International Journal of Modern Developments in Engineering and Science
Volume 3, Issue 1, January 2024
https://www.ijmdes.com | ISSN (Online): 2583-3138

Crop Yield Prediction Using Machine Learning and Deep Learning Models

Archek Praveen Kumar[1], Kotte Venkatesh[2], Yeragorla Sunil[3], Kagidala Mahejabeen[4], Mukkamalla Usman[5*], Onteddu Sreekanta Reddy[6]

[1]*Professor, Department of Electronics and Communication Engineering, Gates Institute of Technology, Ananthapuram, India*
[2,3,4,5,6]*Student, Department of Electronics and Communication Engineering, Gates Institute of Technology, Ananthapuram, India*

***Abstract*: In this project we are using machine learning and deep learning algorithms to predict future crop yield based on weather data such as temperature and rainfall. If farmers know the crop yield before sowing based on historical weather data, then he may take better decision. So, by employing machine/deep learning algorithms we can inform farmers about future crop yield. In proposed method we are using Irish Maize and Potato yield dataset to train all machine learning models and then these models can be used to predict future crop yield. In proposed method we are using random forest, SVR, DNN, CNN, ANN and LSTM. So, we have implemented all 6 algorithms on both datasets. To evaluate performance of each algorithm we are calculating MSE and R2 Score where MSE refers to mean square error (difference between TEST crop yield and predicted yield). R2 refers to correct prediction rate. So, for any algorithm MSE must be lower and R2 must be higher for better crop yield prediction.**

***Keywords*: crop yield prediction, mean square error, R2 value, RF, ANN, DNN, LSTM.**

1. Introduction

The study aims to develop and implement machine learning models to accurately predict crop yields for Irish Potato and Maize crops. The objective is to analyze data to identify key factors influencing crop yields and identify the most significant impact. Machine learning is a fast-growing approach that helps agriculture make viable decisions and improves throughput. Various machine learning classifiers, such as Logistic Regression, Naïve Bayes, and Random Forest, are applied to predict crop yields. The Random Forest algorithm provides the best accuracy when all parameters are combined. The study focuses on supervised learning techniques for crop yield prediction, focusing on the Random Forest algorithm for accurate predictions.

Agriculture is a vital economic activity, accounting for 14% of the Indian economy and a significant source of income for 70% of the population. It plays a vital role in the country's economic development, with research focusing on improving soil fertility, irrigation facilities, fertilizer usage, and pesticide usage. Crop yielding can be improved through genetic development and crop selection based on favorable conditions. Techniques like cropRS and CSM have been developed to improve crop yields.

2. Literature Review

[1] S. S. Sannakki and V. S. Rajpurohit, proposed a "Classification of Pomegranate Diseases Based on Back Propagation Neural Network" which mainly works on the method of Segment the defected area and Color and texture are used as the features. Here they used neural network classifier for the classification. The main advantage is it Converts to L*a*b to extract chromaticity layers of the image and Categorisation is found to be 97.30% accurate. The main disadvantage is that it is used only for the limited crops.

[2] P. R. Rothe and R. V. Kshirsagar introduced "Cotton Leaf Disease Identification was using Pattern Recognition Techniques" which wses snake segmentation; here Hu's moments are used as distinctive attribute. Active contour model used to limit the vitality inside the infection spot, BPNN classifier tackles the numerous class problems. The average classification is found to be 85.52%.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic". K-means clustering used to segment the defected area; GLCM is used for the extraction of texture features, Fuzzy logic is used for disease grading. They used artificial neural network (ANN) as a classifier which mainly helps to check the severity of the diseased leaf.

[4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa, proposed" Automated Vision-Based Diagnosis of Banana Bacterial Wilt Disease and Black Sigatoka Disease "Color histograms are extracted and transformed from RGB to HSV, RGB to L*a*b. Peak components are used to create max tree, five shape attributes are used and area under the curve analysis is used for classification.

[5] Uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases," Color features are represented in RGB to HIS, by using GLCM, seven invariant moment are taken as shape parameter. They used SVM classifier which has MCS, used for detecting disease in wheat plant offline.

