



Application of Orange Data Mining Approach of Argiculture Productivity Index Performance in Tamilnadu

G. Manimannan^{1*}, R. Lakshmi Priya², C. Arul Kumar³

^{1,3}Department of Statitics, TMG College of Arts and Science, Chennai, INDIA

²Department of Statitics, Dr. Ambedkar Govt. Arts College, Vyasarpadi, Chennai, INDIA

***Corresponding Author:** G. Manimannan, Department of Statitics, TMG College of Arts and Science, Chennai, INDIA

Abstract: This research paper attempts to identify the classification of Agriculture Productivity Index (API) with the help of fourteen major crops of Tamilnadu using four hierarchical clustering methods, viz complete, single, weighted and average linkage methods. The secondary database was collected from the Department of Economics and Statistics for a period of ten years from 2003 to 2012. The crops are categorized as cereals, pulses, oilseeds and cash crops. The states consist of thirty two districts with various agro climatic zones. Agriculture Productivity Index (API) is calculated with the use of Eneedy's method. In this paper, the Orange data mining software is used which is an open source data visualization and data analysis for learner and experts. Interactive work flows with a large toolbox using Python programming language. In addition, to cross validate the API, the orange data mining hierarchical clustering techniques is used. The results of cluster analysis generated five clusters based on API and they are labelled as Very High, High, Moderate, Low, Very Low. They are plotted in dendrogram and are highlighted with different colours of five clusters.

Keywords: Classification, Agriculture Productivity Index, Orange Data Mining, Cluster Analysis and Dendrogram.

1. INTRODUCTION

Agriculture plays a vital role of Indian economy in past decades. Indian economy has been considered as an agrarian economy with majority of its population in villages having agriculture and allied activities as their primary occupation. So many scholars have stated, in India most of the population gives importance to agriculture because it not only provides food but also the livelihood to more than half of the population. It also supplies lots of goods and raw materials required by the non-agriculture sectors [3]. In short, agriculture plays an important role in the economic development process of India.

The economic development depends on the net agricultural productivity of an area. The productivity refers to the ratio of index of local agricultural output to the index of total input used in farm production [5]. The Agricultural productivity of an area is indicated by Agricultural productivity index (API). Based on this API the agricultural zones are classified into different categories. This classification of zones by API gains importance for its help in policy making. Among the different methods used for classifying the entire districts of Tamilnadu on the basis of API, in this research paper, traditional hierarchical clustering technique with various methods using orange data mining software is used. This software is developed by python programming language and to visualize the clustering methods with help of drawing schema of orange data mining. The Main objectives of the present research paper is (i) to estimate the Agriculture Productivity Index (API) of Tamilnadu and (ii) to classify the API performance using hierarchical clustering methods with the help of Orange data mining.

2. MATERIALS AND METHODS

2.1. Data Collection

The secondary source data was collected from the Department of Economics and Statistics, Chennai during the period of 2002-03 to 2011-12. Initially 31 districts were considered for this study, unfortunately the data available for ten years period consists only of 28 districts and the remaining three districts split up from three main districts for administrative purpose. The data of three districts

are not available for ten years period and are discarded as outliers. In this research paper, fifteen major crops are chosen based on their area, production and yield. The rest of the crops data not available for the study period. The crops are categorized as cereals, pulses, cash crops and oil seeds. Each category consists of various crops and totally fifteen major crops are cultivated in the districts of Tamilnadu and are chosen carefully. The crops are listed as follows, namely, Paddy, Cholan, Cumbu, Ragi, Maize, Bengal gram, Red gram, Green gram, Black gram, Horsegram, Tapioca, Ground nut, Gingely, Coconut and Sugarcane[6] and [7].

2.2. Agriculture Productivity Index

Many researchers have applied in their studies to calculate API index based on Eneedy'l method [2]. In this research paper also, API index for twenty eight districts of Tamilnadu is calculated and used. The districts and API index is shown in Table 1. The API is calculated by considering the area of cultivation (Hectare) and production (Tons) of a particular 14 major crops in all twenty eight districts. The following proposed algorithm is used to calculate API:

Step 1: The first part is calculated to give an input matrix of production and area of all the crops in a unit area and are labelled as Y , then divided by production of the selected crops in the entire area (Y_n)

Step 2: The second part is calculated to give an input matrix of cultivation area of the selected crops within the districts (T) and are divided by the cultivation area of the selected crops within the entire zone.

