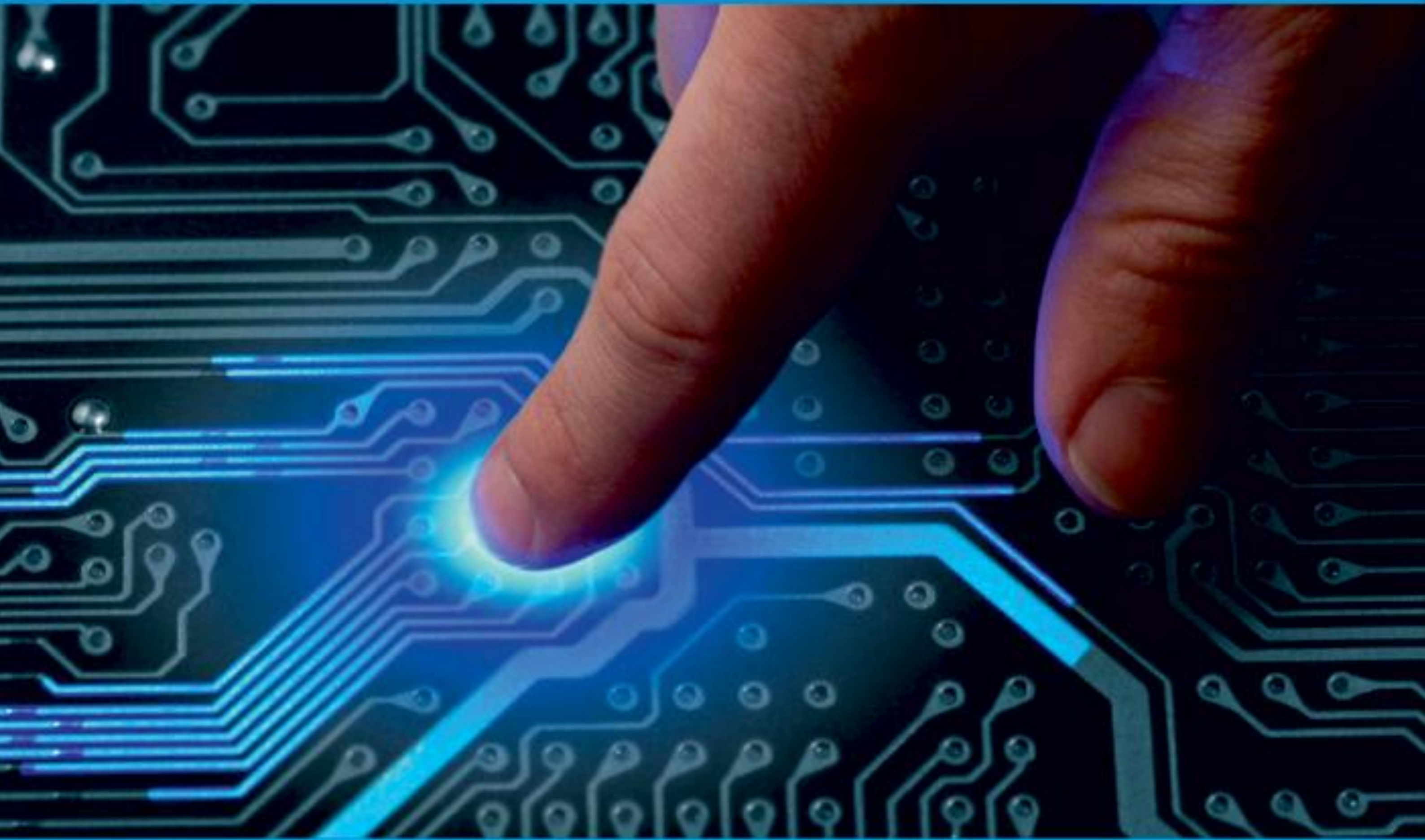
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Special Issue 2, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

| International Conference on Recent Innovations in Engineering and Technology (ICRIET'23)|

| Sharadchandra Pawar College of Engineering, Pune, India |

|| Volume 11, Special Issue 2, March 2023 ||

Using the Random Forest Algorithm for Effective Yield Prediction

Rahul Thorat, Rushikesh Bhangare, Anisha Bhor, Dr. Monika Rokade

Student, Dept. of Computer Engineering, SharadchandraPawar College of Engineering, Otur, India

