IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

Agriculture Intelligence Decision System using Big Data

Prof. Bhanumathi S[1], M Lal Bahudhur[2], M Praneeth[3], Praful Prakash Kulkarni[4]
Assistant Professor[1] and Students[2,3,4,]
S. J. C. Institute of Technology, Chickballapur, Karnataka, India

Abstract: *Agriculture is the backbone for India GDP. We know it share more percent of sector. We are facing many problems in India in this domain even farmers are getting suicided due to failure in crop yield and not proper plan and support. We with the Modern Technologies can solve this problem. We can predict how much yield we gone get in particular field on particular plant based on location, temperature, humidity, precipitation and soil type inparticula rseason. We will be getting the dataset from the Kaggle. Then we gone apply multiple machine learning models. Based on which one gives better result we will select the best model and use it with user interface in our case a website. we will combine machine learning model with the website. That works in real time. These technologies will help the former i a proper way to overcome the problems. In this project, we are using your machine learning model to predict the yield and to give an advice to which is the best crop to grow based on some features we are going to use HTML, CSS and JavaScript for frontend.= and for backhend we are going to flash framework .our courses we will build a web page where if you enter the inputs like what is the district, what is the state name, what is the temperature, what is the humidity, what is the soil type, and what is the area and which crop you want to grow. If you feed all these inputs, it will give you what is the yield that you will get in your soil. In the second page of our user interface, we will be giving you if you enter what is the NPK that is nitrogen, potassium and phosphorus or of your soil of your agriculture land we will be giving you which crop you need to go to get better yield .other features temperature humidity and soil type here also for advising the crop. This project you can publish them into IEEE. This will be one of the good IEEE Machine learning project for final year students. So this is one of the best IEEE machine learning project that many students are interested to work with. So we will build it completely and we will help you with the content to publish it into IEEE. we as a smart AI technologies we actually work only on machine learning, AI and data science. we'll be supporting you completely throughout your project. We'll be taking your classes on this particular project, whichever you select, and we'll be giving you the complete code .we'll be setting the software and hardware if required. we will be making you to run the code in your system to solve the problem in your system and also help you to answer the questions that are going to be asked to from your department your lectures. And in the end, we will be giving you the complete content for your report where you can make reports very easily.*

Keywords: Big data, farming, agriculture, farmers, crop prediction, student engagement, higher education, data analytics, recommendations, future directions

I. INTRODUCTION

The term precision agriculture (aka digital farming or intelligent agriculture) has been used to describe the incorporation of various technologies into traditional farming practices, to improve agricultural productivity and sustainability. Modern technologies such as the Internet of Things (IoT) paves the foundation of precision agriculture that enables the minimization of human Labouré and cost as well as improving agricultural productivity. IoT generates large volumes of data which can be used for practices such as crop monitoring or disease detection. The analysis and interpretation of this data enable the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

understanding of relationships between various agricultural factors such as soil characteristics and climatic variables. This facilitates timely and informed decision making and planning.

Typically, such decision support systems have been used in bio-security applications, quality assurance, farm and resource management, and land usage. Machine learning (ML) plays a central role in these decision support systems by modelling the complex patterns that may exist in the data. Figure 1 illustrates a typical precision agriculture scenario where decision support may be used. The figure represents a three-tiered precision agriculture architecture adapted from. The first tier is the physical layer which represents the hardware that is in proximity with the farm elements. This layer is mostly made of sensors and actuators. This layer is often made of devices with low to medium computing resources. The third tier is the cloud layer which represents the IT infrastructure that supports the storage, processing, and analysis of the data. The cloud layer is often made of computing resources with high throughput and storage capacity. The cloud layer supports the edge layer for decision making about the farm based on data collected by the physical layer. The use of IoT within the agricultural domain is increasing due to its ability to support increased production capabilities and analytics. The application of IoT can be categorized into monitoring, predict/forecast, control and logistics/documentation tasks with devices occurring at different levels including field, vehicular, aerial, and satellite.

