HBRP PUBLICATION

Journal of Advances in Computational Intelligence Theory
Volume 3 Issue 2

Prognosis of Yield of Crop using Machine Learning Techniques

***Jeevitha J*[1], *Megha G S*[2*]**
[1]*Student,* [2]*Assistant Professor*
Department of Computer Science and Engineering
East West College of Engineering, Bangalore, Karnataka-560064, India
Corresponding Author
E-mail Id:- *meghags@ewce.edu.in*

ABSTRACT

India is the country which is dependent on agricultural filed where the economic status entirely or partially depends on this. Selection of crops is the major phase in agricultural planning or for cultivation. But selection of crops depends on various factors such as climatic conditions, market-price, production rate etc. We can improve the agriculture production in our country by using machine learning models which is applied on different sectors of farming. With improvement in the machine learning models the ability to improve the prediction of crop yield is also high. In this paper we use techniques of advanced regression such as Kernel Ridge, Lasso, E-net and polynomial regression algorithms to detect yield and stacked regression conception to enhance the algorithms used for better prediction. This paper mainly uses squared error loss. In addition a user friendly web application is developed for better predictions.

Keywords:-*Crop yield prediction, lasso, kernel ridge, e-net, polynomial and stacked regression.*

INTRODUCTION

Artificial Intelligence is used in machine learning as an application. Machine Learning is defined as understanding the structure of the given data and apply the machine learning models on data which can be understood and utilized by people. Machine learning is used in various fields such as image and voice recognition, online fraud detection, email spam and malware filtering, traffic prediction, product recommendation, medical diagnosis etc. Agriculture is the principal work which is vital for one's living and makes a significant contribution in employment. As the growth rate of population is increasing the need for production of crops is also more which is giving a great pressure nowadays on the agriculture system. But due to many internal and external factors farmers are facing difficulties in every phase of production of crops which is leading to the decrease in the yield of all crops.

Traditional methods of agricultural production has great disadvantages as the time passed and resulted in great loss of not only crop yield but also resulted in fertility od soil and many other factors. And the unnatural techniques developed to increase the yield have great side effects on the global warming. The problem faced by Indian agriculture is the lack of proper technology usage to get desired output. There are many techniques which can be applied in agricultural sectors such as artificial intelligence, deep learning, IoT, data mining, machine learning etc. Machine Learning is the advanced techniques used in modern agriculture industry. So we use advanced machine learning models to overcome the productivity of crop and several other problems. Machine learning in agriculture is used in various aspects such as for

disease and weed detection, species identification, quality of the crop and mainly to detect the yield of a crop.
The previous papers used the simple machine learning techniques such as Support Vector Machine, Random Forest, Linear regression etc., where these models had more percentage of errors and less accuracy of prediction. The features which were used are climatic conditions such as rainfall, sunlight and agricultural factors such as soil type, fertility of soil, nutrients possessed by soil etc., which are heavy features to apply for the machine learning models and gives us more errors. The issue is that we need to collect the data from all throughout the country and give it to third party to do the forecast and to get clarified to the farmer, which necessitates a great deal of labor on the farmer's part but where the farmer does not understand science and technology. So to make it easy, this paper we use factors such as state, district, crop, season area and year for the prediction of that crop and develop a web application which can be used directly by the farmers. The data for this research is acquired from the Indian Government Repository which contains about 2.5 lakhs observations telling which crop is suitable in which state and for the particular season. Using this data the crop yield can be predicted for all types of crops [vegetables, fruits, pulses, cereals, paddy etc.] for every state and for all types of season. We use techniques of advance regression- Lasso, E- net, Kernel ridge, Polynomial Regression and further stacking is done on all these models combined together for enhanced accuracy and to minimize error.

LITERATURE SURVEY

Pavan Patil (2020, February) [1] have created a system by adding more attributes and ameliorate the results, which improves the yield and can recognize several patterns for predictions by using classification algorithms Decision Tree and K- Nearest Neighbour (KNN) algorithms.

Sujatha Terdal (2019, August) [2] have used three methods- K-Nearest Neighbour(KNN), Support Vector Machine(SVM) and decision tree algorithms and have used three datasets- clay dataset, precipitation dataset and production dataset. Python programming and spyder tool are used to implement these algorithms. The mean absolute error is used to compare performance along with cross validation and precision to get the expected result.

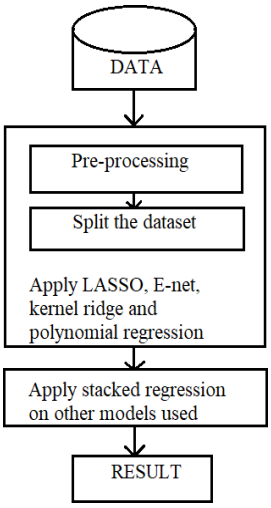
Shivani S. Kale (2019, December) [3] describes the development of a different crop yield prediction model with ANN (Artificial Neural Network), with 3 layer neural network. This paper uses Rectified Linear Activation unit and uses forward and backward propagation techniques. This yield prediction helps the farmers for better decision making about crop harvesting. The model also suggests the success rate for the crops inputted by farmers.

