**The Effects of Water Stress on Crop Yield and Yield Estimation Using Machine Learning**

****

**INDIAN INSTITUTE OF REMOTE SENSING, DEHRADUN**

**By:**

**Pradip Patelia**

**ID: daiict19018**

**(M. Sc. Agriculture Analytics)**

**Rachit Patel**

**ID: daiict19028**

**(M. Sc. Agriculture Analytics)**

**Submitted To:**

**Dr. N. R. Patel**

**(Agriculture & Soils Department)**

**Dr. Prasun Kumar Gupta**

**(Geo-Informatics Department)**

**Table of Contents**

1. Abstract …………………………………………………....………03

2. Introduction ……………………………………………..…….…...04

3.Water Stress Affecting the Plant Yield…………………..……..........05

4.Literature Review ………………………………………..………....06

5. Study-Area…………………………………….……………….…..07

6. Research Question…………………………………….….…….….07

7. Data-Collection…………………………………………..………..07

8. Data Pre-processing……………………………………………….09

9. Insights…………..………………………………………….…….10

10. Workflow…...…………………………………………………….14

11. Machine-Learning Algorithms ……………….………………….15

12. Performance Matrix …………………………….……………….17

13. Conclusion ………………………..…………….……………….18

14. References ……..…………………………….………………….19

**Abstract:**

Since the beginning of human civilization, agriculture has played an important role in our lives. Water is the major component for the growth of the plant. The scarcity of water can affect the growth of the plant, which can lead to yield degradation. So, water stress has a negative impact on plant growth and yield.

Soil moisture plays an important role in crop development and productivity. Because it confirms that the water and nutrients needed for growth are available in quantity, which is enough for plants, optimum soil moisture levels are important for the plant's growth and development. On the other hand, more soil moisture can cause water logging conditions, which can be responsible for crop yield loss and crop failure. Therefore, for enhancing agricultural practises and ensuring food security, it is essential to comprehend the connection between soil moisture and crop output.

Soil moisture affects the crop depending on its type, but it affects the crop the most during its critical growth stage. If the crop cannot get enough water during that period, it affects the yield by about 50–60%.. So the plants need enough water during their critical growth stages. We take Punjab District as our study area, and we take the wheat crop because wheat is particularly sensitive to water stress as the plant grows, peaking in the early stages of dough development.

Our project investigates how this soil moisture impacts the crop yield. Our study involves collecting data related to the soil moisture. We take different parameters that are related to the soil moisture and use them to help us determine how soil moisture affects crop yield. The experiment is conducted in a controlled environment where the amount of water supplied to the plants can be regulated. We choose our study area of Punjab, and we take all parameters, which are enough to describe the environmental conditions of that area so that we can conduct our analysis conveniently.

The results of our study will provide deeper insights into how soil moisture affects yield. The study will demonstrate how different soil moisture levels affect plant growth, yield, and quality. In order to understand how meteorological factors like temperature, evapotranspiration, potential evaporation, T-max, and T-min affect crop production and soil moisture levels, we analyse all those datasets.

Depending on the crop, soil type, and weather factors, different soil moisture levels are ideal for maximising crop yield. Increasing moisture can cause waterlogging, which is a major factor that can affect crop growth. Water stress at critical crop growth stages can significantly lower yield. Hence, in order to maximise the crop yield and lower the chance of crop yield loss, it is very important to understand the ideal moisture levels for various crops and weather situations.

**Introduction:**

Since the beginning of time, agriculture has been the foundation of human civilization. Food security is now a top priority for policymakers, experts, and farmers alike due to the growing worldwide population. Soil moisture is one of the key elements that affects how productive and long-lasting agricultural systems are. A major factor influencing crop growth, development, and yield is soil moisture. It is necessary for sustaining photosynthesis, nutrient absorption, and water balance in plants.