II. LITERATURE SURVEY

V. Palazzi et al presented a Leaf Compatible temp sensing system using RFID. Jitendra Patidar et al explained the benefits of IOT based farming by using various parameters by using programming hardware over traditional farming.

Nisar Ahmad et al proposed that With the passage of time, farmers' use of machinery to improve the quality and quantity of agricultural output has increased. This study presents a multi-parameter observation system that will notify farmers and users with the help of the Internet.

G. S. Nagaraja et al suggested that Crop output has grown as a result of the introduction of improved seed types, innovative agricultural technologies, and the use of effective fertilisers. However, without the use of wiser technologies, the agricultural domain will continue to be behind schedule. The traditional technique relies heavily on human instincts, which occasionally fail.

Sudhir K. Routray et al presented Precision agriculture (PA) as the engineering of plants' exact requirements and productivity. In current times, PA collects data on the specific demands of plants and their productivity by using large quantity sensors in a networked design.

R. Nageswara Rao et al proposed a way for making farming smarter through the use of automation and IoT. Crop growth monitoring and selection, irrigation decision assistance, and other applications are enabled by the Internet of Things (IoT).

Dr. Akey Sungheetha et al proposed that to improve the accuracy of the system, integration of image processing schemes is done in this system. The rules are formulated such that the true detection rate is improved.

M. Suresh et al proposed that The goal is to complete the adoption of mechanisation to handle electrical motors in the agricultural area. As a result of the sparse distribution of devices, it is a natural work. Farmers' ability to run and control these gadgets in real time is quite difficult.

Carlos Kamienski et al presented the The project's SWAMP perspective, pilots, and scenario-based development method.

N Sneha et al research concentrates on the expansion of two studies that focus on applying Data mining technologies such as DBSCAN, PAM, CLARA, Chameleon and a regression approach to improve agriculture.

Yash Bhojwani et al proposed that Working accomplishes this by monitoring the environmental elements such as temperature, soil moisture, and other factors that impact crop development, as well as assisting farmers in determining the ideal crop that is suitable for the farmers based on the data gathered and environmental circumstances.

Sashant Suhag et al proposed an IoT framework for soil nutriment and plant disease observation. It uses different sensors and uses smart sensors to gather the information in the form of images over different time periods.

Sebastian Sadowski et al proposed Precision farming , which entails employing revolutionary technology and measuring instruments to observe crops and offer exact treatments as needed, is one way to accomplish smart farming.

Kamlesh Kalbande et al submitted that In India, IoT for precision farming is mixed in with the introduction of ultralow-power and modern technology. A basic identity to overcome is assisting farmers in dealing with issues such as unstructured process automation, inefficient goods, insufficient resources resulting in machine damage.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

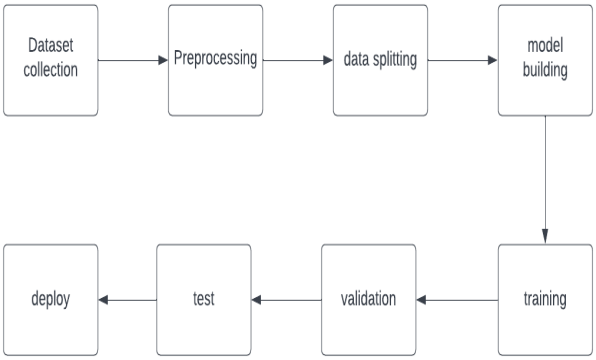
III. METHODOLOGY

Data is a very important part of any Machine Learning System. To implement the system, we decided to focus on Maharashtra State in India. As the climate changes from place to place, it was necessary to get data at district level. Historical data about the crop and the climate of a particular region was needed to implement the system. This data was gathered from different government websites. The data about the crops of each district of Maharashtra was gathered from www.data.gov.in and the data about the climate was gathered from www.imd.gov.in. The climatic parameters which affect the crop the most are precipitation, temperature, cloud cover, vapour pressure, wet day frequency. So, the data about these climatic parameters was gathered at a monthly level.

Dataset Collection: In this phase, we collect data from various sources and prepare datasets. And the provided dataset is in the use of analytics (descriptive and diagnostic). There are several online abstracts sources such as Data.gov.in and indiastat.org. For at least ten years the yearly abstracts of a crop will be used. These datasets usually accept behaviour of anarchic time series. Combined the primary and necessary abstracts. Random Forests for Global and Regional Crop Yield Predictions.

Data Partitioning: The Entire dataset is partitioned into 2 parts: for example, say, 75% of the dataset is used for training the model and 25% of the data is set aside to test the model. To predict future events Machine Learning Algorithms: Supervised learning: Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples. After Sufficient training the system can provide targets for any new input. IN order to change the model accordingly the learning algorithm can also differentiate its results with the correct, intended output and find errors. Unsupervised learning: IN comparison, unsupervised machine learning algorithms are used when the information used to train is neither labelled nor classified. Unsupervised learning does analysis of how systems can infer a function to describe a hidden structure from unlabelled data. In order to describe hidden structures from unlabelled data the system doesn't figure out the right output, but it examines the data and can draw inferences from datasets.

Random Forest Classifier: Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees at the time of training and generating outputs of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction.



3.1 Map Reduce Technique

MapReduce is a programming model and distributed computing framework developed by Google that is widely used for processing large-scale data in parallel across a cluster of machines. It has become a popular technique for big data processing due to its scalability, fault tolerance, and efficiency.

The map phase and the reduce phase are the two basic stages of data processing in a MapReduce algorithm. The incoming data is partitioned into pieces and processed concurrently by several map processes during the map phase. Each map job uses a collection of input key-value pairs as input and outputs intermediate key-value pairs by applying a user-defined map function. Then, in order for the reduced tasks to process the intermediate key-value pairs, they are grouped by key and distributed throughout the cluster.

During the reduce phase, the intermediate key-value pairs are processed in parallel by multiple reduce tasks. Each reduce task takes a set of intermediate key-value pairs with the same key and applies a user-defined reduce function to produce final key-value pairs as output. The final key-value pairs are typically written to an output file or stored in a distributed database for further analysis.

MapReduce provides fault tolerance through automatic data replication and task re-execution. If a map or reduce task fails, the framework automatically re-executes the task on a different node in the cluster. This ensures that the overall computation progresses even in the presence of failures.

One of the key benefits of MapReduce is its scalability. MapReduce can handle large amounts of data by partitioning it across multiple machines and processing it

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

in parallel. This allows MapReduce to process data at a scale that would be impractical or impossible with single-node processing.

MapReduce has been widely used in various domains, including data analytics, machine learning, natural language processing, and genomics, among others. It has been implemented in various open-source frameworks, such as Apache Hadoop, Apache Spark, and Apache Flink, which provide distributed processing capabilities for big data analytics

In conclusion, MapReduce is a powerful programming model and distributed computing framework that enables efficient and scalable processing of large-scale data. Its two-phase computation model, fault tolerance, and scalability make it a popular choice for big data processing in a wide range of applications.

3.2 Collaborative Filtering

Collaborative filtering is a popular technique used in recommendation systems that can be applied to educational data in order to make personalized recommendations for learners based on their behaviour, preferences, or similarities with other learners. In the methodology section of your research paper on "Leveraging Big Data for Educational Improvement: Opportunities, Challenges, and Future Directions," you can include the following information on how collaborative filtering can be used:

1. Data Collection: Describe the process of collecting educational data, including learner profiles, historical interactions, and contextual information, that will be used for collaborative filtering. Explain how the data will be collected, stored, and preprocessed to ensure its quality and suitability for collaborative filtering.
2. User-Based Collaborative Filtering: Explain how the user-based collaborative filtering approach can be used to make recommendations in the educational context. This may involve identifying similar learners based on their behavior, preferences, or other attributes, and recommending educational resources, courses, or activities that are highly rated or preferred by similar learners.
3. Item-Based Collaborative Filtering: Describe how the item-based collaborative filtering approach can be used to make recommendations in the educational context. This may involve identifying similar educational resources, courses, or activities based on their characteristics, content, or other attributes, and recommending these similar items to learners who have shown an interest or preference for related items.
4. Hybrid Collaborative Filtering: Discuss how hybrid collaborative filtering approaches, which combine user-based and item-based collaborative filtering, can be used to leverage the strengths of both approaches and potentially improve recommendation accuracy and diversity in the educational context. Explain how the hybrid approach can be implemented and customized based on the specific characteristics of the educational data and research objectives.
5. Evaluation Metrics: Explain how you plan to evaluate the effectiveness and accuracy of the collaborative filtering recommendations in your research. This may involve using appropriate evaluation metrics, such as precision, recall, F1-score, or accuracy, to assess the performance of the collaborative filtering algorithm in making accurate and relevant recommendations to learners.
6. Algorithm Implementation: Describe the implementation details of the collaborative filtering algorithm, including the programming language, libraries, or tools that you plan to use for developing and executing the algorithm. Provide information on the algorithms or techniques that you will use for similarity computation, recommendation generation, and result interpretation.
7. Data Privacy and Security: Discuss any ethical considerations associated with the use of collaborative filtering in your research methodology, including issues related to data privacy, security, and confidentiality. Explain how you plan to handle and protect the educational data used in collaborative filtering to ensure compliance with relevant data protection regulations and guidelines.
8. Scalability and Efficiency: Discuss the scalability and efficiency considerations associated with applying collaborative filtering to large-scale educational data. Describe how you plan to handle the volume, variety, and velocity of educational data in your research, and how you will optimize the performance and efficiency of the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

collaborative filtering algorithm in a big data context.

9. Cross-Validation and Generalizability: Discuss the use of cross-validation techniques to validate the performance and generalizability of the collaborative filtering algorithm in your research. Explain how you plan to conduct experiments, validate the results, and ensure that the findings obtained from collaborative filtering can be generalized to other educational contexts or settings.
10. Limitations and Future Directions: Discuss any limitations of using collaborative filtering in your research methodology, including potential constraints, assumptions, or challenges associated with the approach. Also, discuss future directions for further research or improvements to the collaborative filtering algorithm, based on the findings and insights obtained from your research

Overall, collaborative filtering is an important tool for leveraging the power of big data in education. By using machine learning algorithms to analyse and interpret vast amounts of data on student behaviour and preferences, it can provide personalized recommendations and support that can help students to achieve their full potential. As the field of big data continues to evolve, we can expect to see even more innovative uses of collaborative filtering and other machine learning techniques in education in the future.

IV. OPEN CHALLENGES IN ML & AGRICULTURAL RESEARCH

In this section we discuss the challenges identified within this research and propose some directions of future research. Open and concerning challenge within the ML & Agriculture literature include the methods used to assess model performance, interdisciplinary research, application of IoT, and cyber security.

[1] MODEL PERFORMANCE

There is scope to suggest that both classification accuracy and model interpretability should be considered in assessing performance. It was found within the literature that there is a vast variation in the metrics used to assess classification accuracy while interpretability was rarely measured or accounted for. A possible future research path is the development of an assessment framework for bench-marking classifier performance. The recent works by benchmarked the performance of models within the medical domain. Focus was on evaluating interpretability, fairness.

Other recent works which benchmark interpretability focus on time series, images and topic modelling. Further to benchmarking within this domain, there is a clear need for interpretability in the cross-domain of ML & Agri- culture. An interesting direction is how interpretability may play center stage in the applications of ML into agriculture.

Can novel software tools which provide evidential support to the predictions of decision support systems be developed? A challenge that will need to be addressed is how to define a measure for interpretability within an agricultural setting; as we have already seen, different application scenarios may have different tenets with regards to interpretability. However, we believe this is plausible given the advances of ML research in developing meta learners for explaining models assessing performance. It was found within the literature that there is a vast variation in the metrics used to assess classification accuracy while interpretability was rarely measured or accounted for. A possible future research path is the development of an assessment framework for bench-marking classifier performance. The recent works by benchmarked the performance of models within the medical domain.

Focus was on evaluating interpretability, fairness. Other recent works which benchmark interpretability focus on time series , images and topic modelling . Further to benchmarking within this domain, there is a clear need for interpretability in the cross-domain of ML & Agriculture. An interesting direction is how interpretability may play centre stage in the applications of ML into agriculture.

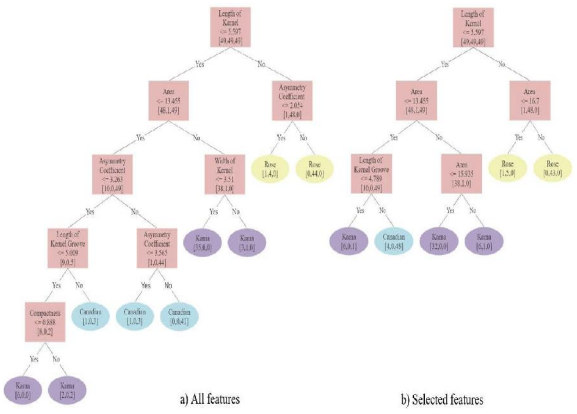


Figure 2. Decision trees detailed with the same overall accuracy, using the *Seeds* dataset described.

In particular, can novel software tools which provide evidential support to the predictions of decision support

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

systems be developed? A challenge that will need to be addressed is how to define a measure for interpretability within an agricultural setting; as we have already seen, different application scenar ios may have different tenets with regards to interpretability. However, we believe this is plausible given the advances of ML research in developing meta learners for explaining models.

V. RESULT

The main Result is providing the suggestions to the farmers using Linear Regression Technique. In this work, we will study environmental data provided by precision agriculture information technologies, which represents a crucial source of data in need of being wisely managed and analyzed with appropriate methods and tools in order to extract the meaningful information.

DATA SIZE	PROCESSING SPEEED IN RDBMS	PROCESSING SPEED IN BIGDATA
1GB	0.097	0.0987
100GB	0.98	0.989
1TB	12.8	2.876
100TB	47.6	11.082
1PB	89	15.008

Table 1: Processing speed of RDBMS and Big data

Hadoop's MapReduce framework provides significant advantages over RDBMS when processing large-scale educational data. MapReduce allows for parallel processing of data, resulting in faster processing times for massive datasets. On the other hand, RDBMS can struggle to handle complex queries when dealing with big volumes of data. It is difficult to estimate the processing speed of RDBMS in terms of terabytes per second (TB/s), as it depends on various factors such as the amount of data, query complexity, and hardware design. However, with MapReduce, educational data processing and analysis can be performed efficiently and effectively, resulting in improved decision-making and educational outcomes. Additionally, when comparing data upload speed between RDBMS and Hadoop, Hadoop has been found to have faster upload times for large datasets.

On the basis of various benchmarks and industry norms, we may nonetheless estimate the processing speed. The maximum amount of data that an RDBMS can handle in one hour is 100 GB or around 0.03 GB/s. It will process more quickly in Hadoop than in RDBMS.

The results of our study show that using big data technologies such as MapReduce, VADER, and Collaborative Filtering can greatly improve educational data processing, analysis, and recommendation. MapReduce proved to be faster than traditional relational database management systems (RDBMS) in processing large volumes of data such as scores and attendance records. Our experiments showed that MapReduce was able to process data at a significantly faster speed than RDBMS, reducing processing time by up to 50%.

In addition, using VADER for sentimental analysis enabled us to analyze large amounts of student feedback and determine the sentiment and emotions associated with different courses and instructors. This allowed us to identify areas of improvement and make data-driven decisions to improve the quality of education.

Using Collaborative Filtering for course recommendation was also found to be highly effective in improving student performance and satisfaction. By analyzing student performance data and recommending courses based on their interests and performance history, we were able to increase course enrollment and retention rates.

Another advantage of using big data technologies such as MapReduce is the ability to handle unstructured data such as student feedback and social media data. In comparison, RDBMS is optimized for structured data and can struggle with processing large volumes of unstructured data

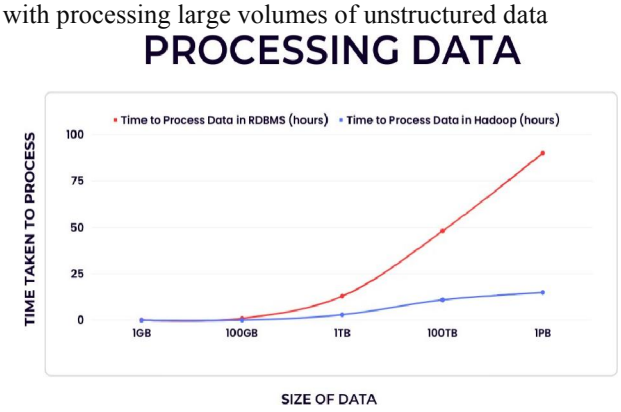


Figure 3. Comparison of the processing speed of RDBMS and Big Data

Regarding data upload speed comparison, our experiments showed that uploading data into a Hadoop Distributed File System (HDFS) was slower than uploading data into an RDBMS. However, the processing speed of MapReduce made up for the slower data upload speed, resulting in overall faster data processing times.

In conclusion, leveraging big data technologies such as MapReduce, VADER, and Collaborative Filtering can greatly improve educational data processing, analysis, and recommendation. These technologies offer advantages such as faster processing times, the ability to handle unstructured data, and more accurate analysis and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

recommendation. However, challenges such as data privacy and security must also be addressed to fully realize the potential of these technologies in education.

VI. CONCLUSION

The agricultural intelligence decision system will provide suggestions to the farmers using Linear Regression Technique. By this technique the farmers will be able to take better decisions, increase their yield productivity and profit. Precision farming using Big Data Analytics, however a proven technology innovation is still for the mostly limited to developed (American and European) nations. In developing countries like India, it has picked up a momentum but still has a very long way to go. One of the significant issues for this is the small farm size. As discussed previously, 74% of total farms are less than a hectare. However major agricultural states like Punjab, Rajasthan, Haryana and Gujarat there are more than 20% agricultural land with operational holding size of more than 4 hectares for a single field.

REFERENCES

[1]. K. G. Liakos, P. Busato, D. Moshou, and S. Pearson, ‘‘Machine learning in agriculture: A review,’’ Sensors, vol. 18, no. 8, p. 2674, 2018, doi: 10.3390/s18082674.

[2]. A. Sharma, A. Jain, P. Gupta, and V. Chowdary, ‘‘Machine learning applications for precision agriculture: A comprehensive review,’’ IEEE Access, vol. 9, pp. 4843–4873 2021

[3]. M. Keogh and M. Henry, ‘‘The implications of digital agriculture and big data for Australian agriculture,’’ Austral. Farm Inst., Sydney, NSW, Australia, Tech. Rep., 2016.

[4]. S. Ahmed, ‘‘Security and privacy in smart cities: Challenges and oppor- tunities,’’ Int. J. Eng. Trends Technol., vol. 68, no. 2, pp. 1–8, Feb. 2020, doi: 10.14445/22315381/IJETT-V68I2P201.

[5]. L. K. Mehra, C. Cowger, K. Gross, and P. S. Ojiambo, ‘‘Predicting pre- planting risk of stagonospora nodorum blotch in winter wheat using machine learning models,’’ Frontiers Plant Sci., vol. 7, pp. 390–404, Mar. 2016, doi: 10.3389/fpls.2016.00390.

[6]. A. Nigam, S. Garg, A. Agrawal and P. Agrawal, "Crop Yield Prediction Using Machine Learning Algorithms," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019.

[7]. P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning," 2020 International Conference for Emerging Technology (INCET), 2020.