Student, Dept. of Computer Engineering, SharadchandraPawar College of Engineering, Otur, India

Student, Dept. of Computer Engineering, SharadchandraPawar College of Engineering, Otur, India

Professor, Dept. of Computer Engineering, SharadchandraPawar College of Engineering, Otur, India

ABSTRACT: Agricultural development faces a hurdle with accurate production predictions. As the demand for accuracy and analysis of agricultural output under various weather circumstances rises, farmers, governments, and traders continue to face significant challenges related to crop production, which is dependent on various climatic variables such as dry seasons and rising temperatures. The random forest technique in this machine learning system may analyse crop development in relation to the prevailing meteorological conditions and biophysical changes. Data sets on crop growth were gathered from a range of sources. Both testing and training use these datasets. The random forest classifier discovered amazing yield prediction abilities. Many findings demonstrate that Random Forest is a very accurate data mining technique and an effective learning algorithm for yield analysis in the current climate.

KEYWORDS: Harvest prediction and Random Forest classifier.

I.INTRODUCTION

The goal of the extremely large branch of computer science known as artificial intelligence (AI) is to develop things that can observe and reason about the world similarly to how people do. The idea of machine learning is widely used in artificial intelligence, which aims to do away with independent programming of machines. Rather than hard coding each step in the We will provide the computer with as many samples as necessary, and the computer will be able to determine what to do when presented with a fresh sample that it has never seen before. To start, I'll give you a quick rundown of the various subjects so you can better grasp what they are and how they function. The operation of various techniques, algorithms, and architectural ideas will next be covered in more detail. Let's start by discussing reasonable theories for how artificial intelligence handled issues prior to the development of machine learning. This will be the final time we cover classical AI; although they are still utilised for different jobs, more potent algorithms have taken their place. This ought to motivate people to adopt cutting-edge machine learning techniques. A quick introduction to neural networks is provided in a separate section near the conclusion. You'll later take the neural network apart to discover its precise workings, and you'll write code to create, train, and test through imprinting.

II. LITERATURE REVIEW

Many methods and strategies have been developed to produce effective crop forecasting. Many of the references describe approaches that can be used to anticipate crops in various ways and are drawn from various case studies. The majority of the concepts focus on using sensors, support vector machines, and big data to grow crops. Crop forecasting makes good use of these concepts.

In this paper, we effectively predict yield using a random forest classifier. Many actions taken by random forest algorithms increase test speed and training accuracy. This article's major goal is to critique other writers and skew the systems they suggest in order to provide you with the greatest solution. A random forest classifier functions like a corpus and is made up of numerous distinctive trees. The Random Forest Classifier divides each distinct decision tree into class predictions, and the class receiving the most votes determines the forecast for the model. The intelligence of crowds is the core idea behind the straightforward but effective random forest classification method. The random forest classifier approach performs well in data science because: Any one of the individual constituent models will perform worse than an enormous number of very similar models (trees) working together as a committee. The key is the weak correlation between models. When assets with low correlations are combined to form a portfolio, their combined value is greater than the sum of their separate sectors. Corresponding algorithms can produce portfolio predictions that are more distinctive than any individual prediction. This best effect exists to shield trees from their unique faults. The tree group is moving in the proper path because some of the trees are destroyed and others of the trees are fine.

I.Using predictive analytics, forecast harvest

The idea of soil fertility analysis is developed using the suggested approach. Based on the data from the sensors, the method then suggests crops to grow. The localised harvest information is subsequently presented in graph form. This article includes a discussion about agriculture where farmers can describe and illustrate this strategy and learn from professionals. The technique then makes recommendations for which fertilisers should be added to boost yield.This enables farmers to examine their crops and cultivate better ones in order to increase yields and profitability. Afterwards, we provide you with more recommendations for fertiliser to apply to your soil as well as additional details regarding your fertiliser reservoir.