Step 3: The API is calculated using *step 1* divided by *step 2* and multiplied by 100.

The formula for API is :

$$API = \frac{Y}{Y_n} \div \frac{T}{T_n} \times 100$$

2.3. Data Mining

Data mining is used for data cleaning, data integration, removal of outliers and knowledge discovery. In the recent years data mining applicatios is used for finance, authorship attribution, text mining, medical science and biological sciences etc. In this study, orange data mining software based on python programming language is used. This software consists of many statistical analysis and give results visually and efficiently. In addition, coss validation of the API using hierarchical clustering methods is used with the help of the following steps. From the oragne data mining software, a schema is drawn with utmost care as per the research requirement. The step by step conststruction of the schema is given below and represented in figure 1.



Figure1. Hiarchical clustering Schema for API

- Step 1: Select file wideget and open your database in the form of file like .tab and.xls format.
- Step 2: Select a data table widget and conect the file widget, then file widget is connected to distance widget.
- Step 3: Select hierarchical widget and connect to distance widget.
- Step 4: select one more table widget for assigning cluster output and connect to hierarchical ciluster widget.

Step 3: Finally, Double click file, data table widget 1, distance, hierarchical cluster and table widget. All these widget assign their output and display the result to report window.

Step 4: Open distance map widget which shows the two dimensional hierarchical clustering map with the label of all districts (Figure 1).

Step 5: Open hierarchical widget, it shows the first, second, third and fourth method of clustering map based on eculidean distance matrix, each method chosen manually. Finally four visualaization of API clustering methods is achieved (Figure 2 to 5).

After these procedures, the reseults are displayed, that are splitted into five clusters, two singleton cluster and rest splitted into three clusters based on their API.

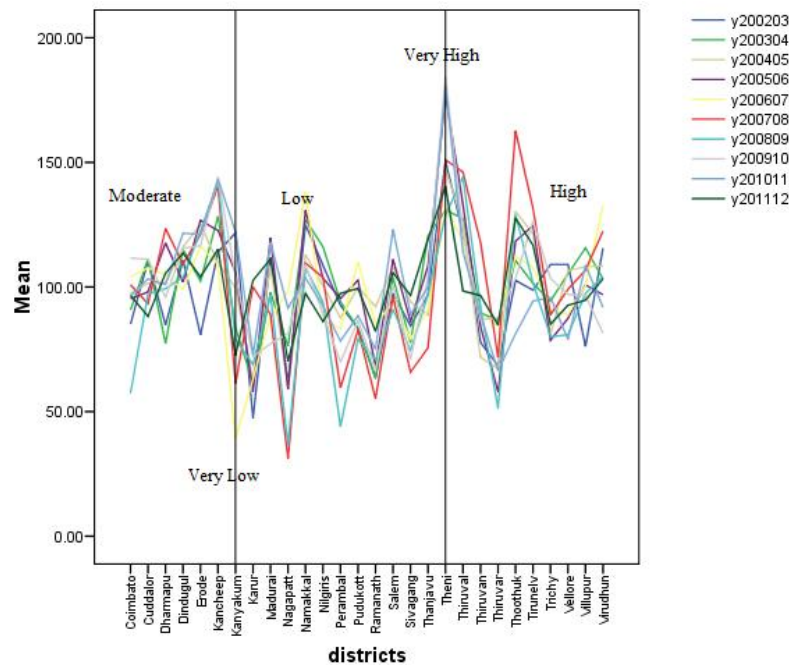


Figure2. Performance of API for the study period

2.4. Clustering Methods

The widget supports four ways of measuring distances between clusters: (i) the first method of single linkage computes the distance between the closest elements of the two clusters, (ii) The second method describe average linkage method that computes the average distance between elements of two clusters, (iii) Third method explains the weighted linkage uses the WPGMA method (Weighted Pair Group Method of Arithmetic averages), the similarity between any single and some group is simply the average of all the involved pair wise similarities and (iv) Finally, the complete linkage method computes the distance between clusters' most distant elements [5].

