DOI:10.26524/jms.2017.2

Journal of Management and Science ISSN: 2249-1260 | e-ISSN: 2250-1819

Application of support vector machine for evaulation of agricultural productivity in the state of tamilnadu

[1]G. MANIMANNAN, [2]C. ARUL KUMAR AND [3]R. LAKSHMI PRIYA

[1] ASSISTANT PROFESSOR, DEPARTMENT OF STATISTICS, DRBCCC HINDU COLLEGE, CHENNAI-72

[2]ASSISTANT PROFESSOR, VIVEKANANDHA COLLEGE OF ARTS AND SCIENCE FOR WOMEN, TIRUCHENGODE, NAMAKKAL, TAMILNADU

[3]R. LAKSHMI PRIYA, ASSISTANT PROFESSOR, DEPARTMENT OF STATISTICS, DR. AMBEDKAR GOVT. ARTS COLLEGE, CHENNAI-39

Abstract: This research paper attempts to identify the agriculture productivity performance in the state of Tamilnadu as agriculture sector is facing so many challenges in the past decades. Most of the agricultural lands are converted into real estate business and also occupied by corporate sector people. Most of the farmers and allied deparment population migrated to other state and even to other countries to live their livelihood, they work as daily wages. In this connection, this research paper attempts to promote agricultural sector as it is the mainstay and backbone of the Indian and Tamilnadu economy. Agriculture plays a vital role in the development of a country as well the state of Tamilnadu. It contributes nearly fifteen percent of Gross Domestic Product (GDP) of India. Seventy percent of the population depends on agriculture for their livelihood. In the past decade agriculture production had faced an increasing trend in districts of Tamil Nadu in all the crops. But nowadays the yield rate has a decreasing trend in Tamilnadu. However, agriculture productivity differs from region to region, which needs a detailed investigation. The main objective of this research paper is to analyze the agriculture productivity of fifteen major Crops in Tamilnadu using Support Vector Machine for district wise classification of entire state of Tamilnadu and Mosaic graph to visualize the performance of agricultural database. The secondary sources of database were collected from Department of Economics and Statistics, Tamilnadu during the period of 2003 to 2012. In this study yield deviation, visualization and classification of fifteen major crops are considered. The results attained three different methods of classifications and are labelled as High, Moderate and Low based on their Enyedi's index method of various crops.

Keywords: Enyedi's Method, Agriculture Productivity, Productivity Districts, Crops Yield, Support Vector Machine and Mosaic Graph.

1.0 Introduction

Agriculture sector plays a very important role in Indian economy from prior to present time period. Indian economy has been considered as an agrarian economy with majority of its population in rural areas having agriculture and allied activities as their main occupation. Most of the 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. And also the agriculture fields 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. This research paper mainly concentrates in Tamilnadu Agriculture productivity and growth.

Tamilnadu Agriculture productivity have a downward trend in the recent years. The recent research established three meaningful clusters and are labeled as High, Low and Moderate agriculture productivity index in Tamilnadu.(G. Manimannan *et.al*.) [9].

The economic development depends on the net agricultural productivity of an area, yield and production. 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, SVMs classification technique with help of data mining software version 3.3.7 is used in this research paper. This software is developed by Python programming language and also this software is mainly used for evaluation, Classification, Regression, unsupervised classification and Visualization of data. Orange data mining software is very user friendly and has powerful scripted language. The main objectives of this research paper is, (i) to identify Agriculture Productivity Index (API) using Enedy'l method, (ii) to classify the productivity performance of various districts for the study period using SVMs Classification data mining techniques with the help of API index.

2.0 Database

The secondary source data source was collected from the Department of Economics and Statistics, Chennai during the period of 2003 to 2012. Initially, 31 districts were considered for this study. After using data mining techniques three districts excelled in analysis. The three districts was separated from three main districts for administrative purpose. Those three districts data are not available for the study period and are discarded as outliers. The remaining 28 districts were used for analysis with yield of various crops. In this research paper, fifteen major crops are chosen based on their area, production and yield. The crops are categorized as cereals, pulses, cash crops and oil seeds. Each category consists of different crops which are cultivated in the districts of Tamilnadu and are chosen carefully. The crops are listed as follows, namely, Paddy, Cholam, Cumbu, Ragi, Maize, Bengal gram, Red gram, Green gram, Black gram, Horse gram, Tapioca, Groundnut, Gingelly, Coconut and Sugarcane [6] and [7].

