iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4071

www.ijraset.com

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

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Agriculture Information Extraction Using Data Analytics in Weka

D. MagheshKumar[1], P. Nivedha[2]
[1,2] Department of Software Engineering, Periyar Maniammai University

Abstract*: With the development of PC based information stockpiling frameworks we have gone over a gigantic measure of storehouse of information. In any case, this information is not extremely accommodating until we realize what we can do with it. We have to make inductions from this huge information with the goal that we can settle on choices driven by learning. Data mining is the procedure of learning revelation in database. From most recent couple of decade's data mining in farming is late research zone. Till now information mining systems were utilized as a part of the organizations and corporate segments, however now these strategies are likewise being utilized for extraction of strong rural information. With the assistance of KDD and data mining we extricate the important informational indexes from the tremendous measure of information. The k-implies bunching is utilized to group the given arrangement of information. This method when connected on the huge arrangement of information then it comes about into enhanced nature of mined information. We have connected this technique to concentrate the creation and utilization of products in different parts of India. The different variables which influence the creation of yields like soil sort and climate are mulled over. For graphically portrayal we have utilized spatial join with the calculation.

Keywords*: KDD (Knowledge Discovery in Database Process), WEKA (Waikato Environment for Knowledge Analysis), DM (Data Mining)

I. INTRODUCTION

Data mining is strategy used to get the concealed data from tremendous databases. It is an effective system which concentrates on the most critical data in their information distribution centres. Data mining devices envision the future patterns and practices for permitting organizations to settle on sensible information driven choices. The real information mining systems are: order, relapse and bunching. Data Mining, likewise prominently known as Knowledge Discovery in Databases (KDD), alludes to as "the nontrivial procedure of distinguishing appropriate, unique, helpful and in the long run sensible type of information". The information is in immense sum which is put away in documents, databases and different vaults. That information is dynamically more essential, in the event that it is not vital to develop controlling means for examination and clarification of information and for the extraction of intriguing learning that could help in basic leadership. The accompanying demonstrates information mining as a stage in an iterative learning revelation prepare.

II. KNOWLEDGE DISCOVERY IN DATA MINING

KDD is a procedure which gets the helpful data from gigantic gathering of information which understand information by utilizing proper systems and techniques. This procedure manage the low-level information into different structures is extremely minimal, conceptual and useful. By utilizing the KDD procedure we can make short reports, displaying the procedure of information and foresee the models that utilization in future.

A. Data is taken from assortment of sources is coordinated into a solitary information store called target information.
B. Then pre-prepared the information and changed into standard arrangement.
C. The information mining calculations used to prepare the information to the yield in type of examples or standards
D. Then those examples and guidelines are changed to new or valuable learning or data.

III. INFORMATION COLLECTION METHODS

There are distinctive systems utilized as a part of the information mining to get the important data. We can utilize information mining to naturally decide huge examples and concealed relationship from enormous measure of information. Information mining gives you experiences and connections that had formally gone unrecognized or been overlooked in light of the fact that it had not been viewed as conceivable to break down them. The information mining process comprises of the accompanying strides:

A. Data Cleaning

In this stage puncturing information and insignificant information are expelled from the gathering.

B. Data Mix

At this stage different information sources like heterogeneous information might be consolidated in a typical source.

C. Data Choice

In this progression, the important information is dissected information chose and recovered from the information accumulation.

D. Data Change

This period of information mining strategy select the information that is changed over into suitable structures for the mining procedure.

E. Data Mining

This is the critical stride in which shrewd methods are connected to concentrate designs conceivably important.

F. Pattern Assessment

This progression is absolutely amazing examples speaking to information are recognized in view of given measures.

G. Knowledge Portrayal

This is the last stage in which accomplish learning is outwardly spoken to the client. This basic stride utilizes disclosure systems to help clients get it.

IV. DATA MINING ALGORITHM

To make the information mining model, an information mining calculation first examine the information then makes the specific sorts of examples .the information mining model utilize this outcome to dissect and characterize the ideal parameters. These parameters are utilized to separate the significance full data. The mining model that a calculation makes from your information can take different structures.

A set of groups that show how the assets in a dataset are connected.

A choice tree that figure the outcome, and clarify how divergent criteria influence that outcome.

