Machine Learning Techniques For Crop Yield Prediction In Indian Agriculture

[1]G Ravi Kumar, [2]Geetesh Challur, [3]Pradhan Babu Tammana

[1]Associate Professor of CSE, Siddartha Institute of Science and Technology, Puttur, Andhra Pradesh, India
[2,3]Indian Institute of Technology (Indian School of Mines) Dhanbad, Jharkhand, India

ABSTRACT: Machine learning techniques contribute an important part in agricultural crop production and this paper analyses to reevaluate the research findings on the application of it. This is a novel approach to agricultural crop production management which is a decision taking tool. Important policy choices, such as import-export, price marketing distribution, etc., require accurate and timely crop production estimates, which are given by the directorate of economics and statistics. However, it is acknowledged that these estimations are not objective since they require a great deal of descriptive evaluation based on several qualitative aspects. Consequently, there is a need for realistic crop output forecasts that are supported by solid statistical evidence. This advancement in computers and data storage has produced vast quantities of data. In this study, a strategy for crop selection based on weather and soil factors is proposed to optimize agricultural output. It also recommends the optimal planting timing for compatible crops based on seasonal weather predictions. For weather forecasting, machine learning systems such as the recurrent neural network and the Random forest classification algorithm are utilized. The results of the suggested approach for weather forecasting are compared to those of a traditional artificial neural network, which demonstrates superior performance for each of the specified weather parameters. Presented are agricultural applications of time series analysis, Markov chain model, k-means clustering, k closest neighbor, and support vector machine.

Keywords: neural network, Machine Learning, Time series analysis, Crop yield prediction,

1. INTRODUCTION

As per our research, every farmer utilizes ambient condition elements such as rainfall, sunlight, and temperature as well as agricultural characteristics such as soil type and nutrient content (Nitrogen, Potassium, etc.). The yield gained is mostly determined by meteorological conditions, as rainfall patterns have a significant impact on growing techniques. In this setting, farmers and agriculturalists need impromptu assistance in anticipating future harvesting occasions in order to optimize crop output.

Agriculture has not yet reached its full potential due to a lack of technological application. Every farmer wants to know the harvest yield he or she may anticipate, making yield prediction a crucial factor for them. In contrast, crop production forecasting is incredibly difficult due to several complicated factors. Climate, soil quality, topography, insect infestations, water quality and availability, genotype, harvest planning, and other factors largely influence crop production.

As a result of intrinsic human intuition, farmers have developed a sense of the yield pattern through time. However, rainfall as a primary factor in agricultural production can significantly disrupt intuitive yield forecast by influencing certain soil and environmental characteristics associated with crop development. In addition, the farmer is only aware of the type of soil that should be used for a particular crop through written guidance. A key purpose for establishing a regulated market is to forestall unfavorable exchanges, reduce costs inside the commercial center, and give reasonable prices to farmers. Several actions have been done to promote rural advertising as a means to grow and sustain national economic development. To benefit agriculture from the emerging global market and get access to possible results, the internal rural marketing system must also be integrated and enhanced.

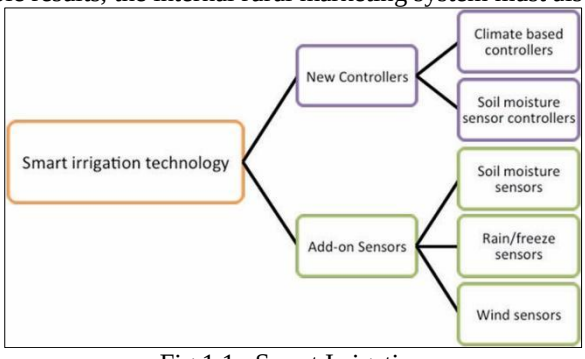


Fig 1.1. Smart Irrigation
(Source: (Gotcher, 2017)

This research aims to examine the impact of combining procedure modeling using machine learning techniques on agricultural production forecasting. These are the particular research objectives: Determine whether a hybrid method (simulation crop modeling + machine learning) would improve corn production estimates. Determine whether hybrid model combinations (different ML x crop model) offer the most precise forecasts.