3.0 Methodology

3.1 Agriculture Productivity Index

Agriculture Productivity Index is calculated By Enyedi's method and this method is most familiar among many research scholars [2]. In this paper the API index is calculated for twenty eight districts of Tamilnadu and it is used for further analysis. The district names and calculated API index are shown in Table 1. The API is calculated by considering the area of cultivation (Hectare) and production (Tons) of particular fifteen crops in twenty eight districts. The following proposed algorithm is used to calculate API:

Table1. Performance of Agriculture Productivity Index using Enyedi's Method

Districts	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
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	97.00	78.94	92.7
Thiruvannamalai	77.78	89.58	71.77	84.23	89.13	118.05	92.00	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	78.32	97.46
Pudukottai	83.28	83.51	99.99	102.73	110.02	82.78	79.14	86.56	88.51	99.42
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	55.23	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
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

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 :

$$API = \frac{Y}{Y_n} \div \frac{T}{T_n} X100$$

3.3 Support Vector Machine (SVM)

The machine learning method of Support Vector Machines (SVMs) is a state of the art pattern recognition technique whose foundations stem from statistical learning theory. However the scope of SVMs goes beyond pattern recognition. Because it can also handle two learning problems, namely, regression estimation and density estimtion. An SVM is a general algorithm based an guaranteed risk bounds of statistical learning theory, so it is called as structured risk minimization principle. The SVM have three boundaries method, viz., linear, non linear boundaries method, SVM classifying overlapping and non overlapping data method (Vapnik-Chervonenkis, 1998).

3.4 Support Vector Machine Classifier

Three- layer neural networks are universal classifiers in that they can classify any labeled data correctly if there are no identical data in different classes. In training multilayer neural network classifiers, usually network weights are corrected so that sum of squared errors between the network outputs and the desired output is minimized. But since the decision boundaries between class aquired by training are not directly determined, the generalization ability depends on the training method. And it degrades greatly when the number of training data is small and overlap among classes is scarce or non-existent.

In addition, in training SVMs, the decision boundaries are deetermined directly from the training data so that the seprating margins of decision boundaries are maximized in the high-dimensional space called feature space. The learning strategy, based on Statistical learning theory developed by Vanik, to minimize the classification error of the training and unknown data. In this connection, the researcher use SVMs classifiers to classify API index database for the study period.

3.5 Support Vector Machine Algorithms

The data mining SVMs algorithms are discussed in the following steps. The Support vector machine (SVM) is a classification technique that separates the attribute space with a hyper plane, thus maximizing the margin between the instances of different classes. The technique often yields supreme predictive performance results. Data mining embeds a popular implementation of SVM from the LIBSVM package. This widget itself is a graphical user interface. The following algorithm will execute SVMs Classification:

Step 1. The learner can be given a name under which it will appear in other widgets. The default name is ‑SVM‖.

Step 2. The next step is to set the classification and test error settings. *C-SVM* and *v-SVM* are based on different minimization of the error function.

Step 3. The next block of options deals with kernel, a function that transforms attribute space to a new feature space to fit the maximum-margin hyper plane, thus allowing the algorithm to create non-linear classifiers with Polynomial, RBF and Sigmoid kernels. Functions that specify the kernel are presented besides their names and the constant values kernel function is 0 by default, degree of the kernel by is 0 by default and gamma constant in kernel function is also 0 by default.

Step 4. Set permitted deviation from the expected value in *Numerical Tolerance*. Tick the box next to *Iteration Limit* to set the maximum number of iterations permitted. After these steps SVM is executed and displays the output for the study period.