M. Kalimuthu (2020) [4] it helps the beginner farmer by guiding them for sowing reasonable crops using a supervised machine learning algorithm that is Naïve Bayes Gaussian classifier with boosting algorithm to detect the crop at high precision of 97%. This model helps to predict the seed as an output for the given input parameters.

Dr. Y. Jeevan Nagendra Kumar (2020) [5] the machine learning models used are- Random Forest, Naïve Bayes Classifier and decision trees. The yield of the crop is predicted by using the factors such as temperature, humidity, pH, rainfall, crop name. Use of Random Forest as main model gives the highest accuracy of predicting the yield of the crop and aids the farmers in deciding which crop to cultivate on the land.
Ramesh Medar (2019, March) [6] it helps in implementing the crop selection method

which helps in solving many agriculture and farmers problems by using Naïve bayes Classifier and K-Nearest Neighbour (KNN). These techniques help in getting maximum crop yield. Sensor technologies are implemented which helps in getting maximum yieldrate of the crops and help in selecting proper crop forcultivation.

METHODOLOGY



Pre-processing

It is the process of preparing a raw data for making the data suitable for the model used. Pre-processing helps in removing or filling the null values and missing values. Robust Scaling is used here for applying the pre-processing on the data because the dataset contains numeric value which uses normalization and put the values in terms of 0 and 1.

Split the dataset

After doing pre-processing on the data split the dataset into two sets- training sets and testing tests. Training sets are used for training the machine learning models and testing sets tests the machine learning model to predict the outcome. For the methodology used here we need to again split the training sets into two sets- train and holdout.

Apply machine learning models

Train the selected base models that is kernel ridge, E-net and polynomial regression with the first part of the training set (train). Then test them with second part(holdout) iteratively for all the dataset. Now the predictions obtained from the second part(holdout) will become the inputs to train the higher level learner meta model. Here the meta model is the lasso regressor.

LASSO Regressor

LASSO stands for Least absolute shrinkage and selection operator. It's a unique method of regression analysis which does variable selection as well as regularization to improve the precision of the forecast. It is used when there are more number of features because it performs feature selection automatically.

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

where,
λ denotes the amount ofshrinkage.

Kernel Ridge Regression

Non-parametric form of ridge regression is known as kernel ridge regression. It combines ridge regression with other techniques using the kernel ruse. This algorithm aims learn a function int the space included by respective kernel and data by minimizing the squared loss witha squared norm regularization.

$$\hat{\beta} = (X^T X + \lambda \mathrm{I})^{-1} X^T y$$

where,
X-the data matrix,
XT-the transpose of X
λ-the conditioning factor I-the identify matrix
y-a vector of values of the dependent

Efficient neural network(E-net)

The E-net (Efficient Neural network) allows for real time pixel-by-pixel semantic segmentation. E-net is up to 12 times faster than current versions, uses 75 times fewer FLOP's, has 79 times fewer

parameters and has equivalent or better accuracy.

Polynomial Regression

It is a sort of statistical regression analysis that models the relationship between the independent variable x and dependent variable y as nth degree dependent variable in x, a polynomial. Polynomial regression denoted E(y|x), matches a noon linear relationship between the value of x and the conditional mean y and function E(y|x) is linear for the unknown parameters estimated in data. So it is also called as multiple linear regression.

Stacking Regression

It is an ensemble machine learning algorithm. It is a technique for creating linear combinations of various predictors to increase prediction accuracy. To evaluate the coefficients in the mixture, cross validation data and least squares under non-negativity constraints are used. Here the predictions obtained by other models are stacked linearly combined together.

Output

Root Mean Square error is the output metric used in this paper. When the models were implemented separately, E-net had a 4% error, lasso had a 2% error, kernel ridge had 2% error and polynomial regression had error around 2.5%. All together the percentage error was more than 1% before stacking and it became less than 1% once the stacking is applied. The farmer can enter the necessary parameter over the web application and can get the prediction as depicted in fig.

CONCLUSION AND FUTURE WORK

Crop yield detection has been a exigent issue faced by farmers as it led to the decrease in economy for India. When stacked regression is used, the results are far superior than when the models were used individually. The output depicted is a web application where specified parameters are passed and crop yield is predicted. The developed web page is user friendly and the accuracy of the predictions are more than 75 percent.

The future work is to construct an application in which agrnomists can use the application and transform the complete structure in their native or zonal language.

REFERENCES

1. Medar, R., Rajpurohit, V. S., & Shweta, S. (2019, March). Crop yield prediction using machine learning techniques. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE.
2. Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019, November). Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.
3. Kalimuthu, M., Vaishnavi, P., & Kishore, M. (2020, August). Crop prediction using machine learning. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 926-932). IEEE.
4. Kale, S. S., & Patil, P. S. (2019, December). A Machine Learning Approach to Predict Crop Yield and Success Rate. In *2019 IEEE Pune Section International Conference (PuneCon)* (pp. 1-5). IEEE.
5. Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., Neha, K., & Devi, V. G. R. R. (2020, June). Supervised

Machine learning Approach for Crop Yield Prediction in Agriculture Sector. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 736-741). IEEE.V. Spandana, V.S. Vaishnavi and K. Neha, ieee.doi: 10.1109/ICCES48766.2020.9137868.