  In recent years, farmers all over the world have faced considerable difficulties as a result of the effects of climate change, such as irregular rainfall patterns, droughts, and floods. Therefore, for sustainable agricultural production, it is now more crucial than ever to understand the connection between soil moisture and crop output. This research intends to investigate how soil moisture affects

  One active factor that affects plant development, occurrence, and production is soil moisture. Water stress significantly lowers the grain yield of Triticum aestivum L. wheat at every stage of growth. It is crucial in determining how much water and fibre is appropriate for plants, which in turn affects their yield and personality. Storms, heat, humidity, and soil type are some of the different material factors that have an impact on soil moisture levels.

  Understanding the relationship between soil moisture and crop yield is essential for reconstructing land practises and attaining food freedom. The significance of soil moisture in farming has existed since the dawn of civilization. In recent times, to meet the increasing demand for bread, fuel, and texture, there has been a growing dedication to something resembling land output and sustainability.

  The effects of global warming (climate change) on agriculture have had a much more important impact on managing soil moisture properly. The relationship between soil moisture and crop production is complex, and various factors play a crucial role in it. Water stress throughout the critical periods of crop development can significantly reduce production, according to many studies. Additionally, excessive moisture can result in waterlogging, which may be harmful to crop tumours. Depending on the crop, soil type, and material conditions, different liquid levels are required for maximum crop output. If plants get the optimum amount of soil moisture, their growth can be exceptionally high and their yield can increase.

**Water Stress Affecting the Plant Yield:**

At every stage of a plant's growth Water is necessary for everything from seed germination to plant maturation, but wheat is particularly sensitive to water stress as the plant grows, peaking in the early stages of dough development. The main environmental element that restricts plant growth and productivity is water stress. It is brought on by a lack of water, which can be the result of both natural occurrences like drought or waterlogging and human activity like overusing water resources, deforestation, and changing land uses. Water stress reduces the amount of water available for photosynthesis, nutrient uptake, and cellular respiration, which has an impact on plant production.

Plants that are under water stress have physiological and biochemical changes that are intended to conserve water and ensure plant viability. Stomatal closure, which lowers water loss through transpiration, and activation of particular genes that control water stress response pathways are some of these modifications. Water stress, on the other hand, can result in tissue damage to the plant and a decrease in production if it is severe and persistent.

The impact of water stress on plant output is influenced by a number of variables, including the type of plant, the stage of growth, the duration and intensity of water stress, and the plant's tolerance to stress. While some plants are more sensitive and need careful management to sustain yield, others are naturally adapted to live in water-limited settings and can retain their yield despite water stress.

Several tactics can be used to lessen the harmful effects of water stress on plant output. These include selecting drought-tolerant plant varieties that can resist water stress and improving water use efficiency through agricultural management techniques like irrigation planning. Crop types can be created through genetic engineering and breeding techniques that are more resilient to water stress.

**Literature Review:**

Soil moisture is a key driver of yield in maize, soybean, and wheat crops. The authors found that even small reductions in soil moisture can cause significant yield losses, with maize being the most sensitive to changes in soil moisture.

In another study, the authors investigated the impact of soil moisture on potato yield. The results showed that a 25% reduction in soil moisture led to a 17% reduction in potato yield. The authors also found that soil moisture had a greater impact on potato yield during the tuber initiation stage than during the vegetative growth stage.

In one study, the authors examined the effects of soil moisture on cotton yield. The authors found that a 10% reduction in soil moisture during the boll development stage led to a 13% reduction in cotton yield. The study also showed that cotton plants were most sensitive to changes in soil moisture during the reproductive growth stage.

There is one paper in which they investigated the impact of soil moisture on wheat yield in arid regions. The authors found that soil moisture was a critical factor affecting wheat yield, and a 20% reduction in soil moisture led to a 50% reduction in wheat yield.

This Aforementioned Studies suggests that the soil moisture plays the significant part in the growth of the crop and its production. If plants don’t get the enough soil moisture at their critical growth stages, then it can affect their yield up to more than 40% so that suggests that optimum plant moisture can be needed for the crop growth and development. Also, the excessive soil moisture can affect the plant yield so soil moisture should be available when plant needs the most that amount of soil moisture plays the important part in determining the growth and yield of the plants.

