Smart Farming System using ML

Madhu Verma, Rajnish Kr. Thakur, Sani Pandey, Shivam Singh

Department of Electronic & Communication Engineering,
Raj Kumar Goel Institute of Technology,
Ghaziabad, Uttar Pradesh, India.

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 3 (May-June 2023)

Abstract

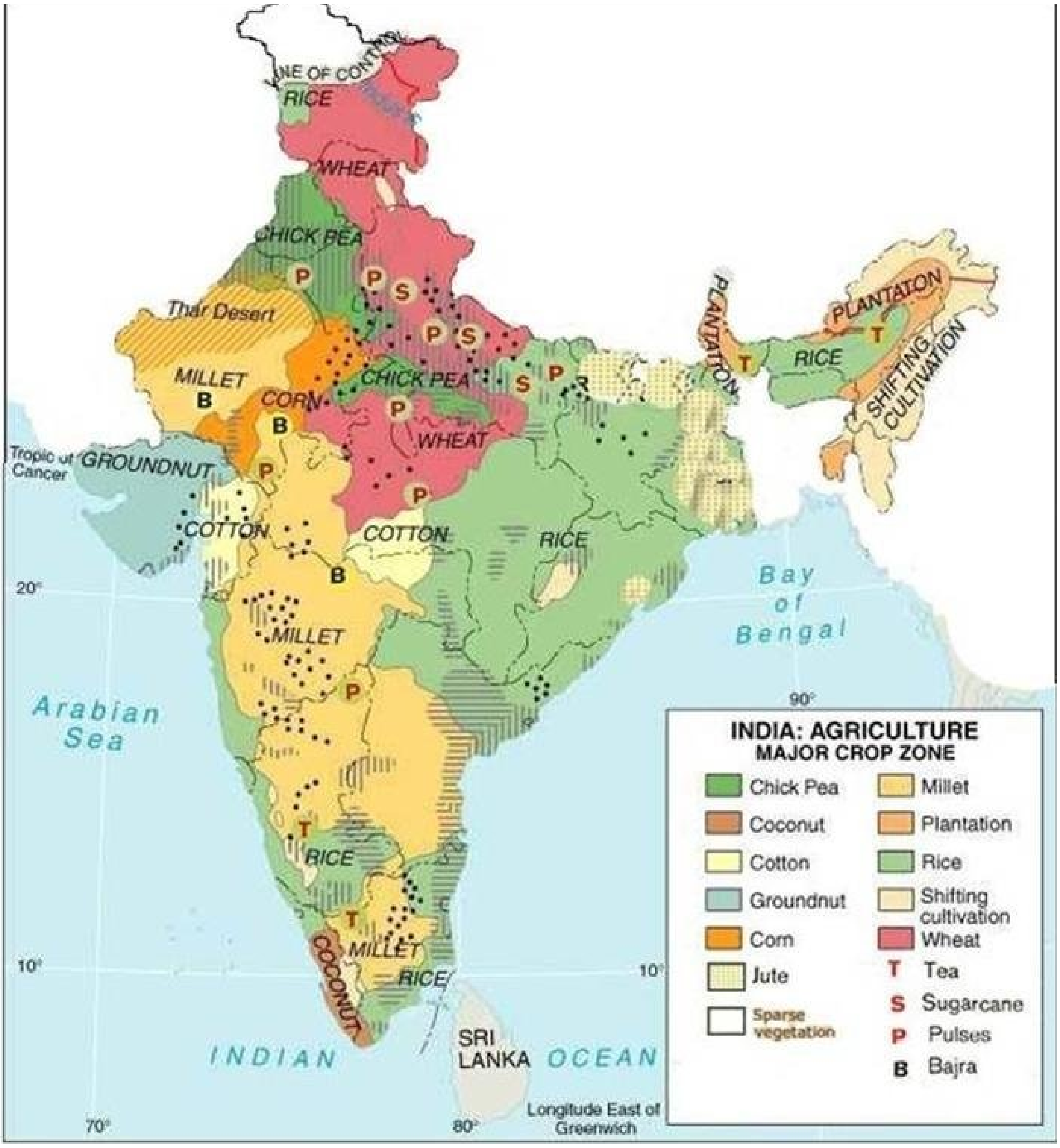
It is widely recognized that agriculture serves as the fundamental pillar of India's economy. This research paper delves into the realm of yield predictions, encompassing the vast array of crops cultivated throughout the nation. What sets this script apart is its unique ability to anticipate agricultural production for any chosen year, employing easily comprehensible variables such as state, district, season, and area. To achieve this feat, the article draws upon an assortment of regression techniques, including the notable Kernel Ridge, Lasso, and ENet algorithms. These sophisticated statistical methods form the bedrock of the paper's prediction methodology, enabling accurate estimations of crop output.

Keywords: Crop Yield Prediction, Lasso, Kernel Ridge, ENet, Stacked Regression, Machine Learning (ML)

1. Introduction

To facilitate understanding and visualization of the vast variety of crops grown in India, they are often grouped into different categories or orders. This categorization helps researchers, policymakers, and farmers to study and analyze crops more efficiently. By organizing crops into orders, it becomes easier to identify similarities and differences between them, understand their cultivation requirements, and develop appropriate strategies for their growth and management. Orders are typically based on common characteristics shared by a group of crops. These characteristics may include botanical features, growth habits, crop duration, climatic preferences, nutrient requirements, or economic significance [1].

Figure 1: Famous Categories of Crops over States in India (based on Season)



The study's dataset, which consists of over 2.5 million data points, includes important characteristics including State, District, Crop, Season, Year, Area, and Product. The areas and ecosystems of India are graphically represented in Figure 1, together with the typical crop sequencing that is seen throughout the year. Advanced regression approaches including Lasso, ENet, Kernel Ridge, and model combining (mounding) were used to improve the precision of yield estimates and reduce mistakes. Improved projections were produced and the analytical process was refined using these advanced approaches. The Literature Review, Methodology, Conclusion, and Unborn Work are the four components that make up the essay's structure. The Literature Review part provides a thorough backdrop for the study by exploring earlier research and intellectual contributions pertinent to the topic.

2. Background Study

The CRY algorithm, introduced by Ananthara M.G. et al. in February 2013, is a predictive model for agricultural datasets that utilizes beehive clustering techniques to forecast crop yields. The researchers investigated various parameters including crop type, soil type, soil pH value, moisture, and crop sensitivity. The study specifically concentrated on predicting rice, sugarcane, and paddy yields in India [2].

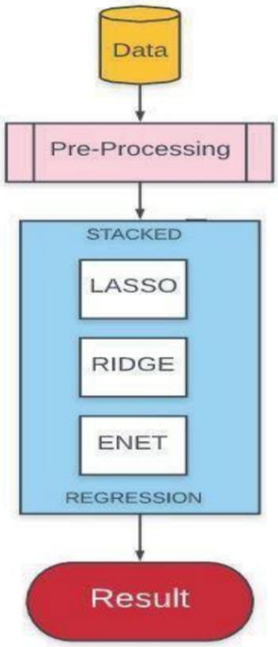
In their research conducted in August 2019, Chawla I. et al. utilized fuzzy logic in conjunction with statistical time series models to predict crop output. The variables used for prediction were similar to temperature and rainfall, which are important factors affecting crop growth. The outcomes of their prediction were categorized into two main categories: "good yield" and "very good yield." In other words, based on their forecast, the crop output was expected to be either satisfactory or exceptionally high. The use of fuzzy logic allowed the researchers to incorporate a degree of uncertainty and imprecision in their predictions, considering the complex and uncertain nature of agricultural systems [3].

In their study conducted in August 2018, Chaudhari A.N. et al. aimed to enhance the accuracy of crop yield prediction by combining three algorithms: clustering k-means, Apriori, and Bayes. By considering factors such as the cultivation area, rainfall patterns, and soil type, their system was designed to determine which crop is most suitable for cultivation in a particular area [4]. In December 2017, Gandge Y. conducted a study focused on colorful crops and applied various machine learning methods to determine the most suitable techniques for different crops. The research aimed to investigate and examine which methods would yield the best results for each specific crop [5]. In July 2016, Armstrong L.J. and colleagues utilized Artificial Neural Networks (ANNs) to forecast rice yield in specific regions of Maharashtra, India. Their study focused on climate variables, including temperature, rainfall, and reference crop evapotranspiration, which fell within a certain range. The researchers obtained historical data from the Indian Government's database, spanning the period between 1998 and 2002 [6]. In July 2016, Petkar O., the same author who employed Support Vector Machines (SVM) and Neural Networks for rice crop yield prediction, introduced a new decision system that serves as an interface for providing input and accommodating errors. In other words, Petkar O. developed a decision support system that allows users to interact with the model by providing input parameters and handling any potential errors or uncertainties in the predictions [7].