6. Haque, F. F., Abdelgawad, A., Yanambaka, V. P., & Yelamarthi, K. (2020, June). Crop yield analysis using machine learning algorithms. In *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)* (pp. 1-2). IEEE.
7. Jain, S., & Ramesh, D. (2020, February). Machine Learning convergence for weather based crop selection. In *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (pp. 1-6). IEEE.
8. Salpekar, H. (2019, October). Design and implementation of mobile application for crop yield prediction using machine learning. In *2019 Global Conference for Advancement in Technology (GCAT)* (pp. 1-6). IEEE.
9. Kavita, M., & Mathur, P. (2020, October). Crop Yield Estimation in India Using Machine Learning. In *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)* (pp. 220-224). IEEE.
10. Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S., & Janani, A. P. (2020, July). An Effective Crop Prediction Using Random Forest Algorithm. In *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)* (pp. 1-5). IEEE.
11. de Freitas Cunha, R. L., & Silva, B. (2020, March). Estimating Crop Yields with Remote Sensing And Deep Learning. In *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)* (pp. 273-278). IEEE.
12. Sanjana, G., Davasam, N. M., & Krishna, N. M. (2020, October). Smart Farming Using IoT and Machine Learning Techniques. In *2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC)* (pp. 1-5). IEEE.
13. Menon, K. D., Raj Jain, A., & Kumar Pareek, D. (2019). Quantitative analysis of student data mining.
14. Pai H, A., HS, S., Soman, S., Pareek, D., & Kumar, P. (2019). Analysis of causes and effects of longer lead time in software process using FMEA SSRN: https://ssrn.com/abstract=3508574 or http://dx.doi.org/10.2139/ssrn.3508574
15. Pai H, Aditya and H S, Sameena and Soman, Sandhya and Pareek, Dr. Piyush Kumar, ROC Structure Analysis of Lean Software Development in SME's Using Mathematical CHAID Model (May 17, 2019).
16. Pai H, A., HS, S., Soman, S., Pareek, D., & Kumar, P. (2019). ROC Structure Analysis of Lean Software Development in SME's Using Mathematical CHAID Model.
17. HS, S., Soman, S., & Kumar Pareek, D. (2019). Fast and efficient parallel alignment model for aligning both long and short sentences.
18. BR, M., Bhavya, B. R., Pareek, D., & Kumar, P. (2016). Education Data Mining: Perspectives of Engineering Students. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN*, 2347-5552
19. Kotagi, M., & Pareek, P. K. (2016). Survey on Challenges in DevOps. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN*, 2347-5552.
20. Soman, S., & Pareek, P. K. (2020). An exploratory analysis on challenges

prevailing in small and medium IT firms. In *Journal of Physics: Conference Series* (Vol. 1427, No. 1, p. 012010). IOP Publishing.

21. Sangeetha, V., Vaneeta, M., Kumar, S. S., Pareek, P. K., & Dixit, S. (2021). Efficient Intrusion detection of malicious node using Bayesian Hybrid Detection in MANET. In *IOP Conference Series: Materials Science and Engineering* .1022, No. 1, p. 012077). IOP Publishing.Pareek, P.
22. K., Swathi, K., & Shetteppanavar, P. (2017, May). An efficient machine translation model for Dravidian language. In *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* . 2101-2105. IEEE.
23. Aditya Pai, H., Pareek, P. K., Narasimha Murthy, M. S., Dixit, S., & Karamadi, S. (2021). An Exploratory Study for Process Optimization in IT Industry. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, 3* . 617-631. Springer Singapore.
24. Soman, S., Pareek, P. K., Dixit, S., Chethana, R. M., & Kotagi, V. (2021). Exploration Study to Study the Relationships Between Variables of Secure Development Lifecycle (SDL). In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3* (pp. 641-649). Springer Singapore.
25. Suhas, G. K., Devananda, S. N., Jagadeesh, R., Pareek, P. K., & Dixit, S. (2021). Recommendation-Based Interactivity Through Cross Platform Using Big Data. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, 3* (pp. 651-659). Springer Singapore
26. Soman, S., Pareek, P. K., Dixit, S., Kotagi, V.. (2020). An Empirical Investigation on Practicing Secure Software Development in Software Development Life Cycle in Small & Medium Level Software Firms in Bengaluru. *International Journal of Advanced Science and Technology, 29*(7s), 5164.
27. Patil, S. S., Pareek, P.,K., Dinesh, H. A., Arlimatti, S.(2017). Review of relay selection techniques in multi-hop wireless sensor network with iot. *International Journal of Creative Research Thoughts (IJCRT), 5*(4).846-850