Grenze International Journal of Engineering and Technology, January Issue

Integrated IoT Environmental Monitoring System for Agricultural Application and Crop Yield Prediction

Bhavnesh Jaint[1], Divyam Gupta[2], Devanshu Verma[3] and Gaurav Gupta[4]
[1-4]Department of Electrical Engineering Delhi Technological University, Delhi, India
Email: bhavneshjaint@dtu.ac.in, divyangupta98@gmail.com, devanshuvermamoh@gmail.com, gguptag18@gmail.com

***Abstract*—The contemporary surge in Internet-connected electronic devices, propelled by the Internet of Things (IoT), has revolutionized data communication across various applications. This paper investigates the increasing application of IoT in monitoring contexts, specifically targeting smart farming. Smart farming aims to deliver necessary resources for defined periods, such as light intensity, ambient temperature, relative humidity, soil moisture content, and soil pH. The process involves sequentially measuring each parameter with a sensor node and sending the data to the cloud for further analysis. The primary objective is to forecast crop production using sophisticated data mining techniques. In contrast to traditional hardware-centric methods, this paper introduces innovative approaches such as Dempster-Shafer theory, Principal Component Analysis (PCA), and Convolutional Neural Network (CNN) for resource monitoring and crop production prediction. Unlike hardware-dependent solutions, these methods leverage computational models to analyze data efficiently. By employing techniques like Dempster-Shafer theory, PCA, and CNN, the accurate assessment of resource parameters is achieved without reliance on specific hardware devices. The proposed system integrates different data mining techniques, including Dempster Shafer, Principal Component Analysis (PCA), and Convolutional Neural Networks (CNN), to anticipate crop production in advance for comprehensive monitoring of agricultural values. This research contributes to modern agricultural efficiency by embracing cutting-edge data mining techniques without the need for specialized hardware. The use of Dempster-Shafer theory, PCA, and CNN enhances the predictive capabilities, and adds modern analysis and mining techniques to modern smart farming practices.**

***Index Terms*—— Internet of Things, Wi-Fi, ESP8266 Wi-Fi module, sensor node, Data Mining, Prediction, Dempster-Shafer, Principal Component Analysis (PCA), Convolutional Neural Networks (CNN).**

I. INTRODUCTION

Agriculture, unfortunately, remains a sector where technological advancements have been slow to gain widespread acceptance. This reluctance is particularly pronounced in developing countries like India, where farmers often face challenging and lethargic conditions. The pressing issues of overpopulation and urbanization contribute significantly to the increasing scarcity of agricultural products. As the population grows, the demand for agricultural resources increases. However, globalization has resulted in significant agricultural land being repurposed for non-agricultural uses by large industries. Consequently, rural farming areas are diminishing, exacerbating the scarcity of agricultural resources. To addres this critical situation, there is an urgent need to

Grenze ID: 01.GIJET.11.1.240

carefully assess and prioritize improvements in crop production. It is imperative to enhance agricultural practices by conservatively utilizing resources and avoiding unnecessary wastage. Smart farming emerges as a promising methodology to achieve this goal. This approach ensures that crops receive precise quantities of resources required for optimal growth within calculated timeframes.

In traditional farming methods, irrigation is often based on a predetermined time schedule, leading to significant water wastage due to farmers' lack of precise knowledge about the ideal irrigation duration. Smart farming addresses this issue by employing soil moisture detection sensors strategically placed at observation sites to measure soil moisture content. These sensors, including those for temperature, humidity, and pH, provide real-time data for informed irrigation decisions, minimizing the exploitation of extra resources compared to traditional farming practices.

This paper proposes the development of a sensor node capable of transmitting data to the Thing Speak cloud. It outlines the methodology employed in smart farming, emphasizing the visualization of the exact percentage of resources required for crops. The paper explores the prediction of crop yield using innovative data mining techniques such as Dempster-Shafer theory, Principal Component Analysis (PCA), and Convolutional Neural Network (CNN). Data Mining, as a crucial tool for extracting valuable insights from enormous datasets, is employed to predict crop production in advance. This predictive capability enables farmers to take immediate measures aligned with the demands of the crop, and it assists the government in determining the exact Minimum Selling Price (MSP) for farmers.

