Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper : 101

Crop yield prediction based on Apriori Algorithm using association rules to enhance agribusiness and optimization by Internet Of Things.

Atul Tripathi[1]*, Bhawani Singh Rathore[1], Dr.Divakar Singh[1]

1 University Institute of Sc. & Technology, Barkatullah University, Bhopal, MadyaPradesh, India

Abstract

Objectives: In order to predict crop productivity, this study will apply data analytics and machine learning techniques to agricultural data and will construct association rules based on fixed attributes and their correlations and optimization using IoT Technique. Methods: Data on various crops is gathered from numerous nations based on predetermined criteria, including area, production, yield, temperature, and rainfall. On a prepared dataset, the pre-processing processes are conducted to run data analytics and machine learning algorithms. An Apriori algorithm is then used to generate association rules using the processed data. Findings: Using the Apriori technique on a crop dataset that had been created, minimum support and confidence levels and lift and association rules were produced. Among them, 'High Yield Production' was predicted by the outcomes of some linked rules then we optimitize by smart farming IoT Devices. Application : The main result of this study is a set of efficient and well-constructed association rules for yield prediction, which will be useful for academics, farmers, and government officials to increase agricultural output.

Keywords: Crop data sets; Association rules; Yield prediction; Apriori Algorithm, IOT

1 INTRODUCTION

In managing and producing food, crop yield prediction is useful. Its impact on the nation's economy cannot be ignored[1]. Due to the rapidly changing weather, water shortages, inadequate irrigation systems, out-of-control costs, and accurate production projection, farmers currently face unknown problems and threats[3].

Rapid weather fluctuations cause low agricultural output, hence it is important to take necessary precautions when predicting yield.

Therefore, it is essential that farmers have access to smart farming tools, particularly those that anticipate crop production using statistical and computational methods. A research study demonstrates how crop output predictions might help farmers take preventative measures to boost productivity. A huge database with a range of metrics, such as area, temperature, rainfall, humidity, and wind speed, can be used to collect past farming experience as well as meteorological data and other influencing elements.

Modern technology can provide a wealth of data on agricultural-related operations, allowing for the extraction and sorting of the necessary information [4]. Data on crops from many nations are gathered for this study from the Agriculture Statistics Department. These data are put through an experiment to determine the yield utilising an association rule-based methodology.

2 LITERATURE REVIEW

The majority of researchers, including those in agriculture, use data science in their work. Surya[5] conducted a survey involving the many aspects of agriculture utilizing big data analytics approaches, and she concluded that adopting big data analytics in agriculture produces better results. Furthermore, researchers who use data mining techniques (DMTs) tend to favour agricultural yield prediction. Rao [6] employed DMTs to predict crop productivity in India. To find the association rules, many data mining algorithms have been documented in the literature. AprioriTid, Apriori, and Eclat are some well-known algorithms that have been used to various variables to discover associations. Ramakrishna[7] utilized DMTs for agricultural land soil and noted an increase in crop productivity. In particular, an Apriori association rules were utilised to produce the advisory reports that support farmers' decisions on crop rotation, fertiliser needs, and harvesting techniques. Supriya [8] created a full system using K-Nearest Neighbour and Naive Bayes techniques for crop yield prediction using DMTs to analyse soil records for crop prediction. Additionally, Additionally, Zingade[9] showed an android system that employed linear regression and data analytics to forecast the most lucrative crop under the current weather and soil conditions. Temperature, precipitation, soil characteristics, and productivity from the previous year are used as selected attributes. Chouhan[2] examines the classification, clustering, and association rule DMTs for analysing crop productivity.

Through the extensive compilation of approaches, such as precision and conservative cultivation to meet obstacles in the field, the Internet of Things (IOT) is revolutionising the agribusiness and engaging the farmers. With one or more technologies listed, researchers have offered many modalities for the agriculture sector, such as the irrigation system described in [10] that is based on soil water measurement to determine irrigation amount. This relies on the Bluetooth communication technology, which has its own drawbacks such a small range and device support? In order to increase energy efficiency, a writer recommended timing the power supply to the sensors in 2016 [11]. An author in a publication [12] mentions the use of IoT in agriculture. However, it demonstrates a lack of interoperability, which is essential when discussing vast agricultural landscapes. In a study [13] that was released in 2017 Jinsoohan offered a method for comparing the energy consumption of two appliances. In their article, N.K. Suryadevara and S.C. Mukhopadhyay employed techniques from pervasive computing, data aggregation, etc. to use Zigbee to monitor environmental conditions.[14] However, as more nodes have been implemented, it may pose concerns about increased power consumption and agricultural automation [15]. The study provides a method for giving farmers real-time information about their fields and crops [16]. Smart farming is demand for future to gain maximum profit by farmers using IoT technique. Use of IoT in agriculture is mentioned by an author in paper [20]. The discovery of smart farming methods that can improve the traditional agriculture sector, which is in decline [23].

