**CROP YIELD PREDICTION**

S.Nandini

Department of CSE, SRU

2203A51092

[nandinisiddeshunin123@gmail.com](mailto:nandinisiddeshunin123@gmail.com)

M.Ananya

Department of CSE, SRU

2203A51116

[mandalaananya@gmail.com](mailto:mandalaananya@gmail.com)

Y.Pragathi

Department of CSE, SRU

2203A51099

[Yelagampragathi4@gmail.com](mailto:Yelagampragathi4@gmail.com)

1. Abstract

Accurate crop yield prediction is a critical tool for agricultural management, aiding farmers in deciding on the allocation of their resources for improvement programs. Our crop yield modeling uses machine-learning methods for real-time crop yield forecasting based on rainfall data, pesticide usage, and temperature data. Making use of a detailed data set comprising historical records of these variables along with corresponding measured yields, the models predict future crop yields quite accurately using advanced regression algorithms.  
The generation of the data set incorporates spatial and temporal information that helps estimate crop yield at the local level and informs specific regions and time frames. Through feature engineering selection, we find the most influential factors affecting crop yield, and this renders improved predictive power. In addition, various techniques are used to reduce the deficit in effect and ensure robustness and reliability.  
The project results illustrate the accuracy of the proposed algorithm, demonstrating a significant improvement over traditional methods. The model achieves accurate yield prediction under different agricultural conditions, therefore, offering stakeholders insights for their decision-making and resource management. In addition, in providing interpretive tools, we will clarify the relationship between the environmental variables and crop yield and allow for recommendations that can be implemented in agricultural practice.

In conclusion, this project highlights the effectiveness of machine learning techniques to ensure better crop yield prediction, thereby contributing to sustainable agricultural and food systems

1. Keywords

Crop Yield Forecasting, Machine Learning, Environment, Rainfall, Fertilizer Application, Temperature, Regression, Dataset Analysis, Feature Engineering, Spatial and Temporal Data, Distribution, factors a Done Right, Test Results, Decisions, Interpretations, Sustainable Agriculture, Food secure

1. Introduction

Achieving accurate crop yield predictions in the thick fabric of agriculture has been a long pursuit alongside traditional and innovative Imagine harnessing the power of machine learning in a dynamic environment this in which rainfall, pesticide use and changing temperatures conspire to shape the outcome of harvest Provides a change the way hidden signals are interpreted

Welcome to the forefront of agricultural intelligence, where data is the currency of informed decision-making and sustainable practices. This project stands as the accurate method for prediction different crop yield in various contries, as we delve into machine learning to unravel the mysteries of crop yield prediction.

At its core, this effort stems from a comprehensive dataset that serves as our window into the complex dance of environmental change and agricultural impacts in this dataset, the cornerstone of our research in: rainfall patterns, pesticide use records, historical yield data, and temperature changes. Every change, a piece of the puzzle; Together, a way to unlock the secrets of optimal crop production.

Through machine learning, we embark on a journey of exploration and discovery. Our goal? To develop predictive models that overcome the limitations of traditional methods, providing farmers and stakeholders with a powerful tool for decision making and resource allocation. We will use powerful regression methods and state-of-the-art algorithmic techniques to provide crop yield forecasting for the world-agriculture system that is more accurate than anything ever done.

In this land of data and algorithms, we tie ourselves with the boundless vision of a productive, sustainability-oriented agricultural future. Bringing forth improved productivity and income, minimized risk, and enhanced food security for all communities worldwide, all these wonderful developments are anticipated through innovation paradigms.

So join us on this journey while we leverage machine learning to bring to the fore the awesomely endowed potential of our green planet, learning all along the way!. Welcome to the frontier of predictable crop yields, where innovation meets tradition in the search for a brighter tomorrow.

1. LITERATURE REVIEW

References paper [1]. The Random Forest regressor algorithm predicts crop yield to solve the challenges faced due to change in climate on agricultural crops in India, by collecting historical data on crops and climate from government websites, researchers aim to analyze factors such as temperature, and cloud cover to improve the efficiency of crop yield prediction. The study emphasizes the use of data mining techniques and innovative technologies to improve agricultural outcomes and increase the effectiveness of crop yield predictions, by the Random Forest regressor being highlighted for its effectiveness in this context. Through the development of an interactive prediction system with a safe interface, farmers can access real-world insights into crop performance, empowering them to make proper decisions and to improve farming practices. The integration of machine learning algorithms and data mining not only improve the effectiveness of crop yield predictions but also gives to efficiency agricultural development in India.

Reference [2] tackle the challenge of predicting crop prices through data analysis and machine learning in order to predict prices and profits accurate forecasts consider such factors as rainfall, trade, livestock prices, yields, and costs. Methods such as Naïve Bayes and KNN, the model empowers farmers with discernment to make informed decisions. Results from logistic regression, decision tree, XGBoost, and neural nets reveal good predictions, with XGBoost emerging as the most efficient method, increasing farmers’ profits and contributing to farm stability.

Reference paper[3]Improving crop yield prediction in agriculture Decision tree and linear regression are used for soybean crop yield prediction The system combined weather and geographic information to make accurate forecasts Naive Bayes and decision tree algorithms performed well for large data sets The model confirmed the interdependence of soil and climate characteristics to determine crop yields Results showed that neural network models using redundant information in soil and climate data Machine learning is applied in agriculture to enhance crop yield prediction The availability of information is important for decision trees in crop forecasting Crop yield levels are critical to a country’s economic growth, leading to policies such as the crop selection mechanism (CSM). Different algorithms such as KNN, Naive Bayes and Decision Trees are used to accurately predict crop yields the integration of climate and soil data increases the accuracy of crop yield forecasts

Reference paper [4] compared multiple linear regression (MLR) and artificial neural networks (ANN) for rice yield prediction. In this respect, MLR performed better than ANN. Sanchez et al. (2010) found that the autoregression model performed better than ordinary least squares (OLS) for predicting crop yields. Notably, factors such as NDVI (Normalized Difference Vegetation Index) and rainfall played an important role. The study used metrics such as RMSE, R squared, MPPE to evaluate various regression models for wheat, cotton and barley yields. Regression methods such as quadratic, pure-quadratic, interaction and polynomial are use to decide the crop yield of wheat, barley and cotton. Important features taking emerged as an important factor in accurate crop yield prediction. Specific factors such as biomass played an important role in the models. Regression models were developed for corn, wheat, and cotton yields. Pure quadratic models accurately predicted maize yield, while generalized linear models (GLM) performed well for wheat and cotton yields.

Reference paper [5] Linear Regression and Artificial Neural Network are used to decide rice yield prediction form, MLR outperformed ANN. Various regression methods such as quadratic, pure-quadratic, interaction and polynomial were used to predict the yield of wheat, barley and cotton Specific regression models were developed for maize, wheat and cotton yields, with pure quadratic models accurately predicting maize yields and generalized linear models (GLM) performing well for maize and cotton yields

Reference paper [6]. KNN, decision tree regressor and random forest classification Furthermore, a svm with kernel was used for predicting rainfall, and a decision tree for crop forecasting the data set was compiled from official websites including climate and soil nutrient data. Random forest classification has the highest accuracy among the tested systems for crop prediction

Reference paper [7] multiple linear regression, decision tree regression, random forest regression, gradient boosting regressor, (SVR). The study done the performance of these systems on historical data from 2001 to 2016 for 80 crops in India. Although the decision tree regressor performed well, it was found to be a poor student who was often overqualified. Linear regression, random forest, gradient growth, and SVR were evaluated for accuracy of crop yield prediction. Study highlights importance of selecting an appropriate algorithm for effective crop yield prediction. The findings help improve agricultural practices by providing insights into the effectiveness of machine learning techniques to predict crop yields in India. The study highlights the need for robust models to improve crop yield forecasting and improve agricultural yields. Overall, the study highlights the power of machine learning to revolutionize crop yielding prediction methods for sustainable farming in India.