3. RESULTS AND DISCUSSION

Two methods have been used to calculate API and they are Enedy'l API and hierarchical clustering methods. In the first method, API is classified as five clusters and they are labelled as Very high, High, Moderate, Low and Very Low productivity region. The very low productivity district is Kanyakumari and high productivity district is Theni. The remaining three major clusters are labelled as high, moderate and low productivity districts. The districts of Coimbatore, Cuddalore, Dharmapuri, Dindugal, Erode and Kancheepuram fall in the moderate productivity region. The low productivity districts are Karur, Madurai, Nagapattinam, Nilgiris, Pudukkottai, Perambalur, Namakkal, Salem, Ramanathapuram, Thanjavur and Sivagangai. Finally, the districts of Thiruvallur, Thiruvannamalai, Thiruvannamalai, Thoothukudi, Tirunelveli, Trichi, Vellore, Villupuram and Virudhunagar fall in high productivity districts. One or two districts are misclassified in all these three major clusters which is due to API and data variation (Table 1).

The second method attempts to identify the hierarchical classification of API which shows the same results as that of API performance. This technique also produces five clusters based on the input

matrix of API. In complete linkage method, the results shows that Theni and Kanyakumari districts are singleton cluster and Theni is labelled as very high productivity district and Kanyakumari is labelled as very low productivity district. The remaining three major clusters are high, low and moderate categories. Majority of the districts fall in high category and they are Trichy, Vellore, Nilgiris, Thiruvannamalai, Coimbatore, Pudukkottai, Nammakkal, Virudhunagar, Dindukul, Cuddalore, Villupuram, Thanjavur, Salem and Madurai. The moderate categories of districts are Kancheepuram, Thiruvallur, Thoothukudi, Dharmapuri, Erode and Tirunelveli. The districts Karur, Thanjavur, Nagapattinam, Perambalur, Ramanathapuram and Sivaganga fall in low productivity (Figure 3).