3 RESEARCH METHODOLOGY

The chosen study endeavour must always be completed in accordance with certain precise, predetermined steps. The primary research methodology used to complete the goal of predicting crop yield is shown in Figure 1. The first step of this research project is to choose the proper variables, wherein yield is chosen as the dependent variable while area, temperature, rainfall, and crop production are chosen as the independent factors. The aforementioned independent and dependent factors serve as the foundation for specifically collecting the necessary data. The datasets are individually produced for the required operations' preparation and subsequent analytical investigations. All dataset(publicly available dataset) here are taken form World Data Bank [21] and FAO (Food and Agriculture Organization) [22]. Pesticides & Yield are collected from FAO. Rainfall & Avg. Temperature are collected from World Data Bank. Yield_df.csv is final dataset processed by cleaning & merging of pesticides, yield, rainfall & avg. temperature.

This research project is show section that deals with the yield prediction. The association rules based approach along with Apriori algorithm is used for crop yield predictionand various association rules are generated.

3.1 *Data collection and processing*

The crop yield data (i.e. Area, Production and Yield) collected from the FAO (Food and Agriculture Organization) and World Data Bank. The other important independent variable data is (i.e. Temperature, Rainfall, pesticides). regarding. The process of crop is usually started from the month of June to November hence the data of these months are preferred for experiments. It is difficult to mention the total number of observations used in the yield prediction due to the average values of datasets. Hence, it is thoroughly stated that approximately 144 observations are experimented and optimized. Many attributes are usually considered collecting the required data, out of which only the relevant attributes are selected and combined to develop

Volume 7- Issue 1, January 2024
Paper : 101

crop data laboratory experiments. The agricultural data and other relevant data variables are organized and combined. The agricultural data includes Area of crop under Cultivation (in Hectares), Production (in Million Tons) and Crop Yield (in KG's per Hectares)., pesticides (in tones). While, other important variables include Temperature (in Degree Celsius), Annual Rainfall (in Milli Meters). All manually collected and published data were transferred into spreadsheet/computerized file, so that crop dataset can be input to the machine for the development of association rules.

