**A logo for a university

AI-generated content may be incorrect.**

**GIS PROGRAMMING COURSE PROJECT**

**Topic: A model for predicting crop yield based on weather conditions and other environmental factors**.

**KATLEGO MOTJOLOPANE – 220016968**

**Topic: A model for predicting crop yield based on weather conditions and other environmental factors**.

Background study:

For successful national planning and to guarantee food security, near accurate results and statistics on the output of crops is necessary. The use of satellite-based remote sensing has become a practical and affordable approach for crop monitoring. The suggested Crop Yield Prediction Model offers an approach to agricultural crop prediction that generates accurate and practical insights for resource management, agricultural planning, and food security optimization by combining geospatial analysis with environmental data.

For sustainable agricultural development, environmental assessment, crop compatibility assessment, and related agricultural practices, geospatial technology is a significant tool. Remote sensing has proven to be an effective technique for tracking the spatial distribution of agricultural croplands and LULC classes. The use of satellite-based spatial imagery enables the rapid, extensive, and continuous monitoring of crop fields.

The following factors justify creating a crop production forecast model, crucial for environmental and agricultural reasons:

* Food Security: By predicting agricultural yields, governments and organizations may prepare for future food surpluses and shortages, which makes planning for the future easier for the government and farmers.
* Adapting to Climate Change: Agriculture is impacted by climate variances. Farmers can create adaptable strategies with the aid of a prediction model.
* Effective Resource Management: Farmers can minimize waste and environmental deterioration by optimizing the use of water, fertilizer, and pesticides by anticipating yields.

**Objectives:**

* Collect and analyse data related to soil type, weather patterns, and environmental factors affecting crop yield.
* Develop a Python-centred model that predicts crop yield with near-accurate results.
* Validate the model using previous agricultural data.
* To provide a user-friendly interface for farmers and environmental stakeholders to access predictive insights.

Weather-related factors impacting crop yield.

Temperature, Rainfall, Humidity, Wind

Other factors to consider include:

Water Availability

Irrigation, water sources, and drought,

Libraries to utilize for data capturing and processing:

**numpy**:

Numerical operations

**pandas**:

Handling structured data in tabular form

Data manipulation and cleaning.

**matplotlib**

Visualization of resultant data and plotting of results

Data collection:

Gather historical crop yield data from agricultural databases.

Collect weather data (temperature, rainfall, humidity, solar radiation.

Acquire soil data (pH, moisture content, nutrient levels, texture.

Integrate remote sensing and satellite imagery where applicable.

Methods to utilize for development:

Geospatial Analysis**:** Use of GIS to enhance predictions.

Linear regression:

Linear regression is an effective method when the relationship between yield and environmental factors is linear. It is suitable for simple relationships between yield and environmental factors.

Data format used for capturing data in a data frame for model testing:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Rainfall(mm) | Pesticides(hg/ha) | Temperature | Crop\_yield) |
| 2010 | 285 | 56 | 20.62 | 19882 |
| 2011 | 285 | 56 | 19.26 | 19297 |
| 2012 | 285 | 56 | 20.11 | 18999 |
| 2013 | 285 | 56 | 20.73 | 18683 |
| 2014 | 285 | 56 | 20.6 | 16000 |
| 2015 | 285 | 56 | 20.52 | 18420 |
| 2016 | 285 | 56 | 20.6 | 20591 |

MODEL TESTING

**Linear Regression:**

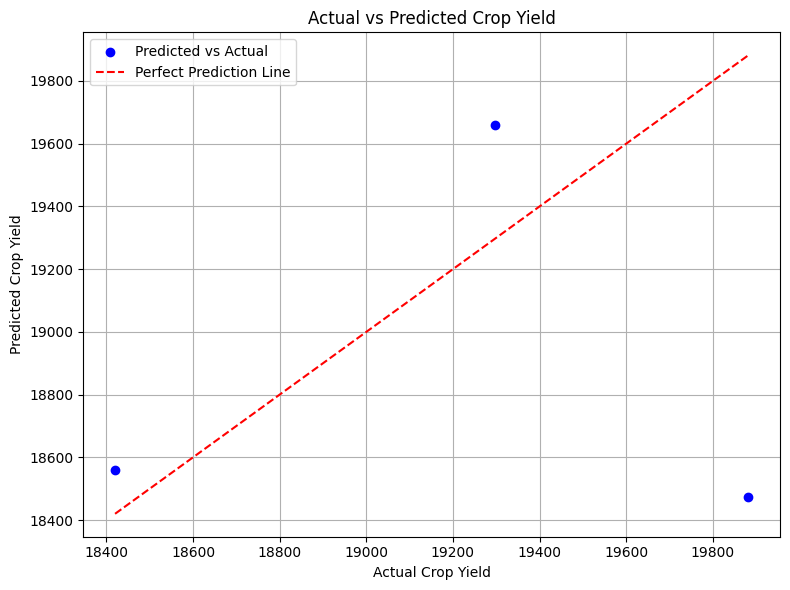
A basic statistical technique called linear regression describes the linear relationship between one or more independent variables (such as temperature, rainfall, and pesticides) and one dependent variable crop yield. It works well when:

The relationships between variables are linear.

Interpretability is important.

A baseline or benchmark model is needed.

Model testing:



Comparing Actual vs Expected Crop Yield:

Plotting the actual and expected crop yield data allows you to see how well the predicted and real values match up.