**Study-Area:**

we take the Punjab State as our study area because Punjab has the greater contribution in wheat production in the India. We take the Wheat Crop because it is major crop of the Punjab District.

**Research-Question:**

* Reduction in the Crop Yield Due to the Water Stress.
* Yield Estimation Using Machine Learning

**Data-Collection:**

For the preparation of the data, we take various parameters to find the relationship between soil moisture and yield. Soil moisture plays an important role in determining the yield, which is why correct parameters should be required for finding the relationship between those two parameters.

We used Google Earth Engine for extracting the data. We used different parameters that were related to the soil moisture. We extracted all those parameters for the 22 districts of the Punjab, and we did that from 2000 to 2020. We used different satellites and sensors and extracted datasets for these parameters. We used different satellites because a single satellite could not provide the range of data for the desired time period that we required.

**Parameters:**

1.NDVI (Normalized Difference Vegetation Index)

2. Soil Moisture

3.Minimum Temperature(T-MIN)

4.Maximum Temperature(T-MAX)

5.Area (Hectare)

6.Production (Tonnes)

7.Total Precipitation(mm)

8.Total Potential Evapotranspiration (PET)

9.Total Evapotranspiration (mm)

10.Yield

**1.NDVI**

NDVI is called the Normalised Difference Vegetation Index. It is the measure of the greenness. It means it is used to assess the vegetation's health as well as its growth. For the NDVI satellite, use the reflectance of the red and near-infrared band wavelengths. NDVI has a range of -1 to +1. For the dense vegetation, it is >0.5, and for the water, it is -1. We used the MODIS (Moderate Resolution Imaging Spectroradiometer) for getting the NDVI data. It has a spatial resolution of 500 metres.

* **NDVI = (NIR - red) / (NIR + red)**

**2.Soil Moisture**

We used GLDAS (Global Land Data Assimilation System) for getting the soil moisture. It gives gridded data for surface variables. It has the spatial resolution about the 28 kms and temporal resolution of the 3 hours. We have collected the soil moisture data for the 10cm depth.it has unit kg/m^2 for the soil moisture.

**3. Maximum Temperature(T-Max)**

We have collected the T-Max from the ERS-5 datasets which is provided by the European Centre for medium range weather forecasts (ECMWF). It has the spatial resolution of 31km and Temporal resolution of Daily. It is the 2-meters air temperature which has the Kelvin (k) unit.

**4.Minimum Temperature(T-Min)**

We have collected the T-Min from the ERS-5 datasets which is provided by the European Centre for medium range weather forecasts (ECMWF). It has the spatial resolution of 31km and Temporal resolution of Daily. It is the 2-meter air temperature which has the Kelvin (k) unit.

**5.Area and Production**

We take the area and production for the 2000 to 2020 of the wheat crops from this website <https://aps.dac.gov.in/APY/Public_Report1.aspx>.

**7.Total Precipitation (mm)**

We have collected the precipitation data from the CHIRPS datasets (Climate Hazards Group Infrared Precipitation with Station data). This Dataset is developed by the Climate Hazards Group. It is the satellite based high resolution infrared data with which is used for drought, flood monitoring and crop yield prediction. It has the spatial resolution about 5.9 km and temporal resolution of 5 days.

**8.Total Potential Evaporation (PET)**

Total Potential-Evaporation is measure of water which is evaporated from the soil under the ideal conditions we assume that there is unlimited supply of the water.it is driven by the atmospheric demand of the water which is determined by the environmental factors such as the temperature, evapotranspiration, solar radiation.

We take data from the ERS-5 datasets which is provided by the European Centre for medium range weather forecasts (ECMWF).it provides the meteorological data at daily temporal resolution and its spatial resolution is the 31 km.

**9. Total Evapotranspiration**

Total Evaporation is the sum of the water which is evaporated from the soil and transpires from the plants into the atmosphere. various environmental variables can affect the rate of the Total Evapotranspiration.it can be useful for the determining the soil moisture. we take the data from the MODIS (Moderate Resolution Imaging Spectroradiometer) which has the spatial resolution about the 500 meters.