The study delves into various regression and classification methods, including regression, CNN, Dempster Shafer, and PCA, within the realm of data mining. These techniques analyze the impact of various parameters, specifically sensor node parameters, on predicting crop production. The comprehensive approach aims to revolutionize agriculture by making it more sustainable and technologically advanced.

II. Literature Survey

In [2], the author delves into the vital role of Wireless Sensor Networks (WSN) in agriculture, aiming to bolster crop productivity. The paper underscores the contemporary need for smart farming techniques, scrutinizing the proposed system's architecture through the meticulous monitoring of environmental parameters. [3] advocates for the integration of the Internet of Things (IoT) in agriculture, elucidating the interconnected nature of agricultural processes and showcasing the utility of IoT technologies. It provides insights into various technologies designed based on IoT concepts. [4] presents a system employing GPS features to survey agricultural fields. The primary objective of the Global Positioning System (GPS) is to observe data sensed at sensor nodes and transmit it to the central node via RS232. [5]'s proposed system meticulously captures and stores light intensity data around crops in an Excel sheet, focusing solely on this parameter for controlling crop production. In [6], the author addresses the challenge of uneven rain distribution in crop production by utilizing soil moisture and pH sensors, with Arduino serving as the control unit and facilitating communication with the cloud. Regarding [7], the paper highlights the issue of food scarcity in Africa and suggests a schematic approach to predicting the total crop production required to meet the needs of African citizens using an Artificial Neural Network model. This model helps forecast annual crop requirements up to 2021. In [8], a comprehensive definition of comparison algorithms for predicting kharif crops in suburban areas of India is provided, with these algorithms executed on the open-source platform WEKA. The resulting calculations and individual error comparisons offer valuable insights. Paper [9] proposes a smart agriculture system based on IoT and cloud computing, which involves building an agriculture information cloud with dynamic resources to ensure load balancing. This system handles large amounts of data collected through Radio-Frequency Identification (RFID) and wireless communication within the agricultural information cloud model. In [10] explores the limitations of IoT, specifically the communication delays between the cloud server and end-user applications. To address this issue, the paper suggests an AI-enabled edge computing model to enhance speed and range for IoT applications. Finally, [12] delves into data mining techniques applicable to IoT, providing an overview and emphasizing the efficient use of various techniques when dealing with IoT concepts to ensure maximum efficacy. The paper not only provides an overview of data mining but also emphasizes the need for a nuanced approach in choosing and implementing these techniques to ensure maximum efficiency in handling the vast and diverse datasets associated with IoT applications. This holistic exploration contributes to the development of robust and effective data-driven strategies in the domain of smart agriculture.

III. Proposed Work

There are, in particular, two broad categories of our proposed work. First, sensor data from a sensor node should be received and uploaded, with each sensor value stored in an object for monitoring and further analysis. Secondly, once the sensor data values are obtained, they can be used to analyze crop yield production, providing an advantage in predicting future crop production.

A. Working with sensor node and thingspeak

The central idea of this work is shown in Fig. 1, featuring sensors for relative humidity, ambient temperature, soil moisture, pH, and CO2, along with a microcontroller unit (MCU) with a Wi-Fi module, a Wi-Fi router, and the ThingSpeak cloud. A website has been created to monitor these details. Soil moisture is measured using an electrode, where the resistance between its terminals changes with moisture levels, affecting both the resistance value and the moisture percentage. A chip inside the sensor converts these resistance changes into an analog voltage. To produce an output voltage between 0 and 3.3 volts, the sensor is supplied with a 3.3-volt input. The soil moisture sensor's voltage signal is directed to the MCU, operating at less than 5V DC with a current limit of 20 mA. The DHT11 sensor, a multivariable sensor, provides relative humidity and atmospheric temperature readings. It measures humidity using capacitors and temperature using temperature-dependent resistors. Changes in output voltage from the resistors indicate temperature fluctuations. The sensor produces an analog signal, requiring noise removal after rectification to maintain signal accuracy.

