2020 International Conference for Emerging Technology (INCET)
Belgaum, India. Jun 5-7, 2020

Crop Yield Prediction based on Indian Agriculture using Machine Learning

Potnuru Sai Nishant[1], Pinapa Sai Venkat[2], Bollu Lakshmi Avinash[3], B. Jabber[4]
[1,2,3,4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, A.P., India.
[1]sai.nishant8@gmail.com, [2]pinapasaivenkat98@gmail.com, [3]avichowdary999@gmail.com, [4]jabberhi5@gmail.com

***Abstract*—In India, we all know that Agriculture is the backbone of the country. This paper predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction.**

Keywords— Crop yield prediction, Lasso, Kernel Ridge, ENet, Stacked Regression.

I. INTRODUCTION

In our research, which we found in the previous research papers is that everyone uses climatic factors like rainfall, sunlight and agricultural factors like soil type, nutrients possessed by the soil (Nitrogen, Potassium, etc.) but the problem is we need to gather the data and then a third party does this prediction and then it is explained to the farmer and this takes a lot of effort for the farmer and he doesn't understand the science behind these factors. To make it simple and which can be directly used by the farmer this paper uses simple factors like which state and district is the farmer from, which crop and in what season (as in Kharif, Rabi, etc.).

In India, there are more than a hundred crops planted around the whole country. These crops are categorized for better understanding and visualization. The data for this research has been acquired from the Indian Government Repository [1]. The data consists of attributes – State, District, Crop, Season, Year, Area and Production with around 2.5 Lakh observations. The fig. 1. depicts the states and territories of India which visualize that which category of crops are famous in which season. We used advanced regression techniques – Lasso, ENet and Kernel Ridge and further we used stacking of these models to minimize the error and to obtain better predictions. This paper is set out as follows: Literature Survey, Methodology, Conclusion and Future Work.

II. LITERATURE SURVEY

Ananthara, M. G. et al. (2013, February) proposed a prediction model for datasets pertaining to agriculture which is called as CRY algorithm for crop yield using beehive clustering techniques. They considered parameters namely crop type, soil type, soil pH value, humidity and crop sensitivity. Their analysis was mainly in paddy, rice and sugarcane yields in India. Their proposed algorithm was then compared with

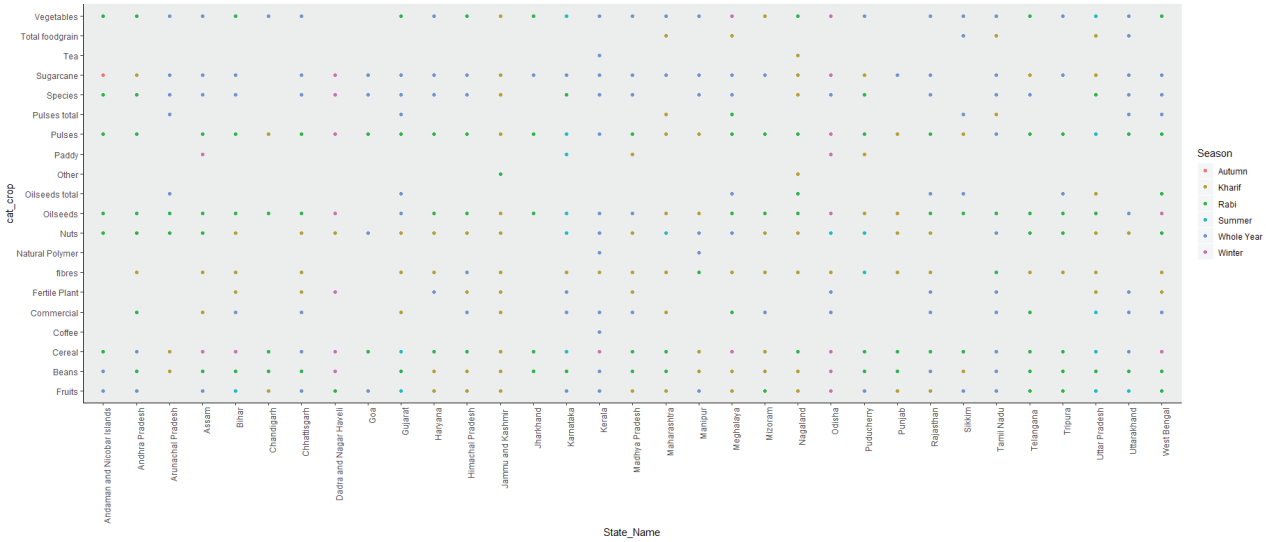

Fig. 1. Famous Categories of crops over states in India (based on Season)

978-1-7281-6221-8/20/$31.00 ©2020 IEEE

C&R tree algorithm and it outperformed well with an accuracy of 90 percent [2]. Awan, A. M. et al. (2006, April) built a new, smart framework focused on farm yield prediction clustering kernel methodology and they considered parameters like plantation, latitude, temperature and precipitation of rainfall in that latitude. They had experimented weighted k-means kernel method with spatial constraints for the analysis of oil palm fields [3]. Chawla, I. et al. (2019, August) used fuzzy logic for crop yield prediction through statistical time series models. They considered parameters like rainfall and temperature for prediction. Their prediction was classification with levels 'good yield' , 'very good yield' [4]. Chaudhari, A. N. et al. (2018, August) used three algorithms namely clustering k-means, Apriori and Bayes algorithm, then they hybridized the algorithm for better efficiency of yield prediction and they considered parameters like Area, Rainfall, Soil type and also their system was able to tell which crop is suitable for cultivation based on the mentioned features [5]. Gandge, Y. (2017, December) used many machine learning algorithms for different crops. They studied and analyzed which algorithm would be suitable for which crop. They have used K-means, Support vector Regression, Neural Networks, C4.5 Decision tree, Bee-Hive Clustering, etc. The factors implying were soil nutrients like N, K, P and soil ph. [6]. Armstrong, L. J. et al. (2016, July) used ANNs for the prediction of rice yield in the districts of Maharashtra, India. They considered climatic factors namely (considering range) temperature, precipitation and reference crop evapotranspiration. The records were collected from Indian Government repository from 1998 to 2002 [7]. Tripathy, A. K. et al. (2016, July) were same authors who used support vector machines to predict the rice crop yield with same features as the previous paper mentioned [8]. Petkar, O. (2016, July) were also the same authors who applied for SVM and neural networks for rice crop yield prediction proposed a new decision system which is an interface to give the input and get the output [9]. Chakrabarty, A. et al. (2018, December) analyzed crop prediction in the country of Bangladesh where they majorly cultivate three kinds of rice, Jute, Wheat, and Potato. Their research used a deep neural network where the data had around 46 parameters into their consideration. Few of them were soil composition, type of fertilizer, type of soil and its structure, soil consistency, reaction and texture [10].

Jintrawet, A. et al. (2008, May) used SVR model for crops like rice to predict the yield where the model was divided into three steps- predicting the soil nitrogen weight followed by prediction of rice stem weight and rice grain weight respectively. Their factors were solar radiation, temperature and precipitation along with those three steps [11]. Miniappan, N. et al. (2014, August) used artificial neural network in modelling multi-layer perceptron model with 20 hidden layers for prediction wheat yield which considered factors like sunlight, rain, frost and temperature [12]. Manjula, A et al. built a crop selection and to predict the yield which considered various indexes like vegetation, temperature and normalized difference vegetation as factors. They distinguished between climate factors and agronomic factors and other disturbances caused in the prediction for better understanding [13]. Mariappan, A. K. et al. analyzed the data regarding rice crop in the state of Tamil Nadu, India. They have considered factors like soil, temperature, sunshine, rainfall, fertilizer, paddy, and type of pest used and other factors like pollution and season [14]. Verma, A. et al. (2015, December) used classification techniques like Naïve Bayes, K-NN algorithm for crop prediction on soil datasets which constituted nutrients of soil like zinc, copper, manganese, pH, iron, Sulphur, Phosphorous, Potassium, nitrogen, and Organic Carbon [15]. Kalbande, D. R. et al. (2018) used support vector regression, multi polynomial regression and random forest regression for prediction of corn yield and evaluated the models using metrics like errors namely MAE, RMSE and R-square values [16]. Rahman, R. M. et al. (2015, June) used mainly clustering techniques for crop yield prediction. The paper explained the analysis of major crops in Bangladesh and divided the variables into environmental and biotic variables. The algorithms applied were linear regression, ANN, and KNN approach for classification [17]. Hegde, M. et al. (2015, June) used multiple linear regression and neuro fuzzy systems for predicting crop yield by taking biomass, soil water, radiation and rainfall as input parameters for the research and their majorly concentrated crop was wheat [18]. Sujatha, R., & Isakki, P. (2016, January) used classification techniques like ANN, j48, Naïve Bayes, Random Forest and Support vector Machines. They have also included both climatic parameters and soil parameters as features in their modelling [19]. Ramalatha, M. et al. (2018, October) used a hybrid approach of combining K-means clustering and classification based on modified K-NN approach. The data was collected from Tamil Nadu, India where the majorly concentrated crops were rice, maize, Ragi, Sugarcane, and Tapioca [20]. Singh, C. D. et al. (2014, January) developed an application to advise crops which works on selected districts of Madhya Pradesh, India. The user would give input cloud cover, rainfall, temperature, observed yield in the past and the system would predict the yield and Depending on the trigger values set, the crop will be labeled and obtain the results [21].

III. METHODOLOGY

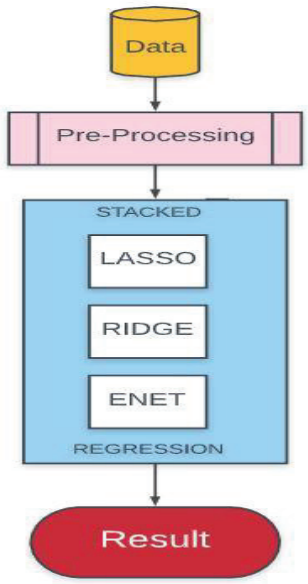

Fig. 2. Process chart of the research project

A. *Pre-processing*

For the given data set, there are quite a few 'NA' values which are filtered in python. Furthermore, as the data set consists of numeric data, we used robust scaling, which is quite similar to normalization, but it instead uses the interquartile range whereas normalization is something which normalization shrinks the data in terms of 0 to 1.

B. *Stacked Regression:*

This is a kind of ensembling but a little of enhancement of averaging. In this, we add a meta model and use the out of fold predictions of the other models used to train the main meta model.

Step-1: the total training set is again divided into two different sets. (train and holdout)

Step-2: train the selected base models with first part (train).

Step-3: Test them with the second part. (holdout)

Step-4: Now, the predictions obtained from test part are inputs to the train higher level learner called meta-model.

Iteratively, the first three steps are completed. For example, if we take a 5-fold stacking, we divide the training data into 5 folds first. We'll then do 5 iterations. We train each base model on 4 folds in each iteration and predict the remaining fold (holdout fold). So, after 5 iterations, we'll be confident that all the data will be used to get out - of-fold predictions that we'll use as a new feature in Step 4 to train our meta-model. We average the predictions of all base models on the test data for the predictive portion and used them as meta-features on which the meta-model is finally predicted. Here, our meta model is Lasso Regressor and that's the reason for being placed at the top in fig. 2. The stacked regression working can be understood from the fig. 3.

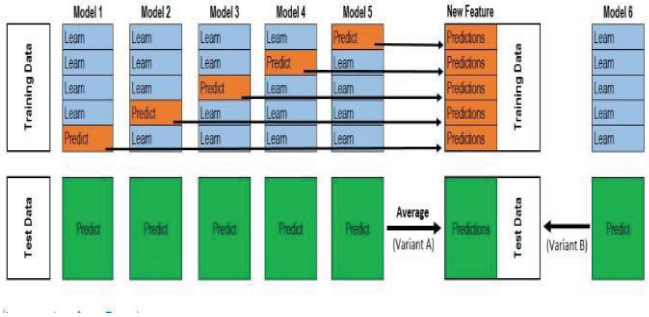

Fig. 3. Stacked Regression

C. *Output:*

The performance metric used in this project is Root mean square error. When the models applied individually, for ENet it was around 4%, Lasso had an error about 2%, Kernel Ridge was about 1% and finally after stacking it was less than 1%. The user or the farmer can enter the following details over the web application to get the prediction as depicted below in the fig. 4.

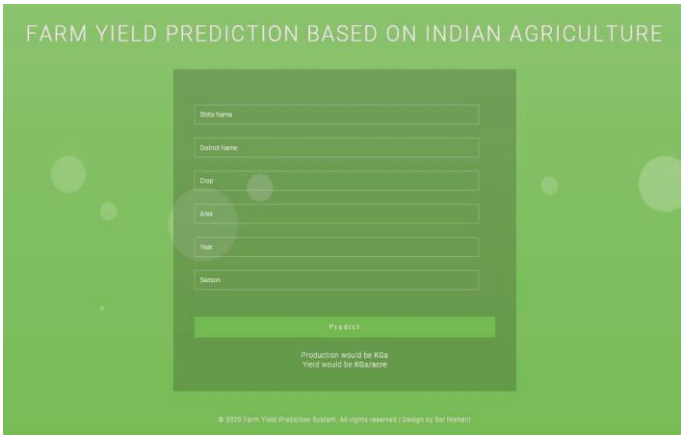

Fig. 4. Interface of Web App

IV. CONCLUSION AND FUTURE WORK

When we apply stacked regression, the result has been so improvised than when those models were applied individually. The output which has been shown in figure is currently a web application, but our future work would be building an application where the farmers can use it as app and converting the whole system in their regional language.

REFERENCES

[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 (ICEECCOT)* (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] Shastry, 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 Syst ems (ICCES) (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.