3. 6 Proposed Algorithm for SVM

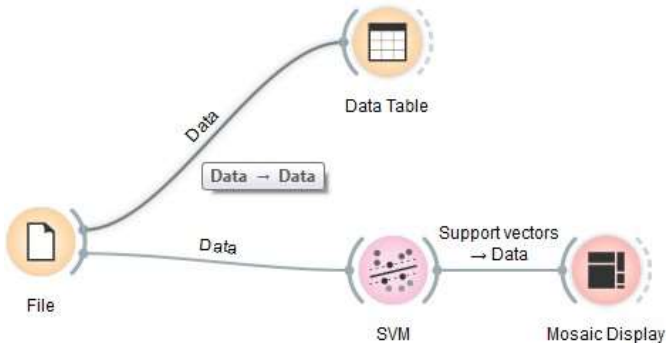
The workflow of SVMs are depicted in the following steps:

Step 1: Select file widget and load your database in the form of file like .tab and.xls format.

Step 2: Select a data table widget and conect to file widget, then file widget is connected to SVM widget.

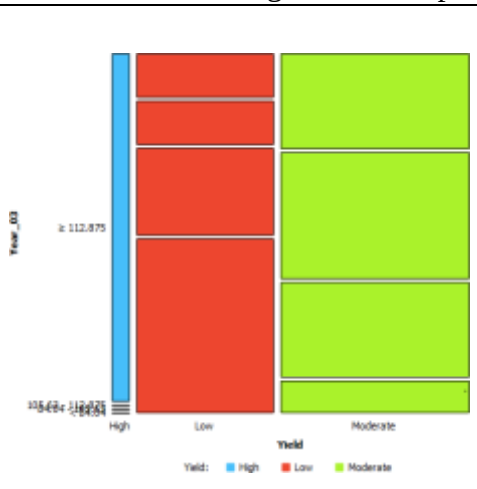
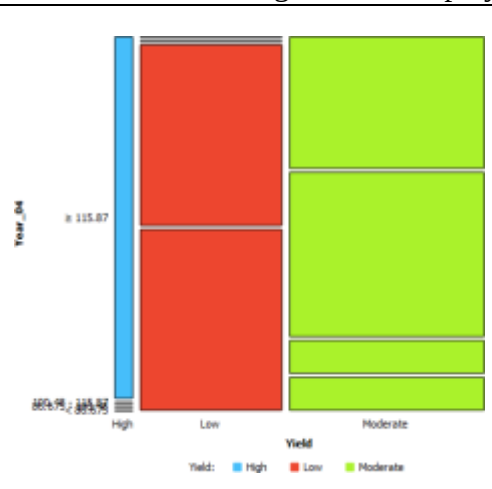
Step 4: Finally, Double click file widget, data table widget, Mosaic graph widget and SVM widget one by one. All these widgets are executed and display their classification tables and mosaic graph for the study period (Table 2 and Figure 1).

