SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
Ramapuram Campus , Chennai – 600 089
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

18CSP109L - MINOR PROJECT

CROP YIELD PREDICTION FOR ENHANCING AGRICULTURE USING COMPARATIVE ANALYSIS & ENSEMBLE LEARNING

BATCH NUMBER - 1

Team Members	Supervisor
Badrinath S [RA2111027020005] Suresh S [RA2111027020047] Irfaan M. A. [RA2111027020006]	NAME: Ms. Adlene Ebenezer P, AP/CSE

Agenda

- Abstract
- Scope and Motivation
- Introduction
- Literature Survey
- Objective
- Problem Statement
- Proposed Work
    - Architecture Diagram
    - Novel idea
    - Module Description
- Software & Hardware Requirements
- Implementation
- Intermediate Results
- References

Abstract

- Develop an advanced machine learning model to predict crop yields more accurately using comparative analysis and ensemble learning.
- Implement four regression algorithms: Linear Regression, Lasso Regression, Ridge Regression, and Decision Tree Regression.
- Evaluate each algorithm's performance based on their Mean Squared Error (MSE) scores to identify the most effective models.
- Select the two algorithms with the lowest MSE scores for further analysis.
- Combine the strengths of the top two performing algorithms through ensemble learning to enhance predictive accuracy.
- Demonstrate that the ensemble approach significantly improves crop yield predictions, providing valuable insights for optimizing agricultural practices and ensuring food security.

Scope and Motivation

Scope

- **Data-Driven Agriculture**: Integrate machine learning techniques into agricultural practices to provide precise and reliable crop yield predictions.
- **Ensemble Learning**: Implement and test ensemble methods to combine the strengths of multiple algorithms for improved prediction accuracy.
- **Scalability**: Develop a model that can be adapted to different types of crops and geographical regions.

Motivation

- **Food Security**: Address the global challenge of food security by improving the accuracy of crop yield predictions, which can help in planning and resource management.
- **Sustainability**: Promote sustainable farming practices by enabling precise and targeted interventions that minimize environmental impact.
- **Economic Benefits**: Support farmers in maximizing their yields and profitability by reducing uncertainties associated with crop production.

Introduction

- Agriculture is vital for feeding the world's population, but predicting crop yields is challenging due to unpredictable factors.
- This project aims to improve the accuracy of crop yield predictions to help farmers make better decisions and increase productivity.
- Different methods are examined to determine the most effective approach for predicting crop yields.
- The project supports farmers in managing their crops more efficiently, leading to better planning and resource allocation.
- Accurate predictions help increase food production and support global food security.
- The project promotes sustainable farming practices by providing precise information, reducing waste, and minimizing environmental impact.
- The goal is to create a reliable tool for farmers to optimize crop yields, ensuring a stable and sufficient food supply.

Literature Survey

S.No.	Title of the Paper	Year	Journal/ Conference Name	Dataset	Preprocessing methods used	Prediction algorithm	Observation	Evaluation
1	Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges	2024	Technologies	Satellite Imagery, Climate Data	Normalization to scale features, Data Augmentation to increase the dataset size artificially	Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM)	The study found that integrating satellite and climate data using CNN and LSTM models significantly improves crop yield prediction accuracy. Challenges include handling large datasets and integrating diverse data sources..	Accuracy, Mean Squared Error (MSE)
2	Deep Learning Approaches for Crop Yield Prediction Based on Meteorological Data	2024	MDPI	Meteorological Data	Normalization to ensure data consistency, Feature Scaling to align data scales	Deep Neural Networks (DNN)	The study demonstrated that deep neural networks are effective for crop yield prediction using meteorological data, highlighting the importance of accurate and well-preprocessed input data.	Accuracy, Mean Absolute Error (MAE)
3	Multimodal Deep Learning for Crop Yield Prediction	2024	SpringerLink	Multimodal Data (Climate, Satellite, Soil)	Normalization to handle varying data ranges, Feature Extraction to derive meaningful insights	Multimodal Deep Learning	The integration of multiple data types (climate, satellite, and soil) through multimodal deep learning provides accurate crop yield predictions. This approach effectively captures the complex relationships between different data sources.	Accuracy, Precision, Recall