II.Analysis of Indian rice yield using the Support Vector Technique

Our nation relies heavily on grains like rice, maize, wheat, and different legumes to produce food. In developing nations, the ability to produce rice depends on the availability of suitable weather. Seasonal climate variations have a negative impact, with occurrences occurring during the dry season lowering production. Farmers and other customers may be better able to control agricultural research and crop selection with the aid of improved methods for crop analysis in a variety of weather situations. The area of the rice field that will be developed is predicted by the proposed study using a supervised learning methodology. In the case at hand, the sequential minimum optimization method was used to achieve the desired results, which were then used to resolve the job disagreement. Low, medium, and high precipitation, impact on crop production, yield, and harvest region are the variables that are examined from June to November.

III.Using big data to estimate profit rates

Evaluation of crop growth is crucial for maximising food security programmes. In contrast to the old conventional approaches utilised for data processing in model training and testing tools, this suggested study provides a novel idea of yield analysis that is based on big data notions. Phase 1 of the proposal makes use of already-existing agriculture-related big data sets and the information that is accumulating quickly thanks to practical big data transformation techniques. A pre-balanced outcome based on accuracy and analysis time is provided in the second phase by the "nearest neighbour" structure, which uses the results from the prior data processing model. Samples from real agricultural datasets demonstrate that the proposed job performs better than existing tasks in terms of both performance and forecast accuracy.

III.EXISTING SYSTEM

The Bayes algorithm used in this task is naïve. A segmentation technique known as a Naive Bayes classifier is a form of machine learning idea. It is simple to utilise this Naive Bayes technique for process analysis as well as separation. As of now, this approach produces a closed-form expression that is utilised to train the data set on the premise that all class conditions are independent of one another. In training, this build-up process discovers high probability.

Lower operating costs are the effect of this. This dataset for the Naive Bayes classifier has a total of n attributes, with F-tuples standing in for the n values of the attributes A1, A2, and An. After that, compare the tuple to each class.The current model employs two phases. When any data are passed in the first stage, it makes recommendations regarding what crop should be planted there. It is advised to plant crops like paddy and maize in an environment with a downfall of 150 mm, ground wetness equal to 39 percentage, degree of 22, and environmental pressure of 954 mbar. In addition, it suggests planting cotton, rice, and chilli peppers in an environment with a downpour of 150 mm, a ground moisture percentage of 32, a temperature of 25 degrees, and a pressure of 947 mbar. But, if the downpour is 175 mm, the ground is 39 percentage wet, the temperature is 23, and the atmospheric pressure is 949 mbar, it suggests that it is not ideal for anyfrom among these three crops. The analysis of crops can benefit from these connected structures.The yield is maximised since the second step demonstrates the ideal application period. where n is the number of tuples in the data set, 1f, MF is a function of the data set, and 1f, 2f, 3f... mf is the median of a specific culture in the data set. To determine seed spread, production, and the ideal weather conditions for harvesting, averages were validated over the previous five years of data. We test proposed tasks and minimise functionality using the screaming card programming approach based on priority.The Hadoop distributed file system has a variation called Map Reduce that streamlines data processing across several nodes and enhances parallel operations on various variables obtained from various agricultural silos. To process huge amounts of data, this approach distributes and operates in parallel. Figure 1 depicts the architecture diagram's Map Reduce structure. This structure explains the evaluated findings that were used to recommend appropriate crops for various weather scenarios.