The sensor must connect to a digitally active MCU input because it produces a digital output pulse. Data updates occur every two seconds because of this digital signal. The DHT11 sensor can measure temperatures from 0 to 50 degrees Celsius with an accuracy of 2 degrees, and humidity levels from 20 to 80% with an accuracy of 5%. It samples data at a rate of 1 Hz and operates on a current between 0 and 2.5 mA. Additionally, an MQ-2 sensor can be utilized to measure atmospheric CO2 levels. This sensor is a type of gas sensor based on metal oxide semiconductor technology, which changes resistance when in contact with gas. The gas concentration can be determined from the voltage using the voltage divider rule. The sensor can detect gas concentrations from 200 ppm to 1000 ppm and operates with a 5V DC input.

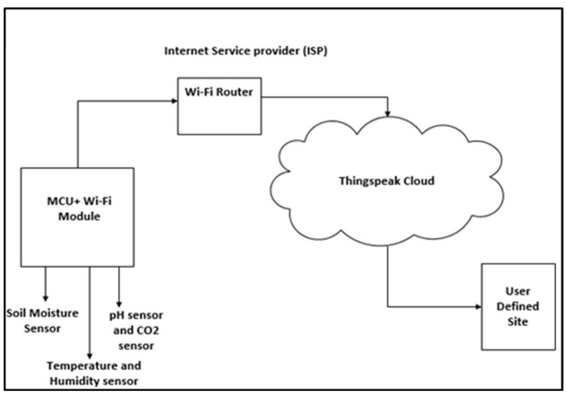

Figure 1. Block Diagram of proposed System

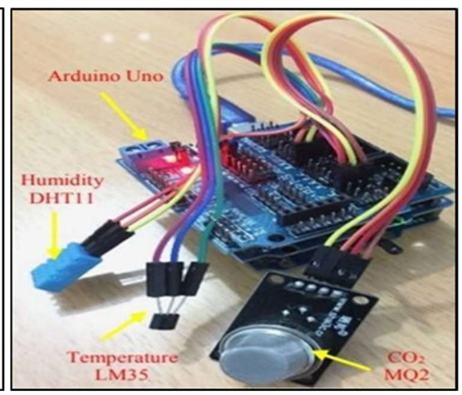

Figure 2. Interfacing of sensors

NodeMCU is a microcontroller unit that operates using the ESP8266 Wi-Fi module board. This board integrates all the components into a single piece, including general input/output pins and additional features like pulse-width modulation and analog-to-digital conversion. Each board component is housed in its own section. The NodeMCU board stores sensor data using its GPIO pins. Humidity and temperature sensor data are transmitted as analog signals to GPIO 0, while soil moisture sensor data is sent to one of NodeMCU's analog input pins. These pins are equipped with a 10-bit embedded ADC, converting the 0-3.3V analog signal range to 0-1023 units for precise measurements. The NodeMCU operates on a 32-bit Tensilica RSIC microcontroller with an 80 MHz clock speed. The device has 80 KB of SRAM memory capacity. Software is used for programming the MCU unit with Arduino IDE. The hotspot of a mobile phone could also be used to connect the NodeMCU board with WiFi, which can act as a router. To enable the connection to the Internet, an ID and password must be defined in accordance with the hotspot, ThingSpeak is used as a cloud platform for the described system. This platform ensures that the users can analyze and visualize data. This platform uses multiple channels to process the raw data that we receive from the sensor. To begin the data analysis process, the following five fields are created: pH, CO2 levels, relative humidity, ambient temperature, and soil moisture content. To define the system on the board, the system must automatically produce an API key in ThingSpeak. In order to provide the most accurate graphical displays of the sensor data, all values are provided as analog data values.

B. Implementation of Data mining

After the data is created, the measurement of each parameter is stored as a dataset. Following data storage, several data mining techniques are employed to accurately forecast future agricultural production levels. Crop yield forecasting is done using three main methods. Below are the specifics of the prediction model, data preparation, and datasets.

Datasets

Each dataset was constructed by beginning with the first value that the employed sensors picked up. Temperature, humidity, soil and atmospheric moisture content, CO2 percentage, and pH value are among the main datasets used in the analysis. The variables in the monthly rainfall pattern dataset are from government of India records that are made available to the public online. In addition to all of these values, the crop's yield is also considered so that it can be compared with the yield per year that is theoretically expected (ha).