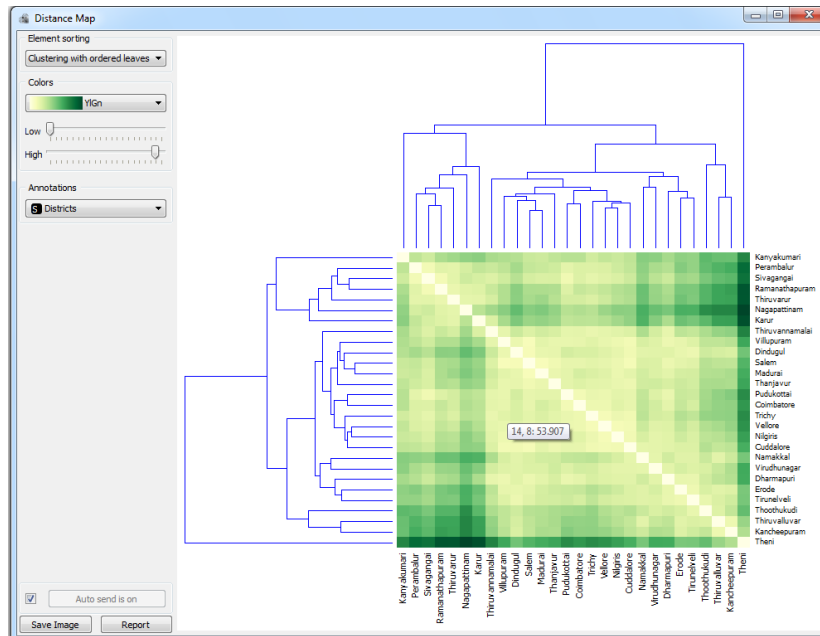


Figure3. Hierarchical Clustering two dimensional distance map

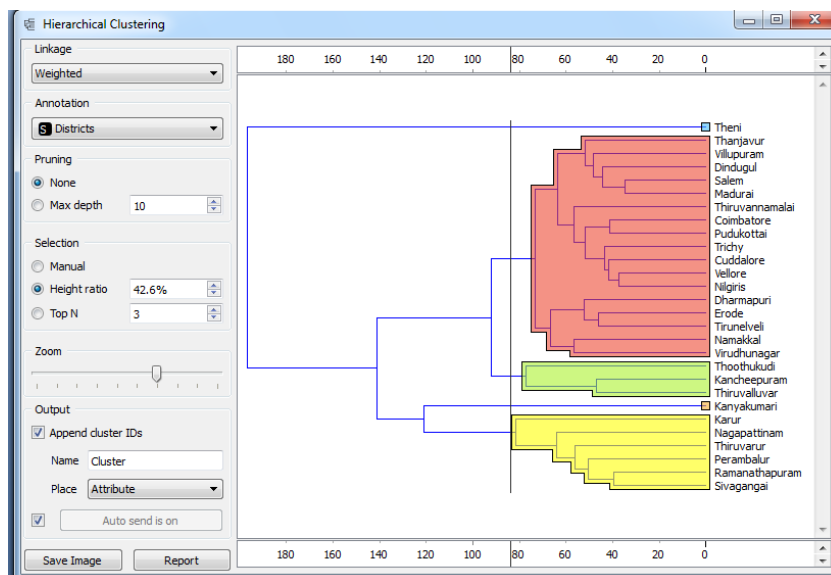


Figure3. Complete Linkage Method

The weighted linkage method is classified with little variation over the study period. The first two clusters are singleton cluster and they are classified as very high and very low productivity namely Theni and Kanyakumari. Thoothukudi, Kancheepuram and Thiruvallur districts productions of API are in moderate category. The districts of Karur, Nagapattinam, Thiruvallur, Perambalur, Ramanathapuram and Sivaganga fall in high productivity districts. The remaining sixteen districts fall in low productivity regions (Figure. 4).

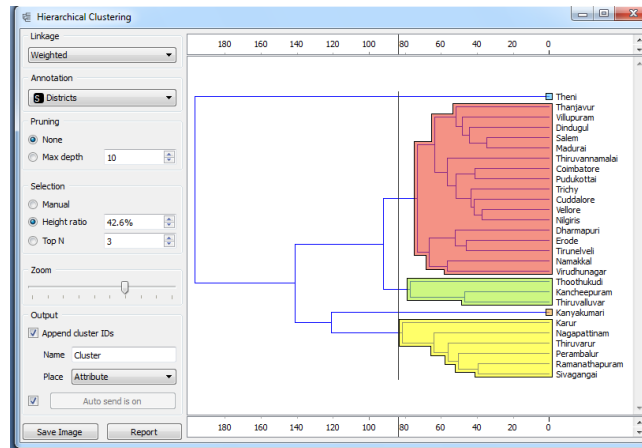


Figure4. Weighted Linkage Method

The average linkage method classified as five clusters, two singleton cluster and three major clusters and they are labelled as very high, high, moderate, low and very low categories. Theni fall in very high productivity districts and Kanyakumari in very low productivity districts in this study period. The high productivity districts are Thiruvannamalai, Villupuram, Dindugul, Salem, Coimbatore, Pudukkottai, Trichy, Cuddalore, Vellore, Dharmapuri, Namakkal and Virudhunagar. The districts of Thoothukudi, Kancheepuram and Thiruvallur fall in moderate production of API. The remaining districts of Karur, Nagapattinam, Thanjavur, Perambalur, Ramanathapuram and Sivaganga fall in the low categories of agriculture productivity region (Figure 5).