A numerical model that anticipate the outcome.

A set of directions that portray how merchandise are gathered together in an exchange, and the probabilities that items are bought together.

Grouping calculations conjecture at least one particular factors, in light of alternate traits in the dataset.

Relapse calculations gauge at least one steady factors, for example, benefit or misfortune, in light of different characteristics in the dataset.

Division calculations parcel information into gatherings, or bunches, of things that have related properties.

Affiliation calculations find connections between unique qualities in the dataset. The fundamental use of calculation is making affiliation manages and utilized as a part of a market wicker bin investigation. Arrangement investigation calculations audit the various successions in information, for example, a Web way stream.

A. K-Means Algorithm

The *k*-Means algorithm is distance-based clustering algorithm which separate the data into a prearranged number of clusters (provided there are enough distinct cases). Distance-based algorithms rely on a distance metric (function) to measure the similarity between data points. The distance metric is measured as either Euclidean or Fast Cosine distance. Data sets are assigned to the nearest cluster according to the distance metric used. The algorithm randomly selects k points as the initial cluster for centres.

1) Each point in the dataset is assigned to the nearest cluster which is based upon the Euclidean distance between the each point of data and each cluster centre.

2) Then cluster centre is recomputed by the find the average of the points in that cluster.

3) Repeat the steps 2 and 3 until the clusters converge. Convergence of clusters may be defined in a different way depending upon the implementation.

4) It is normally mean that no observation can change the clusters when steps 2 and 3 are repeated.

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

V. RESULT

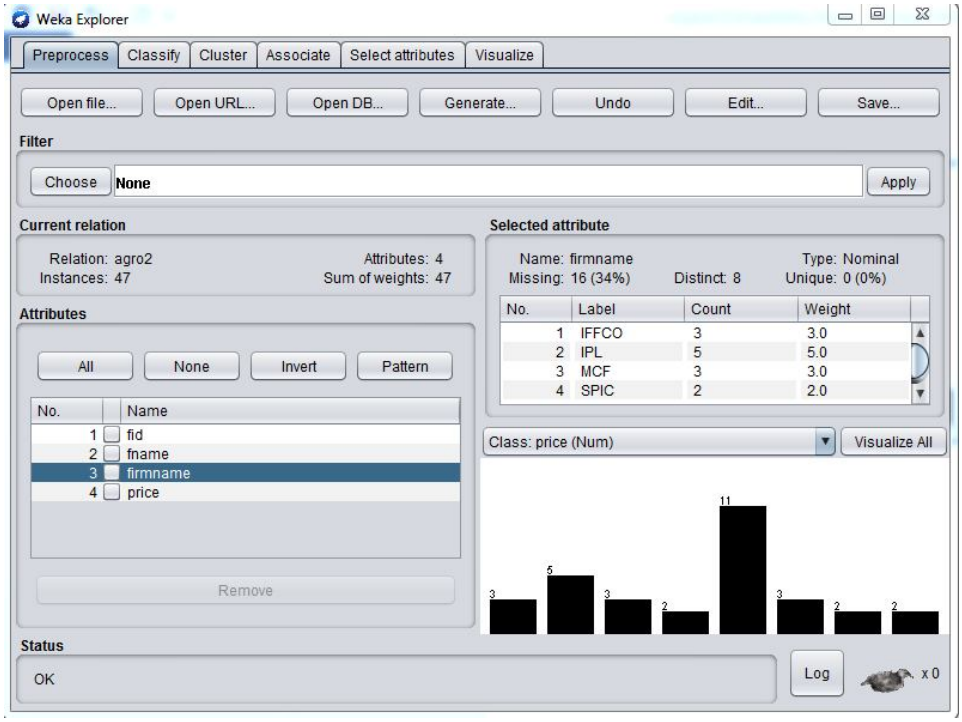


Figure 1.1 shows the pre-process.