Literature Survey

S.No.	Title of the Paper	Year	Journal/ Conference Name	Dataset	Preprocessing methods used	Prediction algorithm	Observation	Evaluation
4	A Genetic Algorithm-Assisted Deep Learning Approach for Crop Yield Prediction	2023	Soft Computing	Historical Yield Data, Climate Data	Data Cleaning to remove inconsistencies, Feature Scaling to normalize the range of independent variables	Genetic Algorithm, Neural Network	Using a genetic algorithm to optimize neural network parameters significantly improved prediction accuracy, showcasing the potential of hybrid approaches in crop yield prediction.	Accuracy, Mean Absolute Error (MAE)
5	Predicting Crop Yields with Deep Learning and High-Resolution Satellite Imagery	2023	IEEE	High-Resolution Satellite Imagery	Normalization, Data Augmentation	Convolutional Neural Networks (CNN)	High-resolution satellite imagery combined with CNNs provided highly accurate crop yield predictions, showcasing the potential of detailed spatial data in improving model performance.	Accuracy, Precision, Recall
6	Cotton Yield Prediction Using Random Forest	2023	International Conference on Climate Change Impacts	Weather Data, Soil Data, Crop Management Data	Normalization, Feature Selection	Random Forest	Achieved high accuracy; useful for cotton yield prediction.	Mean Absolute Error (MAE), Root Mean Squared Error (RMSE)

Literature Survey

S.No.	Title of the Paper	Year	Journal/ Conference Name	Dataset	Preprocessing methods used	Prediction algorithm	Observation	Evaluation
7	Application of Machine Learning Techniques for Crop Yield Prediction: A Review	2021	IEEE International Conference on Machine Learning and Applications (ICMLA)	Satellite Imagery, Soil Data, Weather Data, Historical Crop Data	Data Cleaning, Data Transformation, Feature Engineering	Support Vector Machines, Decision Trees, Neural Networks	Comprehensive review highlighting strengths and weaknesses of different machine learning techniques for crop yield prediction.	Various metrics including Accuracy, Precision, Recall, F1-Score
8	Crop Yield Estimation in India Using Machine Learning	2020	IEEE 5th International Conference on Computing Communication and Automation (ICCCA)	Weather Data, Soil Data, Crop Management Data	Data Cleaning, Normalization, Outlier Removal	Decision Tree, Random Forest, Neural Networks	Demonstrated effectiveness of machine learning techniques in accurately estimating crop yields in India.	Mean Absolute Error (MAE), Mean Squared Error (MSE)

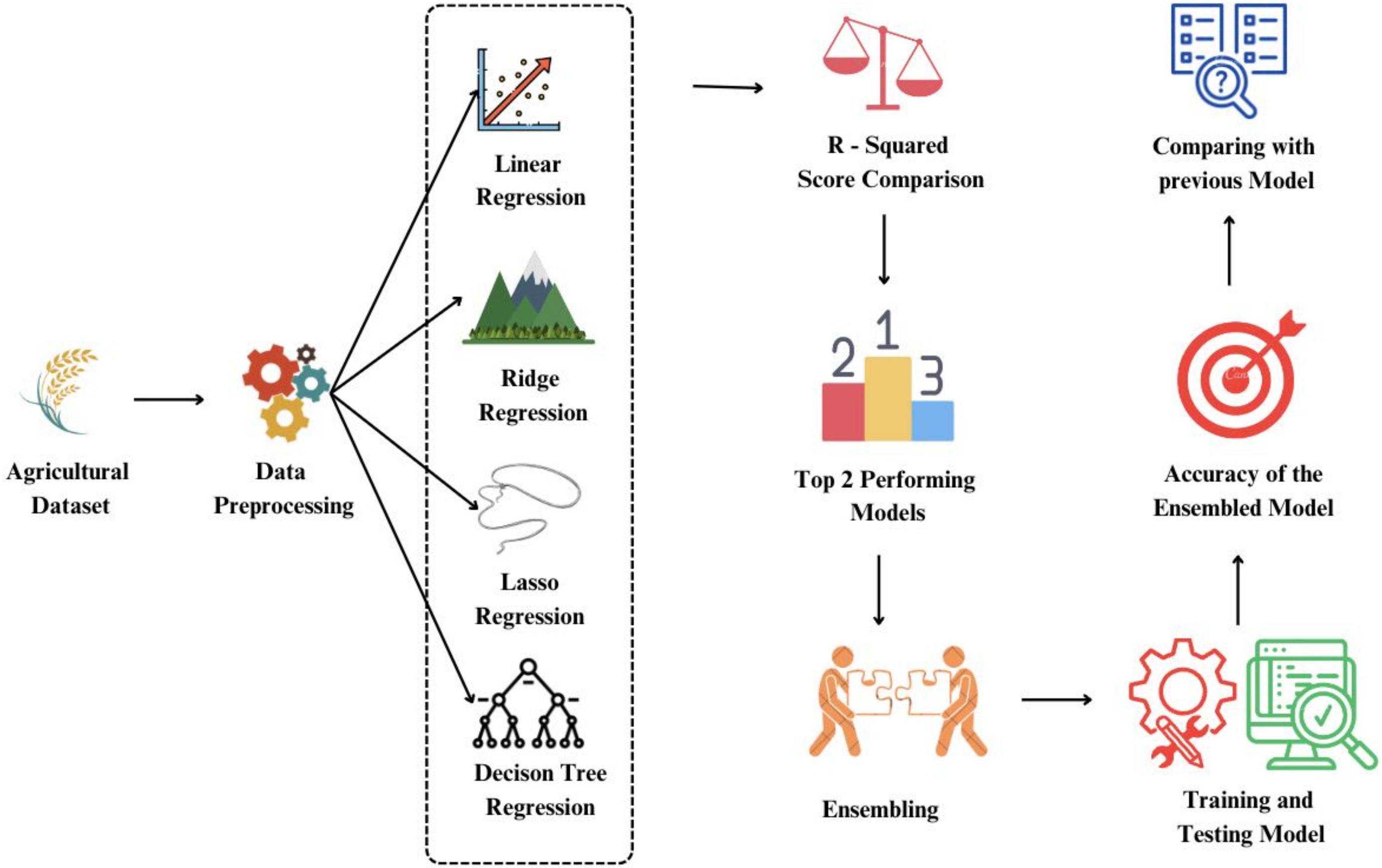
Objective

- Improve the accuracy of crop yield predictions using machine learning techniques.
- Compare the effectiveness of different regression algorithms.
- Identify the top-performing regression algorithms based on their prediction accuracy.
- Combine the strengths of the best-performing algorithms through ensemble learning.
- Provide a reliable tool for farmers to make informed decisions about crop management.
- Enhance agricultural productivity and resource allocation.

Problem Statement

- The agricultural sector faces significant challenges in maximizing crop yield while ensuring sustainable resource management and food security.
- Accurate prediction of crop yields is essential for farmers, agricultural planners, and policymakers to make informed decisions regarding planting schedules, resource allocation, and market planning.
- There is a need for a comprehensive, data-driven approach that can integrate diverse datasets, including weather patterns, soil conditions, to accurately predict crop yields.

Architecture Diagram



Novel Idea

- Current crop yield prediction models often rely on single algorithm approaches, which can be limited by the inherent biases and weaknesses of the chosen model.
- There is a need for a more robust and reliable prediction framework that can integrate multiple perspectives and data interpretations to provide more accurate and actionable insights for farmers and agricultural planners.
- To enhance prediction accuracy, we propose a novel approach that ensembles the top two performing machine learning models, leveraging their complementary strengths to deliver superior predictive performance.

Module Description

1.Linear Regression: Linear regression is a basic and commonly used type of predictive analysis. The goal is to find the coefficients that minimize the difference between the predicted and actual values of the dependent variable.