Identify the most useful crop modeling characteristics for using machine learning to estimate corn yield.

1.1 Aim & Objective

The suggested approach takes into account soil, weather, and previous year's production data and recommends the most lucrative crops that can be grown under the proposed environmental factors. As the system lists out all possible crops, it helps the farmer decide which crop to cultivate. In addition, this approach takes into account historical production data, which provides farmers with a better understanding of crop demand as well as market prices. As this technology will address maximal crop kinds, farmers may learn about crops that have never been farmed.

Objective

1. To increase the accuracy of crop yield prediction as well as to provide an easy to use User Interface.
2. Can examine various climatic characteristics (cloud cover, rainfall, temperature).
3. To develop simple decision models that illustrate how to use stochastic yield estimates to achieve stated agricultural policy objectives.
4. To produce estimates of entire sample industrial production in yields of different crops in certain states, taking into account various technical elements as well as a newly designed meteorological climate index as inputs.

When it comes to crop productivity, the majority of us are preoccupied with its harvest appearance. Predicting crop output is an essential agricultural issue. Each and every farmer endeavors to determine the yield that will be according to his expectations. Historically, yield projections were determined by assessing a farmer's previous experience of a particular crop. The agricultural output is mostly reliant on the weather, pests, but also harvest process layout. For agricultural risk evaluation, it is vital to have reliable data on crop yields of the past.

The scope of the project is to calculate the agricultural yield of a region by analyzing a dataset including key or relevant crop production parameters, such as temperature, humidity, precipitation, and historical crop output. To forecast a continuous value, regression models are utilized. It is a supervised procedure. It focuses on the Construction of a predictive model for agricultural production output.

Shruti Kulkarni (2018) conducted a paper titled “Predictive Analysis to Improve Crop Yield using a Neural Network Model.” In this research, a data-driven model that analyzes and predicts crop production across many districts based on historical soil and precipitation data has been constructed. For this particular crop, Rice is considered. The developed hybrid neural network model detects optimum soil parameter combinations and combines them with the rainfall pattern in a particular location to predict crop production.

This same Time-Series method in Supervised Learning is just the basis for the model’s predictive analysis of precipitation. Recurrent Neural Networks, a kind of Machine Learning model, are utilized to make the ultimate forecast of crop production.

Sonal Jain (2020) conducted a paper titled “Machine Learning convergence with weather-based crop selection inside this categorization and prediction methodologies”. Throughout this article, methods such as Decision tree, Naive Bayes, and Random forest are used for the selection of suitable crops based on weather and soil parameters. A strategy utilizing Naive Bayes, CHAID, Random forest, and K-nearest neighbor that leverages classification criteria, including root depth, surface, soil color, drainage, etc., is provided.

However, this method can only choose one crop, even though multiple crops may be viable for a given location and climate. A naive Bayes classification technique is presented in to give suitability of crops such as chili, rice, maize, and cotton using weather parameters like rainfall, soil moisture, temperature, and atmospheric pressure. This approach also specifies the optimal harvesting and planting times for a specific crop.

According to Mayank Champaneri et.al (2020) Conducted study on “crop yield prediction using machine learning.” This program will aid farmers in estimating their crop production prior to planting, enabling them to make educated decisions. Creating a prototype with only an interactive predictive model is an effort to resolve the problem. This system will be implemented using an intuitive web-based graphical user interface and machine learning algorithms.

The findings of the forecast will be shared with the farmer. Consequently, for this form of data analytics throughout crop prediction, a variety of methodologies or algorithms, as well as with the aid of these algorithms, crop production may be predicted.

In accordance with research conducted by Potnuru Sai Nishan et al. in 2020, “Crop Yield Prediction Based on Indian Agriculture using Machine Learning.” This paper predicts the yield of almost all kinds of crops that are planted in India. The user can anticipate the agricultural production in whatever year they choose by using this script’s novel use of simple parameters like State, district, season, and area.