Datasets Preprocessing

The dataset preparation step consist of gathering data from all sensor datasets and adding more parameters in Microsoft Excel.The dataset preparation step includes gathering data from all sensor datasets and adding more parameters in Microsoft Excel.A distinct set of columns is created for each crop, including monthly rainfall, percentage of CO2, temperature, pH, soil moisture, and humidity. By dividing the value of production by the area being produced, a figure that is almost equivalent to the crop yield can be found. Consequently, by editing the file with a.csv extension, removing the production column and adding the yield column, or vice versa.

Data Mining Techniques

It is possible to ascertain the extent to which a dependent variable depends on one or more independent factors by using statistical modeling techniques such as regression. Both categories of variables are referred to as predictors and dependent variables, respectively. For this method, all other variables that affect crop yield—such as area, temperature, soil moisture, etc.—should be viewed as independent, and production should be viewed as the dependent or target variable. Regression and classification techniques are similar in this context. In this study, Dempster Shafer, CNN (Convolutional Neural Networks), and PCA (Principal Component Analysis) are the classification algorithms used.

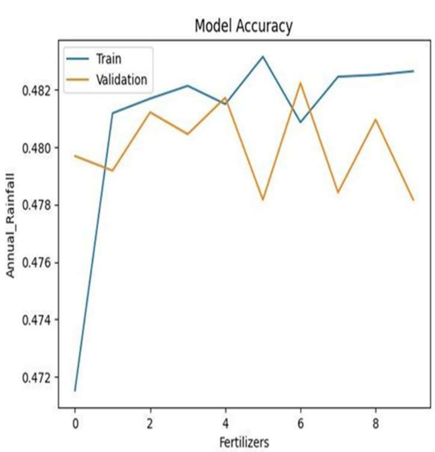

Figure 3. Annual Rainfall v/s Fertilizers (PCA)

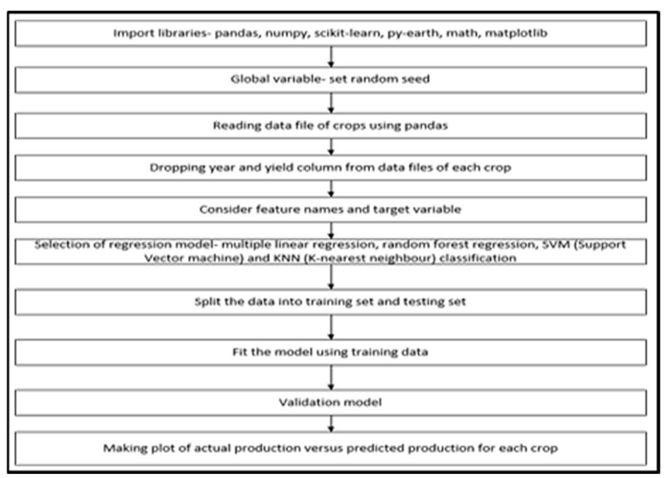

Figure 4. Procedure to construct the prediction model

Constructing the prediction model using sci-kit learn libraries

This section explains how to create prediction models for different data mining techniques, such as the previously discussed Dempster Shafer method, PCA and CNN were used, with implementation facilitated by the sci-kit learn and py-earth libraries. Sci-kit-learn, which interfaces with Python, is an open-source framework for data mining and analysis. The platform can be employed to carry out classification, clustering, and regression techniques. After the import of each library into the prediction model, a random seeding technique is used to develop global variables. Our dataset includes a production column that shows either yield or area in addition to the year and yield columns. Subsequently, the 2 previously indicated columns are no longer included in the dataset. The codes for the first three categories have now been added, and the data has been divided into training and testing sets. Now, the training data is used for evaluation and after applying several algorithms to the data values, a plot of the actual values vs the predictions is generated.

IV. Results and Observations

The mentioned study illustrates the results generated by the employed method, presented in Part A of the proposed research. These results showcase the fluctuation of values over time, which are then monitored using the ThingSpeak cloud server. Data Collected from the agricultural field is visualized in MATLAB, organized into distinct fields within a channel. For instance, Figure 4 displays the procedure to construct the prediction model, while subsequent figures from 5-8 various characteristics graphs are plotted.