A. *Classifier: Trees: J48 -C 0.25 -M 2*

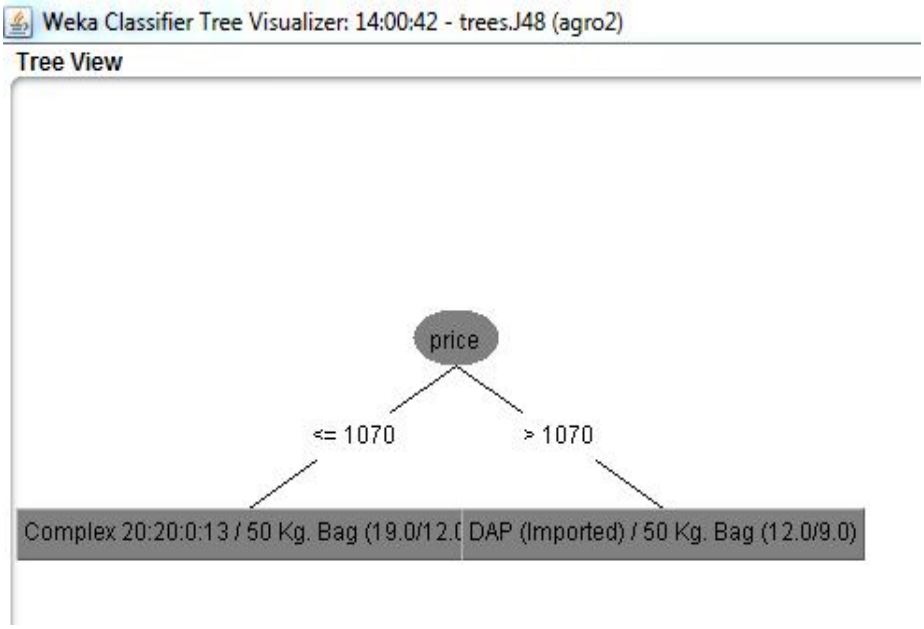


Figure 1.2 shows the classifier

=== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: agro2
Instances: 47
Attributes: 4

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

fid
fname
firmname
price

Test mode: 5-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
---
price <= 1070: Complex 20:20:0:13 / 50 Kg. Bag (19.0/12.0)
price > 1070: DAP (Imported) / 50 Kg. Bag (12.0/9.0)
Number of Leaves: 2
Size of the tree : 3
Time taken to build model: 0.2 seconds
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	7	22.5806 %
Incorrectly Classified Instances	24	77.4194 %
Kappa statistic	0.0133	
Mean absolute error	0.0965	
Root mean squared error	0.2257	
Relative absolute error	97.7761 %	
Root relative squared error	101.3875 %	
Total Number of Instances	31	
Ignored Class Unknown Instances	16	

=== Detailed Accuracy by Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.000	0.000	0.000	0.000	0.000	0.207	0.021	DAP - 18:46:00 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.152	0.021	Complex NPK - 16:16:16 / 50 Kg. Bag
0.000	0.069	0.000	0.000	0.000	-0.069	0.422	0.054	Complex NPK - 10:26:26 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.422	0.054	DAP - 18:46:00 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.152	0.021	Complex NPK - 28:28:0:0 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.207	0.021	Complex NPK - 17:17:17 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.152	0.021	Complex NPK - 12:32:16 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.152	0.021	Complex NPK - 14:35:14 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.152	0.021	SSP - 16.0 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.109	0.021	Urea (Ind) - 46.0 / 50 Kg. Bag
1.000	0.917	0.241	1.000	0.389	0.142	0.614	0.220	Complex 20:20:0:13 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.364	0.057	DAP (Imported) / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.417	0.059	MOP (Imported) / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.087	0.021	Complex 15:15:15:9 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.120	0.021	MOP (Imported) / 50 Kg. Bag DAP (Imported) / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.087	0.021	Complex 20:20:0:13:03 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.120	0.021	MOP / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.120	0.021	Complex 16:20:0:13 / 50 Kg. Bag
0.000	0.000	0.000	0.000	0.000	0.000	0.109	0.021	Complex 24:24:0:8 / 50 Kg. Bag
Weighted Avg.	0.226	0.211	0.055	0.226	0.088	0.028	0.331	0.077

=== Confusion Matrix ===

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

a b c d e f g h i j k l m n o p q r s <-- classified as
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | a = DAP - 18:46:00 / 50 Kg.Bag
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = Complex NPK - 16:16:16 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 | c = Complex NPK - 10:26:26 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 | d = DAP - 18:46:00 / 50 Kg. Bag
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | e = Complex NPK - 28:28:0:0 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | f = Complex NPK - 17:17:17 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | g = Complex NPK - 12:32:16 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | h = Complex NPK - 14:35:14 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | i = SSP - 16.0 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | j = Urea(Ind) - 46.0 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 7 0 0 0 0 0 0 0 0 | k = Complex 20:20:0:13 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 | l = DAP (Imported) / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 | m = MOP (Imported) / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | n = Complex 15:15:15:9 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | o = MOP (Imported) / 50 Kg. BagDAP (Imported) / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | p = Complex 20:20:0:13:03 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | q = MOP / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | r = Complex 16:20:0:13 / 50 Kg. Bag
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 | s = Complex 24:24:0:8 / 50 Kg. Bag

B. Clustering

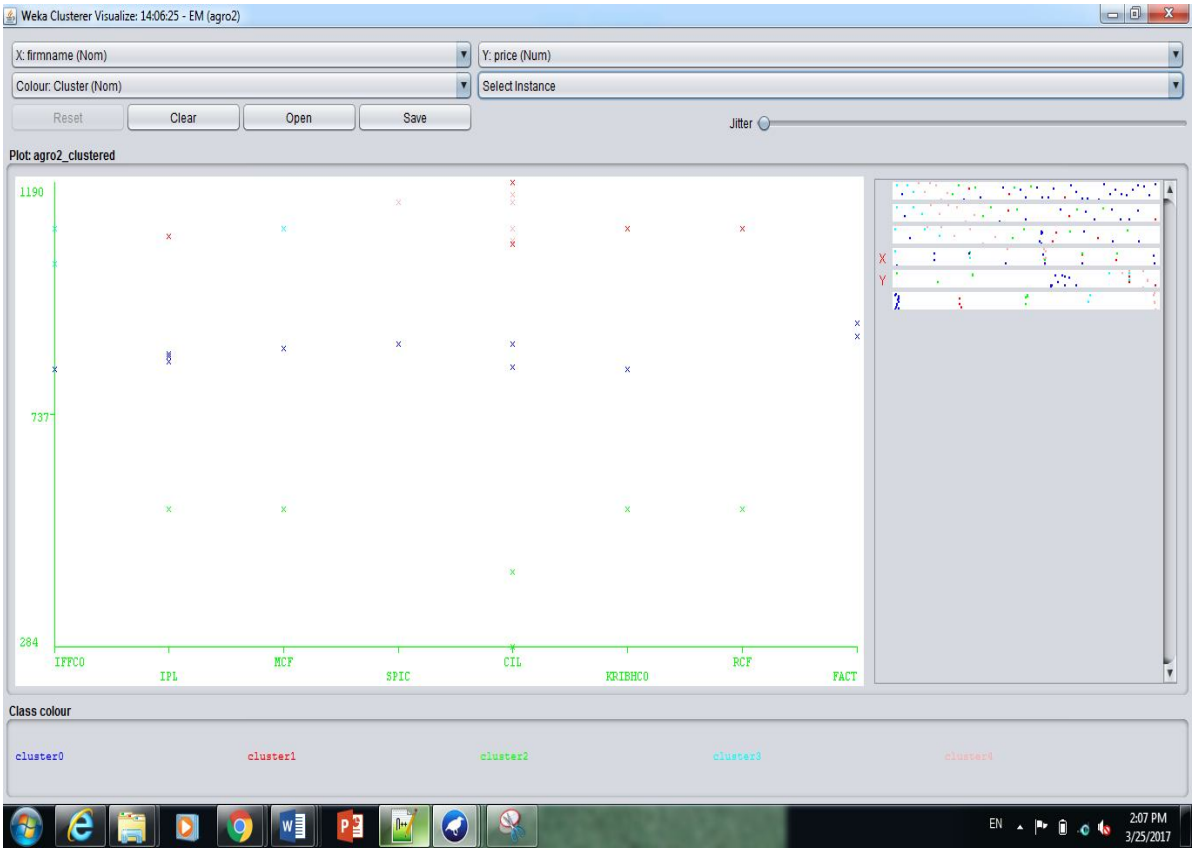


Figure 1.3 shows the clustering results

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

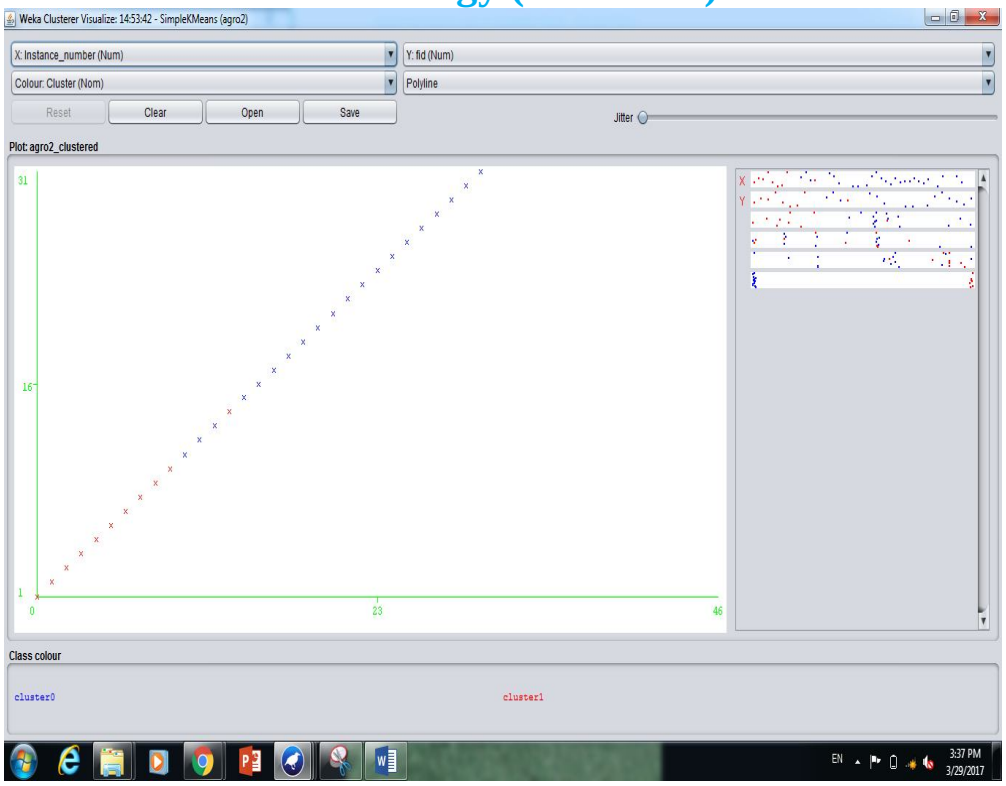


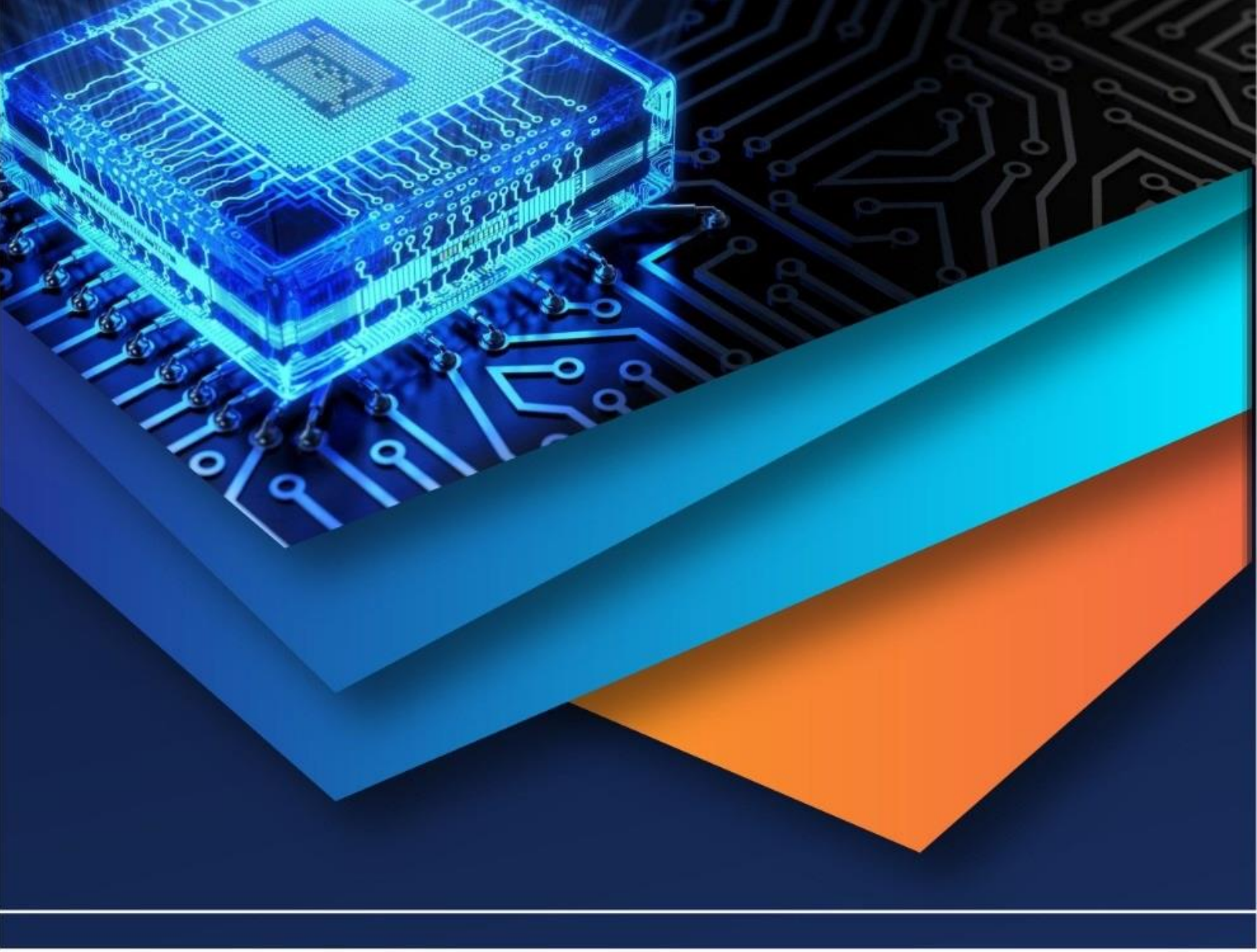
Figure 1.4 shows the clustering results

VI. CONCLUSION

The work concentrate on recognizing the required designs from extensive datasets in horticulture and furthermore finding the little examples for recognizes the generation of products. It contains the essential usefulness of KDD and data mining. This work incorporates the strategies which, take after to separate the data, examination on that data, take after the digging process for the distinguishing proof of examples. In this we utilized the information mining devices and calculations to mine the information and furthermore connected the conditions on different elements. Later on work we can include a few spatial information and calculations to distinguish the required examples and analyse the past calculations and plan the tree like R-tree, R* tree to how the conditions. We can likewise utilize this procedure in therapeutic, college administration framework, showcasing and so on.

REFERENCES

[1] Hetal Patel, Dharmendra Patel, ”A Brief survey of Data Mining Techniques Applied to Agricultural Data “International Journal of Computer Applications, Volume 95– No. 9, June 2014.
[2] Dehua Zhang,” Analysis on the Influencing Factors of Farmers’ Income in Heilongjiang”, International Conference on Civil, Materials and Environmental Sciences (CMES 2015).
[3] Yanxin Zhu, Di Wu and Sujian Li,”Cloud Computing and Agricultural Development of China: Theory and Practice” IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 1, January 2013.
[4] Seena Kalghatgi, Kuldeep P. Sambrekar,” Review: Using Cloud Computing Technology in Agricultural Development”, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 3, March 2015.
[5] Mr. Mahesh D. S1, Ms. Savitha S2, Dr. Dinesh K. Anvekar3.,” A Cloud Computing Architecture with Wireless Sensor Networks for Agricultural Application”, International Journal of Computer Networks and Communications Security ,VOL.2, NO.1, JANUARY 2014, 34–
[6] Prashant Satpute, Omprakash Tembhurne,”A Review of: Cloud Centric IoT based Framework for Supply Chain Management in Precision Agriculture”, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 11, November 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY