Predicting Crop Yields in Indian Agriculture UsingMachine Learning

Teja Ram Mohan Sai Vemulapalli[1], Kalyan Vanam[2], S.Rupak Balaji[3], Pavan Sudheer Varma K[4]
Bachu Anush Kumar[5], Vikas Mangotra[6], [1,2,3,4,5,6]Department of Computer Science and Engineering,
Lovely Professional University,Jalandhar Punjab, India.
[1]tejarammohansai@gmail.com, [2]vanamkalyan050@gmail.com, [3]rupakbalaji0510@gmail.com, [4]sudheervarama7677@gmail.com,
[5]anushrao.326@gmail.com, [6]mangotra.vikas@gmail.com.

***Abstract*—** Agriculture provides the backbone of India's economy both within and between. The paper below provides a brief insight into the creation of a fresh model is currently in progress to estimate the variability in the crop yields across various regions in India. Via this simple setting-based approach, such parameters as year, district, season and area can be used to predict the amounts of crop outputs for certain years. By employing sophisticated regression techniques, Kernel Ridge, Lasso, and ENet algorithms are used, to improve the yield prediction precision. Additionally, it implies a technique of Scaled Regression which is specifically designed to improve the accuracy of the data resulting as long-term predictions.

Key terms: Crop yield forecasting, Lasso algorithm, Kernel Ridge algorithm, ENet algorithm, Stacked Regression.

I. Introduction

While we were going through the data from the earlier prints, we detected common tendency of other scholars who employ climatic variables such as rainfall, light, and agricultural factors like soil makeup and nutrient level (such as nitrogen, potash) which are commonly used for crop yield prediction. Nevertheless, this situation can be tricky because managers need to collect data and integrate the estimations of experts, which increases the farmers' workload and therefore, understanding of the scientific principles is an essential prerequisite for managers. Our paper handles this issue by assembling comprehensible factors such as the location of the farmer (state and district) and the type of crop also telling us the time of the cultivation for instance (Kharif and Rabi).

India has the capacity to provide with more than hundred type of foods and that is because the crops are spread over the entire country. Organizing the crops in the same way as their categories in the summary would help students comprehend and visualize them. With data coming from the repository of The Indian Government with the details such as location, specific crop grown, time of planting, and the corresponding yield for each year are being considered which is almost 250 thousand observations it is our reference data. In our graph (Figure 1), you can see that crops travel across states and territories depending on the season that they are popular for.

We used different advanced regression techniques like Lasso, Elastic Net, and Kernel Ridge, and even combined them with stacking methods. This helped us make our predictions more accurate by lowering errors. Our paper is organized into sections covering what's been written before, how we did our research, what we've concluded, and where we're heading next.

II. Literature survey

Ananthara M.G. and co worked on the CRY model to give us a better approach for predicting crop yields from beehive clustering techniques. Their key indicators included crop classes, soil classes, humidity, soil pH value and sensitiveness of crop, and they analyzed the yields in the mostly paddy, rice, and sugar cane cultivations.

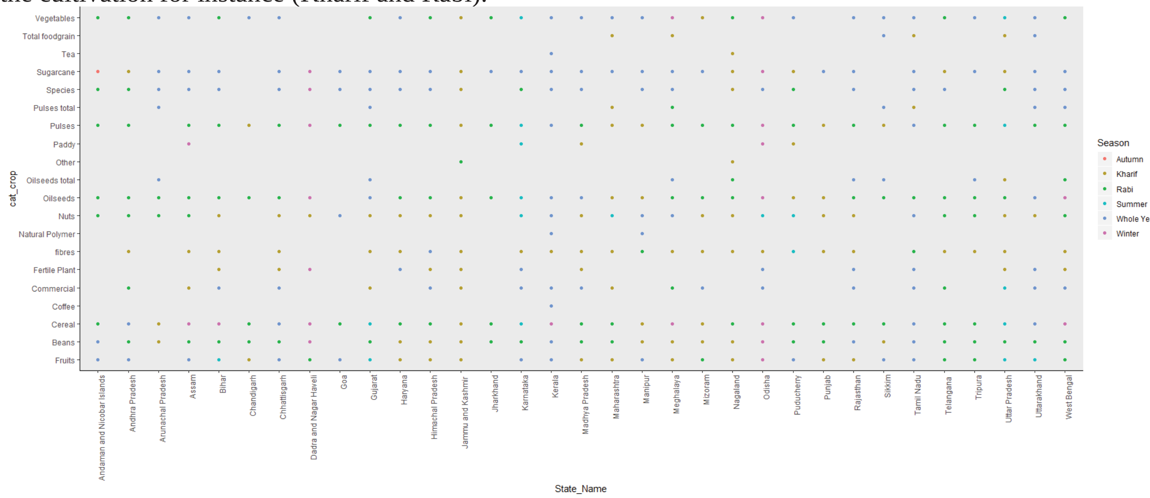


Fig. 1 Popular types of crops grown across different states in India, categorized by the seasons they are typically planted

According to their algorithm analysis in comparison with the C&R algorithm cree, their accuracy was 90 percent. Awan et al. (2006, April) proposed an inelegant architecture of kernel yield clustering methodology on the smart vegetables. Their operating components encompassed plantation enclosure, latitude, temperature, and precipitation, crucial for spatial analysis, utilizing the weighted k-means kernel approach, incorporating spatial constraints [3]. Sensing the crop yield prediction becomes better with the use of fuzzy logic as defined by Chawla, I.et.al, which mainly focuses on the statistical time series models. Among the factors for this insight include rainfall and temperature levels which were categorized as 'good yield' and ' very good yield' [4].

Chaudhari, A. N. [5] introduced a new hybrid model that used clustering k-means, Apriori, and Bayes algorithms in order to predict the yield more efficiently. The criteria chosen for our system includes the area, rainfall and soil type of the field, so the crops suggested by the system are based on these features. Gandge, Y. (2017, Dec.) introduced an application of the machine learning algorithms in crop production, i.e. K-means, SVM, NN, and C4.5 D.T., employing soil-concerned features of nutrient and pH. An effort made Artificial Neural Networks (ANN) use by Armstrong, L.J. [7], We created a model to predict rice yield in Maharashtra, India. We based our predictions on climate-related factors such as temperature, rainfall, and how much water crops typically use in the area. However, the research of [8] was based on Support Vector Machines (SVM) in rice crop yield prediction.

The decision system ran in both C# and python languages with a backend interpretation and representation. In December 2018, the research and used Deep Neural Network besides several other parameters which included the soil composition, the kind of fertilizer and the soil structure. Jintrawet, A. [11] achieved rice yield forecasting through a SVR machine learning process that features categories including regression of soil nitrogen, rice stem, and grain weight, while focusing on environmental indicators such as solar radiation, Precipitation, and temperature [11]. The authors elaborated on the use of a 20-layer artificial neural network to create a model to predict wheat yield, considering factors such as sunlight, rain, frost and temperature using the methodology [12]. Manjula, A. et al. have created a crop picking out and production prediction design, scrutinizing factors including vegetation, temperature, and vegetation indexation, that segregates between the weather and agronomic causes for great precision [13]. Mariappan, A.K, and the research group looked into producing rice crop data sets in Tamil Nadu, India. Their research included soil, climate, sunshine, rainfall, fertilizer, pest, air pollution, and season [14]
Verma et al. (2015, December), classified the crops using the Naïve Bayes as well as by applying the K-NN algorithm based on variables such as soil nutrients [15]. Kalbande, D. R. and others utilized assistance vector regression, multinomial polynomial regression, and the random forest regression for the purpose of corn yield forecasting. They calculated the performance of the models using MAE, RMSE, and R-square error metrics [16]. The opposed researchers grouped environmental and biotic variables into climate conditions and used a multiple regression, artificial neural networks (ANN) and K-nearest neighbors (KNN) approaches, in the study, multiple linear regression and neural fuzzy systems for crop growth prediction involving biomass, soil water, radiation, and rainfall, where wheat, was the main crop of study [18]. Sujatha and Isakki employed classification algorithms like ANN, J48, Naïve Bayes, Random Forest, and help Vector Machines. And the features including climatic and from soil parameters intheir modeling [19].

III. Methodology

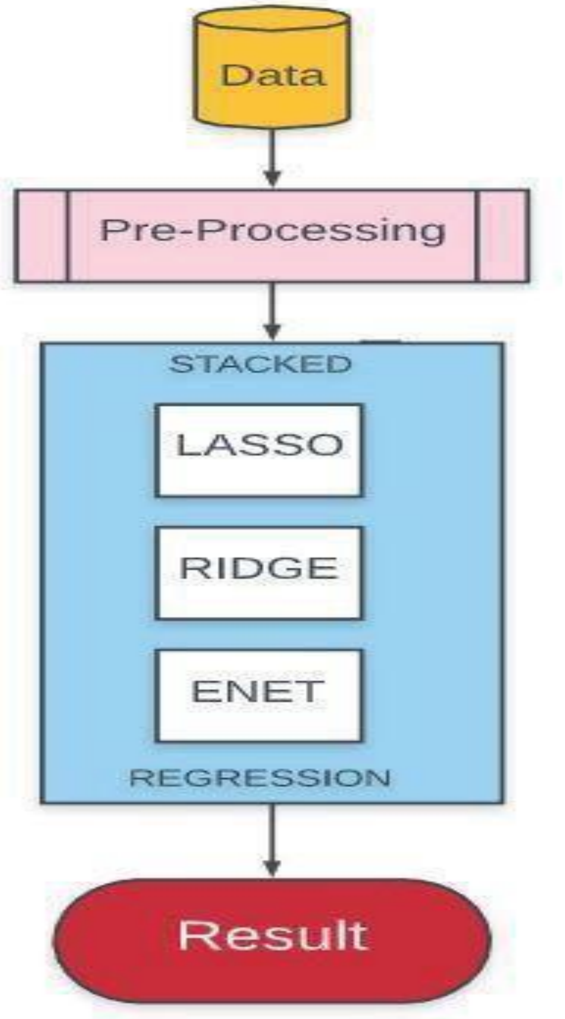


Fig. 2 Research paper Process Chart

A. *Pre-processing*

The data mining on the set of data contains many instances where NA (not available) values have been omitted in Python. Moreover, this dataset being numeric the robust scaling method is used to divide each observation by the average of whole dataset and subtract the average from each observation. It is as if robust scaling corrects for the nominal values, and instead uses the middle and upper values. Normalization (that is often called rescaling) squeezes (normalizes) the data so that it fits into the range of 0 to 1.

B. *Stacked Regression:*

Basically, we combined an AI model with a method called averaging to make it more advanced. First, we split the dataset into two parts: one for teaching the model and the other to check how well it's learned. Next, the base models are moved on to training by incorporating the training set. Then, let's say, the models are applied to the holdout set where they are tested and validated. The next step is to utilize the predictions made from the holdout set in training the upper-level metrics defined as the "meta-model".

This job is being repeated till the outcomes are presentable. This is demonstrated in a 5-fold stacking mechanism consisting of a training dataset divided into 5 folds. A repeated iteration of 5 rounds is performed. In each loop of the learning, base models are evaluated on 4 folds and are also used to predict the remaining one (holdout fold). In 5 iterations of the process all data points will have involved to produce oof fold predictions that, in turn, become features to train the meta-model.

Measures of the preceding forecasts by the base models of the test data form meta-features. These meta-features are average values that will be used as the inputs of the final prediction by the meta-model. Here, the meta model is Lasso Regressor. Consequently, the naming is Lasso regression and the placement is seen at the top of the figure 2. As for the diagram representing the stacked regression in Figure 4, it is as shown below.

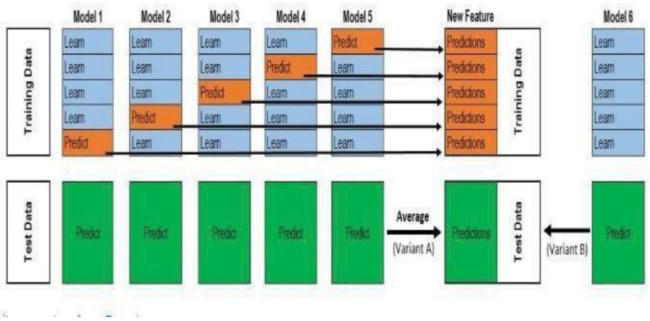
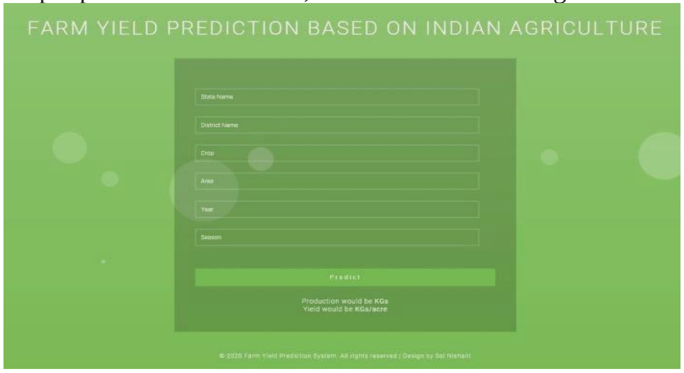


Fig. 3 The Stacked Regression

C. *Output:*

Here the performance parameter to be used is the Root Mean Square Error (RMSE). If utilized separately, the ENet was throwing a model with RMSE at about 4%. The Lasso model had an error rate of approximately 2% and the Kernel Ridge performed with an error of about 1%. After stacking the results which comprised the combined model turned out to have the RMSE of near 0.1%. Farmers will have options to enter necessary data via the application interface and instantly get the output predictions delivered, Visualized in below figure.


FARM YIELD PREDICTION BASED ON INDIAN AGRICULTURE

State Name

District Name

Crop

Area

Year

Season

Predict

Production would be KGs
Yield would be KGs/acre

© 2020 Farm Yield Prediction System. All rights reserved | Design by Sai Nishant

Fig. 4 Web Application Interface

IV. CONCLUSION AND FUTURE SCOPE

Following the performance of stacking regression, it was observed that these models brought improvements compared with separate implementation of the models. At this stage, the output can be seen either via the web application or by importing in the Google Sheets. In future along with this we would like to develop a viable mobile application that farmers can use. The candidate system will equally target the localization of the facility as this will ensure that the farmer understands.

REFERENCES

[1] The sanctioned gate ofdata.gov.in offers an expansive collection of datasets.

[2] Ananthara,M.G., Arunkumar,T., & Hemavathy,R.(2013)." CRY An agrarian yield vaticination model employing freak hive clustering fashion." In Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering(pp. 473- 478). IEEE.

[3] Awan,A.M., & Sap,M.N.M.(2006)." A kernel- grounded intelligent system for soothsaying crop yields." In Proceedings of the Pacific- Asia Conference on Knowledge Discovery and Data Mining(pp. 841- 846). Springer.

[4] Bang,S., Bishnoi,R., Chauhan,A.S., Dixit,A.K., & Chawla,I.(2019)." Fuzzy sense approach for crop yield vaticination exercising Temperature and downfall parameters read through ARMA, SARIMA, and ARMAX models." In Proceedings of the 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)." Algorithmic foretelling of Agrarian Productions Using Data Analytics and the Hybrid System". In Proceedings of the 2018 Fourth International Conference on Computing, Communications, Control and Robotics (ICC1- 5). IEEE.

[6] Gangghe,Y.(2017). 'Different Data Mining Ways for Crop Yield Foretelling in the proceedings of 2017 International conference on electrical, electronics, communication, computer, and optimization ways(ICEECCOT)' (pp.420- 423). IEEE.

[7] Gandhi, N., Petkar, O., and Armstrong, L. J. (2016). "Artificial neural network predictions of rice crop yields". In 2016 IEEE agricultural and rural development technologies conference: Technological innovations in ICT for agriculture and rural development (TIAR) Proceedings, Article no.: (pp.105- 110). IEEE.

[8] Gandhi N, Armstrong LJ, Petkar O and Tripathy AK (2016). "Predicting rice yields in India using the support vector machines". JCSSE: 13th International Joint Conference on Computer Science and Software Engineering, Proceedings (pages 58-61).1- 5). IEEE.

[9] The study conducted by Gandhi, N., Armstrong, L.J., and Petkar, O (2016)." Prediction system for Indian rice crop yields through decision support.", has been published in TIAR Conference on ICT in agriculture.13- 18). IEEE.

[10] Islam,Institutions recognized by the researchers of the study are The R.ST.,Haq,K.K.,&T., Chisty,A.K.(2018)." The Intelligent Approach Based on the Deep Neural Network for the Crop Yield Prediction in Bangladesh." In the Proceedings of the 2018 IEEE Region 10 Humanitarian Technology Conference1- 6). IEEE.

[11] Jaikla, R., Auephanwiriyakul, S., & Jintrawet, A., (2008). Application of support vector machine for rice yield prediction. In Proceedings of the 5th International Conference on Electrical Engineering / Electronics , Computer, Telecommunications and Information Technology(Vol.1,pp.29- 32). IEEE. 12. Kadir, M.K.A., Aob, M.Z., & Miniappan, N. (2014). Predicting potentials of wheat yield through Artificial Neural Networks. In IEEE, 4th International Conference on Engineering Technology and Technopreneurship (ICE2T) (pp.161- 165). IEEE.

[12] Kadir,M.K.A., Ayob,M.Z., & Miniappan,N.(2014)." vaticinating wheat yields using Artificial Neural Networks." In Proceedings of the 2014 4th transnational Conference on Engineering Technology and Technopreneuship(ICE2T)(pp. 161- 165). IEEE.

[13] Manjula, A. & Narsimha, G. (2015). "XCYPF Framework for Smart Farms Boosts Crop Yields" In Proceedings of the 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), pp.1- 5). IEEE.

[14] Similarly to this, Mariappan,A.K. & Das,J.A.B.(2017) state the example of a model for rice yield forecasting in Tamilnadu.It is presented in the 2017 IEEEProceedingsthe technological inventions about ICT for Agriculture and Rural Development.18- 21). IEEE.

[15] Shah, A., Dubey, A., Hemnani, V., Gala, D, and Kalbande, D.R. (2018)," A Smart Farming System Using Regenssion Model to Predict Production Outputs," in the proceedings of the International Conference on Wireless Communication.49- 56). Springer.

[16] Ahemad, A.M.S., Nahid, N.T., Nazim, H., Kabir, M.T., Kumar, D., Fazle, R., and Rokhum, R.M. (2015) 'Revealing Crop Yield Prediction Trends in Bangladesh using Data Mining'. In AI, Springer, pp. 491-51-6). IEEE.

[17] A.,Shastry, H. S., M.,Hegde,. (2015)." An ANFIS model for crop yield prognostication parameters". In the Proceedings of the 2015 IEEE International Advanced Computing Conference(IACC) (pp.253- 257). IEEE.

[18] R. Sujatha and P. Isakki (2016) . "Palmistry of crop harvesting "In Proceedings of the International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp.1-4). IEEE.