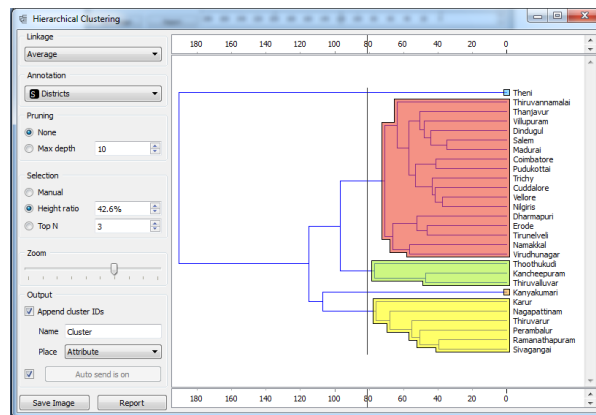


Figure5. Average Linkage Method

The last method is single linkage which is classified only as three clusters. The district of Kanyakumari and Theni is singleton cluster and are labelled as very low and very high productivity region. The remaining Twenty six districts fall in moderate categories. The results of the four methods shows little variation, but the first three methods forms the same cluster and more or less fall in the respective categories and are called as natural clusters (Figure 6).

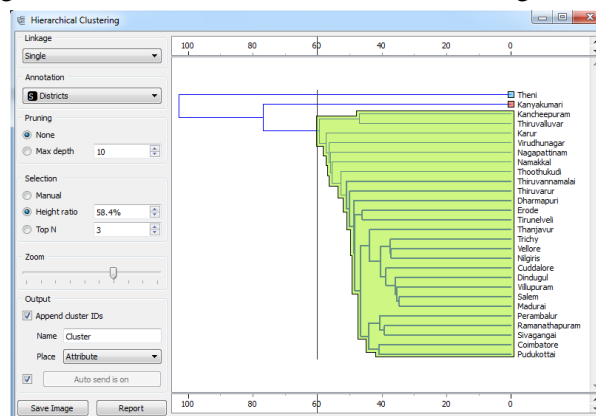


Figure6. Single Linkage Method

Finally the two methods are classified as five clusters based on API and are labelled as Very High, High , Moderate, Low and Very Low categories. Further, the API index make use of Neural Network Classification, Self Organizing Map, Support Vector Machine, Expectation Maximization (EM) Algorithm, DBASCAN algorithm, etc, of data mining tools are applied and the same results might vary little over the results. API is formed as five major clusters based on this study.

The following steps are interpreted and the results are shown from table 1 to table 4.

Step 1. The following table shows how the cases are clustered together at each stage of the cluster analysis in various methods.

Step2. Each method clusters are formed by merging cases and clusters a step at a time, until all cases are joined in one full-size cluster.

Step 3. In each stage, one cluster is joined with another cluster. In this research paper, all the four methods results are shown in Table 2 to Table 5.

Step 4. The Coefficients column indicates the distance between the two clusters (or cases) joined at each stage and the values depend on the proximity measure and linkage method used in the analysis.

Step 5. For a good cluster result, an unexpected jump in the distance coefficient or sudden drops in the similarity coefficient in the result tables are highlighted. This sudden change indicates the optimal stopping point for merging clusters. In this study, a four cluster solution is considered. The subsequent part of the table shows the phase at which each cluster appears first. Single cases existed before the analysis is started, so they are indicated by zeroes.