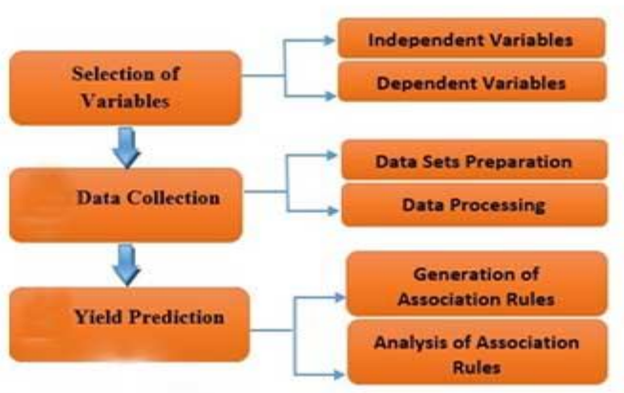


Fig 1. Research methodology

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper : 84

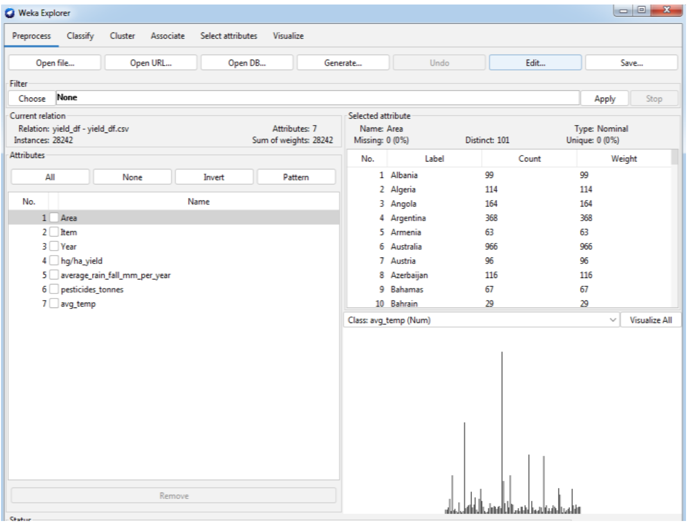
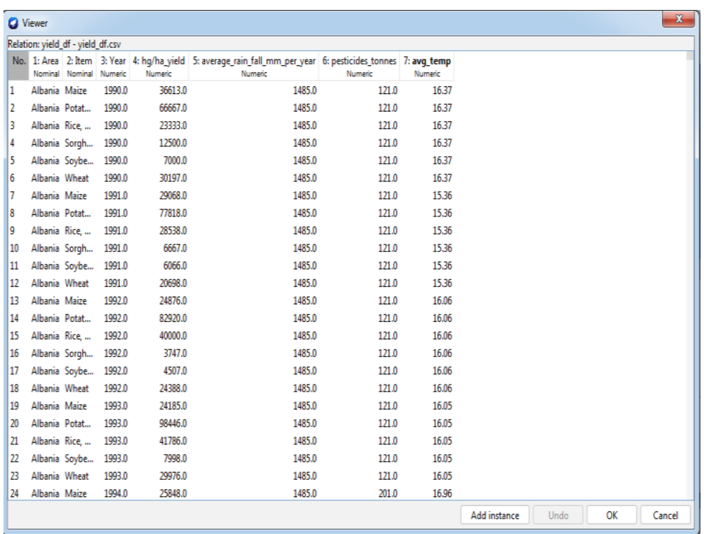


Figure 1 Preprocessing of crop yield prediction dataset in weka.



Viewer

Relation: yield_df - yield_df.csv

No.	1: Area Nominal	2: Item Nominal	3: Year Numeric	4: hg/ha_yield Numeric	5: average_rain_fall_mm_per_year Numeric	6: pesticides_tonnes Numeric	7: avg_temp Numeric
1	Albania	Maize	1990.0	36613.0	1485.0	121.0	16.37
2	Albania	Potat...	1990.0	66667.0	1485.0	121.0	16.37
3	Albania	Rice, ...	1990.0	23333.0	1485.0	121.0	16.37
4	Albania	Sorgh...	1990.0	12500.0	1485.0	121.0	16.37
5	Albania	Soybe...	1990.0	7000.0	1485.0	121.0	16.37
6	Albania	Wheat	1990.0	30197.0	1485.0	121.0	16.37
7	Albania	Maize	1991.0	29068.0	1485.0	121.0	15.36
8	Albania	Potat...	1991.0	77818.0	1485.0	121.0	15.36
9	Albania	Rice, ...	1991.0	28538.0	1485.0	121.0	15.36
10	Albania	Sorgh...	1991.0	6667.0	1485.0	121.0	15.36
11	Albania	Soybe...	1991.0	6066.0	1485.0	121.0	15.36
12	Albania	Wheat	1991.0	20698.0	1485.0	121.0	15.36
13	Albania	Maize	1992.0	24876.0	1485.0	121.0	16.06
14	Albania	Potat...	1992.0	82920.0	1485.0	121.0	16.06
15	Albania	Rice, ...	1992.0	40000.0	1485.0	121.0	16.06
16	Albania	Sorgh...	1992.0	3747.0	1485.0	121.0	16.06
17	Albania	Soybe...	1992.0	4507.0	1485.0	121.0	16.06
18	Albania	Wheat	1992.0	24388.0	1485.0	121.0	16.06
19	Albania	Maize	1993.0	24185.0	1485.0	121.0	16.05
20	Albania	Potat...	1993.0	98446.0	1485.0	121.0	16.05
21	Albania	Rice, ...	1993.0	41786.0	1485.0	121.0	16.05
22	Albania	Soybe...	1993.0	7998.0	1485.0	121.0	16.05
23	Albania	Wheat	1993.0	29976.0	1485.0	121.0	16.05
24	Albania	Maize	1994.0	25848.0	1485.0	201.0	16.96

Add instance | Undo | OK | Cancel

Figure 2 Crop Yield prediction dataset in weka viewer.