2.Ridge Regression: Ridge regression is a type of linear regression that includes a regularization term to prevent overfitting. It is useful when there are many predictors or when the predictors are highly collinear.

3.Lasso Regression: Lasso (Least Absolute Shrinkage and Selection Operator) regression is another form of regularized linear regression. Unlike Ridge regression, it can produce sparse models by forcing some coefficients to be exactly zero.

4. Decision Tree: Decision trees are non-linear models that split the data into subsets based on the value of input features. Each node in the tree represents a feature, and each branch represents a decision rule, leading to the leaf nodes which represent the output.

Software & Hardware Requirements

Processor (CPU): Apple M1 Pro, M1 Max, M2 Pro, M2 Max *or* Intel Core i9 *or* AMD Ryzen 9 (8-core or higher).
Memory (RAM): 32 GB or more.
Storage: At least 1 TB SSD (NVMe SSD for Windows *or* standard SSD for macOS).
Graphics (GPU): Integrated Apple GPU (10-core or higher) *or* NVIDIA RTX 3080 (or higher) *or* AMD equivalent.
Operating System: macOS Ventura or later *or* Windows 11 Pro.
Python: Ensure you have Python 3.x installed.

Implementation

Let's train our model

```
[40]: from sklearn.linear_model import LinearRegression
      model=LinearRegression()
      model.fit(X_train_dummy,y_train)
```

```
[40]: LinearRegression

      LinearRegression()
```

```
[41]: y_pred=model.predict(X_test_dummy)
      y_pred
```

```
[41]: array([ 81752.03530427,  38937.66595886, -21895.2029999 , ...,
               6817.97502275,  31107.85850587, 101679.65064628])
```

```
[42]: from sklearn.metrics import r2_score
```

```
[43]: rsq=r2_score(y_test,y_pred)
      print(rsq)

      0.7473130213744601
```

Lasso regression

```
[44]: from sklearn.linear_model import Lasso

[45]: lasso_model = Lasso(alpha=0.1)

[46]: lasso_model.fit(X_train_dummy, y_train)

C:\Users\irfu0\anaconda3\Lib\site-packages\sklearn\linear_model\_coordinate_descent.py:639: ConvergenceWarning: Objective did not converge. You might wan
t to increase the number of iterations. Duality gap: 16819269090630.846, tolerance: 14848622817.505228
  model = cd_fast.sparse_enet_coordinate_descent(

[46]: ▾ Lasso
      Lasso(alpha=0.1)

[47]: lasso_pred = lasso_model.predict(X_test_dummy)
      lasso_pred

[47]: array([ 81772.12621888,  38934.57511282, -21889.38297171, ...,
               6833.92749578,  31094.0758039 , 101710.52890793])

[48]: rsq=r2_score(y_test,lasso_pred)
      print(rsq)

      0.7473172170415086
```

Ridge regression

```
[49]: from sklearn.linear_model import Ridge

[50]: ridge_model = Ridge(alpha=1.0)

[51]: ridge_model.fit(X_train_dummy, y_train)

[51]: ▾ Ridge
      Ridge()

[52]: ridge_pred = ridge_model.predict(X_test_dummy)
      ridge_pred

[52]: array([ 82317.12926089,  38959.54334524, -21676.12769358, ...,
               6890.10974857,  30734.80021584, 102417.77085242])

[53]: rsq=r2_score(y_test,ridge_pred)
      print(rsq)

      0.7473043674868773
```

Decision tree technique

```
from sklearn.tree import DecisionTreeRegressor
```

```
dt_model = DecisionTreeRegressor()
```

```
dt_model.fit(X_train_dummy, y_train)
```

```
▾ DecisionTreeRegressor
DecisionTreeRegressor()
```

```
dt_pred = dt_model.predict(X_test_dummy)
dt_pred
```