The research utilizes sophisticated regression methods along with Kernel Ridge, Lasso, and ENet methods to estimate yield, thus the concept of Stacked Regression to enhance existing algorithms but deliver a more precise prediction.

Ankit Singh Chauhan e.al (2019) conducted a paper titled “Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models.” According to study Agriculture plays a significant role in the economy of India. Consequently, the crop production forecast is a crucial responsibility for enhancing India’s economic growth. Temperature and precipitation, among other weather factors, may affect crop growth. Therefore, it is essential to incorporate these characteristics when estimating crop production.

Forecasting the weather is a difficult endeavor. In this research, three forecasting methods are employed: ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average), and ARMAX (ARMA with exogenous variables). This model is utilized to anticipate precipitation and temperature, followed by the use of estimate crop yield but uses a fuzzy logic model based on the outcomes of the previous two models.

So according to Kodimalar Palanivel’s (2019) research, “a method for forecasting agricultural production using machine learning but also big data approaches”, water shortages, uncontrolled costs owing to demand-supply imbalances, as well as weather unpredictability needs farmers to be equipped with intelligent agricultural techniques.

Specifically, low agricultural yields due to uncertain climate change, insufficient irrigation infrastructure, a reduction in soil fertility, and the usage of traditional farming techniques must also be addressed. In this study, the utility of several machine learning techniques for predicting agricultural output has been investigated. A method for predicting agricultural production utilizing machine learning methods and the big data computer paradigm has now been presented.

According to the research conducted by Shivani S. Kale (2019), "A Machine Learning Strategy to Predict Crop Yield as well as Success Rate" This study covers the construction of an alternative model for agricultural production prediction using ANN and 3 Layer Neural Network. And using a large quantity of input and output examples, the ANN model generates a formula to establish the relationship in order to develop a model for yield predictions. Rectified Linear Activation Unit (Relu) is used as an activation function. The backward and forward propagation techniques are used.

2. METHODOLOGY

Methodology in research includes the particular procedures or tactics utilized to locate, select, process, and analyze knowledge about a topic. This identical section of something like a research paper's methodology enables the reader to assess the overall validity and dependability of a study.

Data Pre-Processing

Data Preprocessing would be the transformation of unprocessed data into a cleaned data collection. This same data is collected from several sources and in a certain manner which makes analysis impossible. We may change information into a format that is comprehensible by using strategies such as replacing missing as well as null values. The last stage in data preparation is separating training & testing data.

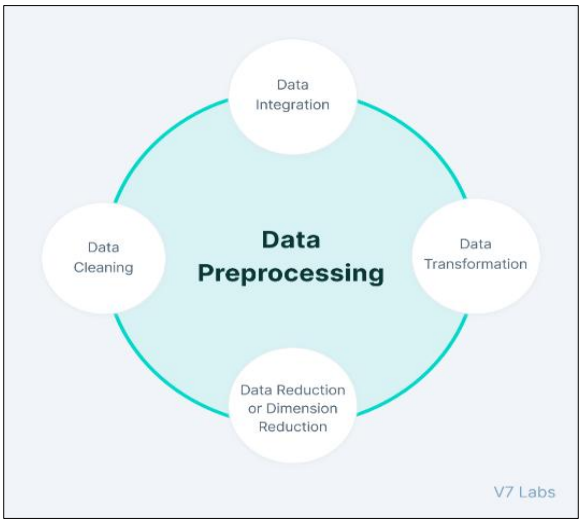


Fig 2.1. Data Processing

Aspects that influence crop yield and production. There are several elements that influence the yield and productivity of every crop. Essentially, these characteristics aid in estimating the annual yield of any given crop. This document includes elements such as Temperature, Rainfall, Area, Relative Humidity, and Wind Velocity.