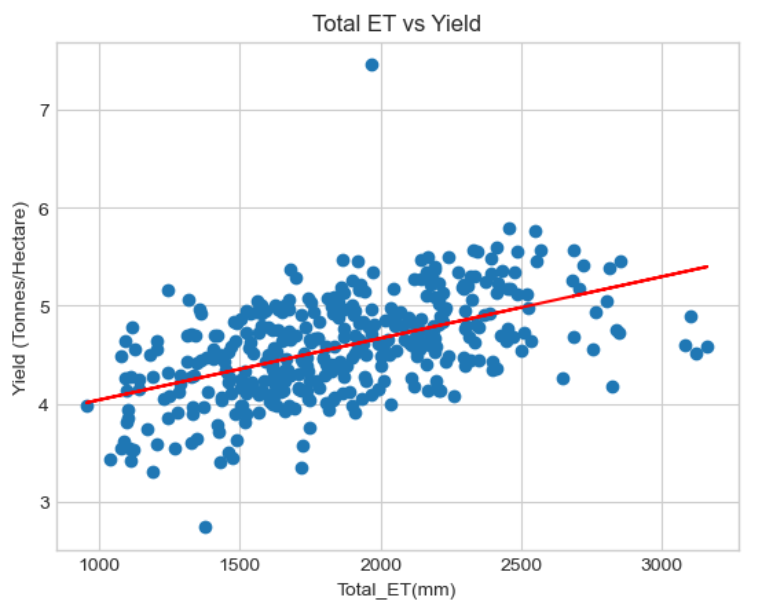
**Pre-processing of the data**

After getting all the datasets for the Punjab, first we combined all the tif files, year and parameter wise, that we got from the Google Earth Engine, then we opened QGIS and ran zonal statistics for each parameter and year on each district. That’s how we got the excel file for all the parameters for our study area, like T-Max, T-Min, soil moisture, precipitation, total potential evaporation, total evapotranspiration, and NDVI.

We combined all the Excel files, we arranged it so that for each year we set all the parameters, and we did that for all the parameters for each year. That’s how we prepare our data, then we check for the nan values, missing values, and duplicate values.

