International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 08 | Aug 2021 www.irjet.net p-ISSN: 2395-0072

Predictive Analysis of Agriculture

Sakshi Manral[1], Ms. Neha Chandela[2]

[1]Student (1716113035), Krishna Engineering College, Branch of Information Technology, 2017-2021
[2]Assistant Professor, Krishna Engineering College

Abstract - *Agriculture, together with its linked industries, is indisputably India's primary source of income, particularly in the country's vast rural areas. It also contributes significantly to the Gross Domestic Product (GDP). In the past few decades, the agriculture field has seen lots of technological changes to improve productivity using prediction methods; farmers can enhance the productivity of crops. Precision agriculture is the process of recognizing previous year yield values, comprehending current environmental requirements, and utilizing data that quantifies variances in soil and crops within agricultural areas. These methods are used to determine the number of crops, seeds, humidity, water level, and other necessary supplements. As a result, this may prevent the use of an excessive number of accessories for cultivation, save money on pesticides and fertilizers, and boost crop production. The main principle behind this notion is that by cultivating the appropriate crop at the correct time and the correct cost, The farmer may attain a larger yield.*

Key Words: Predictive analysis, Regression, Data Collection, Big data.

1.INTRODUCTION

This India, with 1,350,716,488 (1.35 billion) people, is the second-most populous country in the world. [8] Data from World Bank shows that around 60.3 percent of India's land area is agricultural land. [9] India has the world's tenth-largest arable land resource. India has all 15 central climates in the globe, with 20 Agri-climatic areas. In addition, the country has 46 of the world's 60 soil types. For around 58 percent of India's population, agriculture is their primary source of income. [10] Agriculture is a unique industry in which a range of environmental and economic factors influence crop output. Agriculture is influenced by soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall, harvesting, pesticides, weeds, and other factors.

The global population is gradually increasing, while crop production resources are rapidly depleting. The main reason the Indian farmers face catastrophic crop failure is that they do not know how to obtain the best yield. There are many other reasons too for crop failures like pests and adverse weather conditions etc. There is a compelling need to increase crop yield, which can be achieved through agricultural predictive Analysis.

The proper yield calculation for the different crops involved in the planning is a critical issue for agricultural planning. Data mining techniques are a vital component of achieving practical and successful solutions to this issue. Furthermore, yield prediction would also assist in deciding the strategy for a better yield of a specific crop. Tweaking conditions after sowing the crop can result in a better outcome. If the prediction of yield for a particular set of initial conditions is known, then deciding the plan for the growing period will be easier.

2. REQUIREMENT

Economical cause

Agriculture is dying as a desired job, not as a means of producing food. Agriculture is being shunned by both middle-aged and young people nowadays. The country's next generation of farmers may be extinct.

Seventy percent of Indian youths live in rural areas where agriculture is still the primary source of livelihood. Every day, 2,000 farmers abandon their farms. 76% of farmers would instead do something else than farming. A farmer's income is around one-fifth that of a non-farmer. Agriculture is currently put to the bottom of the list when choosing a source of income.

When given the option, about 45 percent of farmers polled by the National Sample Survey Organization said they would leave farming; two out of every five farmers said they would leave farming. According to the 2014 survey conducted by the Delhi-based Centre for the Study of Developing Societies (CSDS), about 60 percent of farmers were ready to quit farming for a better job. The reason is quite simple because anyone's poor yield, whether due to errant rains, pests, etc., affects their livelihood. So technically agriculture is dying in our country the main reason is not getting enough revenue for their crops sometimes due to adverse weather or pests etc.

Technical cause

There have already been many technical advances in this field, but several disadvantages have been listed for the following work.

Data mining techniques are used to analyze agricultural data, resulting in the use of big data.: In this research paper, we can find various techniques like PAM, CLARA (clustering large applications), and DBSCAN, etc. This technique is very complicated. Agricultural Predictive Analysis to Increase Crop Productivity this research paper includes the usage of ZeroR classifiers. This method is a bit expensive because it requires optional hardware like a quadcopter which is fixed in the agricultural field, to calculate the humidity and weather condition of a particular area.

Predictive Mean Matching Method for Missing Data Analysis and Imputation on Agriculture Data: The disadvantage of

this technique is that this technique works only with real-time agricultural inputs and not with historical data.
The following disadvantages will be considered, and we will try to develop a technique which will not consist of these problems.

3. DETAIL SEARCH

3.1 Predictive analysis

In predictive analysis data, statistical algorithms and machine learning techniques are used to produce future outcomes based on percept history, or we can say historical data. Predictive analytics employs a variety of methodologies, including data mining, statistics, and modeling, machine learning, and artificial intelligence. The motive is to achieve a result better than what has already happened and provided the best possible outcome. Predictive Analytics can take past and present data and provide predictions of future trends. The identification of potential risks or opportunities enables us to get more fast and better results.
Decision trees use branching to show possibilities stemming from each outcome or choice.
Regression techniques assist with understanding relationships between variables.
Neural networks utilize algorithms to figure out possible relationships within data sets.

3.1.1 Models Used

Decision tree

Predictive Analysis uses several algorithms to build a prediction model based on historical data. One of the most often used and popular methods for developing prediction models and algorithms is the decision tree. Decision trees are somewhat like flow charts. A conceivable option, outcome, or reaction is represented by each branch of the decision tree. The end results are represented by the tree's farthest branches.
Starting from the beginning, you answer a question that leads us to some questions. We keep on doing this until we reach the terminus and get the final outcome.
For predictive modeling, decision trees are preferred above other methods because they are both simple to grasp and effective. A decision tree's primary purpose is to divide a large amount of data into smaller chunks.
Prediction is divided into two stages. The tree is generated, evaluated, and optimized using an existing set of data in the first stage of the model's development. The model is then used to forecast an unknown outcome in the second stage.

Regression techniques

Another form of predictive modeling is regression analysis. Regression analysis investigates the relationship between a dependent and an independent variable. The target is the dependent variable, while the predictor is the independent variable(s). Forecasting, time series modeling, and determining the causal effect link between variables are all done with this technique.
The use of regression analysis has a number of advantages. The following are the details:
It denotes the existence of significant correlations between the dependent and independent variables.
It expresses the degree to which many independent factors have an impact on a dependent variable.
We can also compare the effects of factors assessed on different scales using regression analysis.
There are different types of regressions available some commonly used regression techniques are:

1. Linear regression
2. Logistic regression
3. Polynomial regression
4. Stepwise regression
5. Lasso regression
6. ElasticNet regression

Neural networks

A neural network is a type of predictive analytics technology that can be used to classify or predict a variable. A Neural Network is a machine learning technique. It attempts to predict values in the same way that a human brain might.
A Neural Network is constructed by connecting a web of input nodes (which serve as the network's starting point and where data is inserted), output nodes (which display the network's results/predictions), and a hidden layer between these nodes.
The Neural Network's unique and efficient feature is the hidden layer between the input and output nodes. Every time the Neural Network is 'fed' input, the algorithm incorporates the data that travels through it by adding 'weights' to the nodes in the hidden layer, which may modify the outcome in the output nodes.

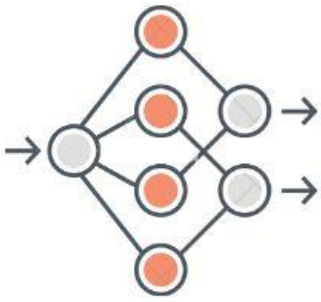

Fig. 1 Neural Network

3.2.1 Data Mining

Data mining is a method of converting raw data into valuable information by searching vast amounts of data for patterns using the software. Effective data collection, warehousing, and computer processing are all required for data mining.
Data mining involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users.

3.2.2 Machine Learning

Machine-learning algorithms use a statistical approach to find patterns in a massive amount of data, and the data here consists of a lot of things—numbers, words, and images, taps, or whatever is the input. It can be fed into a machine-learning system if it can be digitally saved.

There are 14 Types of machine learning

Learning Problems

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

Hybrid Learning Problems

4. Semi-Supervised Learning
5. Self-Supervised Learning
6. Multi-Instance Learning

Statistical Inference

7. Inductive Learning
8. Deductive Inference
9. Transductive Learning

Learning Techniques

10. Multi-Task Learning
11. Active Learning
12. Online Learning
13. Transfer Learning
14. Ensemble Learning.

3.2.3 Big Data

Big data is a word that refers to the vast amount of structured and unstructured data that affects a company on a daily basis. But it's not the quantity of data that matters; its what businesses do with it. Big data may be analyzed for insights that can help you make better decisions and make strategic business movements.

3.2.3.1 TYPES OF BIG DATA.

Structured

One of the sorts of large data is structured data.
Structured data is data that can be handled, stored, and retrieved in a consistent manner. It refers to neatly ordered material that can be stored and accessed from a database with ease using simple search engine methods. For example, the book table in a library's database will be structured so that book details, such as type, order, and so on, are all ordered.

Unstructured

As a result, processing and analyzing unstructured data becomes highly challenging and time-consuming. Email is one example of unstructured data. There are two types of big data: structured and unstructured data.

Semi-structured

Semi-structured data is the third type of big data. Data that has both the structured and unstructured types mentioned above is referred to as semi-structured data. To be more explicit, it relates to data that contains crucial information or tags that can identify individual elements within the data but is not categorized in a database.

4. PRECISION IN AGRICULTURE

Precision agriculture is the process of recognizing previous year yield values, comprehending current environmental requirements, and utilizing data that quantifies variances in soil and crops within agricultural fields [3]. Precision agriculture techniques have played an increasingly important part in the economy during the last decade. Despite the numerous options available, farmers continue to struggle to manage productivity in a consistent manner. It is essential to know the crop condition information at the early stage, and the same task is as crucial at the time of harvesting the crop. Predictive Analysis in agriculture is a crop management concept that aims to raise arable agriculture's environmental, economic, market, and public challenges.
The main objective of the project is to focus on the yield of the crop, and for this, the first and foremost process in prediction is data collection then data analysis, monitoring, and the final result can be plotted. To effectively use this methodology, previous year yield value, cultivated area, irrigation methods, fertilizer and pesticide usage, rainfall level at each season, soil maturity, and weather conditions of a specific location must all be progressed in a systematic manner. The following steps would let us know about the overall concept -:
The first step is to get some data from the previous year. Data is gathered for various weather conditions, soil type, humidity, air quality, crop maturity, labor costs, and previous year's statistics.
Then the second step would be achieved by processing those data using a machine learning algorithm, which would provide us the predicted yield value.
The anticipated value can then be plotted as a graph with matplotlib in Python.
The final result is shown in a graph-based on different parameters.
The quantitative models can produce quantitative results in crop monitoring, which will help to develop crop growth in different conditions.
Predictive Analysis is a technique that uses prior year data to make informed decisions about the future in the present. This model incorporates several patterns derived from previous data in order to analyze future projections.

4.1 Benefits for Farmers

Crop varieties, fertilizer kinds and doses, pesticides and herbicides, and irrigation are all tailored to fit the needs of crops for optimum growth and development.
It may result in higher yields, especially in critical locations where uniform crop management approaches have been used in the past. Farmers can increase labor, land, and agricultural time efficiency by predicting modern

technologies such as machinery, tools, and input information.

Fig. 2 How prediction helps farmers

5. CONCEPT

This project employs the notion of Smart Farming in conjunction with previously collected data.

Data collection

Data collection is the most crucial step in prediction. As the forecast is based on the previous year's statistics, a slight variation in data collection may affect the output. As a result, it's critical to examine the prior year's data and assess the original data value for monitoring.

Soil type, soil maturity, pesticides, fertilizer, weather conditions, humidity, water level, rainfall, labor cost, and availability based on prior-year values are all examples of data. Prediction results are calculated using these data and an algorithm.

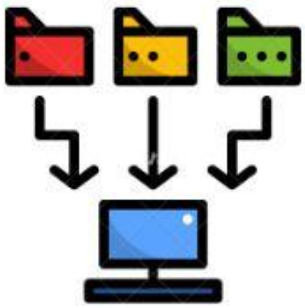

Fig. 3 Data Collection

Graph based model

Predictive agriculture includes predictions based on previous year statistics. The final step is to deliver the result to the end-user. Graph-based approaches are easy to understand the obtained results. So, the output can be easily understood by generating graphs.

Many built-in library functions are available in the Python language. We must use matplotlib, which is an integrated graph charting toolkit, to plot a graph model. Area graphs can be plotted based on expected values by simply applying the values to graph parameters such as yield and productivity. This graph-based method is simple to use and analyze the variance in crop yield from year to year.

6. OTHER USEFUL TECHNIQUES

Multiple linear regressions

Multiple linear regression is a variant of "linear regression" analysis. This model is built to establish the relationship that exists between one dependent variable and two or more independent variables. For a given dataset where x1 ... xk are independent variables and Y is a dependent variable, the multiple linear regression fits the dataset to the model:

The formula for regression is as follows,

Regression Equation

$$y_i = \beta o + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta k x_{ki} + \varepsilon_i +$$

Where,

β0 is the y-intercept and β1, β2... βk parameters are called the partial coefficients. In matrix form

Y=XB+E

Partition around medoids (PAM)

It is a partitioning-based algorithm. It breaks the input data into number of groups. It finds a set of objects called medoids that are centrally located. With the medoids, nearest data points can be calculated and made it as clusters.

The algorithm has two phases:

BUILD phase, a collection of k objects is selected for an initial set S.

Arbitrarily choose k objects as the initial medoids.

Until no change, do.

(Re) assign each object to the cluster with the nearest medoid.

Improve the quality of the k-medoids (randomly select a non-medoid object, O random, compute the total cost of swapping a medoid with O random).

SWAP phase, one tries to improve the quality of the clustering by exchanging selected objects with unselected objects. Choose the minimum swapping cost.

Naïve Bayes algorithm

Bayes theorem provides a technique that we can compute the possibility of a guess given our former knowledge. Bayes' Theorem is declared as:

Where

$$P(h|d) = \frac{(P(d|h) \times P(h))}{P(d)}$$

P(h|d) is the possibility of guess h given the data d. This is called the posterior probability

$P(d|h)$ *is the possibility of data d given that the guess h was correct.*

P (h) is the possibility of guess h being correct (regardless of the data). This is called the prior probability of h.

P (d) is the possibility of the data (regardless of the guess).

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} X = \begin{bmatrix} 1\ x_{11}\ x_{12}\ \dots\ x_{1k} \\ 1\ x_{21}\ x_{22}\ \dots\ x_{2k} \\ . \ . \ . \ \dots \ . \\ . \ . \ . \ \dots \ . \\ 1\ x_{n1}\ x_{n2}\ \dots\ x_{nk} \end{bmatrix} B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_k \end{bmatrix} E = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

7. SUMMARY AND CONCLUSION

7.1 Summary

Crop yield monitoring is a notion of predictive Analysis in agriculture. This concept's implementation should aid farmers in producing larger yields. Input factors can be minimized, and output can be maximized precisely with this notion. The system also handles crop diseases in order to boost yield. The system can locate the required amount of land, water, fertilizer, and pesticides to increase agricultural productivity.

Farmers will be able to plow precise resources for output while reducing costs in a more effective method. This procedure saves a significant amount of money in terms of chemicals and labor. As a result, those farmers get a return on their investment by saving money on fertilizer. Another advantage of this approach is its ability to adapt to changing environmental conditions. Based on various environmental structures, a farmer can cultivate a good crop on a respectable field at the appropriate moment, hence increasing productivity. As a result, precision agriculture aspires to play a significant role in both the economic and environmental fields.

7.2 CONCLUSION

To understand the various attribution features for the precision farming conditions and develop a framework for weighted averaging the features.

To develop machine learning models that can support holistic feature selection-based Analysis of the data and support in more accurate conditions of precision farming conditions.

To develop an interventional process of the machine learning model that can take into account the lifecycle stage of a crop and support in real-time precision farming analysis.

8. FUTURE ENHANCEMENTS

Currently, a larger number of companies focus on predictive analytic techniques to enhance their business. Since agriculture is the backbone of any country, making an intelligent decision in agriculture is becoming a more important task in day-to-day life. As long as the population increases, the demand for food increases, and the arable land level diminish.

These factors force the farmers to do high productivity with fewer resources in a sustainable manner. In the future, sensors can be fixed to predict the yield of any crop. A farmer can take a photo of a crop on a mobile phone and upload the picture in the system, where the system could predict the properties of the crop and can judge the result. Crop properties can be matched with earlier image databases, and by image processing techniques, crop maturity can be identified based on color and weather factors.

As long as the technology develops, prediction techniques may enhance its efficiency. A quadcopter device can be fixed in the agriculture field to calculate the humidity and weather conditions of a particular area. These small rotaries and fixed-wing aircraft can fly at low altitudes and be programmed to fly a specific pattern. The quadcopter can take high-definition images with its camera and produce properties of the crops from the database. Frequently, the quadcopter provides sensor reading and is based on real-time results to make more efficient predictions among the crops. Therefore, predictive Analysis plays a significant role in future agriculture.

9. REFERENCES

[1] Jinubala, V Lawrance, R Analysis of Missing Data and Imputation on Agriculture Data With Predictive Mean Matching Method, International Journal of Science and Applied Information Technology (IJSAIT), Vol.5, No.1, Pages: 01-04 (2016)

[2]T. Rajasekaran1, P. Jayasheelan2 and K.R. Sri Preethaa3, Predictive Analysis in Agriculture to Improve the Crop Productivity using ZeroR algorithm, International Journal of Computer Science and Engineering Communications, Volume.4, Issue.2 (2016): Page.1397-1401

[3]Raunak Jahan, Applying Naive Bayes Classification Technique for Classification of Improved Agricultural Land soils, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1S5, December, 2019

[4]Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, Analysis of agriculture data using data mining techniques: application of big data