

# Crop Yield Prediction using Machine Learning

Lohit V K<sup>1</sup>, L. Vijayalakshmi<sup>2</sup>, Brunda. G<sup>3</sup>, Sanjay M D<sup>4</sup>, Rashmi K T<sup>5</sup>  
<sup>1,2,3,4</sup> CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India

<sup>5</sup> Assistant Professor, CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India

**Abstract:-** In India, we all know that farming is the backbone of the nation. This document predicts yield for all sorts of crops grown in India. This script is unique since it uses simple criteria such as state, district, season, and area to forecast agricultural production in whatever year the user desires. The document is based on regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict output and use the Regression stacking concept for algorithm improvement to provide better foretelling.

**Keywords:-** Crop yield prediction, Lasso, Kernel Ridge, Enet, Stacked Regression.

## I. INTRODUCTION

Everyone uses climatic factors like downfall, sun, and agrarian factors like soil type, and nutrients held by the soil (Nitrogen, Potassium, etc.) in our exploration, which we discovered in former exploration papers. The problem is that we need to collect the data, also a third party makes the vaticination, and also it's explained to the planter, which takes a lot of time and trouble for the planter, and he does not understand the wisdom behind these factors. To make it easy and incontinently applicable to growers, this study use simple parameters similar as the planter's state and quarter, crop kind, and season (as in Kharif, Rabi, etc.). In India, there are over a hundred different crops grown across the country. These crops are divided into orders for easier appreciation and visualization. The information for this study came from the Indian Government Depository (1).

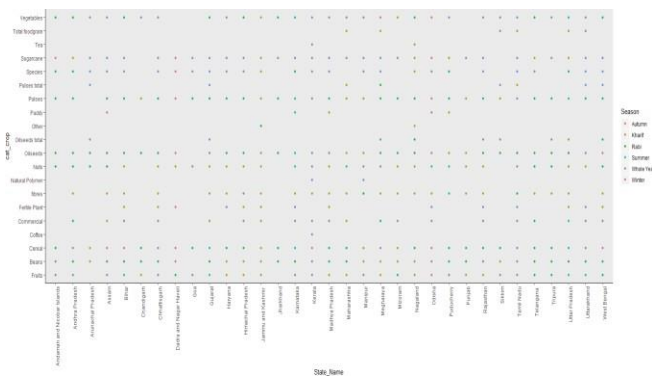


Fig.1.Famous Categories of crops over states in India(basedon Season)

With about 2.5 million compliances, the data comprises of parameters similar as State, District, Crop, Season, Year, Area, and Product. Figure 1 illustrates India's countries and homes, illustrating which crop orders are popular during which seasons. We applied advanced retrogression ways – Lasso, ENet, and Kernel Ridge – and mounding of these models to reduce error and produce bettered prognostications. The ensuing sections comprise this paper Literature Review, Methodology, Conclusion, and Unborn Work.

## II. BACKGROUND STUDY

AnantharaM.G. et al. (2013, February) introduced the CRY algorithm for crop yield employing beehive clustering ways as a vaticination model for agrarian datasets. Crop kind, soil type, soil pH value, moisture, and crop perceptivity were all factors examined. Their exploration concentrated on Indian paddy, rice, and sugarcane yields. Their new fashion was also compared to the C&R tree algorithm, and it outperformed the ultimate with a 90 delicacy. (2). Awan, A.M. et al. (2006, April) developed a novel, smart frame for ranch yield cast clustering kernel methodology, taking into account factors similar as colony, latitude, temperature, and downfall rush in that latitude. For the examination of canvas win fields, they used a weighted k-means kernel approach with spatial limitations. (3) Chawla, I. et al. (2019, August) used fuzzy sense to prognosticate crop yields using statistical time series models. For vaticination, they used variables similar as downfall and temperature.