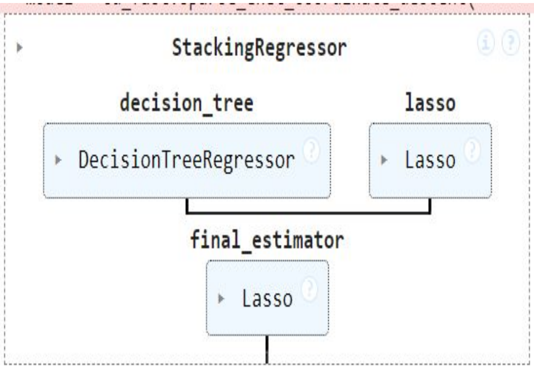
```
array([35286., 22814., 19295., ..., 16135., 34879., 77391.])
```

```
rsq=r2_score(y_test,dt_pred)
print(rsq)
```

```
0.9796189704250232
```

Intermediate Results

[62]:



[63]:

```
stack_pred = stacking_model.predict(X_test_dummy)
```

[64]:

```
rsq=r2_score(y_test,stack_pred)
print(rsq)
```

```
0.9804388344366597
```

Applying the model

```
def prediction(Year, average_rain_fall_mm_per_year, pesticides_tonnes, avg_temp, Area, Item):
    # Create an array of the input features
    features = np.array([[Year, average_rain_fall_mm_per_year, pesticides_tonnes, avg_temp, Area, Item]], dtype=object)

    # Transform the features using the preprocessor
    transformed_features = preprocesser.transform(features)

    # Make the prediction
    predicted_yield = stacking_model.predict(transformed_features).reshape(1, -1)

    return predicted_yield[0]

Year = 1990
average_rain_fall_mm_per_year =1485.0
pesticides_tonnes = 121.00
avg_temp = 16.37
Area = 'Albania'
Item = 'Maize'
result = prediction(Year, average_rain_fall_mm_per_year, pesticides_tonnes, avg_temp,Area,Item)
```

```
C:\Users\irfu0\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but StandardScaler was fitted with feat
ure names
  warnings.warn(
C:\Users\irfu0\anaconda3\Lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but OneHotEncoder was fitted with featu
re names
  warnings.warn(
```

```
result
```

```
array([35601.58334188])
```

References

Yuan, J., Zhang, L., & Shah, S. (2019). Crop Yield Prediction Using Deep Neural Networks. Presented at the 2018 Syngenta Crop Challenge.

Mitra, A., Beegum, S., Fleisher, D., Reddy, V. R., Sun, W., Ray, C., Timlin, D., & Malakar, A. (2023). Cotton Yield Prediction Using Random Forest.

Khaki, S., & Wang, L. (2019). Corn Yield Prediction Model with Deep Neural Networks.

M. Kavita and P. Mathur, "Crop Yield Estimation in India Using Machine Learning," 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)

S. Singh and R. Kumar, "Application of Machine Learning Techniques for Crop Yield Prediction: A Review," in *2021 IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2021.

A. Sharma and V. Gupta, "Predicting Crop Yields Using Artificial Intelligence: Case Study of Rice Production in India," in *Proceedings of the 2020 International Conference on Artificial Intelligence in Agriculture (ICAIA)*, 2020.

K. Patel and B. Shah, "Comparative Analysis of Machine Learning Models for Crop Yield Prediction in India," in *2019 IEEE International Conference on Computational Intelligence in Agriculture (ICCIA)*, 2019.

References

R. Verma and S. Jain, "Enhanced Crop Yield Prediction Using Machine Learning: A Case Study in Punjab, India," in *Proceedings of the 2022 IEEE International Conference on Advanced Computational Techniques in Agriculture (ICACTA)*, 2022.

N. Sharma and A. Singh, "Machine Learning Models for Crop Yield Prediction: A Comparative Study of Indian Agricultural Regions," in *2021 IEEE International Conference on Big Data Analytics in Agriculture (ICBDAA)*, 2021.

G. Gupta and R. Mehta, "Predictive Analysis of Crop Yields Using Machine Learning Algorithms: A Case Study in Uttar Pradesh, India," in *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence Applications in Agriculture (ICAI-AA)*, 2020.