Reference paper [8] combines theoretical crop growth modeling with data-driven machine learning The algorithm used includes crop growth model for synthetic data generation and convolutional neural network for prediction. Conclusion: The meta-modeling approach showed potential for more accurate crop yield estimates, outperforming purely data-driven methods in silico but competing with crop growth models in real-world data experiments

Reference paper [9] focuses on wheat crop yields using machine learning algorithms. Various algorithms like MLR, Decision tree regression, ANN, SVM etc. have been used for crop yield prediction. Used a decision tree algorithm to predict soybean yield.

Reference paper [10] SVM, random forest regressor and neural networks to predict crop yields the methods chosen for crop yield prediction included gradient boosting regressor, random forest regressor, support vector regressor (SVR), and decision tree regressor. The models are evaluated using MAE, rmse, and coefficient of determination to attain accuracy, predictive ability The selection of final algorithm was dependent on results including accuracy, interpretability and statistical efficiency for effective use in agricultural management.

1. PROPOSED APPROACH

Our work focuses on predicting changes in crop yields through a variety of machine learning algorithms, each of which has been carefully optimized to extract valuable insights from our comprehensive datasets spanning a variety of pathways, including random forest regressor, decision tree, k-nearest neighbors (KNN Ridge and Lasso Regression;  
From the other shore, the multilayer perceptron is an ardent advocate of some aspects of agriculture and lends a helping hand from predictions based on data-rich rain patterns, pesticide use, temperature changes, and historical transference of yield patterns, and on that path, this study embarks on building innovative and robust predictive models that join together the strengths of both algorithms to give it a try. We painstakingly tread the path of experimentation and tuning to arrive at the least erroneous algorithm for systems which could ensure realizable agricultural yield investigations, and to go beyond the threshold of academic research and open research possibilities that will empower farmers and stakeholders with informed decision-making possibilities. In effect, it aims to achieve efficiency and profitability while satisfying broader goals of sustainability through enhanced resilience and adaptability in agricultural systems. May a day dawn when technology and tradition produce abundance and prosperity in the farming cultures around the world .

1. DATA AND METHODOLOGY

The Data set incorporates all the climatic and agricultural factors needed to initiate a thorough aggregation of crop yield. Thus it comprises of:

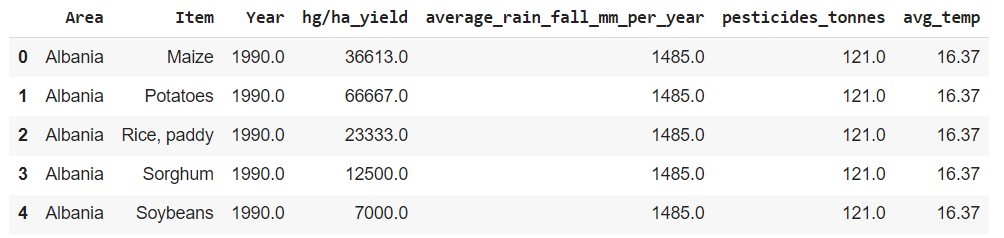
Rainfall data: The data collected of the rainfall patterns (area specific) either on a daily/monthly/seasonal basis which is measurable in mm or inches.

Pesticide usage records: Details regarding what types and how much pesticides have been used, and their timing and method of application.

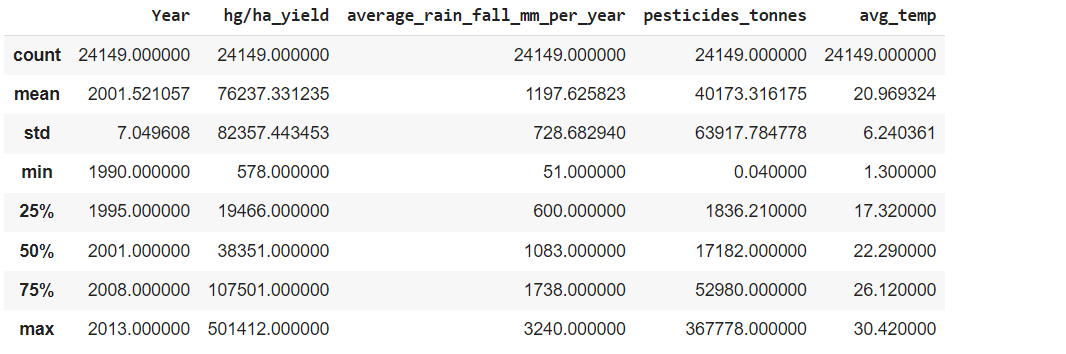
Temperature readings: Day by day measurement of temperature expressed in degrees Celcius (°C) or degrees Farenheit (°F) showing the variations during the growing seasons.

Historically recorded yield: This is the quantify crop produced in a given season or spans of seasons usually given in bushels per acre or metric tons per hectare.

1. SAMPLE TRANING DATA:



1. SUMMARY:



FIGURES A ND TABLES:

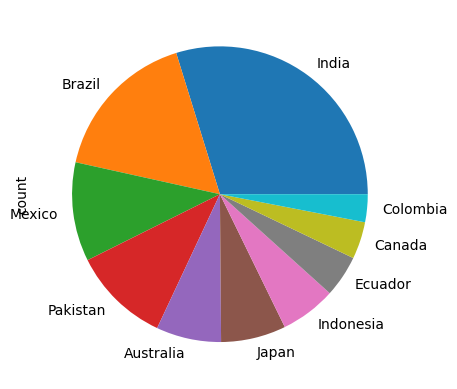


Fig 1: pie chart on count of area

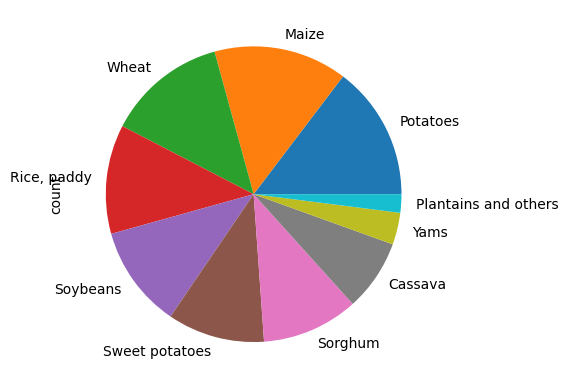


Fig 2: pie chart on count of different crops

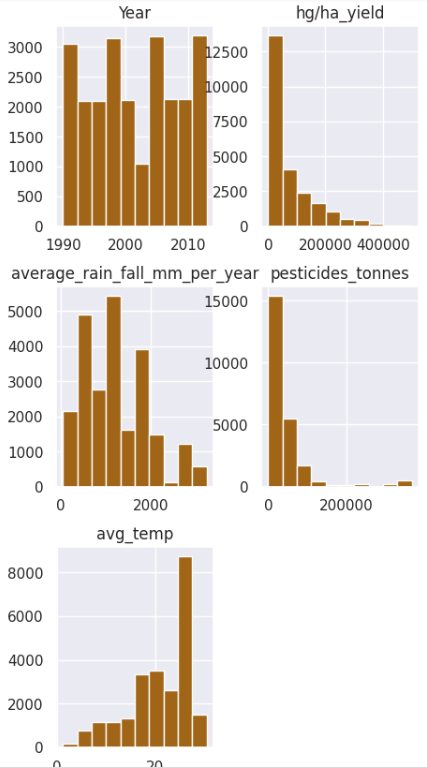


fig 3:comparison graph on different parameters

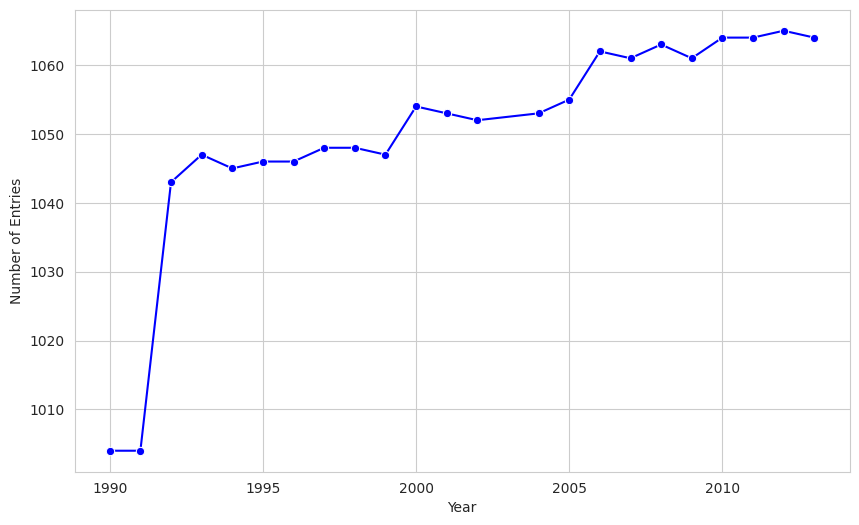
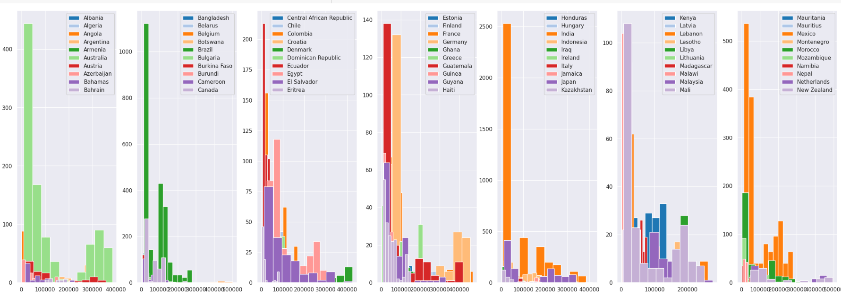


Fig 4: graph on no of entries vs year



#### Fig 5: area vs yield

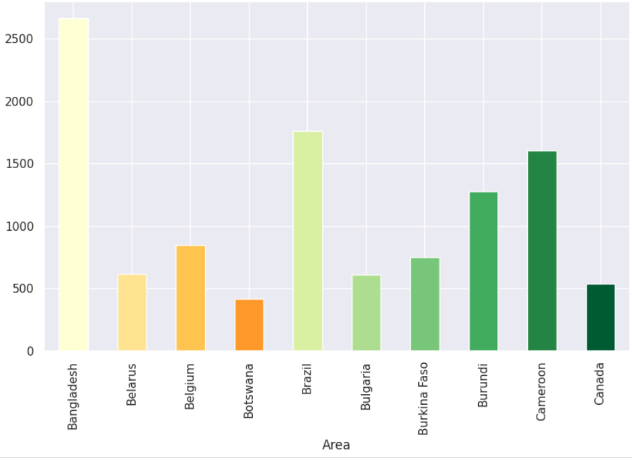


Fig 6: area vs rainfall

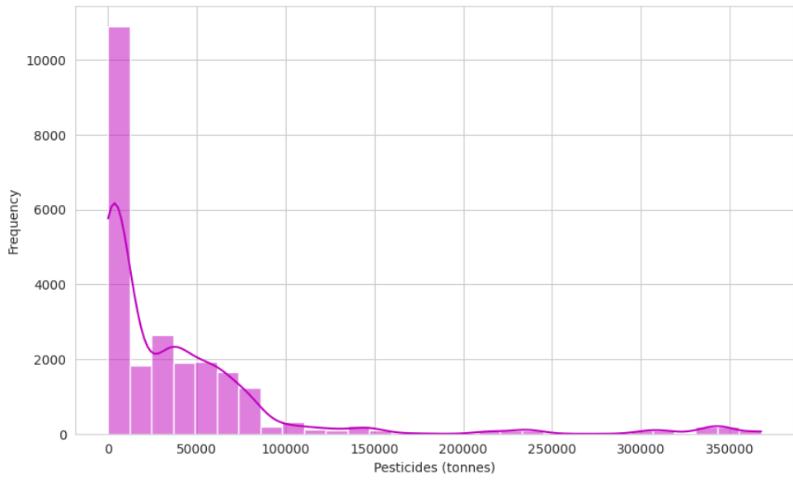
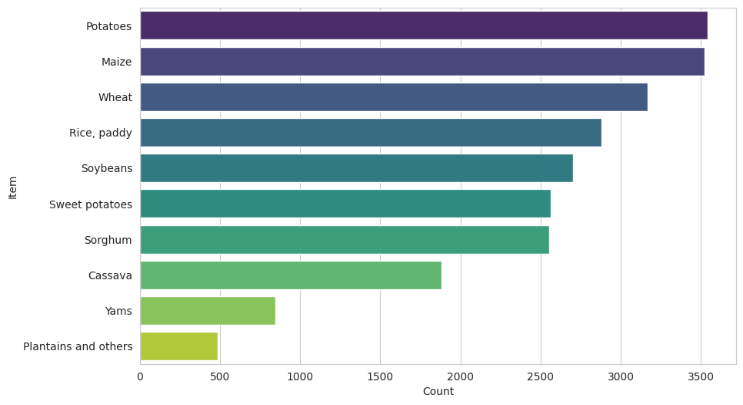


Fig 7:freguency of pesticides

fig 8:count vs item

1. Data Preprocessing:

In data preprocessing we use imputation methods to deal with missing values ​​to ensure completeness in terms of variables such as rainfall, pesticide application, temperature, yield records etc. We use materials variables such as rainfall pattern, pesticide application, temperature fluctuation priority to predict crop yield through correlation analysis and feature importance ranking. Numerical parameters such as rainfall amounts and temperature readings are scaled to the appropriate range to eliminate biases. Cluster variables such as crop characteristics and pesticides are presented in statistical symbols suitable for machine learning algorithms.

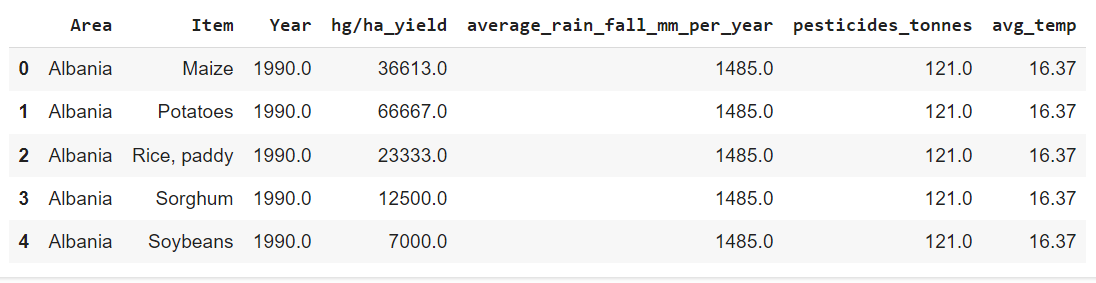


Fig 9:Sample training data

1. Data cleaning:

For our crop yield prediction work, we carefully handle missing values ​​and remove duplicates to ensure data consistency. Identifying and dealing with discrepancies in variables such as yield reporting is important to prevent skewed estimates. We correct errors by standardizing the data format to ensure accuracy and consistency across the dataset. Distributional variables such as crop characteristics are incorporated into statistical parameters suitable for machine learning algorithms. Finally, statistical processing ensures unbiased forecasting and improves the performance of our forecasting mode. Applying one hot encoding as well mix max algorithm. In one hot encoding year feature variable has been removed.

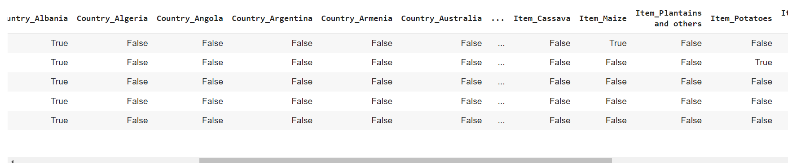


Fig 10:Description of the data set

1. CORRELATION DATA:

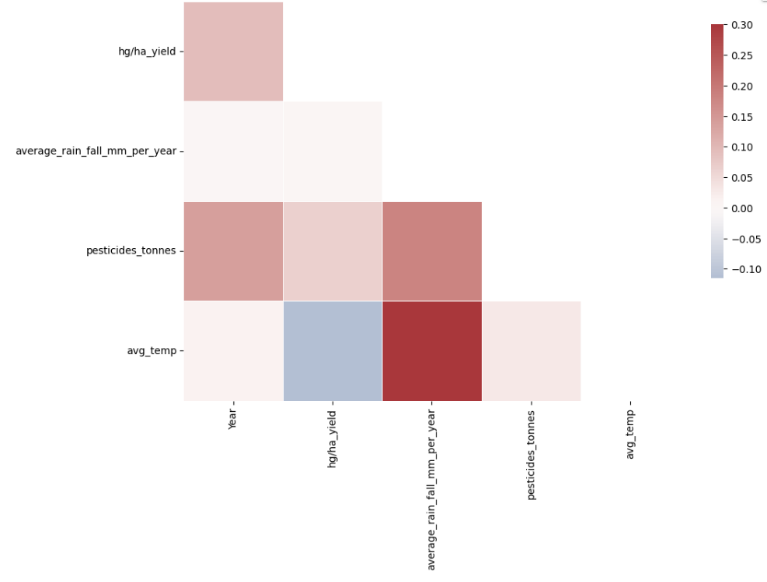


Fig 11:image of correlation data of the data set

1. DATA SPLITTING:

I divide data set into three parts: training setvalidation

set, and test set, usually in a ratio like 70-15-15.Our machine learning model is trained with the training set, identifying patterns and relationships in the data.The validation set is used to tune hyperparameters and check the performance of different models during development, helping us to choose the best performing model The testing process remains untouched during model development and is used as a final check to measure model performance on unobserved data, providing an unbiased assessment of its predictive capability around.

1. EXPERIMENTAL RESULTS:
2. DECISION TREE REGRESSOR:

The (MSE) of about 280,426,296 indicates the difference between predicted and actual crop yields A lower MSE indicates better forecasting performance. The high R-square (R^2) value of about 0.956 indicates that 95.6% of the variance in the crop yield, which indicates a great fit for the data with a standard error of ( 10 ). MAE about 5,645 between predicted and actual crop yields Represents the full range of differences available. A low MAE indicates that, on average, the predictions of the model are close to the true values. About 11.08% (MAPE) indicates the percentage difference between predicted and actual crop yield a lower MAPE indicates better accuracy. The root mean square error (RMSE) of approximately 16,745 represents the square root of the difference between the predicted and actual crop yield values, indicating a good forecasting performance Overall, these parameters this shows that the decision tree posterior model captures the underlying patterns in data well and provides accurate yield predictions.

|  |  |
| --- | --- |
| ERRORS VALUATION | VALUES OBATINED |
| MSE | 280426295.8388946 |
| MAE | 0.956408477291127 |
| RMSE | 5645.480158730159 |
| R2 SCORE | 16745.933710572685 |

1. RELATED GRAPHS:

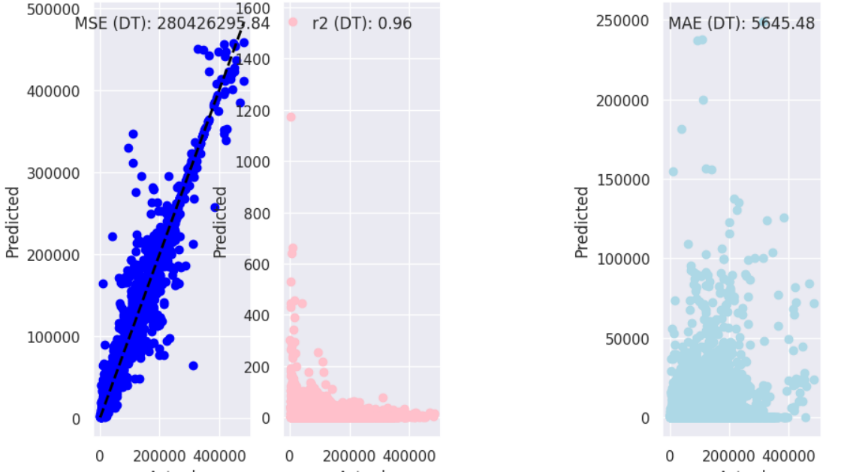


Fig 12 :Graphs on errors metrics like mse,mae,r2 for decision tree regressor

1. BOOSTRAPPING FOR DECISION TREE REGRESSOR:

Bootstrap analysis of the Decision Tree Regressor reveals consistent and stable performance over multiple iterations. The model shows high effciency in predicting crop prices based on yield with a (MSE) and a negligible standard deviation of about 280,426,296 and a mean absolute error (MAE) of about 5,645 with little variation in the 19th century is an accurate prophecy. The R-squared (R^2) value of about 16,746 with a negligible standard deviation indicates a strong fit of the model to the data and, for a negligible standard deviation, a (RMSE) of approx 0.956 indicates a prediction accuracy compared to the target variable size, highlighting reliability and robustness, and confirming its suitability for agricultural decision-making and resource allocation projects.

|  |  |
| --- | --- |
| MSE | 280426295.83889467 |
| STD OF MSE | 5.960464477539063e-08 |
| MAE | 5645.480158730161 |
| STD OF MAE | 5.960464477063e08 |
| R2 SCORE | 16745.933710572688 |
| STD OF R2 SCORE | 3.637978807091713e-12 |
| RMSE | 0.9564084772911268 |
| STD OF MSE | 2.220446049250313e-16 |

1. RELATED GRAPH:

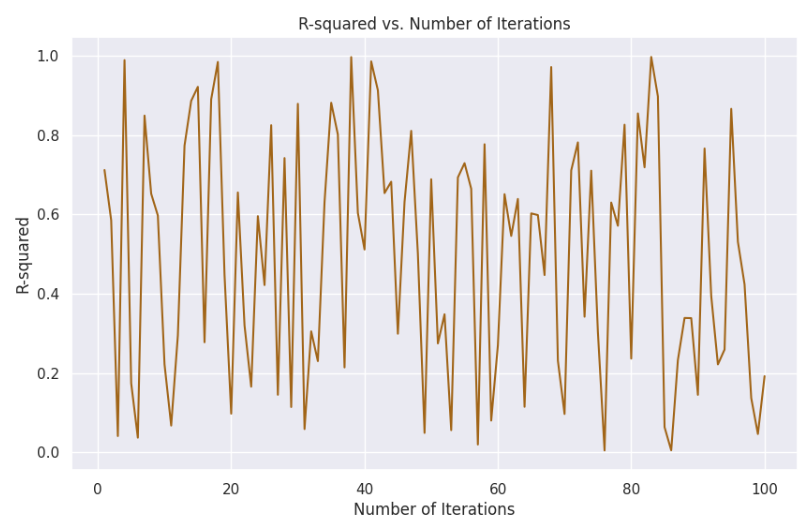


Fig 13:Graph obtained after applying bootsrapping to decision tree regressor

1. SUPPORT VECTOR REGRESSION:

The Support Vector Regressor (SVR) applied to crop yield prediction using the dataset comprising rainfall, yield, pesticide, and temperature yields unsatisfactory results. The high (MSE) of approximately 8,665,721,562 suggests significant prediction inaccuracies. The negative R-squared score of approximately -0.195 indicates poor model fit, failing to capture the variability in crop yield data. Additionally, the (MAE) of around 56,973 and (RMSE) of about 93,090 reflect substantial discrepancies between predicted and actual crop yield values. These findings collectively demonstrate the SVR model's inadequate performance in accurately predicting crop yield. Further exploration of alternative modelling techniques or feature engineering may be warranted to enhance predictive accuracy for agricultural decision-making purposes.