*Corresponding author: mukkamallausman@gmail.com

3. Proposed Method

In proposed paper Random Forest and SVR algorithm have been implemented along with DNN, CNN, ANN and LSTM on two datasets. To evaluate performance of each algorithm we are calculating MSE and R2 Score where MSE refers to mean square error (difference between TEST crop yield and predicted yield). R2 refers to correct prediction rate. So for any algorithm MSE must be lower and R2 must be higher,

A. Dataset

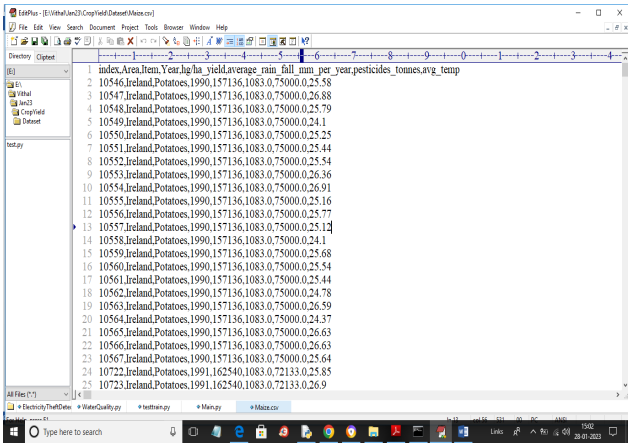

Fig. 1. Dataset

In above dataset screen first row contains dataset column names and remaining rows contains dataset values. We have coded this project using JUPYTER notebook and below are the output screens with code and below colour comments.

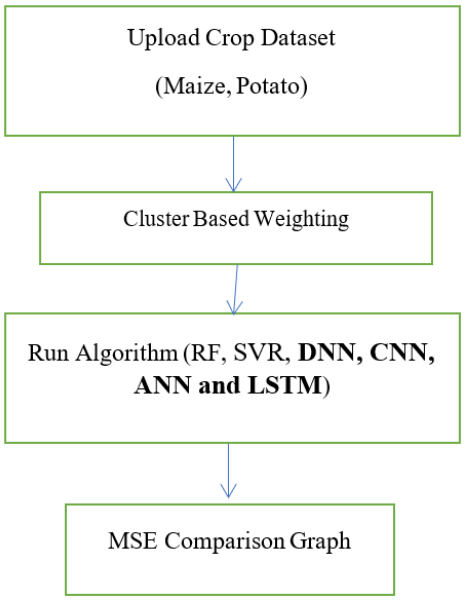

Fig. 2. Flowchart

4. Result

In this paper author employed Random Forest and SVR algorithm to predict future crop yield based on weather data such as temperature and rainfall. If farmers know the crop yield before sowing based on historical weather data then he may take better decision. So by employing machine learning algorithms we can inform farmers about future crop yield. In propose paper author using Irish Maize and Potato yield dataset to train all machine learning models and then this models can be used to predict future crop yield.

As shown in Fig. 1, all the required python packages have been imported.

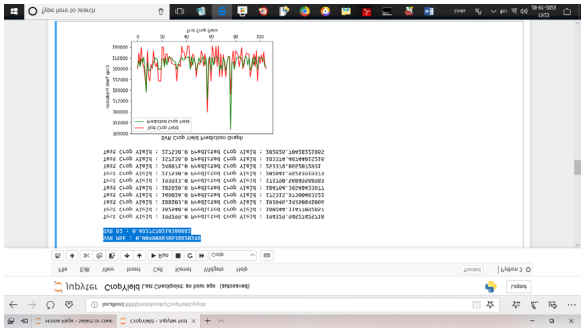

Fig. 3. Graphical representation

In above screen x-axis represents number of test records and y-axis represents Yield values where red line represents TEST yield and green line represents predicted yield and in above graph we can see there lots of gap between red and green line so SVR prediction is not accurate.

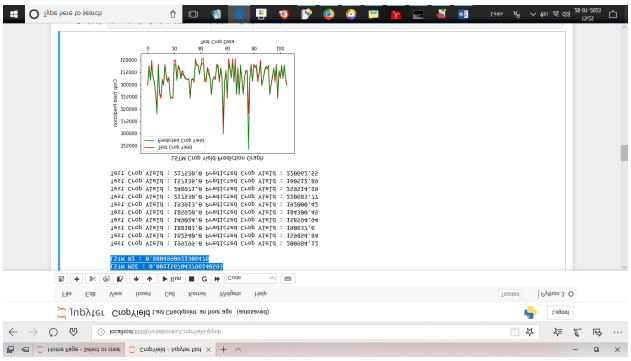

Fig. 4. LSTM MSE and R2 values

In above screen in blue colour we can see LSTM MSE and R2 values and then in next lines we can see TEST crop yield and LSTM predicted crop yield and in LSTM graph we can see both green line and red is fully overlapping so LSTM prediction is accurate

