iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** X **Month of publication:** October 2020

DOI: https://doi.org/10.22214/ijraset.2020.31848

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue X Oct 2020- Available at www.ijraset.com

Review of Machine Learning Techniques and Knowledge Analytics on Crop Production

Pratik Dubey[1], Shashi Kant Panday[2]
[1, 2]*Department of CSE & IT, VNS Group of Institutes Faculty of Engineering, Bhopal (M.P), India*

Abstract: Agriculture is important for basic human beings: 70% of India's population is based on agriculture. Climate change is a problem in 21st century Affecting agriculture, which is a bottleneck for crop production, increasing crop yields is very challenging in today's time.

If you want to increase yields in the future, you should know about the crop. The use of technology in agriculture is increasing day by day. It is difficult to analyze data in yields which data analytics is proving effective in understanding data which Will be helps in solving future agricultural problem.

I. INTRODUCTION

Agriculture is the basis of food grains, 70% of the population of India is dependent on agriculture. A variety of crops are grown in India. India's economic scenario is agriculture. Which is very wide impact of climate change and other factors on farming like temperature, soil, irrigation, rain, fertilizers etc.

Technology like it is big in agriculture, which has proved to help in increasing the yield of agriculture. Data analytics and data mining techniques are being used to increase agricultural production in the present and future. Which is proving helpful in improving crop production, which is also helpful in policy related policies.

Several efforts are being made to increase crop production, using a data set for estimating crop production, analysis of patterns is being done.

Which includes data mining, data analytics, machine learning, artificial intelligence and databases are being used. Data Analytics is a procedural phrase that focus on the numerous types of data analysis. Basically, It is the technique behind analyzing the raw data in order to generate a conclusion about the knowledge. The main objective of the analysis vary depending to the requirement. The requirement of information that needs to be extracted from the raw data makes data analytics purpose specific.

II. REVIEW OF LITERATURE METHODS OF CROP YIELD PREDICTION

data related to crop production in the agriculture sector is exclusively increasing day by day, therefore optimized crop production in agriculture field has become an important research area. There are a variety of data mining techniques have been used by researchers, many of researches for predicting crop production in agriculture. In some researches, Data mining techniques are integrated with various classifying and soft computing techniques to gain an efficient and optimized result in agriculture data set.

III. DATA ANALYTIC TECHNIQUES IN AGRICULTURE

A. Data Analytics

Data Analytics is a procedural phrase that focus on the numerous types of data analysis. Basically, It is the technique behind analyzing the raw data in order to generate a conclusion about the knowledge. The main objective of the analysis vary depending to the requirement. The requirement of information that needs to be extracted from the raw data makes data analytics purpose specific.

Steps of Data Analytics

1) First step is data accumulation. Data can either be taken from single source or from more than one sources according to the aim behind the analysis.
2) Next, data should have properly cleaned. Missing values and redundant values can also be handled properly.
3) When data is nearly ideal to process, a model is designed for efficient analysis. The model can also be integrated along with various techniques and approaches purely determined by the objective.
4) Finally, results should be in readable form. The proper visualization is required to understand the results.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue X Oct 2020- Available at www.ijraset.com