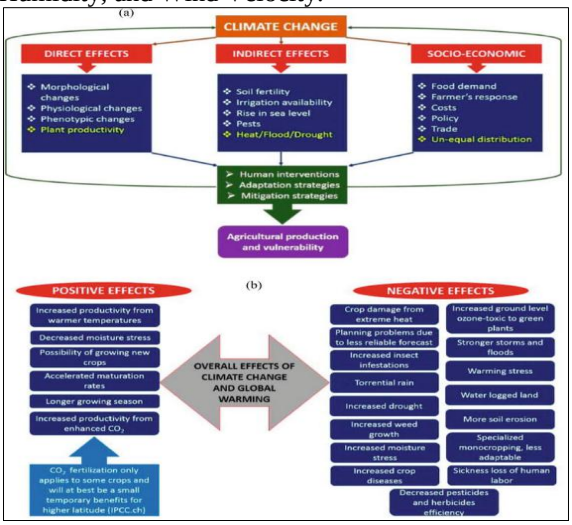


Fig 2.2. Factors affecting Crop Yield and Production Comparison and Selection of Machine Learning Algorithm

Before picking an algorithm to use, we must first analyze and compare them and then choose the one that most closely suits this particular dataset. Machine learning is the best technology for offering a more realistic response to the agricultural productivity challenge. Several techniques for machine learning are applied to anticipate agricultural productivity. Some of the following machine learning methods for selection and accuracy comparison:

Logistic Regression: Logistic regression provides a method for estimating the probability of a target variable that employs supervised learning and categorization. The nature of something like the dependent or targeted variable is dichotomous implying that there are two possible classes.

Naive Bayes: This Naive Bayes classifier assumes that the presence of one feature within a classification is independent of the presence of other features. The Naive Bayes model seems easy to develop and particularly useful for massive data sets.

andom Forest: Random Forest is capable of analyzing crop growth in connection to current climatic conditions but also biophysical change. This same random forest approach produces decision trees from various data samples, predicts that data of each subset, as well as then delivers a more accurate response to the network via voting.

3. PROBLEM STATEMENT

Crop yield prediction is critical for planning and making policy choices. Many nations still rely on ground-based inspections and reports for crop monitoring and production forecasts. These procedures are arbitrary, expensive, and time-consuming. The following are some of the most prevalent issues with present crop production forecast methodologies.

A. Seasonal weather prediction

That data utilized for weather forecasting comes from the NRSA Hyderabad station and contains five years' worth of meteorological characteristics like temperatures, moisture, sun hours, wind direction, and wind direction, among others.

Initial data processing for training your prediction model consists of filling in incomplete data, data transformation, and data standardization. The missing data components are substituted using Linear Interpolation. Continuing with this process, hourly data is translated into daily data, including minimum and maximum temperatures obtained from temperature data. After that, min-max normalization is used to adapt the information to a scale based.

B. Identification of suitable crop

Recurrent neural network (RNN) has been used to forecast seasonal weather. A recurrent neural network (RNN) is a neural network that uses sequential or timed series data. Recurrent neural networks are quite prevalent machine learning approaches that excel in signal analysis, natural language processing, and voice recognition. In contrast to convolutional networks, RNNs effectively handle time series data dependencies since the output of something like the neuron at time t-1 is combined with the subsequent input to feed the neuron at time t. The figure depicts a typical neural RNN structure diagram.

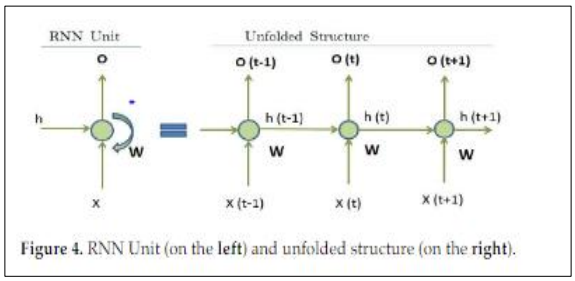


Fig 3.1. RRN Unit and unfolded structure

The advantage of RNN over standard NN is that the preceding step's output is now transferred to the subsequent phase. Consequently, its hidden state in RNN retains the sequence, while still, in conventional neural networks, each value is evaluated separately.