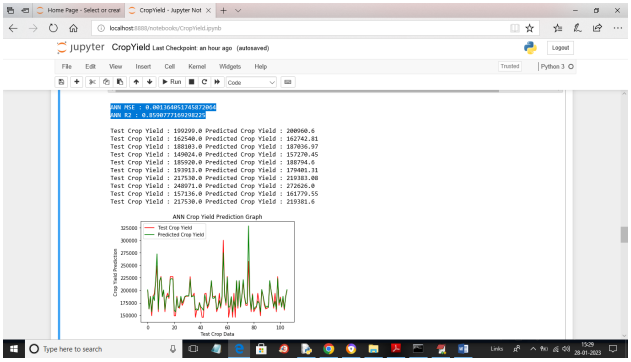

Fig. 5. ANN MSE and R2 values

In above screen in blue colour text we can see ANN MSE and R2 values and then in next lines we can see TEST and predicted crop yield for ANN and then in ANN graph we can see both lines are fully overlapping so ANN prediction also accurate

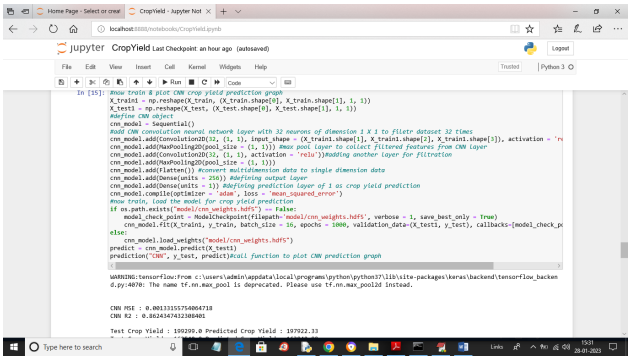

Fig. 6. Training with CNN

In above screen we are training with CNN and after executing above block will get below output

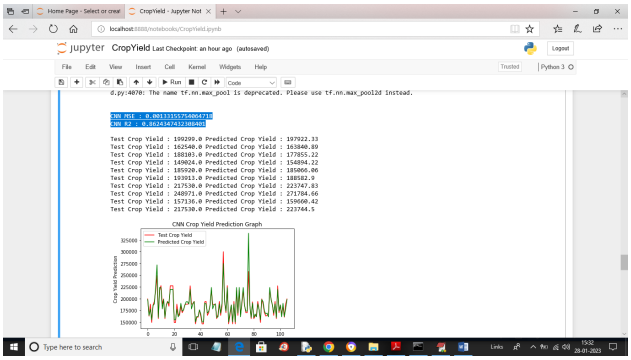

Fig. 7. CNN MSE and R2 values

In above screen we can see CNN MSE and R2 values and then we can see TEST and predicted yield for CNN and then in graph we can see both lines are fully overlapping so CNN prediction is also accurate

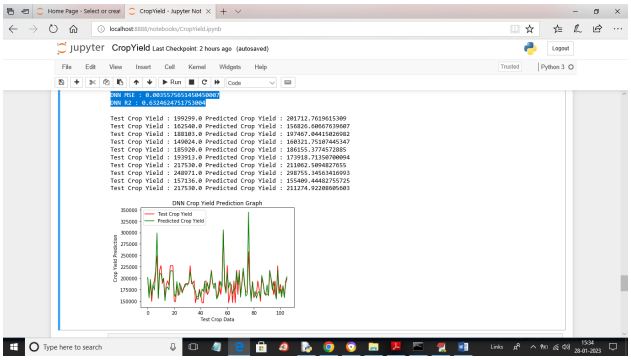

Fig. 8. DNN graph

In above DNN graph there is little difference in red and green line as its contains some gap so DNN prediction is good but not accurate

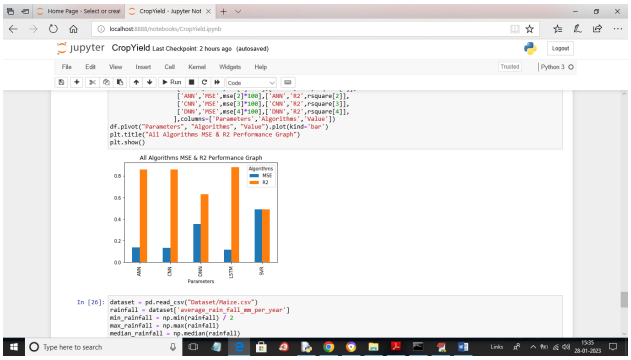

Fig. 9. Graph representation

In above graph blue bar represents MSE and orange bar represents R2 and x-axis represents algorithm names and y-axis represents values and in all algorithms we can see LSTM got high R2 and less MSE compare to all algorithms so we can say LSTM is good at crop yield prediction.