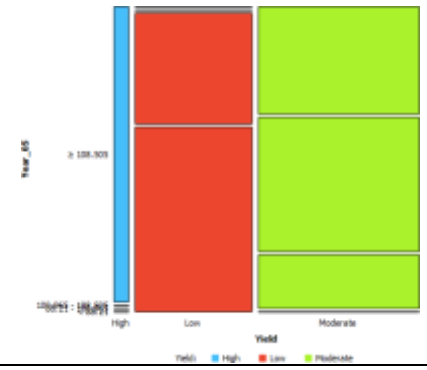
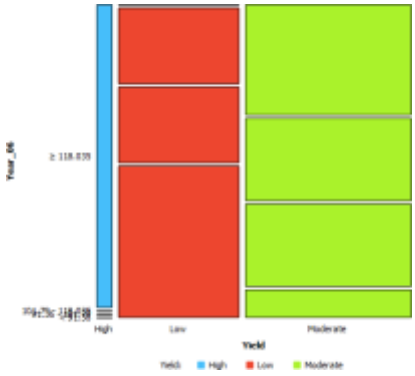
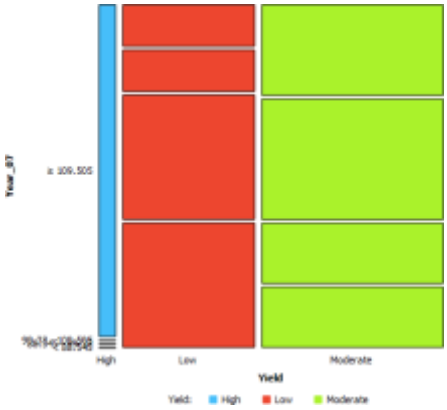
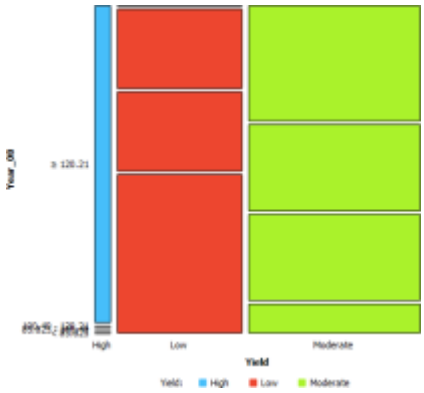
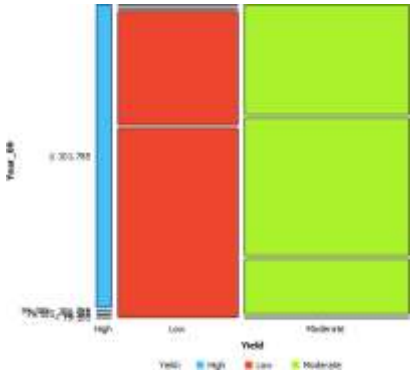
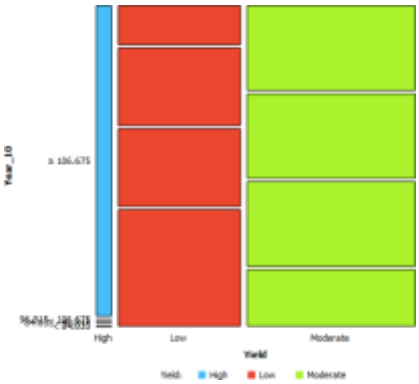
Figure 1. Support Vector Machine Claasification (SVM) Work Flow Diagram

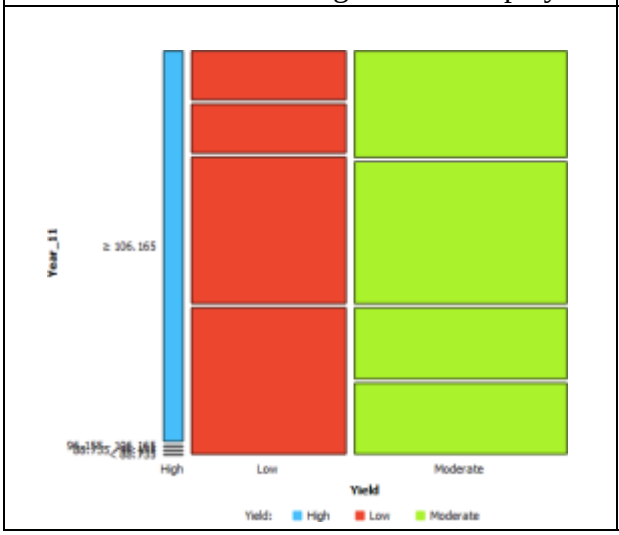
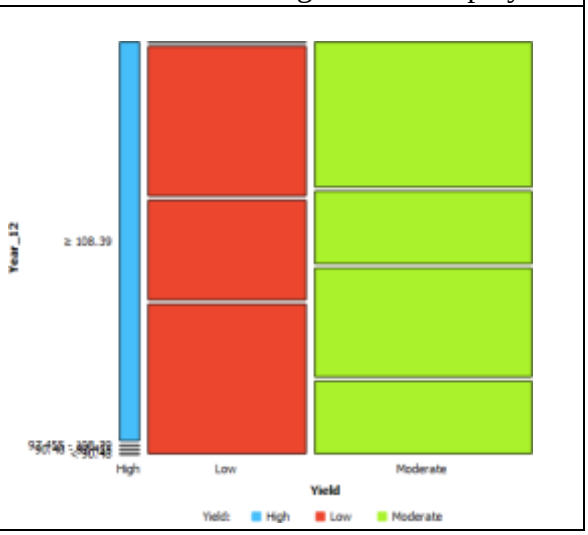


The SVM classifer results shows three clusters and are higlighted . The color blue is high, red is low productivity and green is moderate agricultural productivity index for the study period of ten years. The classified districts are shown in the same color in SVM classification Table 2.