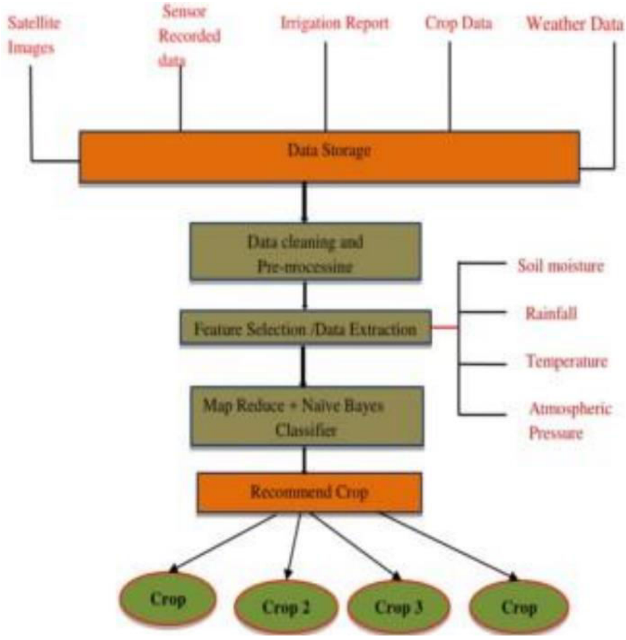


Fig. 1. Architecture Diagram of Existing System

IV. PROBLEM STATEMENT

Compared to other machine learning algorithms, Naive Bayes offers probability-based predictions with a lesser level of accuracy. It takes more time to compute predictions based on probabilities calculated during training.

V. PROPOSED SYSTEM

India is an agricultural nation, and much of its prosperity depends on agricultural extension services and closely associated agricultural industry outputs. In Indian agriculture, the high level of inconstancy causes the farming to collapse. An agricultural development also depends on various soil factors and atmospheric gases, such as nitrogen gas, phosphorus, potassium, rotation of crops, ground moisture, exterior degrees, and climatic aspects such as temperatures and precipitation. The names of the states and districts, the crop-growth year, the region, the mass production, and the crop that has to be cultivated are all collected in the online dataset. The dataset is next subjected to the random forest classifier in order to validate the data.Instead of boosting, the random forest classifier is a potent bagging method. Random forest classifiers use decision trees in simultaneously. These decision trees don't interact with one another. The discrepancy between the actual value and the projected outcome is then graphically shown. It gathers data sets from several sources and keeps them in a data warehouse. After that, the data is sent to the preprocessing procedure, which entails three steps: cleaning, reduction, and normalisation of the data. Inaccurate and partial data are removed throughout the data cleansing process. The information is subsequently transformed into a more straightforward format using data compression. Last but not least, data normalisation is used to change the range of values in numerical columns to a similar scale without altering them. After that, data is chosen using feature selection and data extraction. The process of manually or automatically choosing features appropriate for our project's output is known as feature selection. The process of removing data from a data warehouse so that it can be processed further is known as data extraction. The degree of prediction accuracy is then raised using the random forest classifier, yielding.
Testing module are among the modules that the project uses. The dataset module loads the original dataset using the pandas library file. The visualisation module also includes the matplotlib package, which is used to create plots and arrays. The model is then run using the training module's outputs and inputs. Lastly, the dataset used to evaluate the final model built using the training input set is known as the test input in the test module.These modules are used by the suggested system to produce the production schedule.