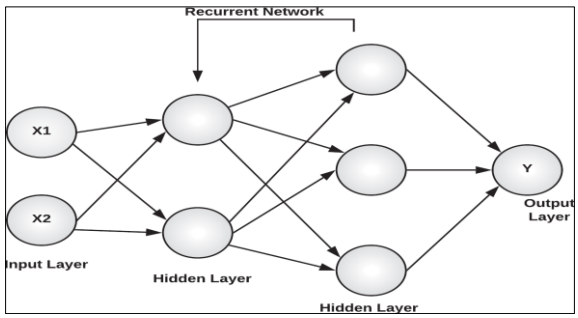


Fig 3.2. Hidden Layer

4. RESULT

The following are the several outputs of this research paper shown:

```
<class 'pandas.core.frame.DataFrame'>
Index: 13 entries, Rice to nan
Data columns (total 8 columns):
2004-05     12 non-null float64
2005-06     12 non-null float64
2006-07     12 non-null float64
2007-08     12 non-null float64
2008-09     12 non-null float64
2009-10     12 non-null float64
2010-11     12 non-null float64
2011-12     12 non-null float64
dtypes: float64(8)
memory usage: 936.0+ bytes
None
```

Fig 4.1. Output 1

As shown above figure 4.1 For implementation of crop yield prediction using Machine learning python language is used. In this we have taken the data from year 2004 to 2012 and the data is stored into float 64 data type.

```
print(crops_prod_data)
```

Crop	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	\
Rice	100.0	101.0	99.0	105.0	112.0	121.0	
Wheat	100.0	101.0	112.0	115.0	117.0	127.0	
Coarse Cereals	100.0	107.0	110.0	115.0	113.0	123.0	
Pulses	100.0	108.0	134.0	124.0	124.0	146.0	
Vegetables	100.0	109.0	103.0	118.0	113.0	124.0	
Fruits	100.0	99.0	99.0	98.0	102.0	104.0	
Milk	100.0	97.0	98.0	98.0	98.0	112.0	
Eggs, Fish and Meat	100.0	102.0	101.0	100.0	99.0	116.0	
Oilseeds	100.0	86.0	85.0	97.0	104.0	103.0	
Sugarcane	100.0	96.0	91.0	87.0	80.0	81.0	
Fibers	100.0	92.0	91.0	96.0	109.0	107.0	
All Agriculture	100.0	99.0	101.0	104.0	106.0	115.0	
NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Fig 4.2. Output 2

As the above figure 4.2 shows, Python is used to implement Machine Learning for predicting crop yields. We used the crop data from 2004 to 2012 to do this.

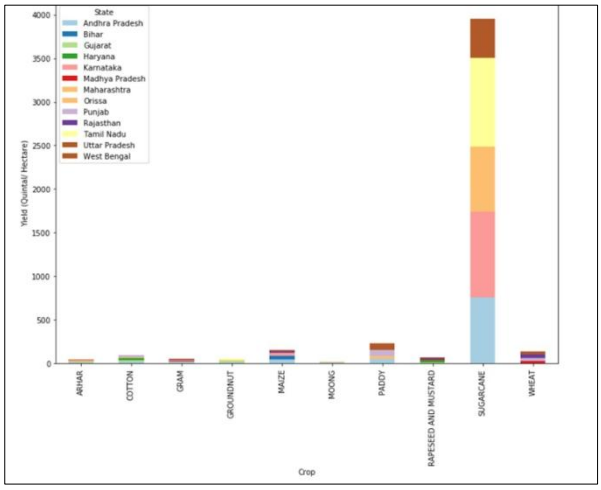


Fig 4.3. Crop

As seen in figure 4.3 above, Python is utilized to create agricultural production prediction using machine learning. We used data from years 2004 to 2012, and the data was recorded as float 64 data.

```
Data columns (total 6 columns):
Crop                                 49 non-null object
State                                49 non-null object
Cost of Cultivation (`/Hectare) A2+FL 49 non-null float64
Cost of Cultivation (`/Hectare) C2    49 non-null float64
Cost of Production (`/Quintal) C2     49 non-null float64
Yield (Quintal/ Hectare)              49 non-null float64
dtypes: float64(4), object(2)
memory usage: 2.4+ KB
None
```

Fig 4.4 Output 3

As seen in figure 4.4, Python is utilized to construct agricultural production prediction using machine learning. We have compiled the data from 2004 to 2012 for this report. As seen in the above figure, cost of cultivation, non-null object, non-null float64.