3.2 *Association Rules*

Association rules are an essential concept in machine learning, utilized by researchers to tackle problems related to big data analytics, such as crop yield prediction. This technique enables the discovery of all the relationships between items in massive databases generated from various sources. To obtain practical results, the Apriori algorithm is commonly used. Association rules are if-then statements that highlight the probability of relationships between data items within large data sets stored in databases. These rules consist of two parts: an antecedent (if) and a consequent (then), both of which are lists of items. The antecedent is an item found within the data, while the consequent is an item found in combination with the antecedent. Various metrics are available to assess the strength of the association between the antecedent and the consequent.[25] When it comes to data mining approaches, association rule mining is always receiving more and more attention from academics looking to examine correlations between individual items or item sets.[26]

3.2.1 *Experiments with Apriori algorithm*

The discovery of widespread patterns is crucial, and the Apriori algorithm allows for significantly improving this task, especially among the items already stored in a database. This algorithm is highly effective and has been successfully used to produce association rules digitally. Given a pre-processed and discretized crop dataset from various countries, the preferred algorithm is applied to obtain frequent item sets and strong association rules. The process starts by creating the basic item set, which consists of a single item paired with all available item sets. The strong association rules obtained from the experiments are

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper : 84

then examined to understand the correlation between the variables, i.e. area, temperature, rainfall, production, and yield.

3.2.2 *Measurements of support and confidence levels*

A list of association rules was created using the Apriori algorithm's output. By establishing minimal support and confidence levels, an effort was made to acquire the greatest number of combinations and stronger connection rules among the components in the crop dataset. The variable X served as a representation of the item sets, and the support was determined as the percentage of all transactions across all data sets that contained the computed item set. It was time to calculate the level of confidence using equation 1 utilising the different support levels that had been computed.

$$conf(X \rightarrow Y) = supp(X \cup Y)/supp(X) \qquad (1)$$

It should be noticed that the computation of supp(X∪Y) truly represents the support for both transactions X and Y's occurrences. Additionally, confidence is typically understood as the probability estimate P (Y|X) used to construct the left and right sides of the deduced rules. Then, to get the support % of the variables X and Y, the left side of the rule is computed using the Eq. 2 given below.

$$lift(X \rightarrow Y) = supp(X \cup Y)/supp(X) \, X \, supp(Y) \qquad (2)$$

Rules	Antecedents	Consequents	Support	Confidence	Lift
R1	average_rain_fall_mm_per_year=1083, ==> Area=India 4048	High Yield= Yes	0.1	1	6.98
R2	Area=India 4048 ==> average_rain_fall_mm_per_year=1083	High Yield= Yes	0.1	1	6.98
R3	average_rain_fall_mm_per_year=1761 2277 ==> Area=Brazil 2277	High Yield= Yes	0.05	1	12.4
R4	average_rain_fall_mm_per_year=758 1472 ==> Area=Mexico 1472	High Yield= Yes	0.05	1	19.19
R5	Area=Pakistan 1449 ==> average_rain_fall_mm_per_year=494	High Yield= Yes	0.05	1 (1)	19.49

Table 1 Association rules generated in weka by applying association rule mining in crop yield dataset.

The established association rules auto generated by using Apriori algorithm on crop Dataset.
When Instances 2824 for R1 Rule and R2 to R5 instances are 1412. In generation of best rule our method takes execution time19.24 seconds as compare to regression method applied by other researchers takes 48.6 seconds execution time in same dataset [24].

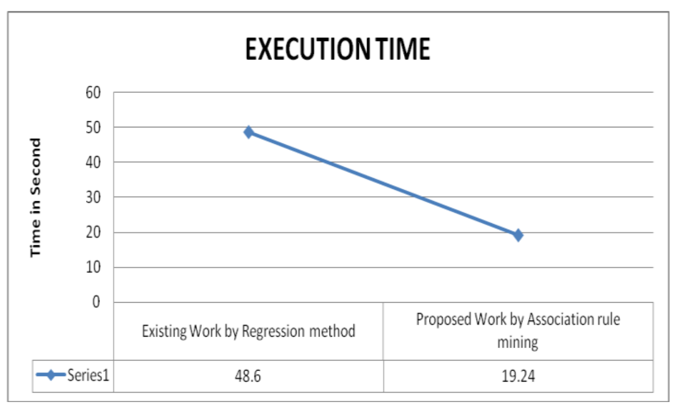


Fig 3 Result Analysis of Execution Time of Existing

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7- Issue 1, January 2024
Paper : 84