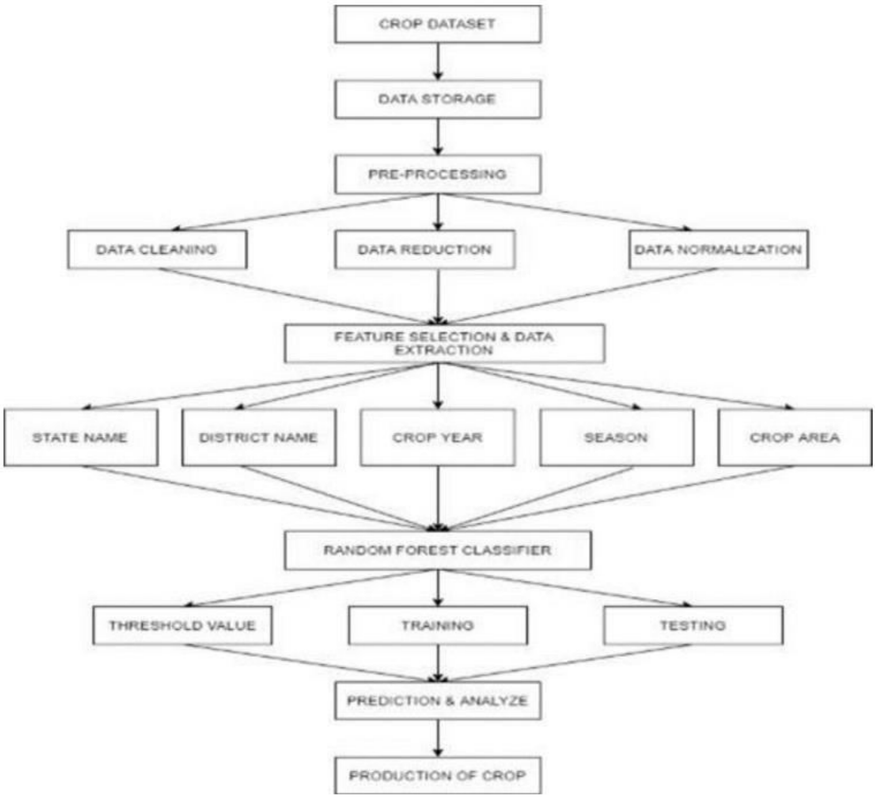


Fig. 2. Architecture Diagram of Proposed System

VI. RESULTS AND DISCUSSION

The preprocessing method in our project goes through a feature selection procedure and chooses appropriate features. The random forest classifier is then given the selected feature. In this manner, the random forest algorithm categorises the attributes, forecasts if the crop is appropriate for the agricultural land, and outputs the crop yield. Evaluation of the random forest algorithm's output.

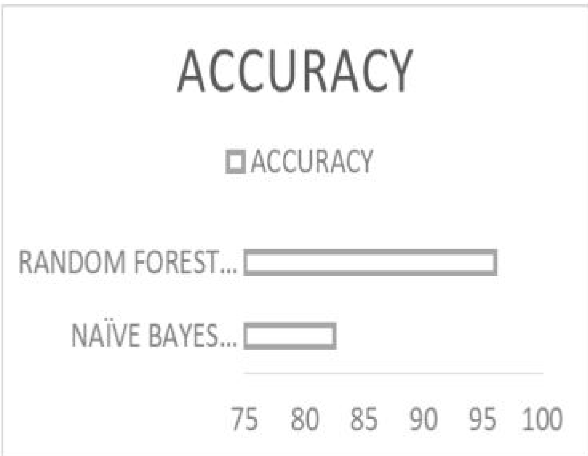


Fig. 3. Accuracy Level Graph for Existing System and Proposed System

The proposed system has a higher accuracy than the existing system. Because it uses packing, the random forest algorithm has a high level of accuracy. The complexity of models that don't fit the training set is reduced using this technique. Here, decision trees are boosting techniques that make the model more complex, leading to an imperfect fit of the training set. A random forest algorithm was developed to address this. Because it is a probability-based method, the Naive Bayes classifier in the current system has a poor level of accuracyBelow are the graphs representing the actual and expected values as well as the change analysis for the region and production.

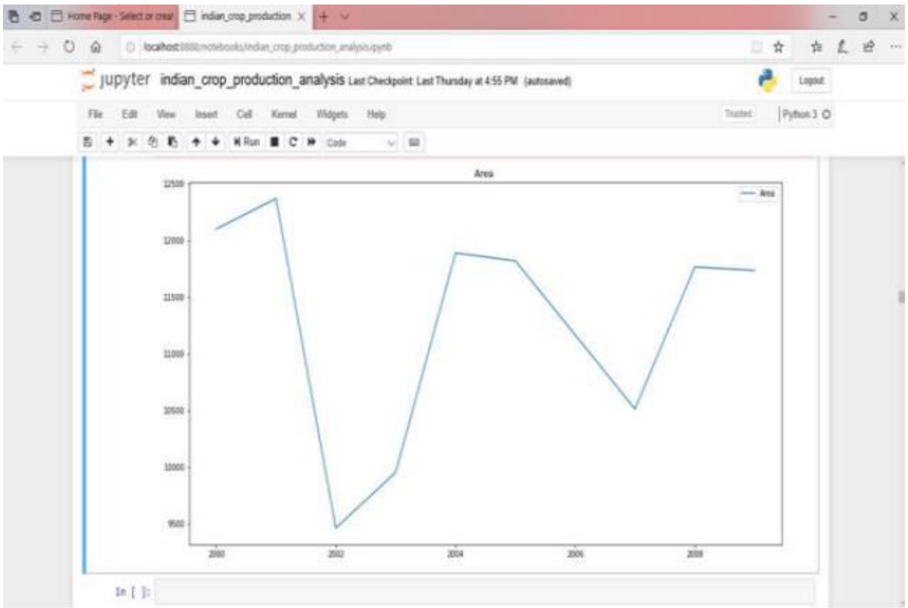


Fig. 4. Analysis of Variations in Area

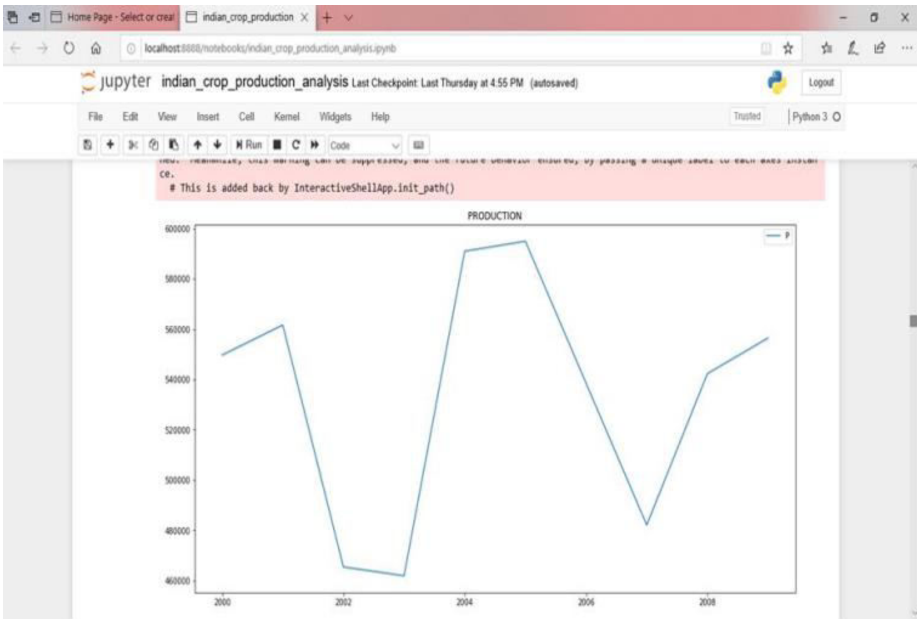


Fig. 5. Analysis of Variations in Production

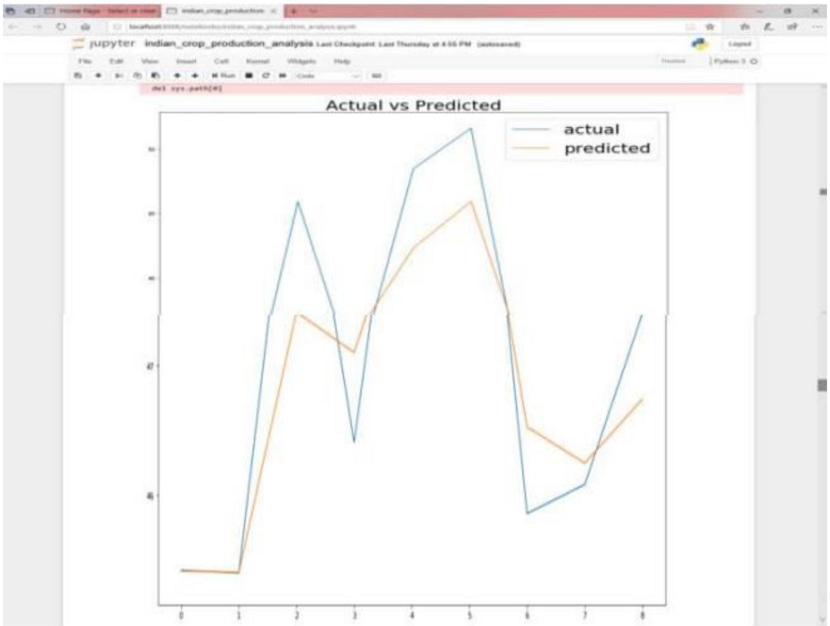


Fig. 6. Actual and Predicted Values.

VII. CONCLUSION AND FUTURE SCOPE

Huge efforts have been made recently to increase crop productivity and forecasting. Using data and communication technologies to produce precise crop growth estimation ideas will benefit agricultural workers and stakeholders, employees, and will raise their accountability to food imports and exports and food security. The agricultural dataset is examined using the random forest classifier. Finally, it can be said that the Random Forest Algorithm can be a real method of increasing crop yield with greater accuracy in agricultural areas. The crop dataset is predicted using a hybrid combination of machine learning algorithms in the future.

Under the framework, image datasets are also processed to forecast crop illnesses and choose crops that would grow well in a given soil.

REFERENCES

1.Monika D.Rokade ,Dr.YogeshkumarSharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic."IOSR Journal of Engineering (IOSR JEN),ISSN (e): 2250-3021, ISSN (p): 2278-8719

2. Monika D.Rokade ,Dr.YogeshkumarSharma"MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE

3.Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. *International Journal of Advanced Science and Technology*, *29*(9s), 2324 – 2331.

[4] . D Ramesh ,B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology.(IJERT),2015.

[5] . F. Baheri, F. Davardoost and V. Ahmadzadeh, "Data mining with learning decision tree and Bayesian network for data replication in Data Grid", Proceedings of the Third International Conference on Contemporary Issues in Computer and Information Sciences, pp. 51- 51, 2012..

[6] .S.Veenadhari, Dr Bharat Misra, Dr CD Singh.2019."Machine learning approach for forecasting crop yield based on climatic parameters.". International Conference on Computer Communication and Informatics (ICCCI), 2014.

[7] .P.Priya, U.MuthaiahM.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology(IJESRT),2018.

[8] . Abdullah, A., Brobst, S, Pervaiz.I.,UmerM.,and A.Nisar.2004. Learning dynamics of pesticide abuse through data mining. Proceedings of Australian Workshop on Data mining and Web Intelligence, New Zealand, January,2004.

[9] . C.T. Dhanya and D. Nagesh Kumar, "Data mining for evolution of association rules for droughts and floods in India using climate inputs", J. of Geo. Phy. Res.114: 1-15, 2009.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

| International Conference on Recent Innovations in Engineering and Technology (ICRIET'23)|

| Sharadchandra Pawar College of Engineering, Pune, India |

|| Volume 11, Special Issue 2, March 2023 ||

[10] . VS Boddu, BNK Reddy, MK Kumar, "Low-power and area efficient N-bit parallel processors on a chip," 2016 IEEE annual India conference (INDICON), pp. 1-4, 2016.

[11] . M. Kaur, H. Gulati, and H. Kundra, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications," International Journal of Computer Applications, vol. 99, no. 12, pp.1-3, August.2014.

[12] . F. Baheri, F. Davardoost and V. Ahmadzadeh, "Data mining with learning decision tree and Bayesian network for data replication in Data Grid", Proceedings of the Third International Conference on Contemporary Issues in Computer and Information Sciences, pp. 51- 51, 2012.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi crossref

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details