

A Study to Maximize the Agricultural Production and Minimize the use of Fertilizers

Kavya Gupta¹, Mrs Nidhi Senger², Mrs. Vasudha Bahl², Dr. Amita Goel²

¹Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, GGSIP University, Delhi, India

²Faculty, Department of Information Technology, Maharaja Agrasen Institute of Technology, GGSIP University, Delhi, India

ABSTRACT

Article Info

Volume 7, Issue 6

Page Number: 225-231

Publication Issue :

November-December-2021

Article History

Accepted : 02 Dec 2021

Published : 10 Dec 2021

A large number of people have attempted to create mathematical models in order to develop and optimize agricultural production; however almost none of them address the factors such as use of fertilizers (N, K, P). More use of fertilizers can deteriorate the fertility of soil and can lead to the low production for years to come. In this study a predictive model is formulated that aims to predict the best crop to be grown in available weather conditions and nutrients present in the soil thereby reducing the use of chemicals and taking a step forward towards organic and chemical free agriculture.

Keywords : Node Deployment, Routing, Artificial Bee Colony, Wireless Sensor Networks.

I. INTRODUCTION

Agriculture plays a vital role in the global economy. India is home to more than 130 crores people and to cater to their food demand is a major challenge. According to FAO estimates in 'The State of Food Security and Nutrition in the World, 2020 report, 189.2 million people are undernourished in India. By this measure 14% of the population is undernourished in India. With the ever increasing human population, understanding worldwide crop yield is central to addressing food security challenges and reducing the use of fertilizers. The demand for food crops have been increasing but is very difficult to increase the area of land under agriculture with so much pressure

on land.[3] So there should be a methodology that increases the crop production on the same amount of land that is available with us. Crop yield prediction is an important agricultural problem. The Agricultural yield primarily depends on weather conditions (rain, temperature), pesticides and water availability. Accurate information about history of crop yield is important to get the maximum output of crop yield on the same dimensions of available land and to reduce the use of fertilizers so we can move towards organic farming. This problem not only persist in India but Worldwide as estimates show that, in 2020, over 237 crore people were grappling with food insecurity globally, an increase of about 32 crores from 2019. With such a negative report there should

be some step taken to overcome the above problems. For this purpose, data was studied that There are three major nutrition requirements for any crop to grow that is nitrogen, phosphorus and potassium also known as NPK. we extended the methodology of [1] and [2] so that a predictive model can be created that majorly work on 8 factors namely Nitrogen ratio in soil, phosphorous ratio in soil, potassium ratio in soil, pH of the soil, rainfall, temperature and humidity. With the study of all this factors a predictive model is created that can predict which crop can be grown best with available weather and nutritional value of soil. This model is based on the pre-existing factors that urges farmers to reduce their dependency on external factors such as using of chemicals and rainfall and being more dependent on what they already have. After the testing of soil in lab, NPK ratio can be found out easily and this model can be applied.

II. METHODOLOGY

This paper develops a predictive model that effectively addresses the above problem. This model not only considers the weather and climatic condition but also the NPK ratio and pH of the soil

DATA SET

Data set has been created by going through the data for agriculture through various website

- <https://www.fao.org/3/a0257e/A0257E05.htm>
- <https://agricoop.nic.in/sites/default/files/Horticulture%20Statistics%20at%20a%20Glance-2018.pdf>

Dataset contains 2200 rows containing weather requirement and NPK requirement for different crops and different fields with 8 columns namely N- Ratio of nitrogen content in the soil, P- Ratio of phosphorus content in the soil, K- Ration of potassium content in the soil, Temperature- Temperature required for the crop to grow in degree Celsius, Humidity- Relative

humidity in percentage, pH- pH value of the soil, Rainfall- Rainfall in mm and Label- Name of different crops.

This dataset consists the data for various fields from all over India. in all total 22 crops were taken and for each crop 100 observation were taken as follows Kidneybeans, Cotton, Banana, Orange, Coconut, Mothbeans, Chickpea, Mango, Mungbean, Lentil, Pigeonpeas, Apple, Rice, Muskmelon, Pomegranate, Coffee, Maize, Blackgram, Jute, Papaya, Grapes, Watermelon. There was equal distribution of crops

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

Given above is the snippet of dataset that shows the first 5 rows .it conatins the data for rice that is grown in different parts of india and gives the information regarding N, P, K, temprature, humidity, ph and rainfall.

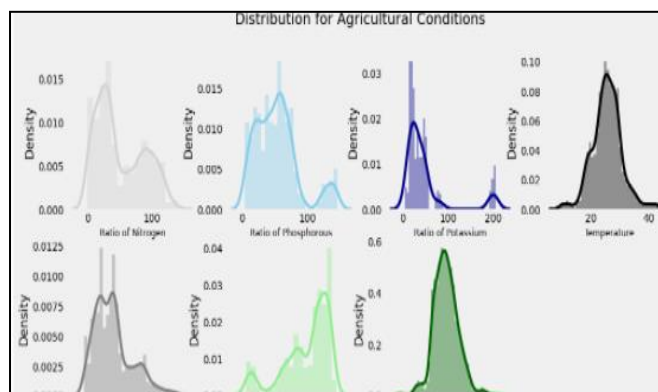
DESCRIPTIVE STATISTICS

The average ratio of nitrogen, phosphorous and also the average temperature, humidity, pH value and rainfall for each crop was checked. After the summary statistics for individual crop was found, an average function was used and minimum, maximum and average for each factor for each crop was found. It was inferred that rice requires good rainfall and banana requires good proportion of potassium and so on for all other crops. After this average requirement of each crop was compared with average conditions as well using interactive functions. Some crops require less nitrogen such as coconut, apple and some require more nitrogen such as cotton, muskmelon etc and such comparisons were taken for each factor.After

this crops which require more than or less than the average given condition to grow were found.

ANALYZING AGRICULTURE CONDITION

Distribution of different agricultural conditions were found using distplot function of seaborn library of python. Below is the screenshot of the graph obtained



After studying the above graphs, potassium and humidity factor were found to be different. Potassium graph was right skewed and humidity was left skewed and all other conditions are normally distributed. Next step was to look for particular crops were found that requires high and low amount of NPK and other factors and after that Crops were classify on the basis of season summers, winters and rainy on the basis of temperature and humidity filters.

CLUSTERING

K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of

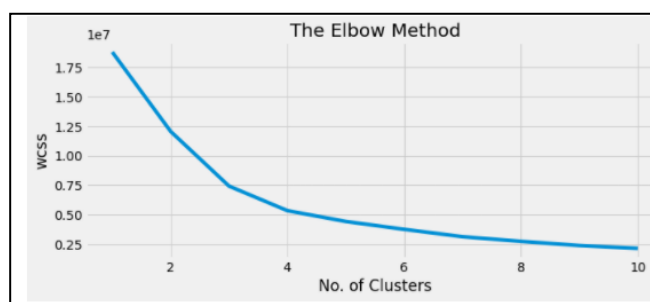
income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points

again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster

performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification if unreasonable, we will modify it(calculate each cluster focal points again), iterate repetitively till we get a reasonable classification K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B.MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points(usually, we choose the first K sample dots of

income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification is unreasonable, we will modify it (calculate each cluster focal points again), iterate repetitively till we get a reasonable classification. K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B. MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points (usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification is unreasonable, we will modify it (calculate each cluster focal points again), iterate repetitively till we get a reasonable classification. K-Means algorithm based on dividing [4] [5] is a kind of cluster algorithm, and it is proposed by J.B. MacQueen. This algorithm which is unsupervised is usually used in data mining and pattern recognition. Aiming at minimizing cluster performance index, square-error and error criterion are foundations of this algorithm. To seek the optimalizing outcome, this algorithm tries to find K divisions to satisfy a certain criterion. Firstly, choose some dots to represent the initial cluster focal points (usually, we choose the first K sample dots of income to represent the initial cluster focal point); secondly, gather the remaining sample dots to their focal points in accordance with the criterion of minimum distance, then we will get the initial classification, and if the classification is unreasonable, we will modify it (calculate each cluster focal points

again), iterate repetitively till we get a reasonable classification. In clustering target number k is defined, which refers to the number of centroids needed in the dataset. A centroid is the imaginary or real location representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is a process of grouping data points with similar traits and assigning them to clusters. K-means clustering is used to determine the optimal number of clusters. The elbow method was used. Clusters of crops were made based on the factors from the data set so that they can be grouped and a predictive model can be created.

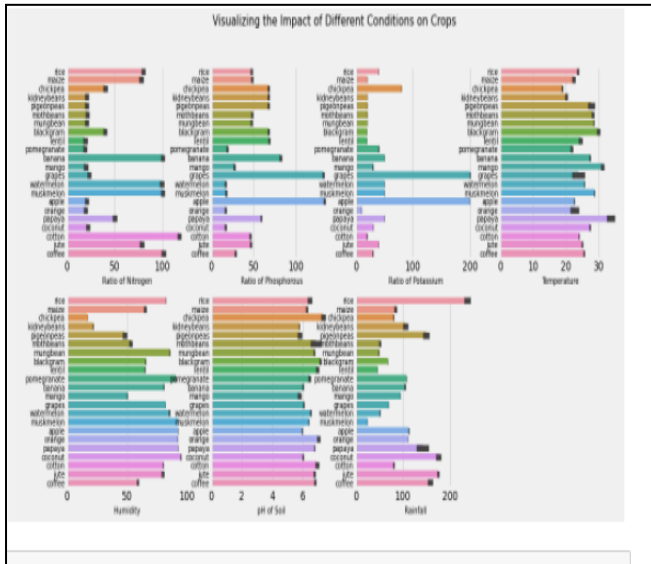


The average is between 3 and 4 so 4 clusters were considered. KMeans algorithm was implemented and 4 clusters were as follows.

But some of the crops were in more than one clustering so hard clustering was done instead of soft clustering so that a crop can be contained in one cluster only

```
Crops in Cluster 1: ['kidneybeans', 'mothbeans', 'blackgram', 'pomegranate', 'lentil', 'orange', 'mungbean', 'mango', 'chickpea']
-----
Crops in Cluster 2: ['cotton', 'maize', 'banana', 'watermelon', 'muskmelon']
-----
Crops in Cluster 3: ['grapes', 'apple']
-----
Crops in Cluster 4: ['rice', 'coconut', 'papaya', 'jute', 'pigeonpeas', 'coffee']
```

SOME PATTERNS



The above graph was created to find out some patterns like cotton require high quantity of nitrogen and grapes and apple requires high amount of phosphorus and other factors for other crops

PREDICTIVE MODEL

Dataset was split into independent and dependent dataset. Dependent variable was label column(crop) Later it was split into training and testing data. Finally, model was created using logistic regression. Logistic regression was imported from sklearn.linear model and the object was created and was training data set was fil into the model Test dataset was predicted with the help of predict function

```
# create a Predictive Model
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
```

III. RESULT

After creating the model real time data was used in which a person has to enter N RATIO,P RATIO, K RATIO, RAINFALL, HUMIDITY, PH ,

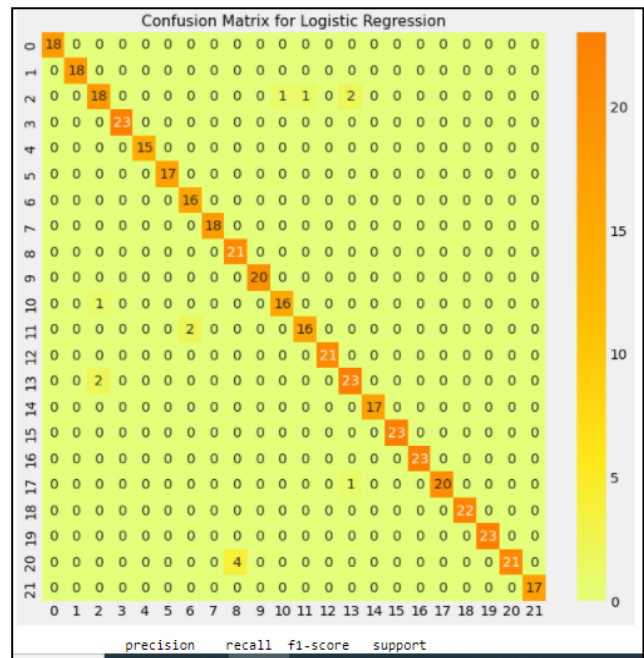
TEMPRATURE so the model can predict which crop can be best grown in the available conditions

Below case predicts that rice is the most suitable crop to be grown with given conditions

```
prediction =model.predict((np.array([[90,
                                     40,
                                     20,
                                     15,
                                     80,
                                     7,
                                     200]])))
print("The Suggested Crop :", prediction)
```

The Suggested Crop : ['rice']

Heat map and confusion matrix was created and Good results for precision recall and f1-score was found and the accuracy was about 97% for the training data set



omegranate	1.00	1.00	1.00	23
rice	1.00	0.84	0.91	25
watermelon	1.00	1.00	1.00	17
accuracy			0.97	440
macro avg	0.97	0.97	0.97	440
ighted avg	0.97	0.97	0.97	440

IV. CONCLUSION

This help the farmer to grow the right crop with given condition and as that NPK requirement will be fulfilled this will considerably reduce the use of chemical fertilizers and improve the quality of food and perfect climatic condition will reduce the chances of crop failure and hence improving the food security for the country. This model aims to create fertilizer free farming and reducing the chances of crop crop failure and thus optimizing agricultural production for India.

V. REFERENCES

- [1]. Basil Manos, Parthena Chatzinikolaou, Fedra Kiomourtzi, Sustainable Optimization of Agricultural Production, APCBEE Procedia, Volume 5, 2013, Pages 410-415, ISSN 2212-6708, published
- [2]. Agricultural Production Structure Optimization: A Case Study of Major Grain Producing Areas, ChinaLU Sha-sha1, 2, LIU Yan-sui1, LONG Hua-lou1 and GUAN Xing-liang1, published
- [3]. S. Saranya and T. Amudha, "Crop planning optimization research — A detailed investigation," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), 2016, pp. 202-208, doi: 10.1109/ICACA.2016.788795
- [4]. F M Zenis , S Supian , E Lesmana IOP Conf. Series: Materials Science and Engineering 332 (2018) 012021 doi:10.1088/1757-899X/332/1/012021
- [5]. J. Brenes and G. Raventos, "Research on Optimization Techniques Applied to Precision Agricultural Processes Supported by Technology in Latin America: Opportunities and a Route to Follow," in 2020 XLVI Latin American Computing Conference (CLEI), Loja, Ecuador, 2020 pp. 462-470. doi: 10.1109/CLEI52000.2020.00060

- <https://doi.ieeecomputersociety.org/10.1109/CL EI52000.2020.00060>
- [6]. Adams, Darius M.; Alig, Ralph J.; Callaway, J.M.; McCarl, Bruce A.; Winnett, Steven M. 1996. The forest and agricultural sector optimization model (FASOM): model structure and policy applications. Res. Pap. PNW-RP-495. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 60 p
 - [7]. Uguru, H., and Akpenyi-Aboh, O. N. (2021). Optimization of agricultural machines through the pre-harvest treatment of sweet paper (cv. Goliath) fruits.
 - [8]. Direct Research Journal of Agriculture and Food Science. Vol. 9, Pp. 167-173

Cite this article as :

Kavya Gupta, Mrs Nidhi Senger, Mrs. Vasudha Bahl, Dr. Amita Goel, "A Study to Maximize the Agricultural Production and Minimize the use of Fertilizers", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 6, pp. 225-231, November-December 2021. Available at doi : <https://doi.org/10.32628/CSEIT217644>
Journal URL : <https://ijsrcseit.com/CSEIT217644>