work & Proposed work

4 OPTIMIZATION USING IOT TECHNIQUE

The practise of agriculture dates back more than 20,000 years. The emergence of agriculture and civilisation coincided during the Neolithic era. India is a predominantly agrarian country, hence its agriculture depends largely on precipitation, soil, moisture, and environmental challenges.Our farmers shifted to the most modern agriculture methods now accessible. IoT systems have been effectively used in many different industries on a global level. In order to boost crop output, Indian farmers now need to employ smart agricultural systems. By combining data from sensors, actuators, and sophisticated electronic devices, a farmer can keep track of the productivity of his or her agricultural land. Smart agriculture can forecast weather data by activating the pump motor and detecting the dampness of the soil in terms of moisture levels with the help of sensors connected to the processing module Arduino-UNO. The smart farm system can be managed from anywhere with the use of networking technologies. Modern solutions for data collection and resource optimisation may be produced when Smart Agriculture and Artificial Intelligence are integrated. Solutions for insects and pests that prevent agricultural loss and maximise resource use have the potential to be revolutionary. Fig.1 shows the IoT process using different modules.

4.1 SOIL MOISTURE SENSOR

A soil sensor that determines soil moisture monitors the volumetric water content in the soil and outputs the moisture level. The sensor averages both the propelled yield and the water content along the whole length of the soil environment, whether it is wet or dry.With an accuracy of ± 1°C and ± 1%, the sensors can monitor temperature from 0°C to 50°C and humidity from 20% to 90% [17] .

4.2 RAINDROP SENSOR

The fundamental functioning principle of rain sensors is resistance testing, and the sensor has two distinct conduction printed leads on its entire surface. When water droplets touch the sensor's surface, the circuit is completed, creating a resistance that is significantly lower than the sensor's open circuit resistance. The sensed data is then sent to the controlling unit [18] .

4.3 EMPERATURE & HUMIDITY SENSOR

The sensor for measuring humidity The electrodes are placed to the surface of the moisture-retaining substrate DHT11. The relative humidity has an impact on how much the resistance between the two electrodes changes. In order to function, humidity sensors must be able to detect changes in electrical currents or air temperature [17].

Fig-1 **IoT Processing Module Phases**

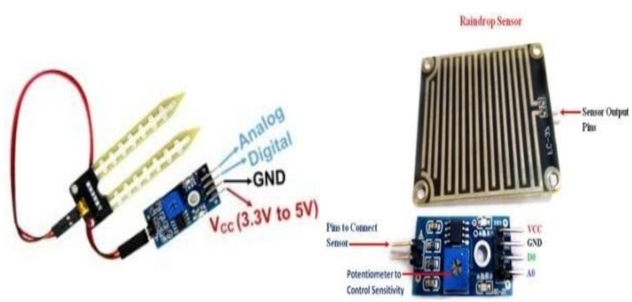
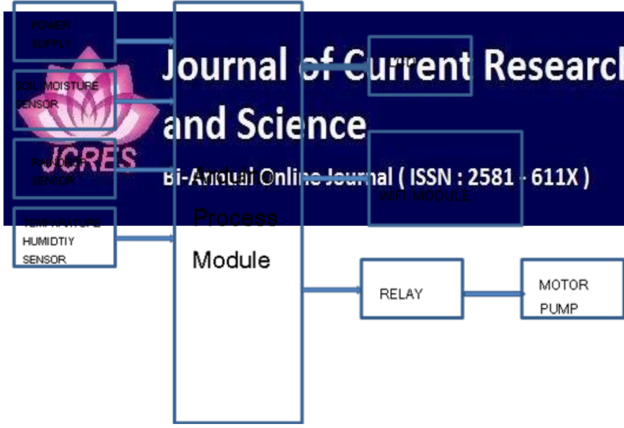


Figure-2 **Soil Moisture sensor, Raindrop sensor and Temperature & Humidity sensor**

4.4 ARDUINO UNO BOARD

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)