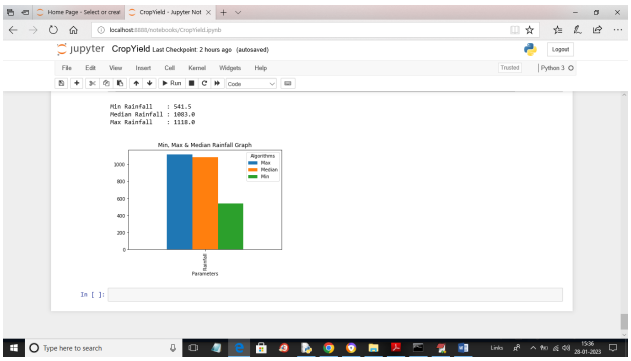

Fig. 10. Plotting MIN, MAX and median

The graph shows the plotting MIN, MAX and median rainfall graph found in the dataset.

5. Conclusion

The research on crop yield prediction using machine learning models holds promise for transforming agriculture. In this project we are using machine learning and deep learning algorithms to predict future crop yield based on weather data such as temperature and rainfall. In proposed method we are using Irish Maize and Potato yield dataset to train all machine learning models and then these models can be used to predict future crop yield. In proposed method we are using random forest, SVR, DNN, CNN, ANN and LSTM. So, we have implemented all 6 algorithms on both datasets. To evaluate performance of each algorithm we are calculating MSE and R2 Score where MSE refers to mean square error.

References

[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal, “Crop Yield Prediction using ML Algorithms,” 2019.
[2] Leo Brieman, “Random Forests”, 2001.
[3] Priya, P., Muthaiah, U., Balamurugan, M., “Predicting Yield of the Crop Using Machine Learning Algorithm”, 2015.
[4] Mishra, S., Mishra, D., Santra, G. H., “Applications of machine learning techniques in agricultural crop production”, 2016.
[5] Jeevan Kumar, “Supervised Learning Approach for Crop Production”, 2020.

[6] Ramesh Medar,Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2019.
[7] Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa, "Machine Learning Methodologies for Paddy Yield Estimation in India: A Case Study", 2019.
[8] Sangeeta, Shruthi G, "Design and Implementation of Crop Yield Prediction Model In Agriculture", 2020.
[9] https://power.larc.nasa.gov/data-access-viewer/
[10] https://en.wikipedia.org/wiki/Agriculture
[11] https://www.ibm.com/weather
[12] https://flutter.dev
[13] https://openweathermap.org
[14] https://builtin.com/data-science/random-forest-algorithm
[15] https://tutorialspoint/machine-learning/logistic-regression
[16] http://scikit-learn.org/modules/naive-bayes
[17] Ebrahimi, M.A., Khoshtaghaza, M.H., Minaei, S., Jamshidi, B., Vision-based pest detection based on SVM classification method. Comput. Electron. Agric., 137, 52–58, 2017.
[18] Chung, C.L., Huang, K.J., Chen, S.Y., Lai, M.H., Chen, Y.C., Kuo, Y.F., Detecting Bakanae disease in rice seedlings by machine vision. Comput. Electron. Agric., 121, 404–411, 2016.
[19] Pantazi, X.E., Moshou, D., Oberti, R., West, J., Mouazen, A.M., Bochtis, D., Detection of biotic and abiotic stresses in crops by using hierarchical self-organizing classifiers. Précis. Agric., 18, 383–393, 2017.
[20] Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., Automatic detection of "yellow rust" in wheat using reflectance measurements and neural networks. Comput. Electron. Agric., 44, 173–188, 2004.
[21] Richardson, A., Signor, B.M., Lid bury, B.A., Bad rick, T., Clinical chemistry in higher dimensions: Machine-learning and enhanced prediction from routine clinical chemistry data. Clin. Biochem. 49, 1213–1220, 2016.
[22] Wildenhain, J., Spitzer, M., Dolma, S., Jarvik, N., White, R., Roy, M., Griffiths, E., Bellows, D.S., Wright, G.D., Tyers, M., Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. Cell Syst., 1, 383–395, 2015.
[23] Kang, J., Schwartz, R., Flickinger, J., and Beriwal, S., Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. Int. J. Radiat. Oncol. Biol. Phys., 93, 1127–1135, 2015.
[24] Craven, B.D. and Islam, S.M.N., Ordinary least-squares regression SAGE. Dict. Quant. Manag. Res., 224–228, 2011.
[25] Friedman, J.H., Multivariate Adaptive Regression Splines. Ann. Stat., 19, 1–67, 1991.
[26] Alonso, J., Villa, A., Bahamonde, A., Improved estimation of bovine weight trajectories using Support Vector Machine Classification. Comput. Electron. Agric., 110, 36–41, 2015.