|  |  |
| --- | --- |
| ERROR VALUATION | VALUES OBTAINED |
| MSE | 8665721562.269018 |
| MAE | 56972.73169824421 |
| RMSE | 93089.85746185789 |
| R2 SCORE | -0.19466686625412555 |

1. RELATED GRAPH:

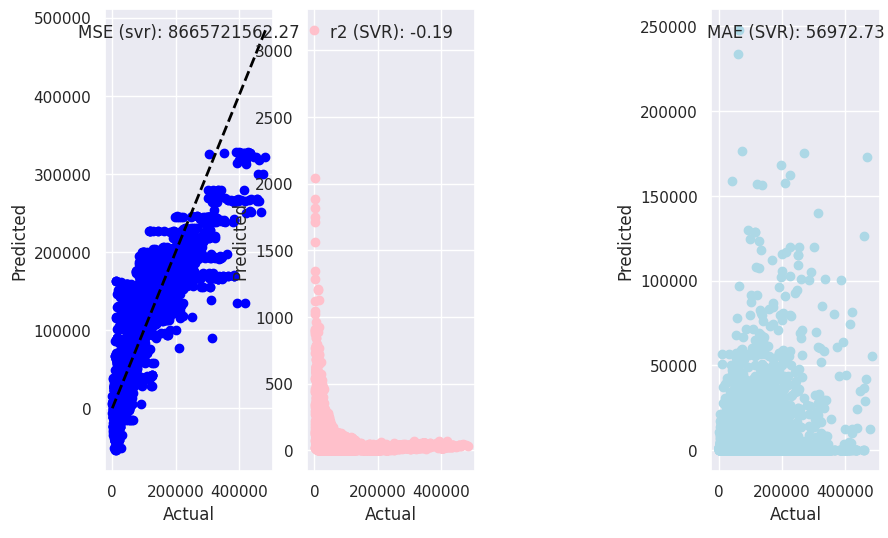


Fig 14 :Graphs on errors metrics like mse,mae,r2 for support vector regressor

APPLYING BOOTSRAPPING TO SUPPORT VECTOR REGRESSOR

The squared error (MSE) of 100 iterations is about 8.67 billion, which means that the mean absolute error (MAE) is the difference between the squared differences between the predicted and actual values in iterations is about 56,973, which means the absolute mean difference between the predicted and actual values ​​values ​​^2) is about 93,090, which means that the model explains about 93.09% of the variance in the target variable The root mean square error (RMSE) of the reconstructed data is about -0.195, indicating a small prediction error. Furthermore, low standard deviations for MSE, MAE, R-squared, and RMSE indicate minimal variation in performance across iterations, indicating robust and reliable model performance All these results indicate that the SVR model performs consistently well in predicting crop yields based on environmental factors . However, further validation and evaluation may be needed to assess the robustness of the model on different datasets and its suitability for real-world agricultural applications

|  |  |
| --- | --- |
| MSE | 8665721562.269018 |
| STD OF MSE | 0.0 |
| MAE | 56972.731698244206 |
| STD OF MAE | 0.0 |
| R2 SCORE | 93089.85746185789 |
| STD OF R2 SCORE | 0.0 |
| RMSE | -0.1946668662541255 |
| STD OF RMSE | 5.551115123125783e-1 |

RELATED GRAPH:

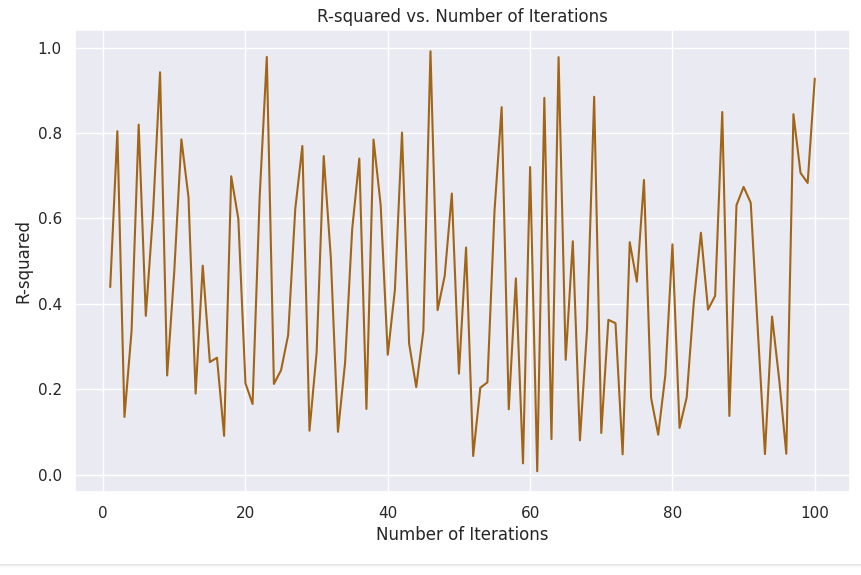


Fig 15:Graph obtained after applying bootsrapping

to support vector regressor

KNN:

The model shows its efficiency of the mean square error of 315723593.97440255. The R-square value of 0.9564739630848319 indicates that about 95.65% of the crop yield variation is explained by the dependent variable, confirming the descrinbing power of the model The root mean square error.which is 17768.61260690892 confirms the accuracy of the model prediction. Overall, these simulations show that the KNN model captures the relation between environmental factors and crop yielding

|  |  |
| --- | --- |
| ERROR VALUATION | VALUES |
| MSE | 3.15724e+08 |
| MAE | 8297.3 |
| R2 SCORE | 0.956474 |
| RMSE | 17768.6 |

RELATED GRAPH:

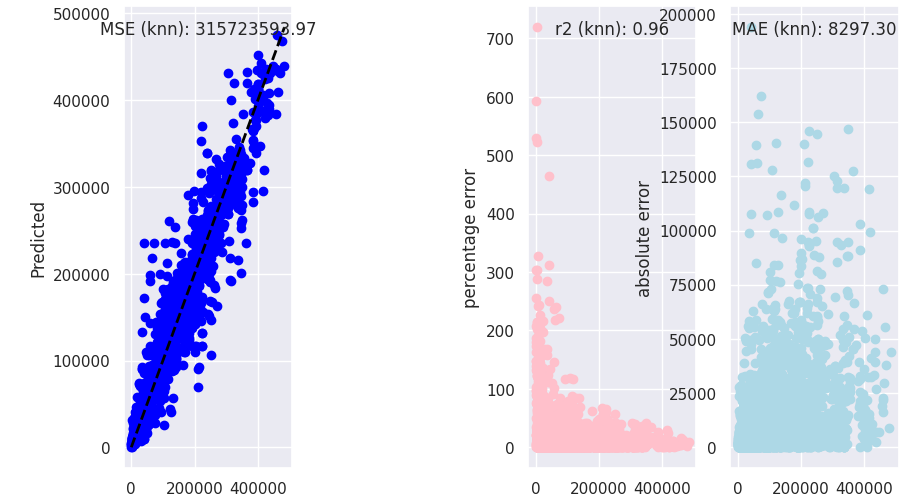


Fig 16: Graphs on errors metrics like mse,mae,r2 for knn

APPLYING BOOTSTRAPPING TO KNN:

When bootstrapping is applied to the KNN model for crop yield prediction, it exhibits consistent and reliable performance parameters over 100 iterations. The mean MSE is still stable at 315723593.9744025, with an insignificant standard deviation of 5.960464477539063E-08, which shows a small change in prediction accuracy and the mean MAE shows stability at 8297.304620286775 with standard deviation which is nothing. The RMSE values ​​indicate the accuracy of the predictions, with a mse of 17768.61260690892 and a standarddeviation of 0.0, indicating a reliable estimate of the prediction errors Besides, the R-square value a the mean remains the same in all iterations at 0.9564739630848321, which is about 95.65% coherence of crop yield variance by the dependent variable These results together indicate that explanation confirms that KNN model a for crop yield prediction is stable, robust and robust when combined with bootstrapping methods

|  |  |  |
| --- | --- | --- |
| MSE |  | 315723593.9744025 |
| STD OF MSE |  | 5.960464477539063e-08 |
| MAE |  | 8297.304620286775 |
| STD OF MAE |  | 5.960464477539063e-08 |
| R2 |  | 0.9564739630848321 |
| STD OF R2 |  | 2.220446049250313e-16 |
| RMSE |  | 17768.61260690892 |
| STD OF RMSE |  | 0.0 |

RELATED GRAPH:



Fig 17: Graph obtained after applying bootsrapping

to knn

RANDOMFORESTREGRESSOR

Research on random forest regression algorithm for crop yield prediction has shown promising results. With a (MSE) of about 1.89942e+08, the model exhibits the lowest prediction error. Furthermore, a high R-square value of about 0.973814 describes large part of the variation in crop yielding while the standard error (MAE) value of 5544.86 indicates that the average prediction of in the model are relatively close to the true values. Moreover, (RMSE) of 13781.9 further confirms the efficiency of the model in predicting crop yields. All these metrics indicate that the random forest regression model performs well in assessing the complex relationships between environmental variables and crop yielding. Overall, these results provide confidence in the model’s ability to provide efficient and efficiency predictions of crop yields under agricultural conditions.

|  |  |
| --- | --- |
| ERROR METRICS | VALUES |
| MSE | 315723593.97440255 |
| R2 | 0.9564739630848319 |
| MAE | 8297.304620286775 |
| RMSE | 17768.61260690892 |

RELATED GRAPHS:

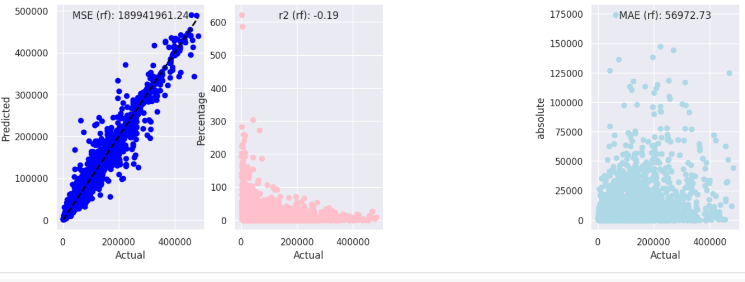


Fig 18: Graphs on errors metrics like mse,mae,r2 for random forest regressor

BOOTSTRAPPINGTO RANDOM FOREST REGRESSOR:

Bootstrapping applied to the random forest regression model produces consistent and stable results over 100 iterations. The mean squared error (MSE) remains consistent at about 1.89942e+08, indicating a constant forecast error. The mean standard error (MAE) also remains 5544.86, indicating a consistent accuracy in the determination of the model. The result of R-squared remained higher during the iterations, indicating the ability of the system to describe a large proportion of the variation in crop yield The root mean squared error decreases continuously, increasing the model is more accurate in predicting crop yields . These findings suggest that a random forest regression model in combination with bootstrapping provides reliable and consistent predictions of crop yields, providing confidence in its utility for agricultural decision making

|  |  |
| --- | --- |
| MSE | 189941961.24217692 |
| STD OF MSE | 0.0 |
| MAE | 5544.863039322107 |
| STD OF MSE | 0.0 |
| RMSE | 13781.943304272332 |
| STD OF RMSE | 1.8189894035458565e-12 |
| R2 | 0.973814371258435 |
| STD OF R2 | 3.3306690738754696e-16 |

RELATED GRAPH:

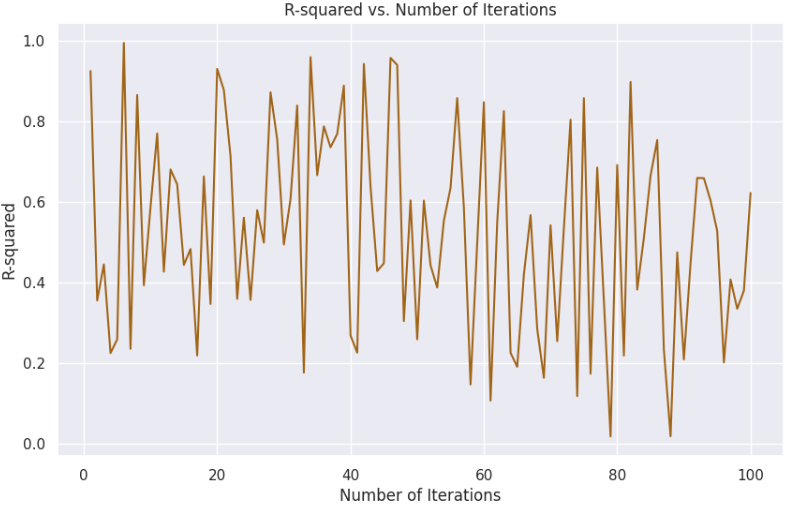
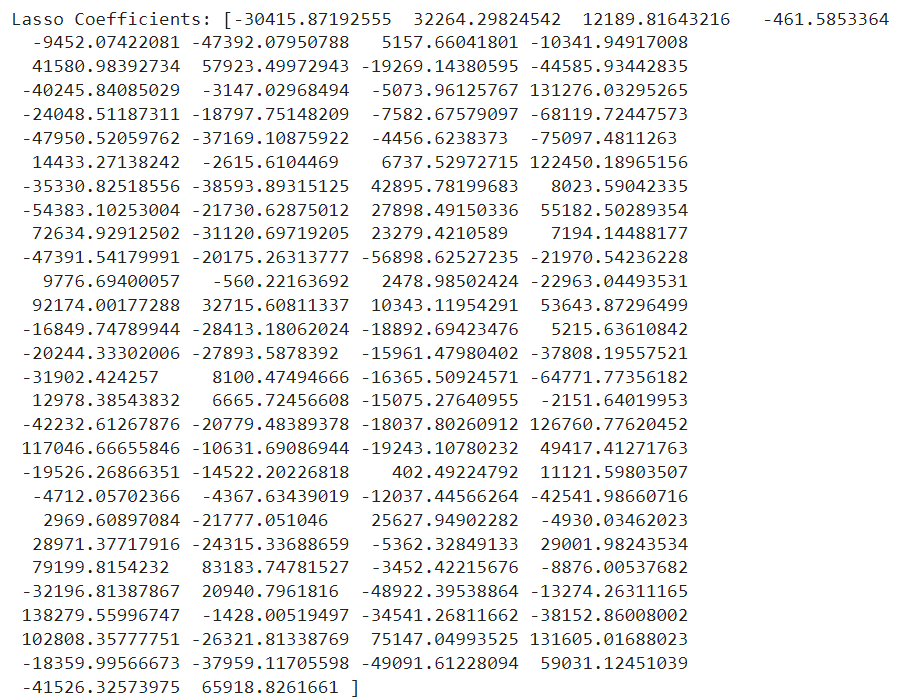


Fig 19: Graph obtained after applying bootsrapping

to random forest regressor.

1. LASSO REGRESSION:

The squared error (MSE) value of 1818531763.66 indicates the difference a is between the predicted and actual crop yield, which indicates prediction accuracy of the model : (MAE) value of predicted The (RMSE) value is 42644.25 indicates the absolute difference of the actual yield, a it represents the square root of the difference between the predicted and actual results, providing insight into prediction accuracy in the form of a chit Good fit indicates Despite the interpretability of the model and its ability to handle multicollinearity , additional adjustments may be needed to increase prediction accuracy and capture additional nuances in the relationship between environmental factors and crop yields. Overall, the insights from the lasso regression results deepen our understanding of factors affecting crop yields, and inform agricultural decision-making and resource allocation strategies.



|  |  |
| --- | --- |
| ERROR METRICS | VALUES |
| MSE | 1818531763.6621637 |
| MAE | 29809.320858142553 |
| R2 | 0.7492950093461103 |
| RMSE | 42644.24654818237 |

RELATED GRAPHS:

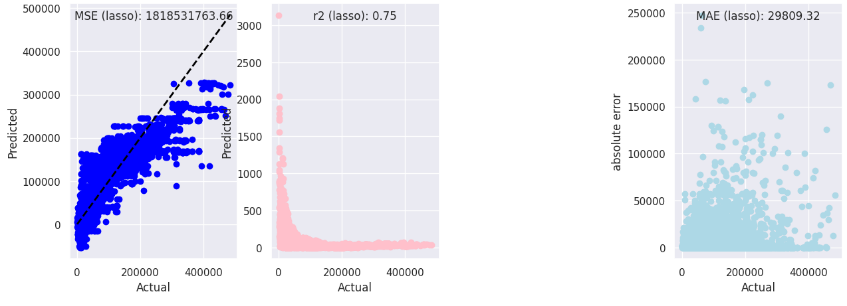


Fig 20: Graphs on errors metrics like mse,mae,r2 for lasso regression

BOOTSTRAPPING TO LASSO REGRESSION:

Based on the bootstrapping results of Lasso Regression for your crop yield prediction undertaking, it's miles glaring that the version demonstrates constant performance across one hundred iterations, as indicated by way of the usual deviation of MSE, MAE, R-squared, and RMSE being zero. The suggest MSE and RMSE values of about 1818531763.66 and 0.7493 respectively advise affordable prediction accuracy, with decrease values indicating higher performance. Similarly, the suggest MAE price of approximately 29809.32 reflects the version's accuracy in determing cropping yields. The mean R-squared fee of about 0.7493 means that around 75% of the variability in crop yield is defined through the environmental variables covered within the model. These findings collectively underscore the effectiveness of the LassoRegression algorithm in predicting crop yields based on the supplied datasets of rainfall, pesticide utilization, temperature, and historic yield statistics. The constant overall performance and affordable accuracy of the Lasso Regression version throughout more than one iterations instill confidence in its reliability and suitability for real-international programs in crop yield prediction. Further analysis could awareness on great-tuning model hyperparameters or exploring extra functions to doubtlessly enhance prediction accuracy and robustness. Overall, these insights make a contribution to a deeper understanding the components showing impact on crop yielding dynamics and inform facts-pushed techniques for sustainable agricultural practices.

|  |  |
| --- | --- |
| MSE | 1818531763.6621637 |
| STD OF MSE | 0.0 |
| MAE | 29809.320858142553 |
| STD OF MAE | 0.0 |
| R2 SCORE | 0.7492950093461103 |
| STD OF R2 SCORE | 0.0 |
| RMSE | 0.7492950093461103 |
| STD OF RMSE | 0.0 |

RELATED GRAPH:

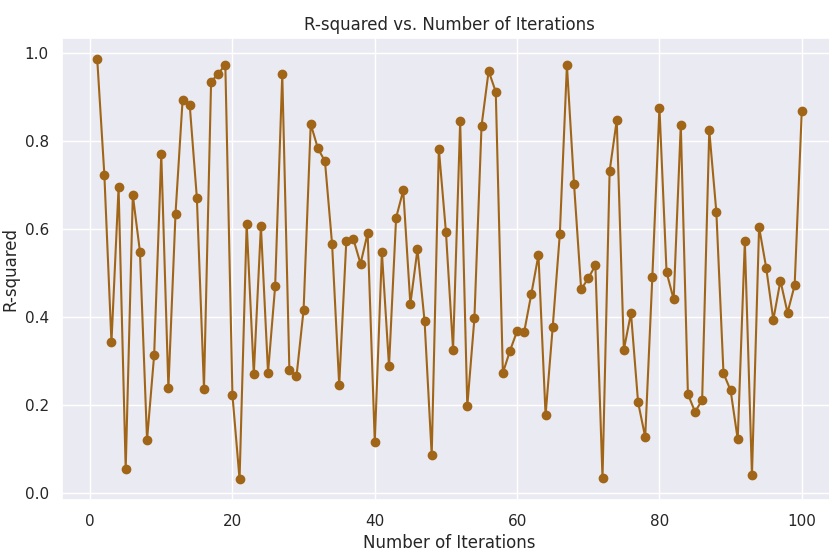


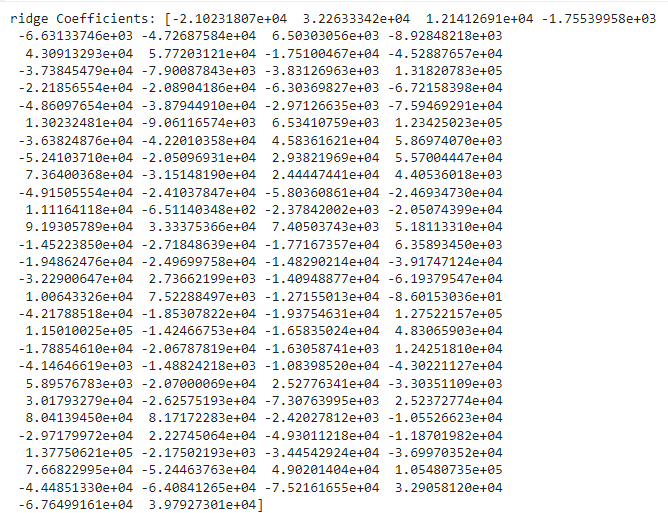
Fig 21: Graph obtained after applying bootsrapping

to lasso regression .

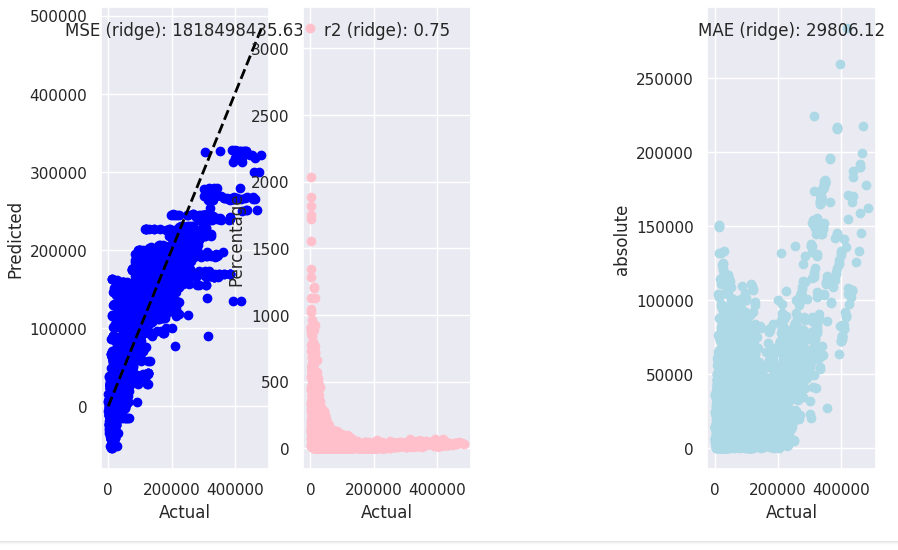
1. RIDGE REGRESSION

The coefficients obtained from the ridge regression model highlight the importance of each environmental variable in predicting crop yields. Positive coefficients indicate a positive relationship with crop yield, while negative coefficients indicate a negative relationship. Mean squared error, mean absolute error , and root mean squared error (RMSE) metrics provide insight into the accuracy of the model, with lowe values ​​indicating better performance with an R-square value of 0.749 indicates that approximately 75% of yield variation can be explained by environmental variables added to the system Shows promise towards predicting crop yields, although additional adjustments may be required and are adopted for real-world applications Considering the possibility of multicollinearity in the variables and the need to improve with additional data sources to improve the robustness of the model. Further research could be to compare the performance of different regression algorithms and examine the interaction effect of environmental factors. These findings highlight the potential of predictive models to optimize agricultural practices and enable sustainable crop management decisions.

|  |  |
| --- | --- |
| ERROR METRICS | RESULTS |
| MSE | 1818498435.6339955 |
| MAE | 29806.118494602797 |
| R2 | 0.7492996039884241 |
| RMSE | 42643.855778224315 |



RELATED GRAPHS:

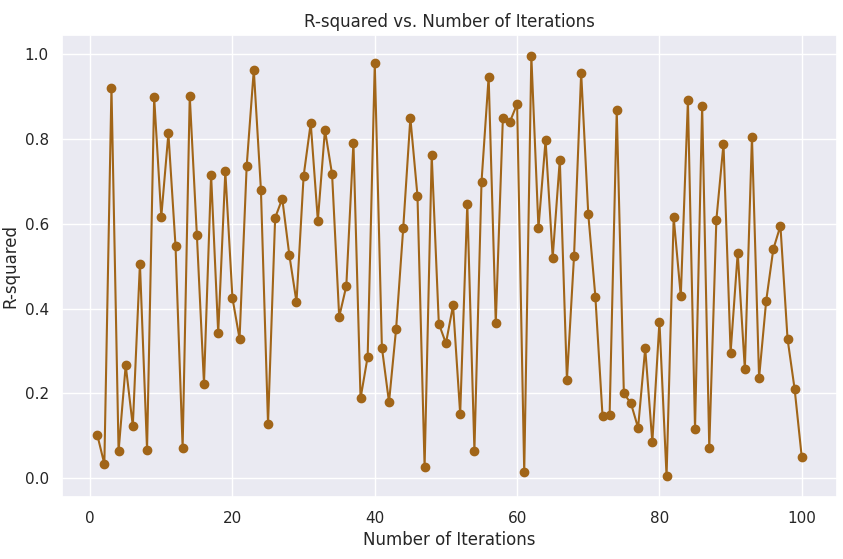


BOOTSTRAPPING TO RIDGE REGRESSION:

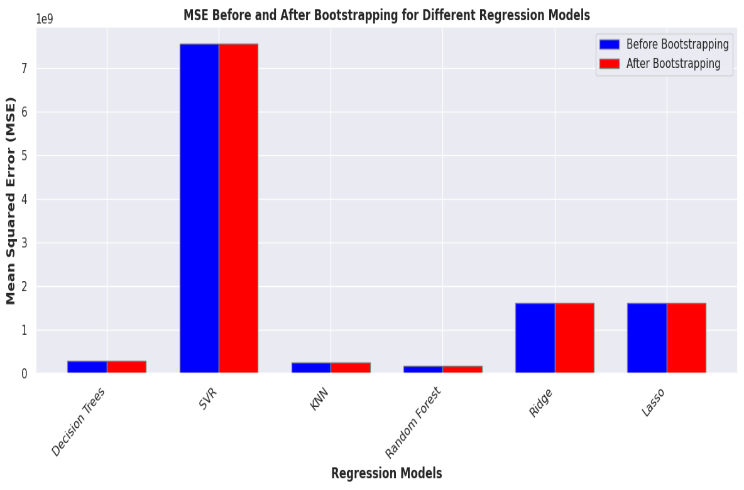
The 100 iterations of bootstrapping to the Ridge Regression model shows uniform results, as shown by the Standard Deviation of 0.0 for MSE, MAE, R-squared and RMSE. This signifies that the version’s performance remained consistent and can be replicated with varying different sample of the dataset. The average values of MSE, MAE, R-squared, and RMSE are used to estimate the predicting quality of the model in general, while MSE and RMSE also serve the purposes of evaluating prediction error and R-squared shows the amount of variation explained by the model. These results intend that Ridge Regression model can reliably estimate crop yield and thus can be used in support of agricultural policy making. Additional assessment should include examining the possible effect of other hyperparameters or adding more elements to improve model performance. Ideally, the trend and stable nature of the Ridge Regression model makes it a useful machine for crop yield forecasts.

|  |  |
| --- | --- |
| MSE | 1818498435.6339955 |
| STD OF MSE | 0.0 |
| MAE | 29806.118494602797 |
| STD OF MAE | 0.0 |
| R2 | 0.7492996039884241 |
| STD OF R2 | 0.0 |
| RMSE | 0.7492996039884241 |
| STD OF RMSE | 0.0 |

RELATED GRAPH:



COMPARISON OF MSE VALUES BEFORE AND BOOTSTRAPPING OF ALL ALGORITHMS USED



1. IMPROVEMENTS:

The fine-tuning of dataset quality will involve monitoring of the dataset quality with a focus on handling of missing data, duplicate entries as well as correction of data inaccuracies in order to enhance reliability and correctness.

Machine Learning: Look into advanced feature engineering strategies that seek to broaden understanding existing variables through input from including complex systems in agriculture.

Methodology: In addition to ridge regression, test other regression methods for example, random forest regression, support vector regression or gradient boosting regression, and fit the best model to your data.

Ensemble learning: Apply such ensemble learning techniques like stacking or voting or regressors in order to combine different models enhancing predictions’

Cross Validation: Employ simple cross-validational techniques in testing your models such as K folds Cross Validation in order to provide adequate assurance that your models will also work with new data.

Domain Collaboration: Coordinate with people who have knowledge in the agricultural industry to better understand the issues encountered in growing crops and their characteristics.

1. CONCLUSION:

Within the wide range of agricultural advancements, it is the crop yield predicting company of ours which shows the compassion between data science and agronomy the best. Through careful data preprocessing and offering data for training right, and placing the right algorithms, we have built an environment full of insights and capabilities which are able to generate insights.

Just as epochs characterize the earth within which our work has faced a number of challenges and ebbs and flows. From putting raw data in the ground, to raising prettily shaped and colored forecasting models, every phase has been an adventure of development and exploration.

As we enjoy the fruits of our labor, it strikes us, though, that the odyssey has no conclusion. The knowledge we accumulate from our predictive models is not a deluge of water but … no, it is a different type of water, more like fertilized synthetic seeds ready for planting that will sustain cultivation practices and agricultural choices for the coming populations.

Amid the transformations of the system of production in agriculture, our work emerges as a source of creativity that darkens the prospects of hopelessness, famine, and deprivation. Such ground we shall tend to by sowing the seeds of a brighter and better tomorrow for everyone where data science is married to agriculture. In the part of our project where we reached the end, it was KNN who was the star algorithm and made its contributions to the story of exploration and creativity in an unforgettable way. What KNN achieved is valuable not only in the precision of forecasting but also in changing our perspective on the processes engaged in agriculture. Really, RM Regression Forest is the big lamp.

1. REFERENCES:

Lobell, D. B., & Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. Environmental Research Letters, 2(1), 014002.

Wardlow, B. D., & Egbert, S. L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. Remote Sensing of Environment, 112(3), 1096-1116.

Huang, J., Yu, H., Guan, X., Wang, G., & Guo, R. (2016). Accelerated dryland expansion under climate change. Nature Climate Change, 6(2), 166-171.

Srivastava, P. K., Han, D., Rico-Ramirez, M. A., & Islam, T. (2019). Application of machine learning algorithms in real-time flood forecasting: A systematic review. Water, 11(5), 974.

Basso, B., Antle, J. M., & Ritchie, J. T. (2015). Simulating crop yield with a multi‐scale, multi‐model framework: A review. Field Crops Research, 186, 64-76.

Shen, J., Krogmeier, J. V., & Anagnostopoulos, G. (2007). Machine learning in agriculture: A review. Computers and Electronics in Agriculture, 63(2), 247-267.

Devi, A. R., Reddy, M. P., & Reddy, D. C. (2019). Crop yield prediction using machine learning algorithms. International Journal of Engineering and Advanced Technology, 9(1), 6067-6071.

Jin, M., & Suh, C. (2018). A comparative study on machine learning-based crop yield prediction models using remotely sensed data. Computers and Electronics in Agriculture, 153, 13-27.