Volume 7- Issue 1, January 2024
Paper : 84

Unwrapped-source electronics prototyping platform Arduino is built on adaptable, user-friendly hardware and software[19]. Arduino can duplicate the provided output required for actuators, motors, etc. using input from a variety of sensors. For individuals who are familiar with basic electronics and the C programming language, it is user-friendly. The core components of the Arduino platform are an Arduino Board hardware board and the Arduino IDE programming software. Additional third-party hardware includes motors, sensor modules, Arduino UNO, and the Arduino Software (IDE)-1.0. A microcontroller board called the Uno is based on the ATmega328P. It has a 16 MHz quartz crystal, 6 analogue inputs, 14 digital input/output pins (of which 6 can be used as PWM outputs), a USB port, a power jack, an ICSP header, and a reset button. Compatible with the microcontroller, it can be powered by a battery or an AC-to-DC adapter, or it can be connected to a computer via a USB cable to get started. With the UNO, you can experiment without being too concerned about making a mistake; in the worst situation, it can be quickly replaced for a low cost. Java was used to create the cross-platform Arduino Integrated Development Environment (IDE), which is available for Windows, Mac OS, and Linux. Writing in the java programming language is done with it. Programmes are loaded and written using it onto the Arduino board [19].

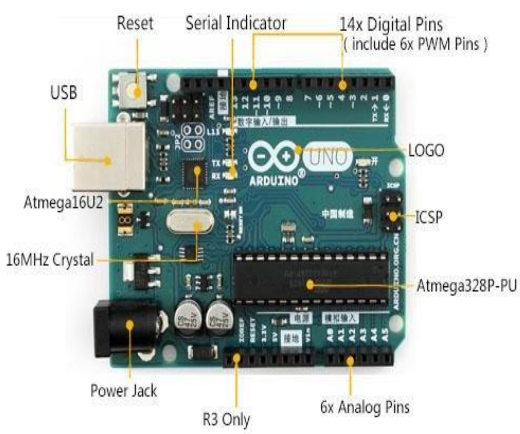


Figure -3 **Arduino UNO board**

4.5 PERFORMANCE OF OPERABILITY

With the use of the item speak.com platform, the system was tested for functionality by checking the soil, temperature, and humidity factors. The image shown makes it possible for the smart irrigation project to function [7].

Fig.4 **LCD data**

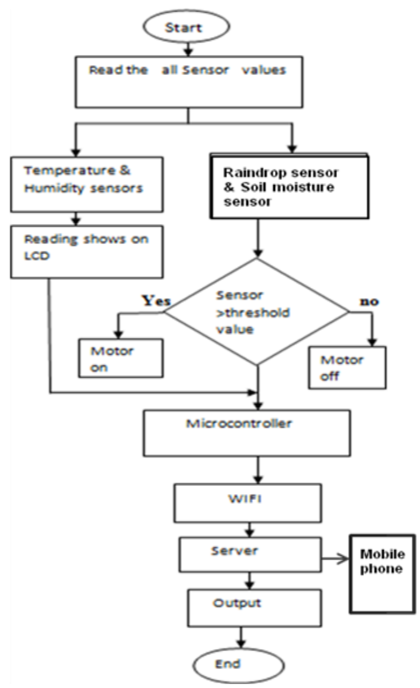


Figure -5 Flow Chart

5 SUMMARY AND CONCLUSION

The aim of this study is to share yield predictions made on previous valuable crop data using analytical techniques and machine learning algorithms. With optimization technique through IoT devices farmers do smart farming. By tracking the state of the soil and the development of the crop in it, these systems enable farmers to address problems with irrigation, temperature, humidity, and other elements.

Volume 7- Issue 1, January 2024
Paper : 84

References

1) Sellam V, Poovammal E. Prediction of Crop Yield using Regression Analysis. Indian Journal of Science and Technology. 2016;9(38):1–5. doi:10.17485/ijst/2016/v9i38/91714.

2) Chouhan S, Singh D, Singh A. A Survey and Analysis of Various Agricultural Crops Classification Techniques. International Journal of Computer Appli- cations. 2016;136(11):25–30. Available from: https://dx.doi.org/10.5120/ijca2016908575. doi:10.5120/ijca2016908575.

3) Park JK, Das A, Park JH. Integrated Model for Predicting Rice Yield with Climate Change. International Agrophysics. 2018;32:203–215.

4) SUPRO IA, MAHAR JA, MAITLO A. Regression Analysis of Rice Data for Yield Prediction Using Python Programming Language. SINDH UNIVERSITY RESEARCH JOURNAL -SCIENCE SERIES. 2019;51(02):183–188. doi:10.26692/sujo/2019.6.32.

5) Surya P, Aroquiaraj IL, Kumar A. The Role of Big Data Analytics in Agriculture Sector: A Survey. International Journal of Advanced Research in Biology Engineering Science and Technology. 2016;2:830–838.