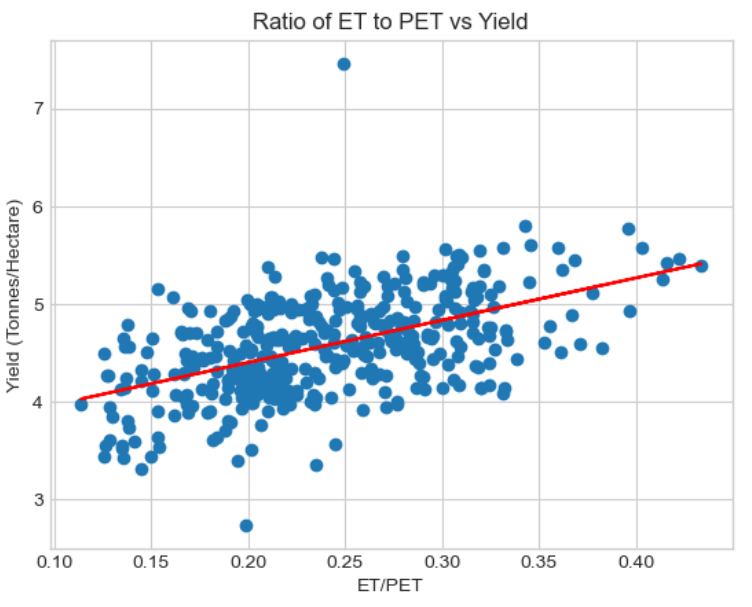
**Insights:**

**Total ET vs Yield:**

****

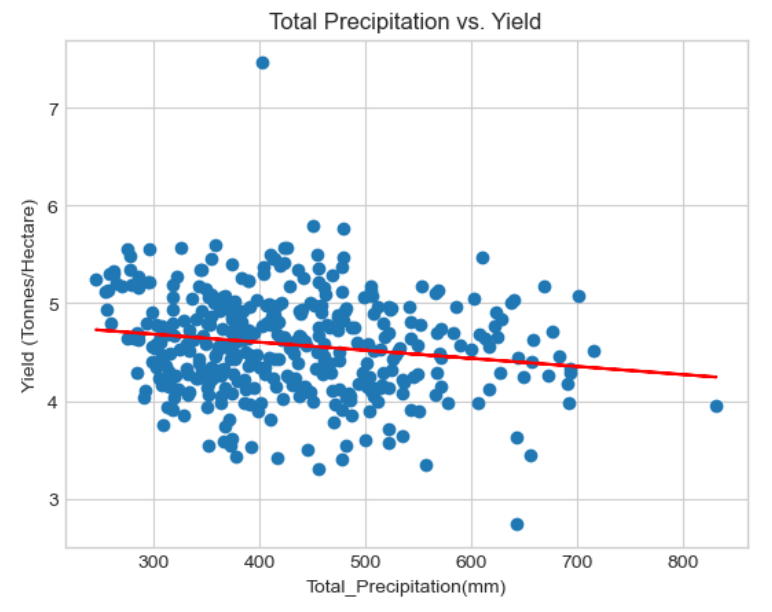
ET, or evapotranspiration, is the combined loss of water from the soil surface through evaporation and from plant leaves through transpiration. ET is an important factor that affects crop yield as it reflects the amount of water that crops used to grow and produce a good yield. In short, if the ET is less, it indicates that your crop is undergoing the Water Stress. As we can see in the graph that as the ET increases, the yield tends to increase which means ET is important factor for crop water management.

**Ratio of ET to PET vs Yield:**

****

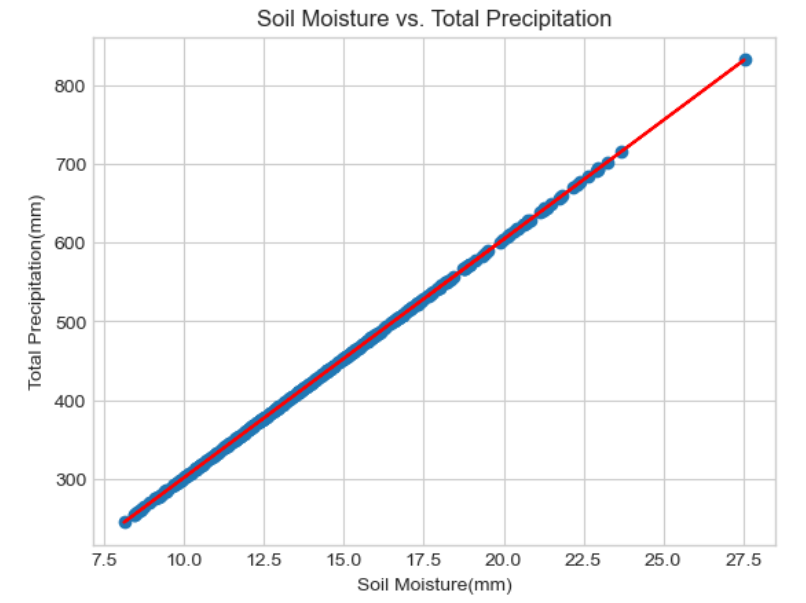
The ET to PET ratio is used to assess the water stress of a crop. If the ET to PET ratio is less, it indicates that the crop is experiencing water stress and not using all the water that it could under optimal conditions. If the ET to PET ratio is greater than one, it suggests that the crop has sufficient water to meet its needs. The ET to PET ratio can also be used to optimize irrigation scheduling by helping farmers to determine when and how much water to apply to the crop. By monitoring the ET to PET ratio, farmers can adjust their irrigation practices to maintain an optimal water balance and reduce the risk of overwatering or underwatering the crop.