Figure 2. Support Vector Machine Claasification (SVM) Mosaic Diagram

Year 2003: SVM using Mosaic Display	Year 2004: SVM using Mosaic Display
	

Year 2005: SVM using Mosaic Display	Year 2006: SVM using Mosaic Display
	
Year 2007: SVM using Mosaic Display	Year 2008: SVM using Mosaic Display
	
Year 2009: SVM using Mosaic Display	Year 2010: SVM using Mosaic Display
	

Year 2011: SVM using Mosaic Display	Year 2012: SVM using Mosaic Display
	

4.0 Results and Discussion

The SVM and Enedy‘l methods have produced same results for API. In the first method, API is classified as three clusters and they are labelled as high, moderate and low Productivity Indices districts. The low productivity indices districts are Sivaganga, Ramanathpuram, Nagapattinam, Thiruvarur and Perambalur. The High productivity Indices is Theni and remaining districts are Thuthukudi, Tirunelveli, Erode, Namakkal, Thiruvallur, Kanchepuuram, etc in the state of Tamilnadu. SVMs classifiers method also achieved three classifications using the API. They are labelled as high, moderate and low productivity Indices districts. Both the method of results is classified as the same. For the years 2003 to 2012, the classification results are shown in the Table 2 and Figure 2. The values underlined in red colours are low productivity indices, blue colour indicates moderate productivity indices and green colour implies high productivity indices districts. The SVMs classification is shown in the form of mosaic graph (Figure 2).

Table 2. Support Vector Machine Claasification (SVM) Method (Year-Wise)

Finally the two methods achieved us three clusters based on API and k-means clustering methods and are labelled as Upward Productivity Indices, Stable Productivity Indices, and Downward Productivity Indices categories. In addition, the API data get the same results over the study period using data mining tools like, Neural Network Classification,

Self-Organizing Map, k-means clustering technique, Expectation Maximization (EM) Algorithm, DBASCAN (***Density-Based Spatial Clustering of Applications with Noise***) algorithm, etc.

5.0 Conclusion

The application of SVM Classifiers shows the result visually in a very effective manner. This study determines an estimation of API and classified the index using SVMs classifiers. In both the methods, results are the same with slight 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 study period of ten years with twenty eight districts of Tamilnadu. The oscillation is due to shortage of rainfall, indecisively dates of opening and closing of river water, overdue monsoons, real estate business, industrialization, government policies and various types of soils. Based on so many constraints, the farmers face so many problems and even lost their agriculture fields. Finally, the research concludes three types of classification achieved using API and SVM techniques during the study period of 2003-2012 with the entire districts of Tamilnadu. Agriculture Productivity indices are labelled as high, moderate and low districts of Tamilnadu. A generalization of the results is under investigation to obtain a set of three 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) Manimannan G. *et.al.* (2016), Classification of agricultural productivity index of Cauvery delta zone using artificial neural network, Agriculture Science Digest, A Research Journal, India, pp. 1-6.
6) Han, J., Kamber, M. 2012. Data Mining: Concepts and Techniques, 3rd ed, 443-491.
7) Shafi M (1983) Agricultural Productivity and Regional Imbalances – A study of Uttar Pradesh. Concept Publishing Company, New Delhi.
8) Season and Crop Report, Tamilnadu (Various Issues), published by Department of Economics and Statistics, Chennai.
9) Statistical Hand Book of Tamilnadu (2012-2013), Department of Economics and Statistics, Chennai.
10) Vladmir N. Vapnik (1998), Statistical Learning Theory, ISBN. 0-471-03-03003-1.

*****