6) Rao PR, Gowda SP, Prathibha RJ. Paddy Yield Predictor Using Temperature, Rainfall, Soil pH, and Nitrogen. Emerging Research in Electronics. Lecture Notes in Electrical Engineering 545. 2019;p. 245–253.

7) Ramakrishna BV, Satyanarayana B. Agriculture Soil Test Report Data Mining for Cultivation Advisory. International Journal of Computer Application. 2016;6(2):11–16.

8) Supriya DM. Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach. International Journal of Innovative Research in Computer and Communication Engineering. 2017;5(5).

9) Zingade DS, Buchade O, Mehta N, Ghodekar S, Mehta C. Crop Prediction System using Machine Learning. International Journal of Advance Engineering and Research Development. 2017;4(5):1–6.

10) M.K.Gayatri, J.Jayasakthi, Dr. G.S. Anandha Mala, (2015). Providing Smart Agricultural Solutions to Farmers for better yielding using IoT. IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR 2015) .

11) Nikesh Gondchawar, Dr. R.Complexion.Kawitkar, "IoT based agriculture", all-embracing almanac consisting of contemporary analysis smart minicomputer additionally conversation planning (ijarcce), vol.5, affair 6, June 2016. Overall Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 5 Issue: 2 177 – 181

12) SoumilHeble, Ajay Kumar, K.V.V Durga Prasad, Soumya Samirana, P.Rajalakshmi, U. B. Desai. A Low Power IoT Network for Smart Agriculture [15] Rajesh M, Salmon S, Dr. Veena .

13) Dr. N. Suma, Sandra Rhea Samson, S. Saranya, G. Shanmugapriya, R..Subhashri, (2017). IOT Based Smart Agriculture Monitoring System. International journal on recent and innovation trends in computing , energy efficiency and communication-IJRITCC volume: 5 issue:

14) PaparaoNalajala, D. Hemanth Kumar, P. Ramesh and Bhavana Godavarthi, 2017. Design and Implementation of Modern Automated Real Time Monitoring System for Agriculture using Internet of Things (IoT). Journal of Engineering and Applied Sciences, 12: 9389- 9393.

15) R. Nageswara Rao, B. Sridhar, (2018). IoT based smart crop field monitoring and automation irrigation system. Proceeding of the second international conference on inventive system and control (icisc2018). [22] Sahitya. Roy, Dr Rajarshi. Ray, Aishwarya Roy, Subhajit

16) PaparaoNalajala, P Sambasiva Rao, Y Sangeetha, Ootla Balaji, K Navya," Design of a Smart Mobile Case Framework Based on the Internet of Things", Advances in Intelligent Systems and Computing,

17) Anand Nayyar, Er. Vikram Puri, (2016). Smart farming: IoT based smart sensors agriculture stick for live temperature and moisture monitoring using Arduino, WSN(Wireless Sensor Networking) systems& solar technology. Internet of things: a review. In Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on (Vol. 3, pp. 648-651). IEEE. [10] Prof. K. A. Patil,

18) B J Bose, K. Schofield, and M. L. Larson, "Rain sensor"

19) S. A. Arduino, "What is Arduino?," Arduino Doc., 2015

20) SoumilHeble, Ajay Kumar, K.V.V Durga Prasad, Soumya

Samirana, P.Rajalakshmi, U. B. Desai. A Low Power IoT Network for Smart Agriculture [15] Rajesh M, Salmon S, Dr. Veena .

21) https://data.worldbank.org/

22) http://www.fao.org/home/en/

23) Adithya Vadapalli,Swapna Peravali ,Venkatarao Dadi 2020 "Smart Agriculture System using IoT Technology" Publisher: International Journal of Advance Research in Science and Engineering (2319-8354) Vol.No.9, Issue No. 09.

24) https://www.kaggle.com/code/kushagranull/crop-yield-prediction

25) Atul Tripathi, Bhawani Singh Rathore, Divakar Singh Article: Survey Paper on Agricultural Dataset for Improving Crop Yield Prediction using Machine Learning Algorithms 184(46):28-34, February 2023 IJCA - Number 46 (ISBN: 973-93-80902-83-8) DOI 10.5120/ijca2023922571

26) Bhawani Singh Rathore, Anju Singh, Divakar Singh Article: Secure Sum based Privacy Preservation Association Rule Mining on Horizontally Partitioned Data International Journal of Computer Applications (0975 – 8887) Volume 134 – No.14, January 2016 DOI 10.5120/ijca2016908084