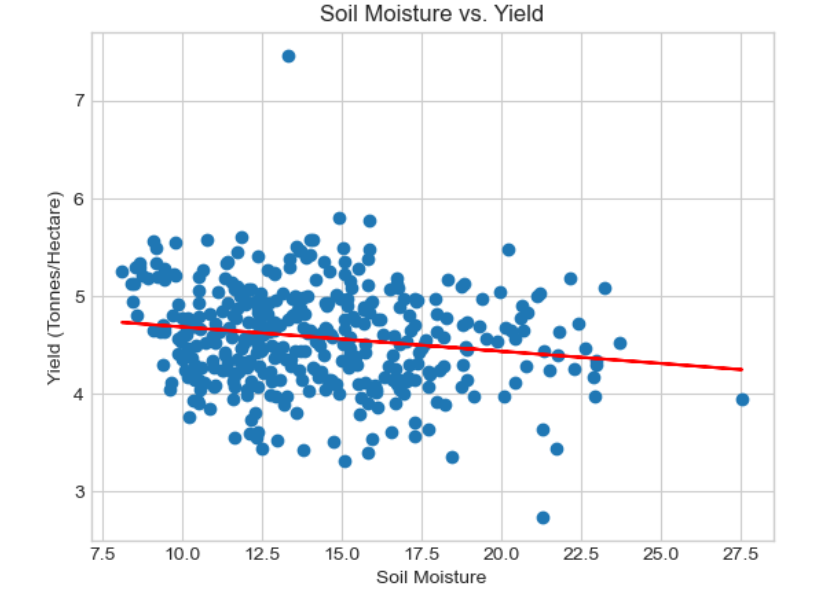
**Total Precipitation vs Yield:**

****

Direct rainfall can cause yield loss in wheat in Punjab, especially if it occurs during the flowering or grain filling stage of the crop. Excessive rainfall during these stages can lead to lodging of the plants, which can result in reduced grain yield and quality. Lodging occurs when the stems of the wheat plants are weakened by heavy rain and wind, causing them to bend or break. This can lead to reduced photosynthesis, limited nutrient uptake, and increased susceptibility to diseases and pests. In addition, high humidity levels after rainfall can also increase the risk of fungal diseases such as rust and powdery mildew, which can further reduce crop yields. Therefore, while rainfall is essential for crop growth, excessive or poorly-timed rainfall can negatively impact wheat yields in Punjab. So, basically, it states that in Rabi season, source of water for crop growth should not be direct rainfall but the irrigation facilities.

**Soil Moisture vs Total Precipitation:**

****

****

As we can see, Soil Moisture does have the linear relationship with the Rainfall. So, hereby we can say that moisture is increasing due to the poorly timed rainfall and thus it causes the drop in crop yield.

**Yield Prediction using Machine Learning**

**Workflow:**

**Requisite**

**Parameters**

**Area,**

**Production**

**& Yield Data**

**Data Collection**

**Data Cleaning &**

**Processing**

**Data Splitting**

**Train Dataset**

**Test Dataset**

**Train Model**

**Fine tune the Model**

**Evaluate the Model**

**Predictions**

**Predictive Model**

**Different ML algorithms used for yield prediction:**

**1.Linear Regression:**

Linear Regression is a statistical modelling technique used to establish a relationship between a dependent variable and one or more independent variables. It is a simple yet powerful technique that is widely used in data analysis, machine learning, and other fields. In Linear Regression, the goal is to find a linear relationship between the independent variables and the dependent variable.

**2.Polynomial Regression:**

Polynomial Regression is a variation of Linear Regression that involves fitting a polynomial equation to the data instead of a straight line. In other words, it is a type of regression analysis in which the relationship between the independent variable X and dependent variable Y is modelled as an nth degree polynomial.

Polynomial Regression is useful when the relationship between the dependent and independent variables is not linear but can be approximated by a polynomial function. It can capture non-linear relationships and can be used to model complex data patterns.

**3.Decision Tree:**

A Decision Tree is a machine learning algorithm that can be used for both classification and regression tasks. It is a graphical representation of all possible solutions to a decision based on certain conditions.

The Decision Tree is constructed by recursively splitting the data into subsets based on the values of the independent variables. Each split is based on the variable that provides the most information gain, which is a measure of the reduction in uncertainty of the target variable (class or regression value) after the split.

The goal of the Decision Tree is to create a model that predicts the target variable by learning simple decision rules inferred from the data features. The decision rules are represented as a tree, where the internal nodes represent the features or attributes, the branches represent the possible values of those features, and the leaves represent the predicted class or regression value

**4.Random Forest:**

Random Forest is a machine learning algorithm that is an extension of Decision Trees. It is a type of ensemble learning method that combines multiple Decision Trees to create a more accurate and stable prediction model.

In Random Forest, a large number of Decision Trees are trained on random subsets of the data and the features. Each tree is trained independently on a different subset of the data, with the subsets sampled with replacement (known as bootstrap aggregating or bagging), and each node in the tree is split using a random subset of the available features. This randomness helps to reduce overfitting and increase the model's generalization ability.

The predictions from the individual trees are then combined by taking the average (in regression problems) or the mode (in classification problems) of the predictions made by each tree. This combination of multiple trees leads to a more stable and accurate prediction model.

**5.XGBoost:**

XGBoost (Extreme Gradient Boosting) is a popular machine learning algorithm that is used for both regression and classification problems. It is an extension of the Gradient Boosting algorithm that enhances the performance and speed by introducing several additional features.

XGBoost works by combining multiple weak learners, typically decision trees, to create a strong prediction model. It trains each tree sequentially by minimizing the loss function using gradient descent, and then combines the predictions from each tree to make a final prediction.

**6.LightGBM (Light Gradient Boosting Machine):**

LightGBM is a popular open-source gradient boosting framework that is used for both regression and classification problems. It is designed to be efficient and scalable, making it suitable for large-scale datasets.

LightGBM works by training an ensemble of decision trees using a boosting approach. It uses a gradient-based approach to find the optimal split points for each node in the decision tree, which helps to improve the accuracy of the model.

One of the key features of LightGBM is its ability to handle high-dimensional data. It uses a technique called 'Histogram-based Gradient Boosting' that converts the numerical features into categorical bins, which reduces the memory usage and makes it faster to train

**Performance Report:**

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **R2-Score** | **MSE** |
| **Linear Regression** | **0.5120** | **0.1335** |
| **Random Forest Regression** | **0.5998** | **0.1095** |
| **Polynomial Linear Regression** | **0.6136** | **0.1057** |
| **LGBMRegressor** | **0.6632** | **0.0921** |
| **XGBRegressor** | **0.6726** | **0.0895** |

**Conclusion:**

In our project, we found out that the water stress has a significant impact on the yield of the crop, and using machine learning, we can estimate the yield accurately. Our study figured out that water is an essential parameter for agricultural crop production. If the amount of water in the soil is less, it can reduce grain production by up to 60%, which is quite a huge number. Innovative technologies like machine learning can help the farmers predict their yield in advance so that they can manage their production.

**References:**

1. Dong, W., Li, C., Hu, Q., Pan, F., Bhandari, J., & Sun, Z. (2020). Potential Evapotranspiration Reduction and Its Influence on Crop Yield in the North China Plain in 1961–2014. Advances in Meteorology, 2020, 1–10. <https://doi.org/10.1155/2020/3691421>
2. India, T. (2022, April 21). Wheat yield in Punjab down by 5 quintal/hectare due to intense heat wave. Tribuneindia News Service. <https://www.tribuneindia.com/news/punjab/wheat-yield-in-punjab-down-by-5-quintal-ha-388043>
3. Joshi, V. (2021, May 29). Punjab reaps record wheat production but per hectare yield goes down by 3%. Hindustan Times. <https://www.hindustantimes.com/cities/chandigarh-news/punjab-reaps-record-wheat-production-but-per-hectare-yield-goes-down-by-3-101622317293853.html>
4. Medhavy, T. T., Sharma, T., Dubey, R. S., Mahey, R. K., & Sharma, P. (1995). Development of a wheat yield model for Punjab using remotely sensed data and historical yield trends. Journal of the Indian Society of Remote Sensing, 23(1), 23–30. <https://doi.org/10.1007/bf02997432>