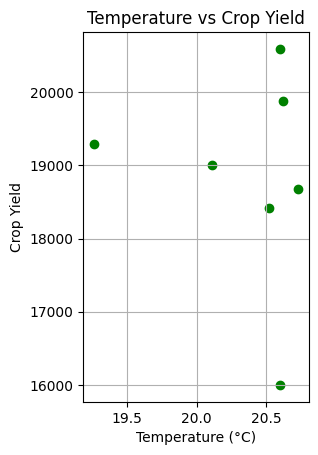
The model is effective if the predicted values closely match the actual values.  
The model's predictions are inaccurate if there is a large discrepancy between the actual and anticipated values.

The scatter plot is visually inspectable:  
Good performance is indicated by a linearly aligned plot with points(blue) near a diagonal line (from bottom left to top right).

The model is correctly forecasting the crop production, according to the scatter points that are closely located to the diagonal line.

**Temperature vs. Crop Yield (tons)**

One regression challenge in machine learning and data science is predicting agricultural production depending on temperature. Understanding how temperature variations impact agricultural productivity, particularly considering climate change, requires this type of investigation.



Testing the model by using the current or desired input temperature with the aid of the codes utilized to test the models, we can apply the following codes in Python to get an estimate of Crop Yield for 2025 based on temperature:

print("\n--- Predict Crop Yield for 2025 ---")

try:

temp\_input = float(input("Enter expected average temperature for 2025 (°C): "))

input\_data = pd.DataFrame({

'Rainfall(mm)': [285],

'Pesticides(hg/ha)': [56],

'Temperature': [temp\_input]

})

prediction = model.predict(input\_data)

print(f"\nEstimated Crop Yield for 2025 (Temp: {temp\_input}°C): {prediction[0]:.2f}")

except ValueError:

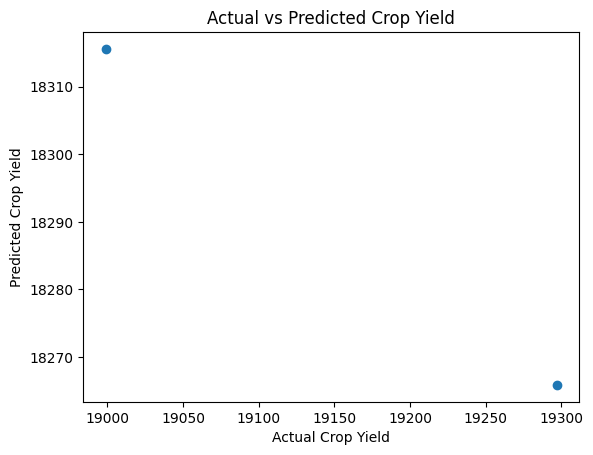
print("Invalid input. Please enter a numeric temperature value.")

The above code allows you to edit the temperature of the desired year along with the rainfall for that specific year and the pesticides used for the crops for the specific year, which outputs a predicted crop yield in tons for the year of choice.

Crop Yield based on the rainfall data alone:

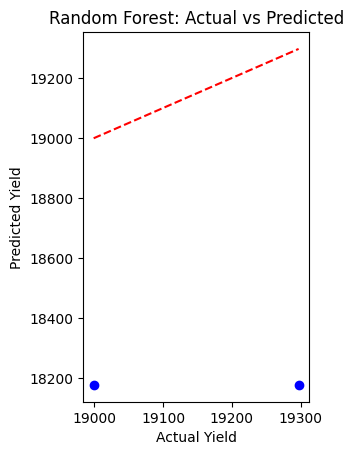
Rainfall for the respective years:

|  |  |  |
| --- | --- | --- |
| Year | Rainfall(mm) | Crop\_Yield) |
| 2010 | 285 | 19882 |
| 2011 | 285 | 19297 |
| 2012 | 285 | 18999 |
| 2013 | 285 | 18683 |
| 2014 | 285 | 16000 |
| 2015 | 285 | 18420 |
| 2016 | 285 | 20591 |



Building a Non-linear model for the crop yield prediction

Random Forest: To generate predictions, the machine learning method Random Forest employs a group of decision trees. It creates a more stable and accurate forecast by constructing several decision trees and integrating their output (voting for classification or averaging for regression).



Gradient Boosting Performance:

R-squared: -9.7955

Mean Squared Error: 239671.57

R-squared (R²): Indicates how effectively the model accounts for the variance in the desired outcome (crop yield).  
  
MSE (Mean Squared Error): This is the average squared prediction error; the lower, the better.

Minimal application of Random Forest due to a lack of datasets, which leads to inaccurate training of data and predictions.

General observations:

The annual rainfall of 285 mm is consistent. Because rainfall is constant, the model may not detect enough variability to forecast crop yield based just on rainfall. Nonetheless, the model can take into consideration the trend in crop output over time by using the crop yield from the prior year as a feature.

Yield Variation from Year to Year: Despite consistent rainfall, agricultural yields vary significantly, particularly in 2014 when yields decrease drastically. This suggests that the crop yield is influenced by variables other than rainfall and the crop output from the prior year. For instance, the model may not account for elements such as farming practices, pest infestations, soil conditions, or temperature.

Model Performance: With reasonable R-squared and MSE values, the model would perform rather well. Although the constant rainfall may limit the amount of variance that the model can explain, the yield from the previous year is a helpful addition.

In conclusion, evaluate the model's performance using both visual inspection (actual vs. projected plot) and numerical metrics (such as R2 and MSE) after testing. You can either use more sophisticated techniques to increase the accuracy of the current model or make improvements based on this evaluation.

References:

* Scikit-learn Library (Machine Learning Framework)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

* Crop Yield Prediction with Machine Learning

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151, 61–69. https://doi.org/10.1016/j.compag.2018.05.012

* Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018).

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151, 61–69. https://doi.org/10.1016/j.compag.2018.05.012