A. Principal Component Analysis (PCA)

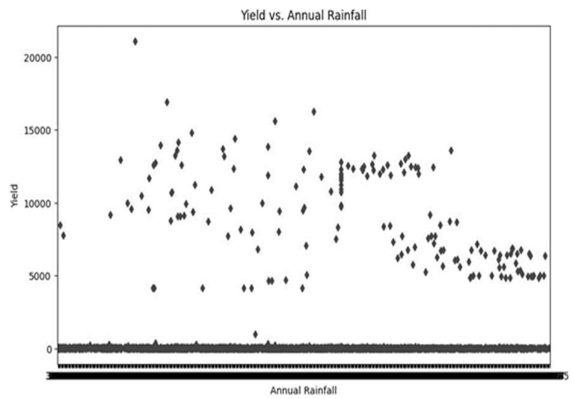

Figure 5. Yield Vs Annual Rainfall

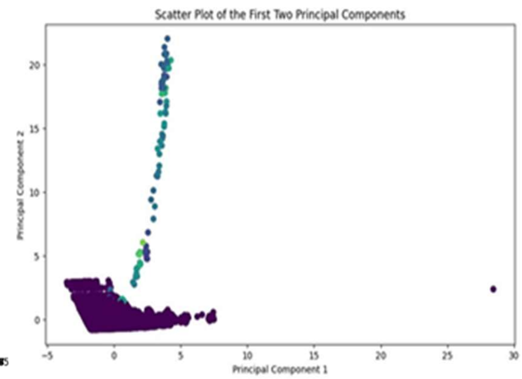

Figure 6. Scatter Plot of the First Two Principal Components

The graph between Yield and annual rainfall is plotted and it is observed that annual rainfall is scattered according to different yield amounts. The scatter plot of the First Two Principal Components 1 vs. 2 is plotted and it is summarized that soil is alkaline in nature because the majority of time the value crosses 7.

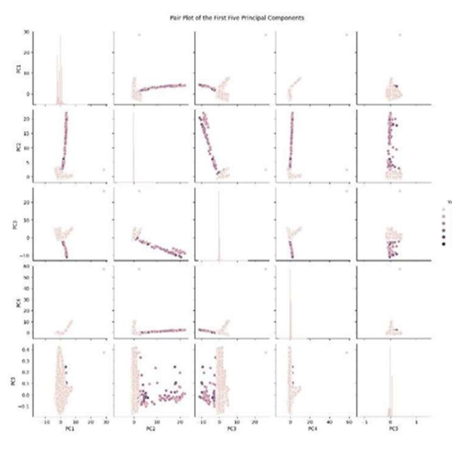

Figure 8. Comparison between Yield and Area

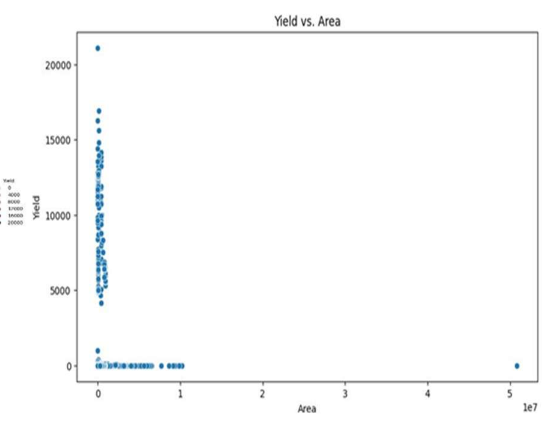

Figure 7. Pair plot of 5 Selected Components

The outcome of the research conducted in Part B, which involves the visualization of forecasted crop production through diverse data mining methods outlined earlier, is presented as Pair plot of the first Five Principal Components in figure 7. Every algorithm is employed on the provided datasets, and a comparison plot corresponding to each is depicted below. The plot shows specific components, highlighting a peak where PCA accurately represents data for over 10,000 values. In Figure 8, there is a graph comparing Yield and area. Although the actual and predicted graphs closely align at lower values, differences arise at higher values.

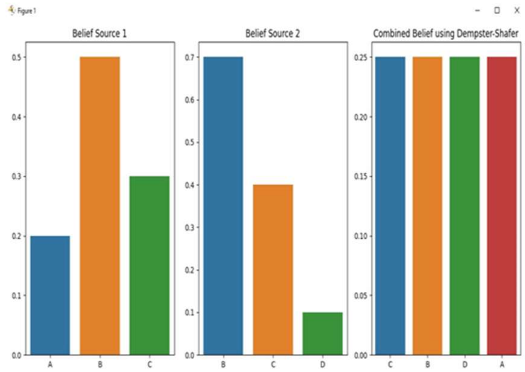

Figure 9. Dempster Shafer classification- comparison between Belief Source 1 and 2

B. Dempster Shafer Method

Based on the graph, it can be inferred that the classification method yields more accurate results for smaller data values. This is attributed to its precision in measuring data values of lower magnitude compared to those of higher magnitude in the comparison of actual versus predicted crop production.

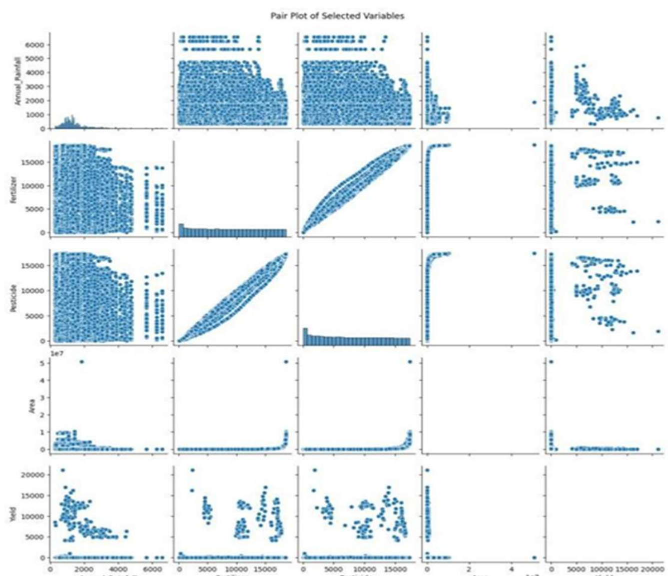

Figure 10. Pair Plot of Selected Variables

After analyzing the outcomes of all three methodologies, an examination was conducted to determine the accuracy percentage of the computed values in relation to the theoretical values. This observation encompassed various data values within the dataset and involved a total of 15,000 records for analysis.

C. Convolutional Neural Network (CNN)

The convergence of training and validation accuracy to a similar level in the last epochs indicates that the model has learned to generalize well on the validation data. The figure 12 shows the bar plot of Average Yield v/s Fertilizer and as we increase the amount of fertilizer, average yield increases.
The table suggests that for a small dataset, the Dempster-Shafer Classification method is the most suitable for prediction, while for a larger dataset, PCA is the recommended choice.

V. Conclusions

This paper demonstrates the evolution of a system designed to gather data from sensor nodes through IoT technology in the agricultural sector. The system effectively collects and transmits data to the ThingSpeak cloud,

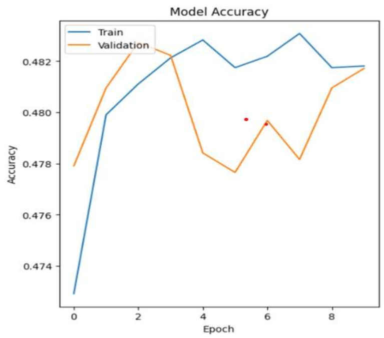

Figure 11. Accuracy over Epochs

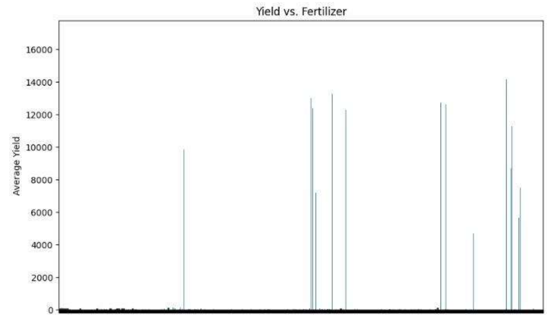

Figure 12. Bar Plot of Average Yield v/s Fertilizer

TABLE I. ACCURACY COMPARISON OF DEMPSTER SHAFER, CNN, PCA

NUMBER OF DATA VALUES USED	Accuracy Percentages of the Demonstrated Algorithms		
	Dempster Shafer	CNN	PCA
0	0	0	0
2000	78.20	69.20	67.01
3000	73.60	66.30	70.49
4000	79.89	59.38	54.90
5000	65.48	65.58	58.60
6000	50.97	63.95	61.90
7000	67.80	67.89	67.00
8000	52.78	68.79	65.00
10000	60.50	66.25	80.78
12000	58.57	67.23	84.50
15000	58.80	69.8	92.90

which users can access via their personalized websites. Data mining techniques are employed to forecast crop production, providing farmers with insights into the accuracy of their farming practices. Various regression and classification methods are utilized to generate outputs and corresponding plots. Future endeavors may involve presenting the data through mobile applications in addition to custom websites. Additionally, the system identifies crop diseases through image detection, enabling the application of appropriate fertilizers and pesticides for disease control. The integration of Artificial Intelligence and neural networks as regression methods holds promise for delivering more precise results.

ACKNOWLEDGMENT

We want to express our thanks to Mrs. Bhavnesh Jaint for her ongoing motivation and guidance during this project. Her timely support was essential for completing this paper. Additionally, we appreciate the publishers of the research papers we referenced.

REFERENCES

[1] W. Klosgen, J.M. Zytkow, Handbook of data mining and knowledge discovery, Oxford University Press, 2002.
[2] Kiruthika M, Shweta T, Mritunjay O, Kavita S, “Parameter monitoring for precision agriculture”, International Journal of The Scientific Research and Innovation 2015.
[3] O.P. Uma Maheshwari, M Savitha, “A Study of IoT (Internet of Things) in Agriculture”, International Journal of Innovative research in computer and communication Engineering.
[4] Dr. N. Suma, S. R. Samson, S. Saranya, G. Shanmugapriya, R. Subhashri “IoT Based Smart Agriculture Monitoring System”, International Journal on Recent and Innovation Trends in computing and communication.
[5] Sandip Khot, M. Gaikwad, “ Development of cloud based Light Intensity Monitoring system for Greenhouse Using Raspberry Pi” , IEEE International conference on computing communication control and automation (ICCUBEA) Aug 2016

[6] Sheetal V, A Bakshi, Tanvi T, “Green House by using IoT and Cloud computing”, IEEE International conference on Recent trends in Electronics, information & communication technology May 2016

[7] Aymen E Khedr, Mona Kadry, Ghada Walid(2015), “Proposed framework for Implementing Data Mining Techniques to enhance Decisions in Agriculture sector Applied case on Food Security Information Center Ministry of Agriculture, Africa”, international conference on Communication, management and Information technology (ICCM).

[8] Kuljit Kaur, Kalnwalpreet Singh Atwal “Effect of Temperature and Rainfall on Paddy Yield using data mining”, 2017 7th international conference on Cloud Computing, Data Science and Engineering Confluence.

[9] F. Tong Ke, “Smart Agriculture Based on Cloud Computing and IoT", in Journal of Convergence Information Technology, Volume 8, Number 2, 2013.

[10] Sivaganesan D. (2019). DESIGN AND DEVELOPMENT AI- ENABLED EDGE COMPUTING FOR INTELLIGENT-IOT APPLICATIONS. Journal of trends in Computer Science and Smart technology (TCSST), 1(02), 84-94.

[11] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM’ S FOR INTELLIGENT COMPUTING SYSTEM. Journal of trends in Computer Science and Smart technology (TCSST), 1(01), 14-24.

[12] Scikit-learn, Machine learning in python retrieved from http://scikitlearn. org.

[13] Crop yield prediction using machine learning: A systematic literature review – ScienceDirect

[14] Agriculture | Free Full-Text | Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize (mdpi.com)

[15] Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes | Neural Computing and Applications (springer.com).