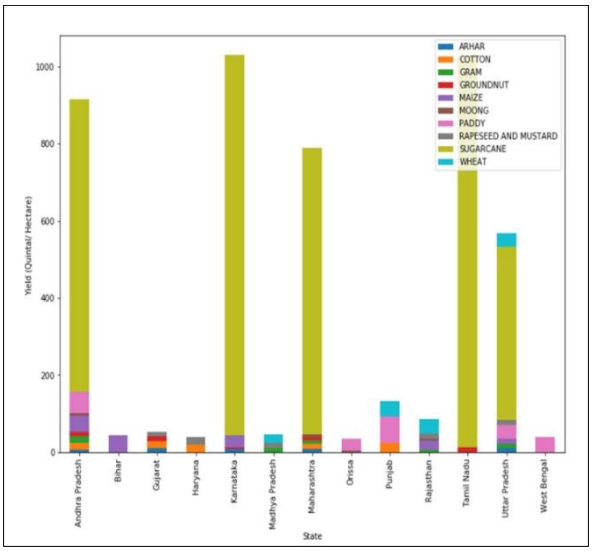


Fig.4.5 Static

As shown in the figure above, figure 4.5, machine learning is used to predict crop yields. The language python is used to make this happen. In this, we took the data from 2004 to 2012 and put it into a float 64 data type.

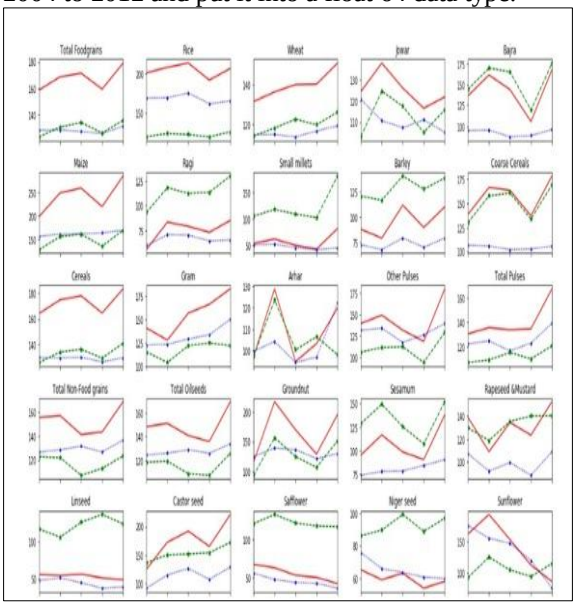


Fig 4.6 Output 4

As seen in figure 4.6, Python is utilized to construct agricultural production prediction using machine learning. The data is recorded as float 64 data type and spans the years 2004 through 2012.we can see the output towards crop yield prediction.

	0	1	2	\
Crop	ARHAR	ARHAR	ARHAR	
State	Uttar Pradesh	Karnataka	Gujarat	
Cost of Cultivation (`/Hectare) A2+FL	9794.05	10593.1	13468.8	
Cost of Cultivation (`/Hectare) C2	23076.7	16528.7	19551.9	
Cost of Production (`/Quintal) C2	1941.55	2172.46	1898.3	

	3	4	\
Crop	ARHAR	ARHAR	
State	Andhra Pradesh	Maharashtra	
Cost of Cultivation (`/Hectare) A2+FL	17051.7	17130.5	
Cost of Cultivation (`/Hectare) C2	24171.7	25270.3	
Cost of Production (`/Quintal) C2	3670.54	2775.8	

	5	6	7	\
Crop	COTTON	COTTON	COTTON	
State	Maharashtra	Punjab	Andhra Pradesh	
Cost of Cultivation (`/Hectare) A2+FL	23711.4	29047.1	29140.8	
Cost of Cultivation (`/Hectare) C2	33116.8	50828.8	44756.7	
Cost of Production (`/Quintal) C2	2539.47	2003.76	2589.99	

	8	9	...	\
Crop	COTTON	COTTON	...	
State	Gujarat	Haryana	...	
Cost of Cultivation (`/Hectare) A2+FL	29616.1	29919	...	
Cost of Cultivation (`/Hectare) C2	42070.4	44018.2	...	
Cost of Production (`/Quintal) C2	2179.26	2127.35	...	

Fig 4.7. Output 5

As seen in figure 4.7, Python is used to build agricultural production forecasts using machine learning. The data spans the years 2004 to 2012 and is of the float 64 data type. We can now view the cost of cultivation.

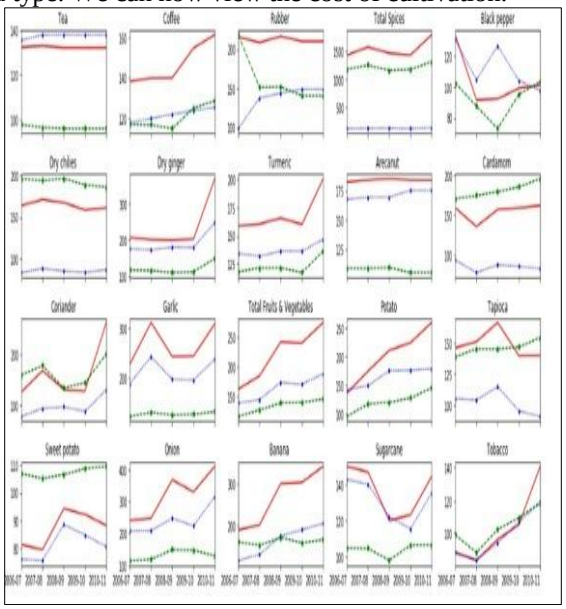
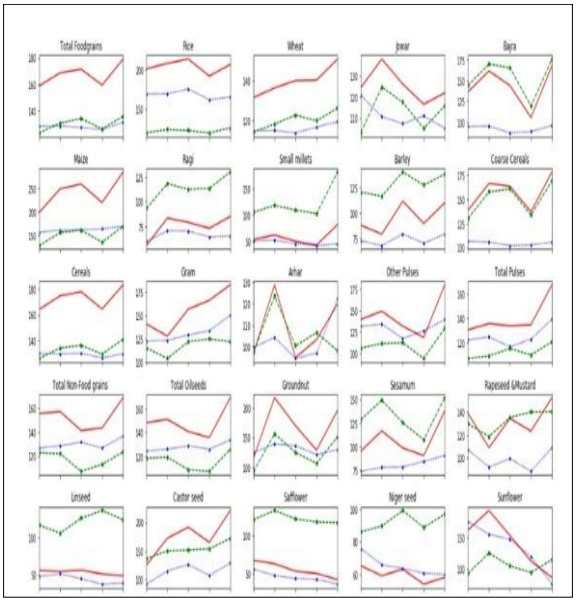


Fig 4.8 Output 6

As can be seen in figure 4.8, Python is used to create machine learning-based predictions of agricultural productivity. The data spans the years 2004 through 2012 and is stored as float 64 data type. The output of crop yield forecast is seen.

Conclusion

The emphasis of this analysis has concentrated on science but also technology, but a wide variety of choices, including socioeconomic and variables such like technology dissemination as well as farmer access to technologies, must also be addressed. Thus, the approach will aid in alleviating the challenges farmers experience and prevent them from trying suicide. It will operate as a conduit to offer farmers with the knowledge they need to increase produce and profitability, hence reducing suicide rates also alleviating his troubles. By analyzing the issues that affect crop yield is that climatic factors such as average temperature, average humidity, average rainfall, and route map for a selected crop for particular features with date specification. This methodology to design and development of crop prediction and crop yield prediction using different machine learning techniques are used such as K-Near Neighbor, Decision Tree,



Random Forest Classifier helps to increase yield and subsequent profit of agricultural production.

References

1. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature 1review. Computers and Electronics in Ag.riculture, 177(January), 105709. https://doi.org/10.1016/j.compag.2020.105709.
2. 2.Jain, S., & Ramesh, D. (2020). Machine Learning convergence for weather based crop selection. 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2020, (February 2020). https://doi.org/10.1109/SCEECS48394.2020.75
3. 3.Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: A review paper. Indian Journal of Science and Technology, 9(38). https://doi.org/10.17485/ijst/2016/v9i38/95032

4. 4.Bouighoulouden, A. (2020). Crop Yield Prediction Using K-Means Clustering. School of Science and Engineering – Al Akhawayn University, (April), 1–35.
5. Nishant, P. S., Sai Venkat, P., Avinash, B. L., & Jabber, B. (2020). Crop yield prediction based on indian agriculture using machine learning. 2020 International Conference for Emerging Technology, INCET 2020, (June). https://doi.org/10.1109/INCET49848.2020.9154036
6. Sangeeta, & Shruthi, G. (2020). Design and implementation of crop yield prediction model in agriculture. International Journal of Scientific and Technology Research, 9(1), 544–549.
7. Kulkarni, S., Mandal, S. N., Sharma, G. S.Mundada, M. R., & Meeradevi. (2018). Predictive Analysis to Improve Crop Yield using a Neural Network Model. 2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018, 74–79. https://doi.org/10.1109/ICACCI.2018.8554851
8. 8. PALANIVEL, K., & SURIANARAYANAN, C. (2019). an Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques. International Journal of Computer Engineering and Technology, 10(3). https://doi.org/10.34218/ijcet.10.3.2019.013
9. 9. Journal, I. (2018). Prediction of Crop Yield using Machine Learning. International Research Journal of Engineering and Technology (IRJET) .
10. Kale, S. S., & Patil, P. S. (2019). A Machine Learning Approach to Predict Crop Yield and Success Rate. 2019 IEEE Pune Section International Conference, PuneCon 2019, 1–5. https://doi.org/10.1109/PuneCon46936.2019.9105741
11. 11.Bang, S., Bishnoi, R., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. 2019 12th International Conference on Contemporary Computing, IC3 2019, 1–6. https://doi.org/10.1109/IC3.2019.8844901
12. Malik, P., Sengupta, S., & Jadon, J. S. (2021). Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms. Proceedings of the Confluence 2021: 11th International Conference on Cloud Computing, Data Science and Engineering, 1004–1007. https://doi.org/10.1109/Confluence51648.2021.9377147
13. 13.Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018). AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA 2018, 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697349
14. 14.Nisha, K. G., & Sreekumar, K. (2017). A review and analysis of machine learning and statistical approaches for prediction. Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, Icicct, 135–139. https://doi.org/10.1109/ICICCT.2017.7975174
15. 15.Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015). (Paper 7 Tafur). May, 138–145.
16. 16.Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151(November 2017), 61–69. https://doi.org/10.1016/j.compag.2018.05.012
17. 17.Pant, J., Pant, R. P., Kumar Singh, M., Pratap Singh, D., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings, 46(xxxx), 10922–10926. https://doi.org/10.1016/j.matpr.2021.01.948
18. 18.Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine learning applications on agricultural datasets for smart farm enhancement. Machines, 6(3). https://doi.org/10.3390/machines6030038.
19. Chaudhary, K. (2020). Prediction of Crop Yield Using Machine. 4(9), 153–156. www.irjet.net
20. Ghosh, S., & Koley, S. (2014). Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks. International Journal on Recent and Innovation Trends in Computing and Communication, 2(2), 292.
21. 21.Gupta, J., Chauhan, A., Agarwal, A., Ranghuvanshi, A. S., & Saxena, R. (2018). Machine Learning Approach for Crop Yield Prediction and Crop Variety Recommendation in Android Application. 7(1).
22. 22.Bali, N., & Singla, A. (2022). Emerging Trends in Machine Learning to Predict Crop Yield and Study Its Influential Factors: A Survey. Archives of Computational Methods in Engineering, 29(1), 95–112. https://doi.org/10.1007/s11831-021-09569-8