Table1. Performance of Agriculture Productivity Index using Eneedy'l Method

Districts	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10	2010-10	2011-12
Kancheepuram	114.03	128.3	109.61	122.65	108.99	140.52	142.4	144.04	143.22	115.15
Thiruvalluvar	121.62	127.4	125.82	132.97	118.38	146.06	144.14	121.47	113.93	98.38
Cuddalore	110.49	111.3	101.51	98.03	107.63	93.31	97.18	111.13	103.25	88.26
Villupuram	76.17	115.69	108.05	100.85	100.45	106.75	97.43	96.63	108.74	94.7
Vellore	109.04	105.28	106.63	87.43	89.08	99.13	80.85	97	78.94	92.7
Thiruvannamalai	77.78	89.58	71.77	84.23	89.13	118.05	92	86.89	88.96	96.57
Salem	97.48	103.8	107.88	111.07	93.6	96.21	90.96	100.97	123.13	105.68
Namakkal	124.95	126.99	113.01	130.71	138.18	109.58	107.76	108.88	103.59	97.45
Dharmapuri	84.8	77.36	101.65	117.67	104.8	123.44	99.31	95.73	100.91	105.51
Coimbatore	85.13	90.82	97.41	95.29	104.03	100.86	57.4	111.57	96.47	96.55
Erode	80.77	102.4	125.16	126.71	116.35	121.11	122.46	116.55	120.99	104.04
Trichy	109.04	94.36	80.99	78.59	90.71	88.87	80.02	103.41	95.84	85.09
Karur	47.39	61.39	62.82	57.99	63.09	100.12	68.95	71.98	72.75	102.63
Perambalur	92.66	94.06	87.61	95.39	83.3	59.59	44.04	69.86	78.3	97.46
Pudukottai	83.28	83.51	99.99	102.73	110.02	82.78	79.14	86.56	88.51	99.4
Thanjavur	96.85	120.33	88.81	108.41	90.8	75.6	97.06	103.93	100.01	119.23
Thiruvarur	68.33	86.64	67.26	57.84	84.26	71.69	51.3	87.63	66.36	84.76
Nagapattinam	62.1	76.31	59.99	58.91	98.21	31.19	36.75	81.28	91.52	70.21
Madurai	111.72	97.83	108.44	119.78	88.01	88.97	96.94	77.29	117.86	111.27
Theni	150.45	130.91	147.02	177.16	131.7	151.07	128.91	184.52	180.92	140.37
Dindugul	110.72	115.06	116	101.99	99.07	109.2	103.94	115.58	121.59	113.59
Ramanathapuram	63.7	63.26	92.09	68.34	85.24	55.2	70.87	66.68	74.98	82.25
Virudhunagar	115.54	103.13	108.57	96.91	132.79	122.37	104.26	81.51	91.81	103.31
Sivagangai	84.33	77.74	94.05	86.01	79.2	65.81	74.4	70.97	89.09	96.61
Tirunelveli	98.72	101.4	121.26	124.63	107.95	131.54	101.57	126.5	94.36	116.75

Application of Orange Data Mining Approach of Agriculture Productivity Index Performance in Tamilnadu

Thoothukudi	102.61	110.66	130.37	118.4	112.37	162.64	128.86	103.51	81.21	127.62
Nilgiris	108.65	116.05	100.42	103.49	86.69	103.77	93.83	89.1	92.41	86.15
Kanyakumari	121.69	83.77	99.7	105.91	39.11	61.28	79.07	104.47	122.16	72.53

Table2. Complete Linkage Method

Stage	Cluster Combined		Coefficients
	Cluster 1	Cluster 2	
1	5	27	895.108
2	7	19	1053.786
3	3	5	1427.626
4	22	24	1536.558
5	10	15	1651.247
6	3	12	1758.597
7	7	21	1916.753
8	11	25	1928.179
9	4	7	2193.477
10	1	2	2198.118
11	4	16	2477.804
12	14	22	2547.627
13	9	11	2596.316
14	3	10	2732.713
15	8	23	2947.112
16	14	17	3045.297
17	3	4	3180.016
18	3	6	4048.495
19	8	9	4329.282
20	14	18	4653.400
21	8	26	4947.746
22	13	14	5889.735
23	3	8	5903.125
24	1	3	9670.552
25	13	28	11614.014
26	1	13	14270.224
27	1	20	38243.352

Table3. Weighted Linkage Method

Stage	Cluster Combined		Coefficients
	Cluster 1	Cluster 2	
1	5	27	29.918
2	7	19	32.462
3	22	24	39.199
4	3	21	40.113
5	10	15	40.636
6	11	25	43.911
7	5	12	45.986
8	3	4	46.797
9	1	2	46.884
10	7	16	50.202
11	9	11	52.833
12	8	23	54.287
13	17	18	55.956
14	3	7	56.073
15	5	6	56.427
16	5	10	59.430
17	14	22	59.451
18	9	26	75.207
19	3	8	77.419
20	14	17	78.596
21	13	14	94.654
22	1	9	99.601

Application of Orange Data Mining Approach of Agriculture Productivity Index Performance in Tamilnadu