In July 2016, Petkar O., the same author who employed Support Vector Machines (SVM) and Neural Networks for rice crop yield prediction, introduced a new decision system that serves as an interface for providing input and accommodating errors. In other words, Petkar O. developed a decision support system that allows users to interact with the model by providing input parameters and handling any potential errors or uncertainties in the predictions [8]. In December 2018, Chakrabarty A. et al. conducted a study on crop prediction in Bangladesh, focusing on the cultivation of three main crops: rice (three different types), jute, wheat, and potatoes [9]. Mariappan A.K. and colleagues conducted a study on rice crop statistics in Tamil Nadu, India. The research focused on various fundamental aspects of the climate that influence rice cultivation [10]. In 2018, Kalbande D.R. and colleagues employed Support Vector Regression (SVR), Multi Polynomial Regression, and Random Forest Regression to forecast sludge yield. These regression models were estimated using parameters that closely aligned with Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) values [11].

3. Methodology

Figure 2: Process Chart of the Research Project



A. Pre-processing

The given dataset for sludge yield contains several 'NA' values, which are typically missing or null values. To handle these missing values in Python, a filtering process was employed to remove or handle the 'NA' values appropriately. This process could involve techniques such as dropping the rows or columns with 'NA' values or imputing them with estimated values based on other data points.

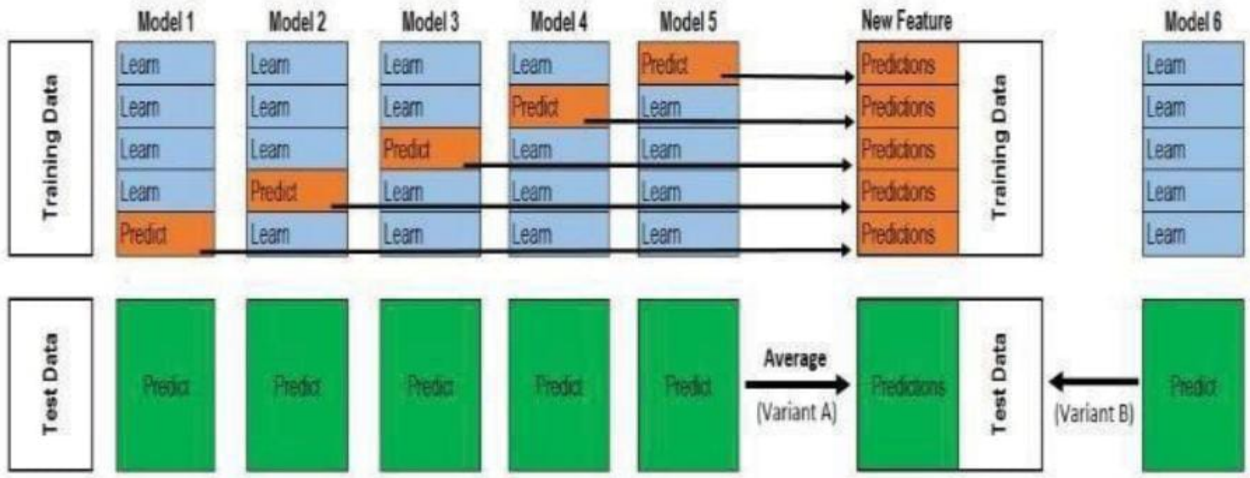
B. Stacked Regression

The process described is a form of ensemble learning technique known as stacking with averaging. It involves combining multiple models and incorporating a meta model to improve the overall predictive performance.

Step 1: The training dataset is divided into two separate sets - the "train" set and the "holdout" set.
Step 2: The base models (e.g., Support Vector Regression, Multi Polynomial Regression, Random Forest Regression) are trained using the training set.
Step 3: The trained base models are then tested using the holdout set.
Step 4: The predictions obtained from the base models on the holdout set are now utilized as inputs for the meta-model. The meta-model, also known as the advanced-position learner, is trained using the holdout set predictions as features and the actual target variable values.

The training data is divided into five different folds or subsets, typically done randomly or using a specific sampling strategy. Each base model is trained iteratively on four out of the five folds (crowds) in each replication. This means that for each fold, four folds are used for training the base models, while the remaining fold acts as the holdout or test set. The base models are then used to make predictions on the holdout fold (the remaining crowd). These predictions serve as the prognostications or output of the base models for that particular fold. The prognostications obtained from all the base models on the test data (the holdout folds from each replication) are combined and used as meta-features. These meta-features are additional input features for the meta-model. The meta-model is trained on the complete dataset, including the original features and the meta-features generated from the base models. The meta-model learns to make predictions based on these combined features and the actual target variable. In Figure 2, the top portion represents the meta model, which in this case is the Lasso Regressor. let's discuss the operation of piled retrogression as shown in Figure 3.

Figure 3: Stacked Regression



C. Output

In this design, the root mean square error (RMSE) is used as the performance metric. RMSE is a measure of the differences between predicted values and actual values in a regression problem. When the models were applied collectively, the ENet model had an error of about 4, the Lasso model had an error of about 2, and the Kernel Ridge model had an error of about 1. After "mounding" (assuming it means combining or stacking the models), the overall error was lower than 1. To obtain the predictions shown in Figure 4.

Figure 4: Interface of the Web APP

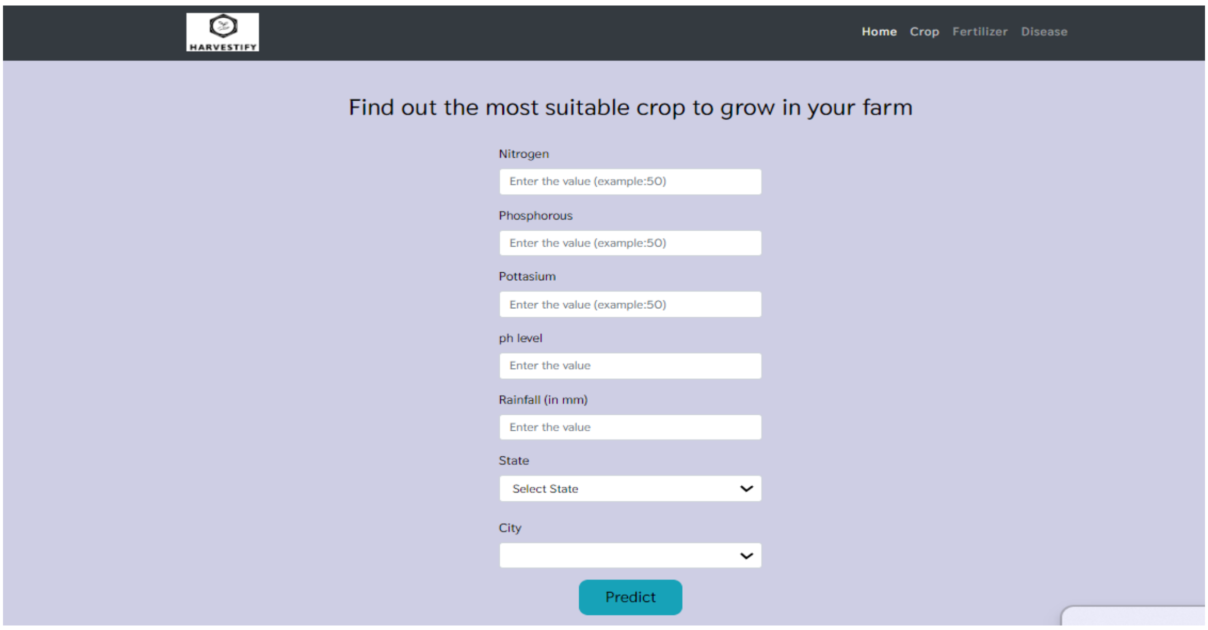
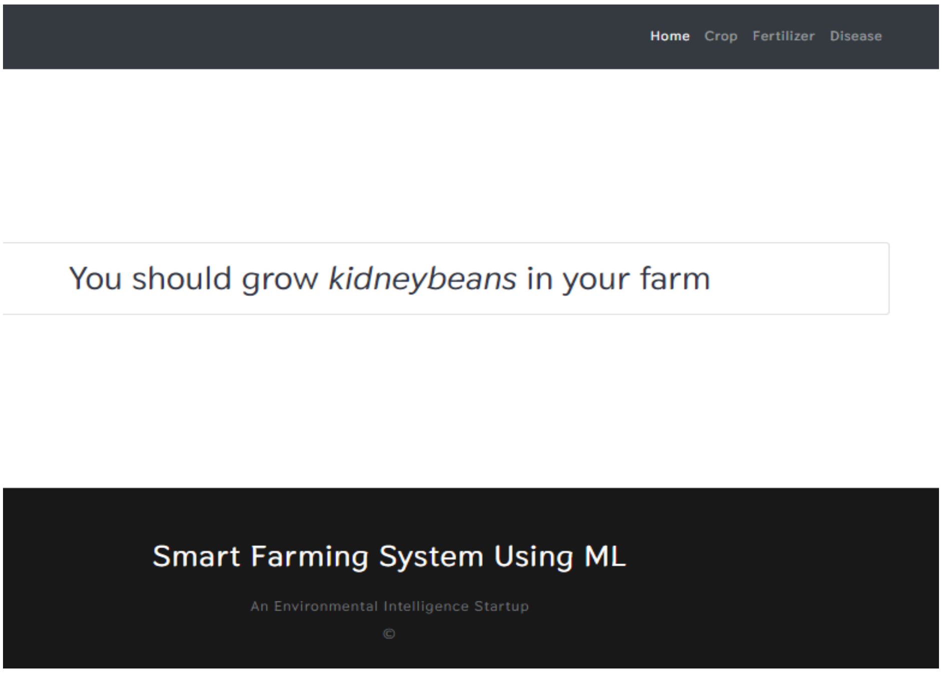


Figure 5: Crop Predicted by the System (Output)



4. Conclusion and Future Work

The implementation of layered regression in our model has yielded significantly improved results compared to using individual models in isolation. When providing inputs such as State, City, pH level, Rainfall, Nitrogen, Phosphorus, and Potassium, our model predicts the crops that are most suitable based on these inputs. Currently, the model is available as an online application, allowing users to input the relevant parameters and obtain crop recommendations. However, as part of our future work, we plan to develop a dedicated mobile application that farmers can easily access and utilize. This will provide a more user-friendly and convenient platform for farmers to benefit from the crop recommendation system. Additionally, we aim to enhance the accessibility and usability of the system by translating the entire application into regional languages. This will enable farmers from diverse linguistic backgrounds to effectively utilize the application without language barriers.

References

1. https://data.gov.in/
2. Ananthara M.G., Arunkumar T., Hemavathy R. (2013, February). CRY—an improved crop yield prediction model using bee hiveclustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, IEEE, pp. 473-478.
3. Bang S., Bishnoi R., Chauhan A.S., Dixit A.K., Chawla I. (2019, August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In 2019 Twelfth International Conference on Contemporary Computing (IC3), IEEE, pp. 1- 6.
4. Bhosale S.V., Thombare R.A., Dhemey P.G., Chaudhari A.N. (2018, August). Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, pp. 1-5.
5. Gandge Y. (2017, December). A study on various data mining techniques for crop yield prediction. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), IEEE, pp. 420-423.

6. Gandhi N., Petkar O., Armstrong L.J. (2016, July). Rice crop yield prediction using artificial neural networks. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), IEEE, pp. 105-110.
7. Gandhi N., Armstrong L.J., Petkar O., Tripathy A.K. (2016, July). Rice crop yield prediction in India using support vector machines. In 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, pp. 1-5.
8. Gandhi N., Armstrong L.J., Petkar O. (2016, July). Proposed decision support system (DSS) for Indian rice crop yield prediction. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), IEEE, pp. 13-18.
9. Islam T., Chitty T.A., Chakrabarty A. (2018, December). A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh. In 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, pp. 1-6.
10. Mariappan A.K., Das J.A.B. (2017, April). A paradigm for rice yield prediction in Tamilnadu. In 2017 IEEE Technological Innovationsin ICT for Agriculture and Rural Development (TIAR), IEEE, pp. 18-21.
11. Shah A., Dubey A., Hemnani V., Gala D., Kalbande D.R. (2018). Smart Farming System: Crop Yield Prediction Using Regression Techniques. In Proceedings of International Conference on Wireless Communication, Springer, Singapore, pp. 49-56.
12. https://github.com/Gladiator07