|| Volume 1 ||Issue 1 ||April 2016||ISSN (Online) 2456-0774

INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

Crop and Yield Prediction Model

Shreya S. Bhanose [1], Kalyani A. Bogawar [2], Aarti G. Dhotre [3], Bhagyashree R. Gaidhani [4]

Student, *Computer Science & Engineering, K. K. Wagh Institute of Engineering and Research, Nashik, India*[1 2 3 4]

***Abstract*—An agricultural sector necessitate for well defined and systematic approach for predicting the crops with its yield and supporting farmers to take correct decisions to enhance quality of farming. The complexity of predicting the best crops is high duet unavailability of crop knowledge-base. Crop prediction is an efficient approach for better quality farming and increase revenue. Use of data clustering algorithm is an efficient approach in field of data mining to extract useful information and give prediction. Various approaches have been implemented so far are worked either for crop prediction.Crop prediction model aiding farmers to take correct decision. This indeed helps in improving quality of farming and generate better revenue for farmers. Traditional clustering algorithms such as k-Means, improved rough k-Means and-means++ makes the tasks complicated due to random selection of initial cluster center and decision of number of clusters. Modified K-Means algorithm is thereby used to improve the accuracy of a system as it achieves the high quality clusters duet initial cluster centric selection.**

Keywords:-crops, quality farming, prediction, k-means, disease, yield, temperature affect, water requirement, evapo-transpiration, plant.

I INTRODUCTION

A crop prediction is a huge problem that occurs. A farmer had an attention in understanding how much produce he is going to expect. Traditionally farmers decide this based on permanent experience for specific yield, plants and weather conditions. Character directly thinks about produce prediction rather than concerning on crop prediction.

If the correct crop is expected then yield will be better. Problem of crop and yield prediction using modified k-means clustering algorithm thereby creating better earnings for berry farmers. Clustering is the process of grouping the data into classes or groupings, so that objects within a cluster have high similarity in agreement to each other but are incredibly dissimilar to objects in option clusters.

A bunch of data objects can be treated collectively during the time that you group and so may be looked at as a classic of data compression. Unlike category, clustering is a powerful means for partitioning the collection of data into organizations based on data likeness and then ascribe labeling to the relatively small number of groups. Clustering is an unsupervised learning as it does not rely on predefined classes and class labeled training examples. Because of this, clustering is a form of learning by observation, rather than learning by examples. Because shown in Figure. 1, the three clusters are

Formed containing data factors depending on center position. The cluster center is shown by + signs. The quality of clusters will depend on how dense it is. So, cluster having more number of points is cluster of good quality.

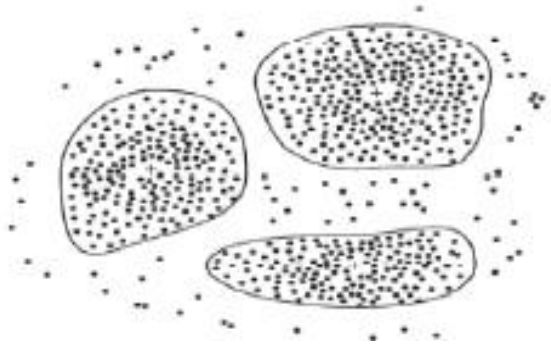


Figure 1 Cluster Analysis

This paper proposes Bee Hive algorithm for predicting crop yield from historical data set. This algorithm handles large data set but it has drawback of having number of tunable parameters and k value

II LITERATURE SURVEY

CRY An improved Crop Yield Prediction model using Bee Hive Clustering Approach for Agricultural data sets (2013)

This paper proposes Bee Hive algorithm for predicting crop yield from historical data set. This algorithm handles large data set but it has drawback of having number of tunable parameters and k value.

An improved Rough K-means algorithm with weighted distance measure (2012):

This paper proposes a solution to search initial central points and combine it with a distance measure with weight. It requires additional parameter such as density threshold and number of cluster.

k-means++: The advantages of careful seeding(2013):

This algorithm suggests K-Means++ clustering algorithm by using randomized seeding technique. It has drawback of number of cluster value and decision of initial center.

An EM clustering algorithm which produces a dual representation (2012):

This paper suggests an EM algorithm which handles real world data set but it randomly selects k-value and becomes sensitive to noise and also highly complex in nature.

III MOTIVATION

A crop prediction is a widespread problem that occurs. During the rising season, a farmer had curiosity in knowing how much yield he is about to expect. In the earlier period, this yield prediction become a matter of fact relied on

Farmer's long-term experience for specific yield, crops and climatic conditions. Farmer directly goes for yield prediction rather than concerning on crop prediction with the existing system. Unless the correct crop is predicted how the yield will be better and additionally with existing systems pesticides, environmental and meteorological parameter related to crop is not considered. Promoting and soothing the agricultural production at a more rapidly pace is one of the essential situation for agricultural improvement. Any crop's production show the way either by interest of domain or enhancement in yield or both. In India, the prospect of widening the district under any crop does not exist except by re-establishing to increase cropping strength or crop replacement. So, variations in crop productivity continue to trouble the area and generate rigorous distress. So, there is need to attempt good technique for crop prediction in order to overcome existing problem.

IV METHODOLOGIES OF PROBLEM SOLVING AND EFFICIENCY ISSUES

A. *Traditional approaches:*
Experience based farming and agriculture

B. *With sentiment analysis:*
[1]CRY: Crop and Yield prediction using BeeHive Algorithm

C. *Efficiency issue:*
CRY: It predicts the yield based on historical data. It randomly selects centroid. It requires additional parameters such as density and threshold

V ARCHITECTURE OF CROP PREDICTION

Architecture is a system that unifies its components or elements into a coherent and functional whole. The architecture of crop prediction is shown in Figure 2 and the block description is as follows.

Crop knowledge base: The crop knowledge base [4] consists of farm knowledge such as crop types, soil types, soil-ph value, crop disease and pesticides, seasonal parameter such as kharif, rabbi and summer crops. The knowledge-base also consists of zones as well as district information, environmental parameter such as maximum and minimum temperature value and average rainfall

*Clustering Approaches:**[5], [7]*** The three clustering approaches is used such as k-Means, k- Means++ and traditional k-Means. The determined value of number of clusters and initial cluster centers is provided to modified k-Means clustering algorithm. Because of the number of clusters (k value) is required at starting for traditional k-Means and k- Means ++, the same calculated value of number of clusters is provided and initial cluster centers are uniformly chosen.[1], [2]All three approaches performed clustering and provide output in the form cluster number and centroid matrix.

Sample Testing and Prediction: There is need to provide input parameters such as zone, district, and selection of seasons, soil type, maximum temperature, minimum temperature and average rainfall for sample testing. Based on the output values of each clustering, the test data calculates the distance measure with clustering output and selects minimum distance as a predicted value. Then, the predicted cluster value is founded in output cluster number (idx) and as per the priority the very first output value of predicted cluster is selected. Then, the similar number of records of output value is founded in expected value and accuracy in terms of its count value is calculated. The accuracy count is shown by pie chart.

VI SYSTEM ARCHITECTURE

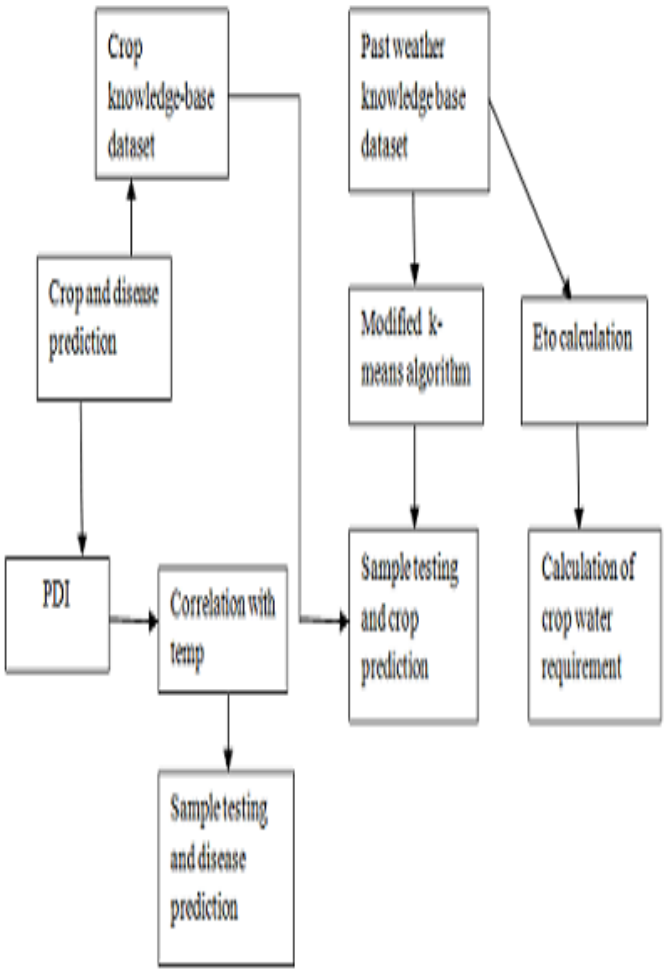


Figure 2 Architecture Design

VII EARLY BLIGHT DISEASE PREDICTION OF TOMATO CROP

Importance:

Early blight is one of the most common tomato diseases, occurring nearly every season wherever tomatoes are grown. It affects leaves, fruits, and stems and can be severely yield limiting when susceptible cultivars are used and weather is favorable. Severe defoliation can occur and result in sunscald on the fruit.

Identification:

** Signs and Symptoms:*

1. LEAVES:

Initially small dark spots form on older foliage near the ground.
Leaf spots are round, brown and can grow up to half inch in diameter.
Larger spots have target like concentric rings and tissue around spots often turns yellow.

Severely infected leaves turn brown and fall off, or dead, dried leaves may cling to the stem.

2. STEM:

Seedling stems are infected at or just above the soil line. The stem turns brown, sunken and dry (collar rot).If the infection girdles the stem, the seedling wilts and dies.

Stem infections on older plants are oval to irregular, dry brown areas with dark brown concentric rings.

3. FRUIT:

Fruit can be infected at any stage of maturity.

Fruit spots are leathery, black, with raised concentric ridges and generally occur near the stem.

Infected fruit may drop from the plant.

BRIEF MODULE DESCRIPTION :

Disease prediction module describes about prediction of percent of tomato crop has been affected by the early blight disease and thereby helping the farmer to immediately provide the pesticides to cure crop to generate better yield and thereby helps in generating better revenue for farmer. It also describes the correlation of weather parameter like maximum temperature affecting in disease development.

METHODS:

1. CALCULATION OF PDI:

PDI is percent disease index used to measure disease intensity.

PDI = sum of all grades of leaves/Total number of leaves observed *100/9

Where,

9-maximum grade of leaf;

Index scale	Percent symptoms
0	No symptoms
1	1% or less leaf area
3	1-10% of leaf area
5	11-25% of leaf area
7	26-50% of leaf area
9	51% or more area

Figure 3 Scale Used to Measure Disease Intensity

Figure 4 symptoms of early blight disease in different grades

2. CALCULATION OF CORRELATION:

Correlation is used for predictive relationship between entities. Disease prediction module uses correlation factor to find the

relevance between maximum temperature with PDI, affecting in disease development.

CORR= [NΣXY - (ΣX)(ΣY) / Sqrt([NΣX2 - (ΣX)2][NΣY2 - (ΣY)2])]

If correlation of maximum temperature with PDI is positive, then tomato crop is affected by the disease and if correlation is negative, then, crop is not affected by the disease.

IMPLEMENTATION REPORT:

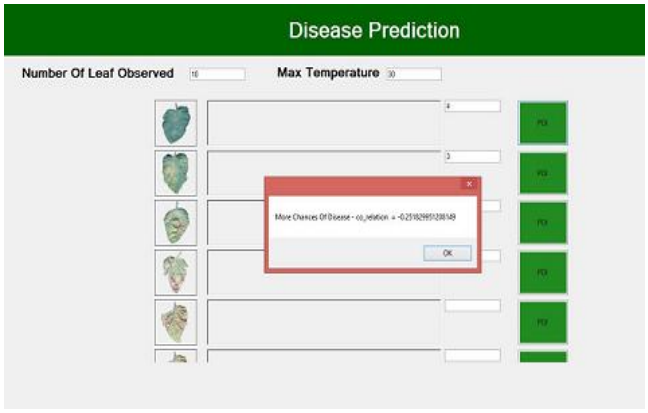


Figure 5 Disease Prediction

VIII WATER REQUIREMENT OF TOMATO CROP

Water requirement is the most important factor for the healthy growth of crops. The amount of water potentially required to meet the evapo-transpiration needs of plant so that plant does not suffers in its growth through short supply of water.

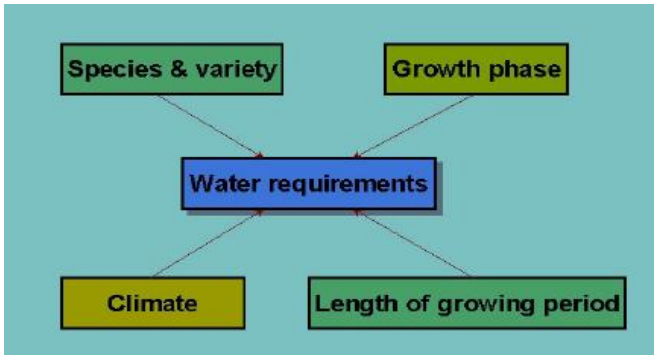


Figure 6 Water Requirements of Crops

The evapotranspiration rate is the amount of water that is lost to the atmosphere through the leaves of the plant, as well as the soil surface. Therefore, in order to estimate the water requirement of a crop we first need to measure the evapotranspiration rate. The evapotranspiration rate, ET0, is the estimate of the amount of water that is used by a well-watered grass surface that is roughly 8 to 15 centimeters in height. Once ET0 is known, the water requirement of the crop can be calculated.

IX MODIFIEDK-MEANS CLUSTERING

The modified k-means algorithm is most well known data clustering approach based on improvement of sensitivity of initial centers(seed point) of clusters. This algorithm partitions the whole space into different segments and calculates the frequency of data point in each segment. The segment which shows maximum frequency of data point will have the maximum probability to contain the centroid of cluster. The steps are:

1. Input:-data set and value of k.
2. If the value of k is 1 then Exit.
3. Else
4. /*divide the data point space into k*k, means k vertically and k horizontally*/
5. For each dimension
{
6. Calculate the minimum and maximum value of data points
7. Calculate range of group(RG) using equation ((min+max)/k)
8. Divide the data point space in k group with width RG
9. }
10. Calculate the frequency of data points in each partitioned space.
11. Choose the k highest frequency group.
12. Calculate the mean of selected group. /* This will be the initial centroid of clussster.*/
13. Calculate the distance between each clusters using equation (3)
14. Take the minimum distance for each cluster and make it half using equation (4)
15. For each data points p= 1 to N0
{
16. For each cluster j= 1 to k
{
17. Calculate d(ZP,MJ) using equation (1)
18. If (d(ZP,MJ) dcJ)

{
19. Then ZP assign to cluster CJ
20. Break
21
{
22. Else
23. Continue;
}
24. If ZP does not belong to any cluster then
25. ZP min(d(ZP , Mi)) where i [1, Nc]
26. }
27. Exit.
28. else
29. Calculate the centroid of cluster using equation (2) of k-means algorithm.

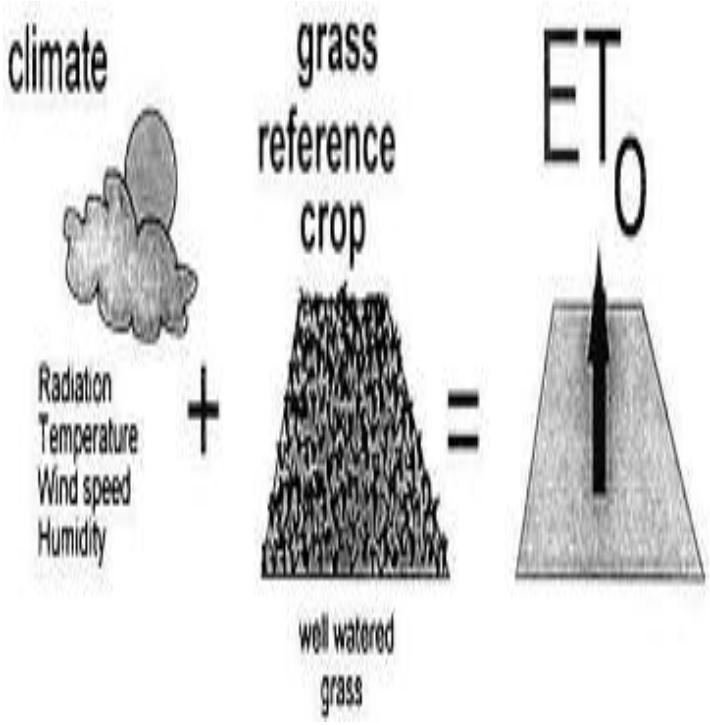
30. Go to step 13.

PENMAN-MONTEITH EQUATION:

The reference rate, ET0, is calculated using the Penman Equation, which takes into account the climatic parameters of temperature, solar radiation, wind speed and humidity.
A variation of this equation, published by the FAO is:

$$ET_o = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

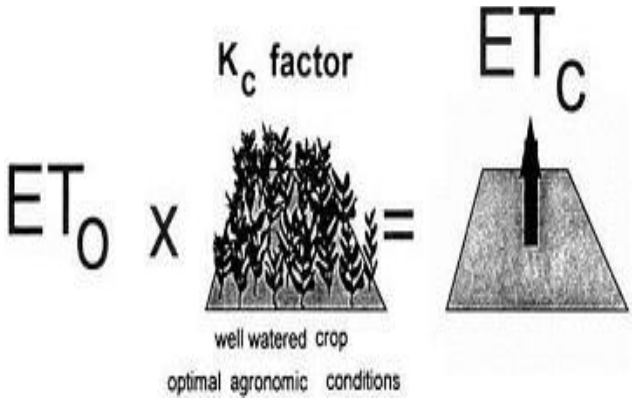
Where,
ETo reference evapotranspiration [mm day-1],
Rn net radiation at the crop surface [MJ m-2 day-1],
G soil heat flux density [MJ m-2 day-1],
T air temperature at 2 m height [°C],
u2 wind speed at 2 m height [m s-1],
es saturation vapour pressure [kPa],
ea actual vapour pressure [kPa],
es - ea saturation vapour pressure deficit [kPa],
D slope vapour pressure curve [kPa °C-1],
g psychometric constant [kPa °C-1].



ESTIMATION OF CROP WATER REQUIREMENT :

ET0 represents the maximum, or potential, evapotranspiration rate that can occur. However, the water requirement of the crop is usually less than ET0, as there are factors of the crop itself that have to be taken into account. These include the growth stage of the plant, the leaf coverage that provides shade to the ground, and other particulars of the crops that make them vary from each other. With these factors taken into account, ET0 is converted into ETc, through the crop-specific coefficient, Kc.ETc represents the evapotranspiration rate of the crop under standard conditions (no stress conditions).When calculating ETc, one must identify the growth stages of the crop, their duration and select the proper Kc coefficient that need to be used.

ETc = Kc*ET0.



Climatic effects are incorporated into ET0, while the effects of the crop characteristics are incorporated into Kc.

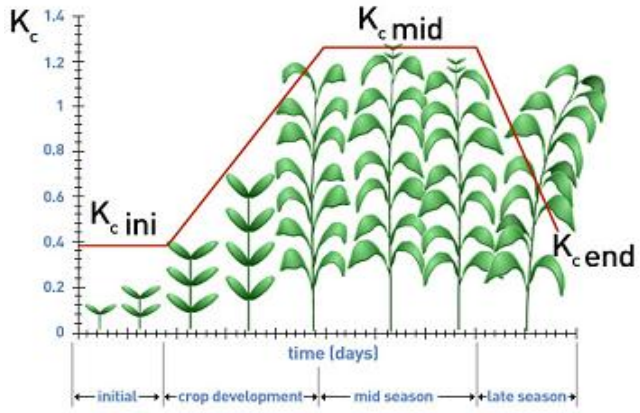


Figure 7 crop stages of tomato and crop coefficients used for water management

Example for calculating water requirements of crop:
crop : tomato
growth stage : initial growth
kc for initial stage : 0.45
et0 (measured by a local meteorological station):
9 mm/day
etc = kc*et0 = 0.45 x 9 = 4.05 mm/day

X CONCLUSION AND FUTURE WORK

Clustering is a data exploration algorithm and plays significant role for extracting knowledge boost of information. Clustering technique applied in harvest dataset has resulted in novel approach that has value

success in predicting harvest. However, the key problem with existing clustering algorithms like random initialization of bunch centers and uniform supply of number of groupings as an input are stated. The drawbacks are overcome by proposing altered k-Means clustering algorithm which used the formulated value to initialize cluster centers also to determine number of clusters. This work shows about modified k- Means clustering in crop conjecture by increasing quality and accuracy count. The altered k-Means clustering algorithm is evaluated by comparing k-Means and k-Means++ algorithms and achieved the most number of high quality clusters, right prediction of crop and maximum accuracy count. Info mining plays an important role in Agriculture sector for better prediction of harvest. The proposed work is done on crop dataset belong to Maharashtra Condition. Our future work includes to consider geographical area using world geographic information system for global harvest prediction.

ACKNOWLEDGMENT

The authors feel a deep sense of gratitude to Prof. Hadole Asst. Professor of K.K.Wagh Agriculture College for his motivation and support during this work. The authors are also thankful to the Prof N.G. Sharma, Asst. Professor of K.K.Wagh College of Engineering and Research Education, Nashik for being a constant source of inspiration.

REFERENCES

[1] M. Ananthara, T. Arunkumar, and R. Hemavathy, Cry: An improved crop yield prediction model using bee hive clustering approach for agricultural data sets," in Pattern Recognition, Informatics and Medical Engineering (PRIME), 2013 IEEE International Conference , pp. 473-478.

[2] U.P. Narkhede and K.P.Adhiya, A Study of Clustering Techniques for Crop Prediction - A Survey, American International Journal of Research in Science, Technology, Engineering Mathematics, vol 1, Issue 5, ISSN no: 2328-3491, pp. 45-48 ,2014.

[3] Department of agriculture, Maharashtra state, Accessed on12-Feb-2014.[On-line].Available: http://www.mahaagri.gov.in/.

[4] D. M. Kiri L. Wagstaff and and S. R. Sain, Harvist: A system for agricultural and weather studies using advanced statistical methods," 2005. [Online]. Available: https://www.agriskmanagementforum.org.

[5] R. A.A. and K. R.V., Review - role of data mining in agriculture," International Journal of Computer Science and Information Technologies(IJCSIT), 2013,vol. 4(2),no. 0975- 9646, pp. 270-272.

[6] M. Kannan, S.Prabhakaran, and P.Ramachandran, Rainfall forecasting using data mining technique," International Journal of Engineering and Technology, vol. 2, no.0975- 4024, pp. 397-401, 2010.

[7] S. Kim and Wilbur, " An EM clustering algorithm which produces a dual representation," Proceedings of 10th IEEE International Conference on Machine Learning and Applications and Workshops (ICMLA), vol. 2, pp. 90-95,2011.