with simple code changes and additional hardware

## **2. RAINFALL IN INDIA FOR AGRICULTURE; A REVIEW**

Numeric data is easier to handle hence soil is converted in numeric value. According to water retainity soil is divided into 10 parts with 10 as max retainity soil and 1 the lowest water retainity.

Water retainity index	1	2	3	4	5	6	7	8	9	10
Example soils	Sandy soil		Sandy loams	Black soil(b asalt)		Red soil		Loam soil		Clay soil

Agriculture is the backbone of Indian economy. Due to global warming and climate change traditional farming in the regular months have been distorted and crops have been ruined is the most common phrase seen today. This not only gives economic losses but also the main reason for farmer suicide. Now agriculture needs support, time has come for technology to take over change. For a crop to grow ,favourable soil conditions, ambient rainfall and temperature is necessary. So as now due to climate change temperature and rainfall cannot be well defined , example rains in December and January or irregular temperatures have made it difficult for farmers and common man to predict months of plantation and yield of the crop due to irregularities. So we have formulated an analysis by prediction of a favourable crop based on temperature and current rainfall with soil conditions. Data science have proved that data which we have plays a vital role in predictions and iot based applications. Data science in agriculture is a growing field and has a wide scope in future.

Accordingly we have developed a prediction where current temp, soil condition and rainfall when entered we tell the crop which will give the highest yield. India is a land of diversity and varied soil conditions are found here.Each crop needs varied soil conditions and hence india is a homeland for production of various crops.

Now 3 parameters are varying hence we should consider a 3d geometry where each axis represents a parameter.

X represents- temperature

Y represents- rainfall

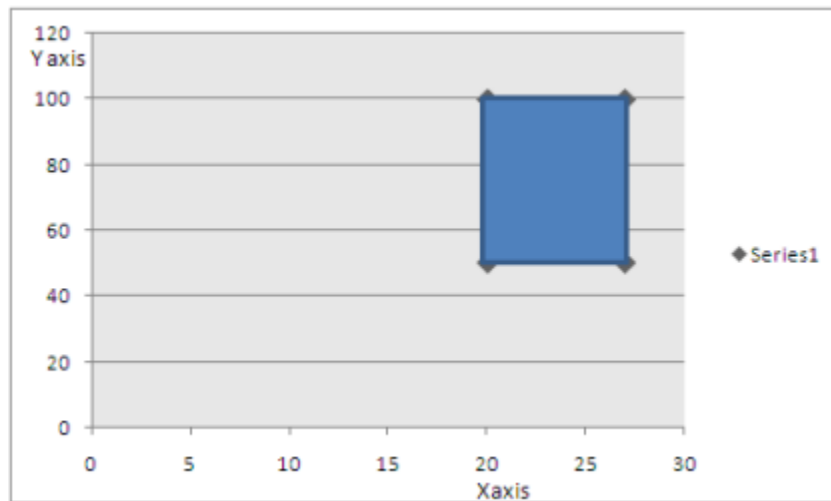
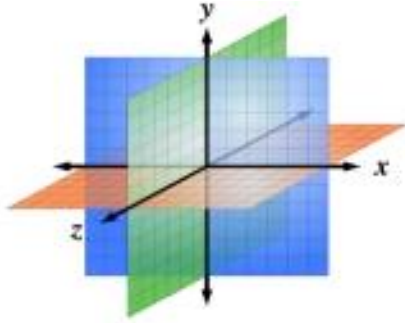
Z represents- soil

Each crop has an area in which the conditions are favourable.

Considering for sugarcane temp:20-27 rainfall:50 to 100 and soil (4)

Hence the rectangle in 3rd geometry is formed.

For convenience if we divide this 3d graph into 3;2d graphs.



Following are various applications of Data Science in agricultural sector:

1. Soil and Crop analysis
2. Weather Prediction
3. Fertilizer Recommendation
4. Disease Detection and Pest Management
5. Adaptation to climate change
6. Automated Irrigation System

In this Project, we are analyzing data of Soil and Crop and weather for crop prediction

## Crops with favourable soil ,rainfall and temp range

Sr no	crop	Temp range	Rainfall range	soil
1	Rice	22-32 degree celsius	150-300 cm	Clay soil
2	Wheat	10-15 degree celsius	75-100 cm	Loamy soil
3	Millets	27-32 degree celsius	50-100 cm	Red soil
4	Grams	20-25 degree celsius	40-45 cm	Loamy soil
5	Sugarcane	20-27 degree celsius	75-100 cm	Blacksoil (basalt)
6	Cotton	21-30 degree celsius	50-100 cm	Blacksoil (basalt)
7	Oil seeds	20-30 degree celsius	50-75 cm	Sandy loams
8	Tea	20-30 degree celsius	150-300 cm	loamy
9	Coffee	15-28 degree celsius	150-250 cm	Loamy

Accordingly we have developed a prediction where current temp, soil condition and rainfall when entered we tell the crop which will give the highest yield. This data will maximize the production yield and no farmer has to depend on their luck for the plantation.

## RANDOM FOREST ALGORITHM :

Understanding Random Forest

Advanced Algorithm Classification Machine Learning Project Python Regression  
Structured Data Supervised

This article was published as a part of the Data Science Blogathon

## Introduction

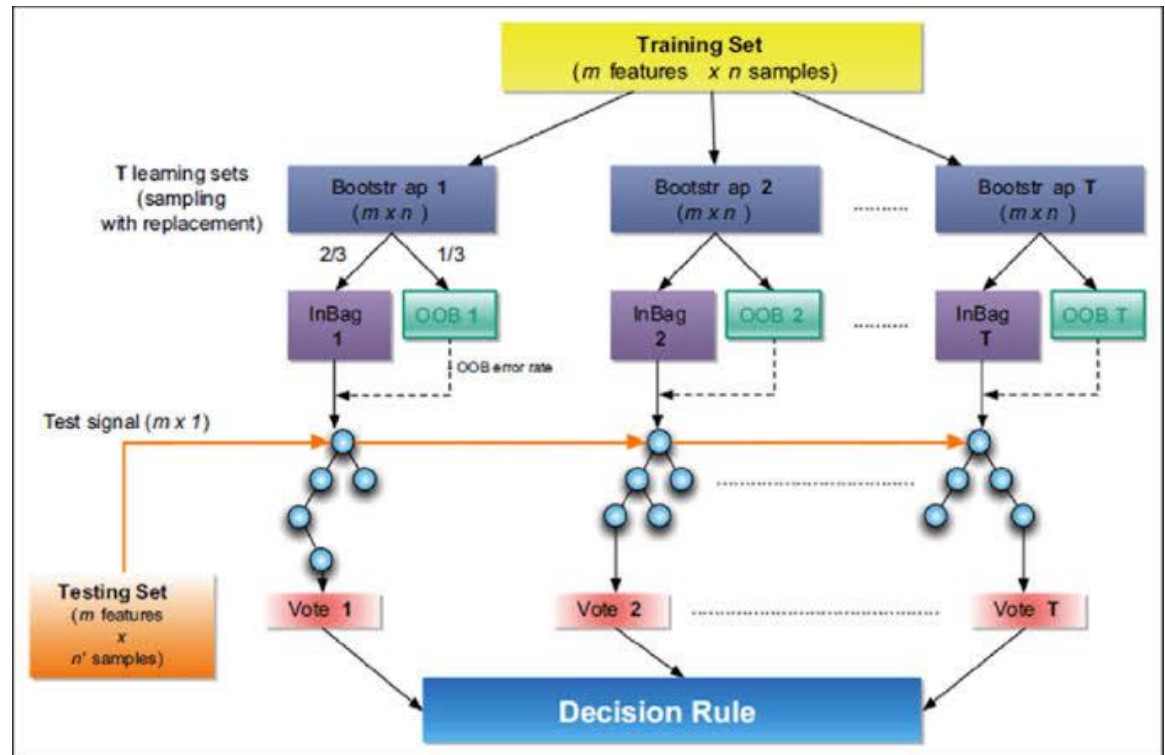
Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.

## **Real Life Analogy**

Let's dive into a real-life analogy to understand this concept further. A student named X wants to choose a course after his 10+2, and he is confused about the choice of course based on his skill set. So he decides to consult various people like his cousins, teachers, parents, degree students, and working people. He asks them varied questions like why he should choose, job opportunities with that course, course fee, etc. Finally, after consulting various people about the course he decides to take the course suggested by most of the people.

## RANDOM FOREST SPLIT:



## XGBOOST ALGORITHM:

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now. Please see the chart below for the evolution of tree-based algorithms over the years.

Evolution of XGBoost Algorithm from Decision Trees

XGBoost algorithm was developed as a research project at the University of Washington. Tianqi Chen and Carlos Guestrin presented their paper at SIGKDD Conference in 2016 and caught the Machine Learning world by fire. Since its introduction, this algorithm has not only been credited with winning numerous Kaggle competitions but also for being the driving force under the hood for several cutting-edge industry applications. As a result, there is a strong community of data scientists contributing to the XGBoost open source projects with ~350 contributors and ~3,600 commits on GitHub. The algorithm differentiates itself in the following ways:

A wide range of applications: Can be used to solve regression, classification, ranking, and user-defined prediction problems.

Portability: Runs smoothly on Windows, Linux, and OS X.

Languages: Supports all major programming languages including C++, Python, R, Java, Scala, and Julia.

Cloud Integration: Supports AWS, Azure, and Yarn clusters and works well with Flink, Spark, and other ecosystems.





## **Support Vector Machine Algorithm:**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane: SVM algorithm can be used for Face detection, image classification, text categorization, etc.

## **Types of SVM**

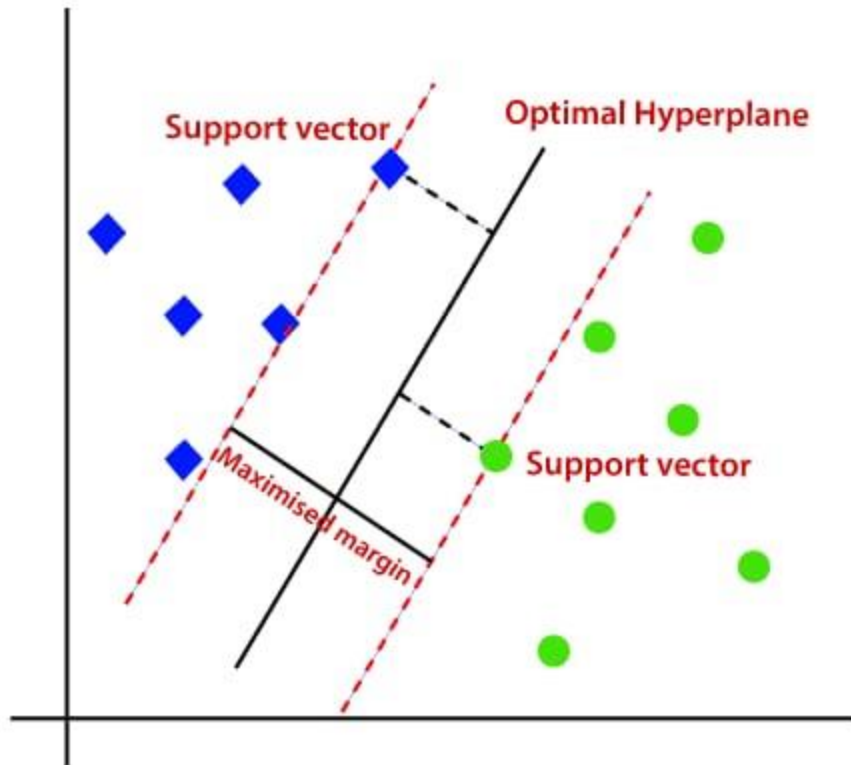
### **SVM can be of two types:**

#### **Linear SVM:**

Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

#### **Non-linear SVM:**

Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.



Weather aberrations can cause physical damage to crops.

With help of this project we can predict in certain environmental conduction which crop should be taken.

From the graph of % of production we can determine sowing and harvesting period of particular crop in given temperature and rainfall.

This data will continue to enhance farmer efficiency by further enabling them to monitor each plot of land and determine the precise input needed for their crops.