Their vaticination was classified as 'good yiel 'or' veritably good yield.' (4). Chaudhari, A.N. et al. (2018, August) used three algorithms clustering k- means, Apriori, and Bayes algorithm, which they also hybridized for better yield vaticination effectiveness, taking into account parameters similar as area, downfall, and soil type, and their system were suitable to tell which crop is suitable for civilization grounded on the mentioned features. as rainfall and temperature. Their prediction was classified as 'good yield' or 'very good yield.' [4]. Chaudhari, A. N. et al. (2018, August) used three algorithms: clustering k- means, Apriori, and Bayes algorithm, which they then hybridized for better yield prediction efficiency, taking into account parameters such as area, rainfall, and soil type, and their system were able to tell which crop is suitable for cultivation based on the mentioned features (5). Gandge, Y. (2017, December) used a variety of machine literacy algorithms for colorful crops. They delved and anatomized which algorithm would be applicable for which crop. They employed K- means, Support Vector Regression, Neural Networks, C4.5 Decision Tree, Bee-Hive Clustering, and other ways. The inferring factors were soil nutrients similar as N, K, P, and soilph. (6). Armstrong, L.J. et al. (2016, July) used ANNs to prognosticate rice yield in Maharashtra sections in India. They took into account climatic factors similar as temperature, rush, and reference crop evapotranspiration (within a certain range). From 1998 to 2002, the records were attained from the Indian Government's depository. (7). Tripathy, A.K. et al. (2016, July) employed support vector machines to read rice crop yield with the same features as the former paper. (8). Petkar, O. (2016, July), the same authors who applied SVM and neural networks for rice crop yield vaticination,

proposed a new decision system that's an interface to give input and admit affair. (9).A. Chakrabarty et al. (2018, December) delved crop vaticination in Bangladesh, where they primarily cultivate three types of rice jute, wheat, and potato. Their study employed a deep neural network to dissect data that included 46 parameters. Among them were soil composition, toxin type, soil type and structure, soil thickness, response, and texture(10). Jintrawet,A. et al. (2008, May) used the SVR model for crops similar as rice to prognosticate yield, with the model divided into three way prognosticating soil nitrogen weight, rice stem weight, and rice grain weight. Along with those three way, their factors included solar radiation, temperature, and rush (11). Miniappan,N. et al. (2014, August) used an artificial neural network to model amulti-layer perceptron model with 20 retired layers for wheat yield vaticination, taking into account factors similar as sun, rain, frost, and temperature (12). Manjula, A et al. developed a crop selection and yield vaticination model that took into account colorful indicators similar as foliage, temperature, and regularized difference foliage as factors. For a better understanding, they distinguished between

factors, agronomic factors, and other disturbances caused by the vaticination (13). Mariappan,A.K., and associates delved rice crop statistics in Tamil Nadu, India. They took into climate account rudiments similar as soil, temperature, sun, downfall, toxin, paddy, and nonentity type, as well as pollution and season (14).Verma,A. et al. (2015, December) used crop vaticination ways similar as Naive Bayes and the K-NN algorithm on soil datasets containing nutrients similar as zinc, bobby, manganese, pH, iron, sulfur, phosphorus, potassium, nitrogen, and organic carbon (15). Kalbande,D.R. et al. (2018) prognosticated sludge yield using support vector retrogression, multi polynomial retrogression, and arbitrary timberretrogression and estimated the models using criteria similar as MAE, RMSE, and R- forecourt values (16). Rahman,R.M., et al. (2015, June) primarily used clustering ways to prognosticate crop yield. The paper described the analysis of major crops in Bangladesh and classified the variables as environmental and biotic. For bracket, direct retrogression, ANN, and the KNN approach were used (17). Hegde,M. et al. (2015, June) used multiple direct retrogression and neuro- fuzzy systems to prognosticate crop yield using biomass, soil water, radiation, and downfall as input parameters for the exploration, with wheat as the primary crop (18). Sujatha,R., and Isakki,P. (2016) used bracket ways similar as ANN, j48, Nave Bayes, Random Forest, and Support VectorMachines.In addition, they've included climatic and soil parameters as features in their modeling (19). Ramalatha,M. et al.(2018, October) combined Kmeans clustering and bracket grounded on a modified K-NN approach. The information was gathered in Tamil Nadu, India, where the most abundant crops were rice, sludge, ragi, sugarcane, and tapioca (20). Singh,C.D., et al. (2014, January) developed a crop- advice operation that works in a many Madhya Pradesh sections. The stonerwould input pall cover, downfall, temperature, and preliminarily recorded yield, and the system would anticipate the yield, marker the

crop, and admit the results grounded on the detector values established (21).

### III. METHODOLOGY

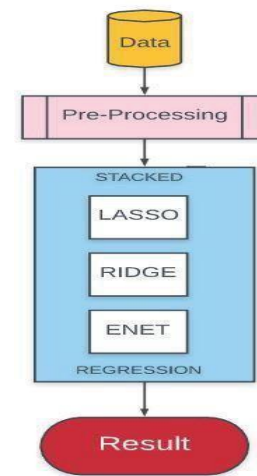


Fig. 2. Process chart of the research project

#### A .Pre-processing

There are numerous 'NA' values in the given data set, which are filtered in Python. Likewise, because the data set contains numeric data, we used robust scaling, which is analogous to normalization but rather uses the interquartile range, whereas normalization shrinks the data in terms of 0 to 1.

#### B. Stacked Regression

This is a form of assembling with a little averaging boost. We add a meta model to this and use the eschewal of fold prognostications from the other models to train the primary meta model.

Step- 1 The entire training is divided in two different sets. (Train and holdout)

Step-2 Using the first part, train the named base models (train).

Step-3 Test them with Part II. (holdout)

Step-4 The prognostications attained from the test section are now inputs to the meta- model, a advanced- position learner.

The first three way are completed iteratively. For illustration, if we want to do a5-fold mounding, we must first divide the training data into 5 crowds We will also go through 5 duplications. In each replication, we train each base model on four crowds and prognosticate the remaining crowds (holdout pack). So, after 5 duplications, we'll be confident that we have used all of the data to induce out-of-fold prognostications, which we'll use as a new point in Step 4 to train our meta- model for the prophetic portion, we equaled the prognostications of all base models on the test data and used them as meta- features on which the meta- model is eventually prognosticated. Our meta model is Lasso Regressor in this Case, which is why it's at the top of Figure 2. Figure 3 depicts the operation of piled

retrogression.

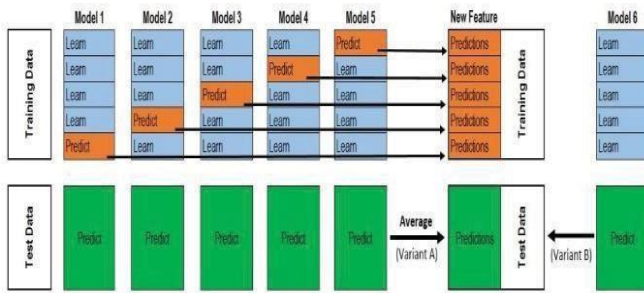


Fig. 3. Stacked Regression

C. Output

Root mean square error is the performance metric used in this design. When the models were applied collectively, ENet had an error of about 4, Lasso had an error of about 2, Kernel Ridge had an error of about 1, and after mounding it was Lower than 1. The stoner or planter can enter the following information into the web operation to gain the vaticination shown in fig. 4 below.

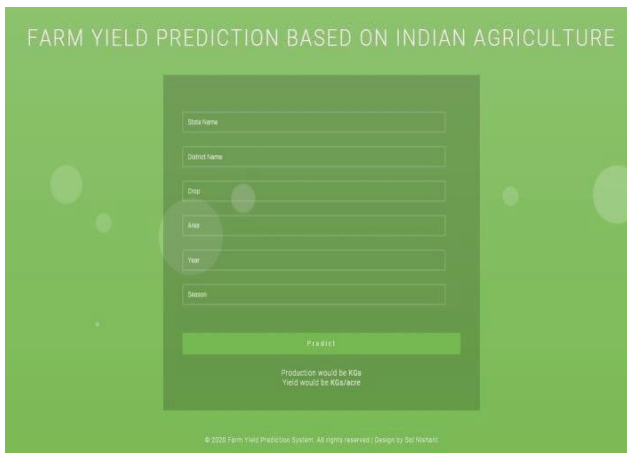


Fig. 4. Interface of Web APP

IV. CONCLUSION AND FUTURE WORK

When we use layered regression, the results are far better than when we use those models individually. When we give input to our model such as State, District, Season and User. Then output gives how much tons of crop can be yielded, how much will be the price of crop per kg. The output presented in the picture is now an online application but over future work would include developing an app for farmers to use and translating the entire system into their regional language.

VI. ACKNOWLEDGEMENT

We would like to thank Dr. Shantharam Nayak and Assistant Professor Rashmi K T for their valuable suggestion, expert advice and moral support in the process of preparing this paper.

VII. REFERENCE

[1] "data.gov.in." [Online]. Available: <https://data.gov.in/>  
 [2] Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile

Engineering (pp. 473-478). IEEE.  
 [3] Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 841-846). Springer, Berlin, Heidelberg.  
 [4] 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) (pp. 1-6). IEEE.  
 [5] 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) (pp. 1-5). IEEE.  
 [6] 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 (ICECCOT) (pp. 420-423). IEEE.  
 [7] 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) (pp. 105-110). IEEE.  
 [8] 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) (pp. 1-5). IEEE.  
 [9] 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) (pp. 13-18). IEEE.  
 [10] Islam, T., Chisty, 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) (pp. 1-6). IEEE.  
 [11] Jaikla, R., Auephanwiriyakul, S., & Jintrawet, A. (2008, May). Rice yield prediction using a support vector regression method. In 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (Vol. 1, pp. 29-32). IEEE.  
 [12] Kadir, M. K. A., Ayob, M. Z., & Miniappan, N. (2014, August). Wheat yield prediction: Artificial neural network based approach. In 2014 4th International Conference on Engineering Technology and Technopreneuship (ICE2T) (pp. 161-165). IEEE.  
 [13] Manjula, A., & Narsimha, G. (2015, January). XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. In 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO) (pp. 1-5). IEEE.  
 [14] Mariappan, A. K., & Das, J. A. B. (2017, April). A paradigm for rice yield prediction in Tamilnadu. In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp. 18-21). IEEE.  
 [15] Paul, M., Vishwakarma, S. K., & Verma, A. (2015, December). Analysis of soil behaviour and prediction of crop yield using data mining approach. In 2015 International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 766-771). IEEE.  
 [16] 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 (pp. 49-56). Springer, Singapore.  
 [17] Ahamed, A. M. S., Mahmood, N. T., Hossain, N., Kabir, M. T., Das, K., Rahman, F., & Rahman, R. M. (2015, June). Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 1-6). IEEE.  
 [18] Shastri, A., Sanjay, H. A., & Hegde, M. (2015, June). A parameter based ANFIS model for crop yield prediction. In 2015 IEEE International Advance Computing Conference (IACC) (pp. 253-257). IEEE.  
 [19] Sujatha, R., & Isakki, P. (2016, January). A study on crop yield forecasting using classification techniques. In 2016 International

- Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp. 1-4). IEEE.
- [20] Suresh, A., Kumar, P. G., & Ramalatha, M. (2018, October). Prediction of major crop yields of Tamilnadu using K-means and Modified KNN. In 2018 3rd International Conference on Communication and Electronics Systems (ICES) (pp. 88-93). IEEE.
- [21] Veenadhari, S., Misra, B., & Singh, C. D. (2014, January). Machine learning approach for forecasting crop yield based on climatic parameters. In 2014 International Conference on Computer Communication and Informatics (pp. 1-5). IEEE.