23	3	5	104.451
24	13	28	127.541
25	1	3	131.617
26	1	13	209.802
27	1	20	282.672

Table4. Average Linkage Method

Stage	Cluster Combined		Coefficients
	Cluster 1	Cluster 2	
1	5	27	29.918
2	7	19	32.462
3	3	5	35.097
4	3	12	38.429
5	22	24	39.199
6	7	21	39.296
7	10	15	40.636
8	4	7	42.677
9	11	25	43.911
10	3	16	44.170
11	14	22	46.052
12	1	2	46.884
13	4	9	47.195
14	3	10	48.637
15	14	17	50.553
16	3	6	51.202
17	4	23	52.154
18	4	11	55.074
19	14	18	57.375
20	4	8	57.499
21	3	28	59.703
22	4	26	61.784
23	13	14	63.392
24	1	4	67.599
25	1	3	73.694
26	1	20	82.461
27	1	13	95.276

Table5. Single Linkage Method

Stage	Cluster Combined		Coefficients
	Cluster 1	Cluster 2	
1	5	27	14.959
2	7	19	31.190
3	22	24	50.790
4	3	21	70.846
5	10	15	91.164
6	11	25	113.120
7	5	12	135.700
8	3	4	158.606
9	1	2	182.048
10	7	16	207.674
11	14	22	234.127
12	9	11	260.754
13	8	23	287.897
14	3	7	315.787
15	17	18	343.765
16	6	10	374.332
17	5	6	408.883
18	9	26	445.903
19	14	17	486.623
20	13	14	530.351
21	8	9	574.451
22	3	5	630.653

Application of Orange Data Mining Approach of Agriculture Productivity Index Performance in Tamilnadu

23	1	8	689.082
24	3	28	749.301
25	1	20	850.101
26	3	13	1019.756
27	1	3	1286.220

4. CONCLUSION

The application of Orange data mining shows the result visually in a very effective manner. This study determined on estimation of API and classified the index using Orange hierarchical clustering methods. In both the methods, results are same with little variation due to database and different methods of classification. Classification results are obtained by Data mining methods and API from various districts of Tamilnadu. In this study, agriculture productivity index data shows slight fluctuation over the periods of ten years with twenty eight districts of Tamilnadu. The oscillation is due to shortages of rainfall, indecisively dates of opening and closing of river water, overdue monsoons, real estate business, industrialization, government policies and various types of soils. Finally, the research concludes five types of classification achieved using API and data mining techniques during the study period of 2003-2012 with the entire districts of Tamilnadu. Agriculture Productivity indices are labelled as Very High, High, and Moderate, Low and Very Low productivity districts of Tamilnadu. A generalization of the results is under investigation to obtain a set of five classification of API for any given year.

REFERENCES

- [1] Anderson TW (1984) An Introduction to Multivariate Statistical Analysis, 2/e, John Wiley and Sons, Inc., New York.
- [2] Jasbir Singh, Dhillon S S (2002) Agricultural Geography, Tata McGraw Hill Publishing Company Limited, New Delhi.
- [3] Kalpana Sahoo, Narayan Sethi (2012) Investigating the impact of Agriculture and Industrial Sector on Economic Growth of India. OIDA International Journal of Sustainable Development 05:05, pp. 11-22.
- [4] Kohonen T (2001) Learning vector quantization. Springer Berlin Heidelberg.
- [5] Han, J., Kamber, M. 2012. Data Mining: Concepts and Techniques, 3rd ed, 443-491.
- [6] Shafi M (1983) Agricultural Productivity and Regional Imbalances – A study of Uttar Pradesh. Concept Publishing Company, New Delhi.
- [7] Season and Crop Report, Tamilnadu (Various Issues), published by Department of Economics and Statistics, Chennai.
- [8] Statistical Hand Book of Tamilnadu (2012-2013), Department of Economics and Statistics, Chennai.

Citation: G. Manimannan, et.al., " Application of Orange Data Mining Approach of Argiculture Productivity Index Performance in Tamilnadu ", *International Journal of Scientific and Innovative Mathematical Research (IJSIMR)*, vol. 7, no. 8, pp. 8-16, 2019. Available: DOI: <http://dx.doi.org/10.20431/2347-3142.0708003>

Copyright: © 2019 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.