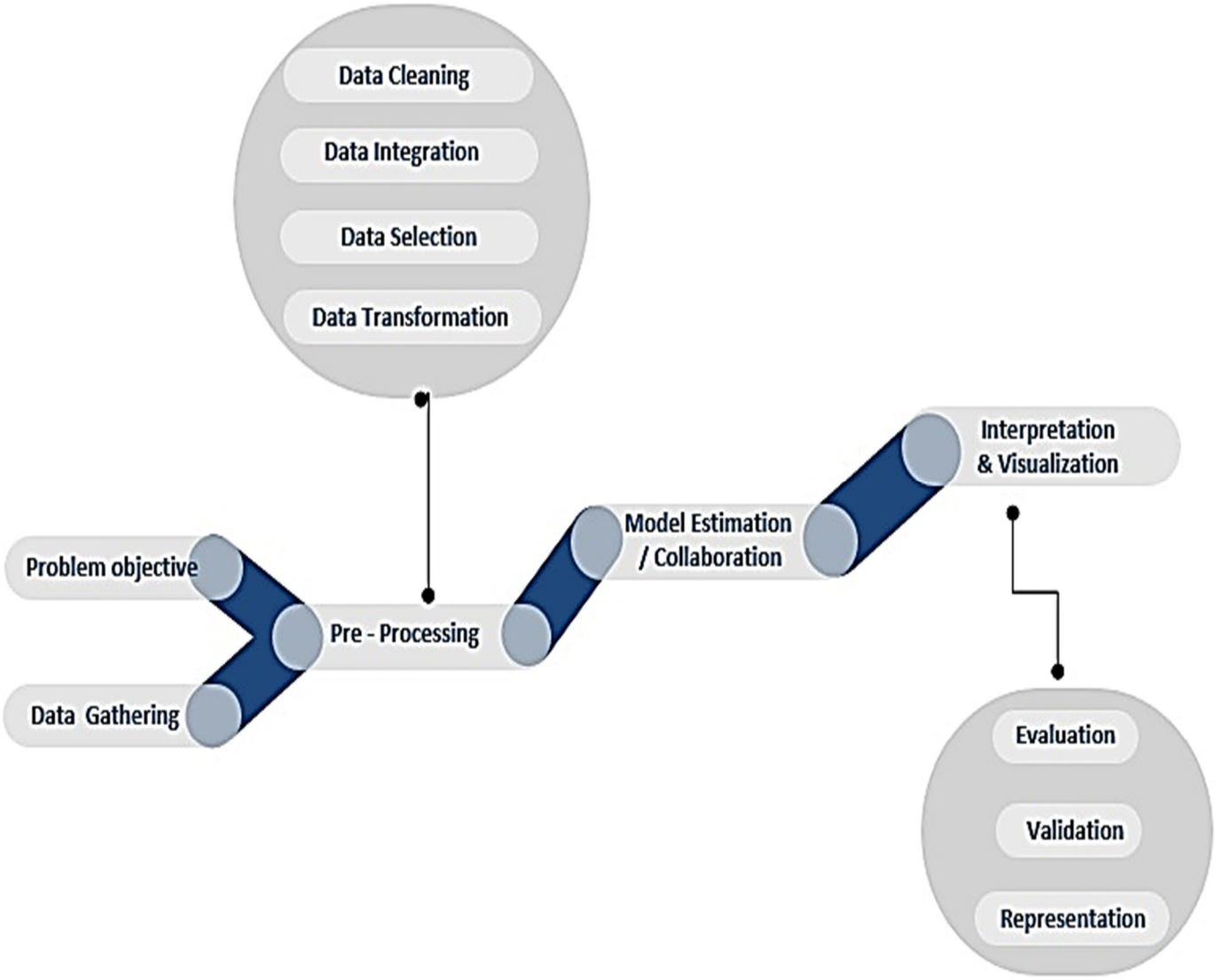


Figure 1.1: Generalized Analytical Model[29]

B. Data Mining

Data mining is simply an approach for extracting intention-based knowledge from the raw data set. It can also be used to explore huge data, the most frequent set of patterns in a dataset. The main aim behind the actual Data mining procedure is to excerpt the information data from a large collection of databases and transform it into an understandable framework for further use. These techniques are mostly used in the scenario where prediction-based problems need to be handled. Traditional approaches of data mining mostly suitable for basic clustering, forecasting and classification problems [17]. Data mining can be used for Prediction, Identification, Classification, and Optimization.

Data mining can be defined as the excerpting of hidden prophetic information from a large database. It can be defined as a conceptual phenomenon which is used to find out in huge data with the objective of finding useful information. The main aim of this technique is to find interesting patterns that were previously unknown as well as novel information. The terms KDD and data mining both are different [13].

KDD (Knowledge Discovery from Database) states to the approach of ascertaining useful information from data. It includes the assessment and elucidation of the patterns to make the decisions of what makes the grade as meaningful information. It also involves a superior description of the coding schemes, pre-processing technique, sampling, and data projection.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue X Oct 2020- Available at www.ijraset.com

C. KDD Process

According to Fayyad, Piatetsky Shapiro, Smyth, the term Knowledge Discovery in Database (KDD), invokes the top-level request for particular data mining methods. Learning, modeling, database, statistics, artificial intelligence, acquisition of systems knowledge and data visualization [26].

The amalgamating objective of this approach is to excerpt useful information from the data warehouse.

In this approach by performing data extraction methods to excerpt (recognize) what will recognize as knowledge, based on measurement and threshold specifications, using a database along with any necessary pre-processing, sampling and transformations of the database.

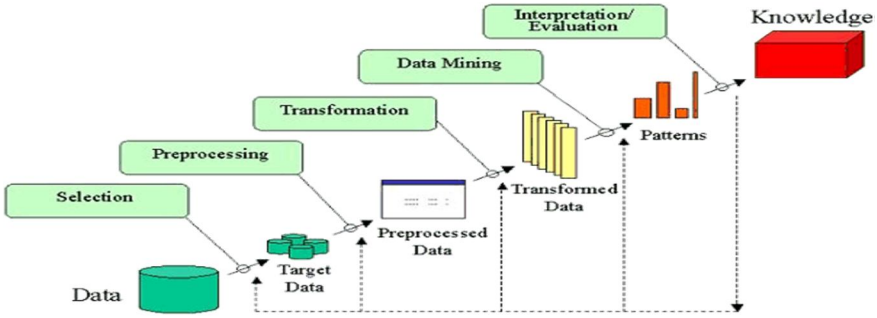


Figure 1.2: KDD Process [27]

In Figure 1.2, according to Fayyad, Piatetsky Shapiro, Smyth, a typical KDD process is described. The general process of research and interpretation of models involves the repeated application of the following steps:

1) Establishing an Understanding

a) The application domain.
b) The relevant information.
c) The aim of the end-user.

2) Create the desired data such as select an appropriate data or focus on a subset of attributes or data samples, in which the analysis will be made.

3) Cleansing and pre-processing of Data

a) Removal of noisy values.
b) Techniques for dealing with missing or inappropriate data values.
c) Accounting for time series statistics and acknowledged changes.

4) Data Reduction

a) Find useful attributes to represent data related to the purpose of the task.
b) Use methods of reducing or transforming dimensionality for reducing the actual number of attributes considered or to find invariable representations for the data.
c) The judgment on the aim of the KDD approach is done by using classification, regression, prediction, grouping, etc.

5) Choosing The Appropriate Data Mining Algorithm

a) Opt for the approach to be used for excerpting patterns from the data.
b) Decide which parameters and models will be suitable.
c) Selecting an appropriate data mining model with the global norm of the KDD approach.

6) Data mining

a) Looks for desired patterns which are efficient or a set of representations such as association rules, grouping, or classification, etc.
b) Understanding extracting Knowledge.
c) Integrating excerpting knowledge.

Data mining models be made up of a multifaceted set of association rules, equations or "transfer functions" that may be used to isolate useful data models, understand and predict behaviors.

They can be divided into two main classes according to their goals, as follows:

- Supervised / Predictive Models
- Unsupervised Models

D. Unsupervised Models

The models in which there is no output field or no target values, only input fields are available. Pattern Recognition is not a direct approach and it is not driven by a particular target attribute. The aim of such models is to uncover patterns in the set of input values [26].

Unsupervised models include

1) Clustering Models
2) Association and Sequence Model

E. Supervised / Predictive Models

In this type of model, the aim is to predict an event or to calculate the values of continuous and numerical attributes. In these models, input fields or attributes and destination field or target values are given. Input attributes are called as predictors because the model uses them to determine a prediction function for the output value [26].

We can assume input attributes as the X part of the function and the target value as the Y part, the model uses the input attributes which are analyzed on the basis of consequences on the output value, pattern recognition is supervised by the output value, associations are recognized among predictors and target values. An input-output attributes mapping function is engendered by a model which compares inputs with the target values and authorize the target values according to the given input values.

F. Data Mining Applications

Data mining can be implemented in many applications such as:

1) Data mining for Agricultural Crop Prediction
2) Data mining in Telecommunication Industries
3) Data mining in Retail Industries
4) Data mining in Healthcare and Biomedical Researches
5) Data mining and knowledge discovery in Science and Engineering

G. Data Mining Techniques

The most popular data mining techniques are:

1) *Estimation:* These types of models are similar to the classification model, but both have one major difference. They are used for prediction continuous field values on the basis of observed values of input attributes [16].
2) *Clustering:* Clustering is the approach in which a similar type of data are grouped together [2]. Typically, it is performed to provide the user with a high-level view of the database [19].
3) *Association:* It is a popular approach to find out interesting patterns and relationships between attributes in DB. The Association rules have been published to find out the regularity between the data of large-scale transactions recorded by POS systems (sales points) in supermarkets.
4) *Visualization:* Visualization is the process of representation of data so that users can easily understand complicated patterns. In this method, the graphical representation can also be used for data. It can be used with other data mining models to provide a better representation of interesting data.
5) *Classification & Prediction:* Forecasting and Classification are two customs of data analysis that can be cast-off to predict future data trends or to extract models that describe important data types. Based on certain criteria, the classification includes categorical (unordered) labels, prediction models, continuous value functions [24]. Bayesian categorization, Bayesian belief networks, rule-based classifiers and Support Vector Machines. Forecasting approaches include Linear Regression, Non-linear Regression, etc.

IV. ARTIFICIAL INTELLIGENCE

Agriculture plays a vital role in India's economy. Over 58 percent of the rural households depend on agriculture as their principal means of livelihood, Agricultural exports constitute 10 percent of the country's exports and is the fourth-largest exported principal commodity category in India. On the back of increased FDI and conducive government initiatives, the agriculture sector is increasingly looking at ways to leverage technology for better crop yield. Many technology companies and startups have emerged in the past few years with targeted agri-based solutions that benefit the farmers.

In this article, we explore applications of artificial intelligence (AI) to provide business leaders with an understanding of current and emerging trends, and present representative examples of popular applications.

Based on our research, the most popular applications of AI in Indian agriculture appear to fall into three major categories:

1) *Crop and Soil Monitoring:* Companies are leveraging sensors and various IoT-based technologies to monitor crop and soil health.
2) *Predictive Agricultural Analytics:* Various AI and machine learning tools are being used to predict the optimal time to sow seeds, get alerts on risks from pest attacks, and more.
3) *Supply Chain Efficiencies:* Companies are using real-time data analytics on data-streams coming from multiple sources to build an efficient and smart supply chain.

V. MACHINE LEARNING

Machine learning is a trending technology nowadays and it can be used in modern agriculture industry. The uses of ML in agriculture helps to create more healthy seeds.

A. Machine Learning Methods

In machine learning agriculture, the methods are derived from the learning process. These methodologies need to learn through experiences to perform a particular task. The ML consists of data that are based on a set of examples. An individual example is defined as a set of attributes. These sets of characteristics are known as variables or features. A feature can be represented as binary or numeric or ordinal. The performance of the machine learning is being calculated from the performance metric.

The performance of the ML model improves as it gains experience over time. To determine the performance of ML models and the machine learning algorithms agricultures various mathematical and statistical models are used. Once the learning process is completed, then the model can then be used to make an assumption, to classify and to test data. This is achieved after gaining the experience of the training process.

B. Machine Learning Functions

It can be divided into two categories, namely supervised and unsupervised learning.

1) *Supervised Learning:* In this machine learning agriculture method, the input data is represented with examples to the corresponding outputs. The primary goal of this function is to create a rule that will map the inputs to the corresponding outputs. In some cases, the inputs might not be available that may lead to missing output. The trained model is then used in supervised learning to predict the disappeared production and then the data is being tested.
2) *Unsupervised Learning:* In this machine learning agriculture technique, there is no difference between the trained models and the test sets, while unlabeled data is being used. The goal of this method is to find the hidden patterns.

C. Uses of Machine Learning (ML) in Agriculture

Artificial Intelligence is being used in various sectors from home to office and now in the agriculture sectors. Machine learning in agriculture used to improve the productivity and quality of the crops in the agriculture sector.

1) *Retailers:* The seed retailers use this agriculture technology to churn the data to create better crops. While the pest control companies are using them to identify the various bacteria's, bugs and vermins.
2) *AI is used to Boost the Yield of Crops:* The AI technologies are used to determine which corn and which conditions will produce the best yield. It will also determine which weather condition will give the highest return.
3) *AI helps to Identify bug Hunters:* One of the companies named Rentokil is using AI to kill all the bugs and vermin. Other companies are making use of Android app which is developed by Accenture to find bugs. The app takes the pictures of the bug and runs the app called as PestID. When a bug is found app will provide an immediate solution which helps the technician to take further actions. It will also recommend the chemical to be used to kill the bugs.

D. Machine Learning (ML) Models Used in the Agriculture Industry

1) The agricultural farmers are now taking advantage of the machine learning models and their innovations. Using AI and machine learning is good for the food tech segments.
2) The Farmers Business Network that is being created for the farmers a social network will make use of the ML and the analytic tools to drive the results of data on pricing.
3) Robots are now managing the crops and also monitoring them.
4) Sensors are helping to collect the data related to crops.
5) According to research if AI and ML are being used in agriculture, then the agriculture sector will grow in the coming years.

VI. CONCLUSION

In agriculture field, there are numerous factors that affects the overall production of a crop. These factors include environmental as well as meteorological factors. In Agriculture, a huge amount of crop-related data is generating throughout the year. This data includes various parameters such as soil nutrient value, rainfall, soil type, etc. Using Supervised Knowledge Analytics the data generated throughout the year can be analyzed for the parameters responsible for the poor productivity of a particular crop. In order to do so, the model assesses each and every record for that year and compare it with the knowledge base. The knowledge base is designed by taking experimentally and scientifically proven values of various parameters for every respective crop.

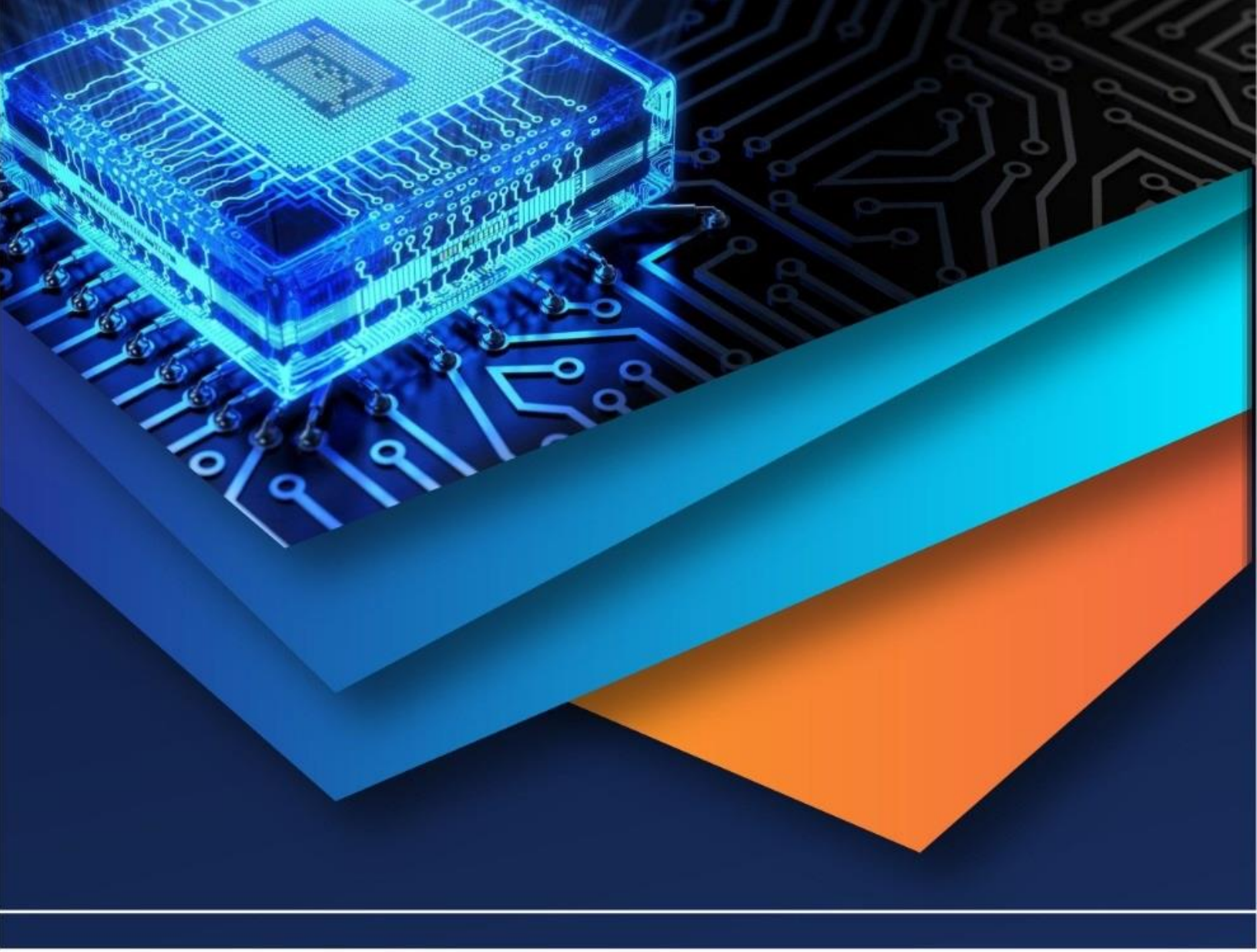
The analytical model evaluates the recommended values for the defected parameters. These parameters can be used for improving the production of the current harvested crop. The recommendations can also be used for future crop planning according to the area and other factors. The recommendations are authentic and trusted because the validation itself is done by experimentally proven values of various genuine research institutes.

Recommended values of fertilizer fairly able to reduce wastage of fertilizer that caused due to lack of optimum values of fertilizer required to be used by any crop.

REFERENCES

[1] S. Mishra, P. Paygude, S. Chaudhary and S. Idate, "Use of data mining in crop yield prediction," 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 2018, pp. 796-802.
[2] A. T. M. S. Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman, "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh," IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), Takamatsu, 2015, pp. 1-6.
[3] B. Imane, B. Abdelmajid, T. A. Mohammed, T. A. Mohammed and T. A. Youssef, "Data mining approach based on clustering and association rules applicable to different fields," International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, 2018, pp. 1-5.
[4] D.Rajesh, "Application of Spatial Data Mining for Agriculture", International Journal of Computer Applications (0975 – 8887) Volume 15– No.2, February 2011.
[5] Miss. Snehal S. Dahikar, Dr. Sandeep V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", IJIREEICE, Vol. 2, Issue 1, January 2014.
[6] Manish Sahane, Balaji Aglave, Razaullah Khan and SanjaynSirsat, "An Overview of Data Mining Techniques Applied to Agriculture Soil Data", International Journal of Agriculture Innovations and Research, Volume 3, No. 2, September 2014.
[7] M. Paul, S. K. Vishwakarma and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach," International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, 2015, pp. 766-771.
[8] N. Gandhi and L. Armstrong, "Applying data mining techniques to predict yield of rice in humid subtropical climatic zone of India," 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 1901-1906.
[9] N. Hemageetha, "A survey on application of data mining techniques to analyze the soil for agricultural purpose," 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 3112-3117.
[10] A. Savla, N. Israni, P. Dhawan, A. Mandholia, H. Bhadada and S. Bhardwaj, "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture," International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2015, pp. 1-7.
[11] P. Surya, Dr. I. Laurence Aroquiaraj, "Crop Yield Prediction in Agriculture Using Data Mining Predictive Analytic Techniques", IJRAR December 2018.
[12] K. Pavya, Dr. B. Srinivasan, "Feature Selection Techniques in Data Mining: A Study", IJSDR, June 2017.
[13] Snehal S. Dahikar, "An Artificial Neural Network Approach for Agriculture Crop Prediction Based on Various Parameters", IJARECS, Volume 4, Issue 1, January 2015.
[14] Gang Liu, Xuehong Yang and Minzan Li, "An Artificial Neural Network Model for Crop Responding to Soil Parameters", Springer 2005.
[15] P. Priya, U. Muthaiah and M. Balamurugan, "Predicting Yield of The Crop Using Machine Learning Algorithm", IJESRT, April 2018.
[16] Georg Ruß, "Data Mining of Agricultural Yield Data: A Comparison of Regression Models", Springer 2009.
[17] Sunita Beniwal, Jitender Arora, "Classification and Feature Selection Techniques in Data Mining", IJERT, August 2012.
[18] Ms. Sonali. B. Maind and Ms. Priyanka Wankar, "Research Paper on Basic of Artificial Neural Network", IJRITCC, January 2014.
[19] K.Bharatha Krishna and S.S.Suganva, "Application of Data Mining in Agriculture", IJRCAR, August 2017.
[20] Dai, Qing-yun & Zhang, Chun-ping & Wu, Hao. Research of Decision Tree Classification Algorithm in Data Mining. International Journal of Database Theory and Application, 2016.
[21] Bhumika Gupta, Aditya Rawat, Akshay Jain, Arpit Arora and Naresh Dhami. Analysis of Various Decision Tree Algorithms for Classification in Data Mining. International Journal of Computer Applications 163(8):15-19, April 2017.
[22] Priyanka Gaur, "Neural Networks in Data Mining", International Journal of Electronics and Computer Science Engineering, ISSN- 2277-1956.
[23] Dr. Yashpal Singh, Alok Singh Chauhan, "Neural Networks in Data mining", Journal of Theoretical and Applied Information Technology.
[24] Nayak, Janmenjoy & Naik, Bighnaraj & Behera, Dr. H. A comprehensive survey on support vector machine in data mining tasks: Applications & challenges, 2015.
[25] Yujun Yang, Jianping Li and Yimei Yang, "The research of the fast SVM classifier method," 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015.
[26] Bharati, M &, Ramageri., DATA MINING TECHNIQUES AND APPLICATIONS. Indian Journal of Computer Science and Engineering, 2010.
[27] Han, J., & Kamber, M., Data mining: Concepts and techniques. San Francisco: Morgan Kaufmann Publishers, 2001.

[28] Lei Shi, Qiquo Duan, Xinming Ma and Mei Weng, "The Research of Support Vector Machine in Africultural Data Classification", IFIP, 2012.
[29] Sharma, Gaurav. (2019). Data Analytics Model for Optimal Crop Management. International Journal for Research in Applied Science and Engineering Technology. 7. 690-693. 10.22214/ijraset.2019.6119.
[30] https://emerj.com/ai-sector-overviews/artificial-intelligence-in-indian-agriculture- an-industry-and-startup-overview/
[31] https://technostacks.com/blog/machine-learning-in-agriculture/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY