Implementation of Cluster Analysis and Artificial Neural Networks as an Alternative for Klassen Typology and LQ: Case of Coconut

Cungki Kusdarjito

1 Faculty of Agriculture, Janabadra University, Yogyakarta, Indonesia.

E-mail: ckusdarjito@janabadra.ac.id

Abstract. Due to the campaign supported by manufacturers of soy-oil in the United States which stated that consuming coconut oil might cause heart disease, coconuts' popularity started to decline. As a result, the coconut was neglected by farmers and productivity was declining. Yet, recent study showed the opposite results. Coconut was good for health. As a result, the demand for coconut products increased. Indonesia, as the world's largest coconut producers cannot maximize these opportunities due to aging coconut which have been neglected for many years. Coconut products may improve the income of farmers and encourage sustainable agriculture as well as diversify farmers' income. Some simple methods, such as Klassen Typology or Location Quotient can be used to classify the potential area to be developed or rejuvenated. Nevertheless, these approaches are not able to generalize and not suitable to be implemented for cases with large data. This research tried to use cluster analysis such as dendrogram, Principle Component Analysis and Artificial Neural Network for classification. The results show that the dendrogram provided good results, whilst the Principle Component Analysis and Artificial Neural Networks required more data for better results.

1. Introduction

Due to the negative campaign supported by manufacturers of soy-oil in the United States that consuming coconut-oil may cause heart disease, coconuts' popularity continued to decline since the 60s. Coconut was abandoned by farmers with no attempt for rejuvenation for a long time. Coconut (Cocos nucifera) is widely distributed in tropical regions of Asia, Africa, Latin America and in the Pacific Islands. Economically, coconut is an important commodity in the Philippines, Indonesia, Papua New Guinea, Sri Lanka, South India, Malaysia, Tanzania and in the Caribbean. In the Philippines, as one of the world's largest exporter of copra (dried coconut flesh) and coconut oil, more than 30 percent of the population is involved directly or indirectly to coconut industries for their lives [1]. Meanwhile, In Indonesia, coconut has long been known and greatly contributes to people's lives, either in terms of economic, social and cultural aspects [2]. Coconut trees are often referred to as the “tree of life” because from the coconut trees many products can be derived [1], from direct consumption as food and beverages to lubricant and surfactant for industrial purposes [3]. Other products that can be derived from coconuts are coconut sugar, Virgin Coconut Oil (VCO), and coconut water.

Recent research shows that coconut products are good for health and promoted as a superfood. With the help of celebrities, coconut is dubbed as a superfood, result in a surge in demand. The increasing demand mainly comes from the coconut water and VCO [2]. The global market is now witnessing the
demand for coconut products increase significantly. Globally, demand for coconut products growing exponentially [4]. With the current trends, the estimated global market for coconut will increase to USD 10 billion by 2030. This increased demand for coconut products mostly originated from the higher added value products which in turn will provide opportunities to increase the income for millions of small-scale coconut producers [4].

Although there is optimism for increasing coconut demand in the future, the previous prediction from the World Bank in the 80s stated that coconut oil compared to the main competitor did not give bode well for the communities who rely on coconut as the source of main income [1]. The latest data show that the share of coconut oil and its growth relatively minuscule [5]. Since the price of coconut products are still volatile, mixed cropping between coconut and the perennial plant may improve and stabilize the farmers’ income. For instance, coconut can be mixed with cacao or coffee.

The aim of the research is implementing clustering technique, ie. Dendogram, Principle Component Analysis (PCA) and Kohonen Self Organizing Map (Kohonen SOM) in place where Klassen Typology and Location Quotient (LQ) are commonly used to identify the base (leading) commodities in certain area. Klassen Typology or LQ can be used to classify the potential area to be developed for coconut. Nevertheless, these approaches are not suitable to be implemented for cases with large data. Klassen Typology and LQ are also difficult to be combined with other data, such as socio-economic data. They only portrait the comparative condition for the area being analyzed. Since LQ and Klassen Typology are categorization processes, this task can also be handled by cluster analysis. This paper uses cluster analysis, either based on a statistical approach or an artificial neural network. Both can be classified as machine learning. Dendrogram, PCA and Kohonen SOM were implemented to identify the coconut industry at the provincial level along with other commodities. Cluster analysis, especially Kohonen SOM has many flexibilities. For instance, data from other sectors can be easily added to the model. They are also able to handle large data by learning the pattern of the data provided and make a generalization.

2. Materials and Methods

2.1. Data Collection

Secondary data were collected from the Indonesian Statistical Bureau (BPS). All calculations were carried out using R and RStudio, an open-source statistical software. Data used in this paper were coconut production and a land area of coconut in 2018. The growth of the coconut land area is based on data in 2015 and 2018. Dynamic Location Quotient (DLQ) and Back Propagation Artificial Neural Network can be used for more detailed and longer time series.

This research implemented the clustering technique and machine learning to automate the process of categorization commonly used in Klassen Typology and LQ. “Klassen Typology” was replaced by a dendrogram whilst “LQ” was replaced using biplot-PCA and Kohonen SOM. The result from Dendrogram was used to classify the coconut condition in each province in Indonesia, whether it is developing or declining and in which provinces it can be developed in anticipating increasing demand for coconut products globally. Further, as an alternative for LQ, all plantation commodities were presented to Kohonen Layers and iterated up to 5000 times to learn the pattern. The cut off $\alpha$ was set from 0.05 to 0.01. Commodities included in this research were palm, coconut, rubber, coffee, cacao, sugarcane, tea, and tobacco.

This paper is not intended to demonstrate how to operate R or how the calculation was carried out; rather this paper shows how machine learning can assist in developing agricultural sectors. Machine learning differs from the classical method such as statistics. It evaluated the pattern, shows the hidden facts and makes a generalization. Recall can be presented by presenting new patterns to the Kohonen layers. List of Provinces in this paper is indexed as shown in Table 1 below:
2.4. Kohonen Self-Organizing Map (Kohonen SOM)

The application of neural networks has gained popularity in data mining since it has the advantage of freeing the process from predetermined models. They may detect nonlinear relationships automatically with no need assumptions on normality [16]. Yet, the substitution of neural network-based techniques in the place of statistical modeling techniques needs justification on grounds other than that of novelty. Statistical techniques provide a wealth of diagnostics that can be used to rigorously evaluate alternative solutions [17].
The Self-Organizing Maps (SOM) was introduced by Teuvo Kohonen in the early 80s [18]. SOM belongs to the general class of unsupervised neural network (NN) models. SOM is an exploratory data analysis that projects a multi-dimensional data set onto space with a small dimension from large data sets. It differs from the backpropagation neural network. The main difference is that network-learning is 'unsupervised' which is equivalent to clustering methods. SOM also differs from traditional cluster analyses in statistics. Traditional clustering methods involve a variety of algorithms but almost invariably build distinct self-contained clusters [19]. The method involves iterative adjustment of weights to capture and preserve the properties of the data. Once the weights have been established, the network operates simply by finding the Kohonen node which is the nearest match to a given input vector, measured in terms of the Euclidean distance between the input vector and the weights of the node [19]. Yet, SOM had suffered from a lack of rigorous results on its convergence and stability [18]. SOM shows neighboring segments that are linked based on many similar characteristics but may differ in one or two [19].

3. Results and Discussion

3.1. Dendrogram
This section illustrates how dendrogram can be used for clustering similar to Klassen typology.

Figure 1. Scatter plot and dendrogram.

Figure 1a plots coconut productivity (ton/ha/year) in the y-axis and coconut growth of land area (%) in the x-axis without any modification. Average productivity is 0.759 ton/ha/year whilst the average growth is -1.684 percent. In other words, the area of coconut is decreasing. Growth was measured from 2015 to 2018, whilst production uses data for the year 2018. K-means uses in the dendrogram is k = 6. Both average values of x and y usually are set as cut off in the manual process. In this paper, both mean values are used for reference only and they are not being implemented in constructing a dendrogram since clustering can be done automatically. The dendrogram shown in Figure 1b is constructed based on Euclidian distances after all values are standardized. The results shown in the dendrogram are consistent with raw data in Figure 1a, even though the cut off is not implemented during the clustering. Based on the results on dendrogram, the value is presented in Table 2 below:
Table 2. Clustering obtained from the dendrogram

| y > 0.759* | Y < 0.759y* |
|------------|-------------|
| 13, 24, 31 | 21, 32, 18, 20, 1, 15, 22, 33 |
| 7, 3, 29, 19, 27, 12, 26, 8, 14, 6, 17, 16, 5, 30, 25, 28 |
| 23, 10, 9, 11 |

*mean value, for reference only

Referring to Table 1, Figures 1a and 2a, point 24 (North Sulawesi) and 31 (North Maluku), are among provinces with high productivity and “high growth rate” for coconut compare. Yogyakarta (13) has high productivity but stagnated in area expansion. Growth in Yogyakarta actually is negative, ie. 1.40% between 2015-2018. On the opposite side, point 23 (North Kalimantan), 10 (Riau Archipelago), 9 (Bangka-Belitung) and 11 (West Java) are characterized by low productivity with fast declining in the coconut area. Point 22 (East Kalimantan) and 15 (Banten) and 1 (Aceh) in Klassen Typology are classified as low productivity and low growth regions. Point 2 (North Sumatera) and point 4 (Riau) are not included in the Table (outliers). North Kalimantan and Riau archipelago both are new provinces. The result in West Java is quite interesting. It seems that land conversion in West Java is intensive due to its proximity to Jakarta. Also, tea is the dominant commodity in West Java.

Figure 2a represents biplot-PCA. The direction of the arrow represents ‘increasing’ while the opposite direction represents ‘decreasing’. This process is simple since only two components are involved. The axes of productivity and the axes of growth are perpendicular to each other. This means both are not correlated. Figure 2b represents the heatmap of Euclidian Distance. Point 2 (North Sumatera) with relatively low productivity and high growth rate in the coconut area is located in the opposite direction with point 4 (Riau) with high productivity but low in growth-rate (negative). In Figure 2b, the Euclidian distance between both points is the largest, indicated by the darkest red color. Yet, point 2 (North Sumatera) and point 4 (Riau), both are outliers. Point 23 (North Kalimantan), are suffering both a low expansion of coconut area and productivity. Point 13 (Yogyakarta) characterized by high productivity but stagnated growth. Point 24 (North Sulawesi) and 31 (North Moluccas) have more balance between growth and productivity. Even though Yogyakarta has the highest productivity, it is not the main player in Indonesia since the area is quite small compared to other provinces (see the next discussion). Point 31 (North Moluccas), 24 (North Sulawesi) also have a high productivity rate. Thus, all two provinces are among the candidates to be developed for the coconut industry; whilst in Riau, the declining area of coconut needs to be halted, even though its productivity is among the highest.

![Figure 2. Biplot and heatmap.](image-url)
3.2. *Kohonen-SOM*

Table 3 shows the list of LQ greater than 1 for each base commodity at the provincial level.

**Table 3.** List of commodities for each province with LQ > 1 (2018)

| Commodity | Province with LQ > 1 (Production) |
|-----------|----------------------------------|
| Palm      | North Kalimantan, Central Kalimantan, East Kalimantan, Bangka Belitung Archipelago, Riau, South Kalimantan, West Papua, West Kalimantan, North Sumatera, Papua, West Sulawesi, West Sumatera, Bengkulu, Jambi |
| Coconut   | North Sulawesi, North Moluccas, Moluccas, Bali, Yogyakarta, East Nusa Tenggara, Gorontalo, West Nusa Tenggara, Banten, Central Java, Central Sulawesi, West Java, South Sulawesi, South-East Sulawesi, East Java, Riau Archipelago, West Papua, Papua |
| Rubber    | Riau Archipelago, South Sumatera, Banten, West Java, Jambi, Bengkulu, Lampung, Central Java, West Kalimantan, South Kalimantan, West Sumatera, Aceh, North Sumatera |
| Coffee    | East Nusa Tenggara, Bali, South Sulawesi, Lampung, Aceh, West Java, Bengkulu, West Nusa Tenggara, East Java, South Sumatera, Central Java, Banten |
| Cacao     | South-East Sulawesi, South Sulawesi, Central Sulawesi, East Nusa Tenggara, West Sulawesi, Moluccas, Bali, Papua, Gorontalo, North Moluccas, Banten, West Sumatera, Aceh, Lampung, East Java, West Papua, North Sulawesi, Yogyakarta, West Nusa Tenggara |
| Sugarcane | East Java, Lampung, Central Java, Gorontalo, Yogyakarta, West Java, South Sulawesi |
| Tea       | West Java, Central Java, East Java, West Sumatera |
| Tobacco   | West Nusa Tenggara, East Java, West Java, Yogyakarta, Bali, East Nusa Tenggara, South Sulawesi |

If we look closely at the production-based biplot-PCA analysis (Figure 3), the base commodities in each island are different. Java is dominated by tobacco and sugarcane (except tea for West Java), Kalimantan is dominated by rubber and palm, Sulawesi is dominated by cacao. Sumatera is also dominated by palm and rubber, and in some areas, coffee can be found. Cacao with rubber and palm almost located in the opposite direction in the biplot; therefore, it is difficult to find cacao in the region which dominated by palm and rubber. This also indicates that the agronomic and climate environment needed by cacao differs from rubber and palm. In terms of production, tobacco, sugarcane, and coconut are closely related, but both are not related to rubber and palm tree. If we evaluate the growth of land, the result may differ. Coconut is more loosely related to coffee and cacao. This indicates the mixed cropping need to be intensified for coconut, cacao, and coffee.

![Figure 3. PCA for base commodities and relative position of provinces.](image-url)
Unsupervised learning can be used for classification and generalization. In this paper, Kohonen SOM is used to classify similarity among provinces, especially provinces that can be selected to develop coconut industries. The results are shown in Figures 4a and 4b.

Figure 4. Kohonen-SOM

Figure 4a shows a fan-like diagram radiating from the center. The radiating diagram representing the weight of Kohonen layers. Nine clusters are set in advance. Cluster 1 is the circle in the most left bottom of the diagram, second cluster located next (or in the right) of the first cluster and so on. In Figure 4b, circles are populated with many small dots. Dots in Figure 3b shows the number of provinces fitted into specific clusters. For example, cluster 1 contains 5 provinces that have “similar” property. Cluster 1 represents the provinces with specialized in cacao (see also the radiating diagram in Figure 4a). These provinces also produce coconut but not dominant. List of provinces which fitted to each cluster is presented in Table 4.

| SOM Clustering | Province | Commodity |
|----------------|----------|-----------|
| **Major Producers** | | |
| Cluster 1 | West Sumatera, Central Sulawesi, South Sulawesi, South-East Sulawesi, West Sulawesi | Cacao, (Coconut) |
| Cluster 2 | North Sulawesi dan North Moluccas | Coconut |
| Cluster 3 | Aceh, Bengkulu, Bangka-Belitung, Riau Archipelago, Yogyakarta, Banten, Bali, East Nusa Tenggara, North Kalimantan, Gorontalo, Moluccas, West Papua, Papua | see minor producers |
| Cluster 4 | West Java | Tea, (Coconut, Coffee) |
| Cluster 5 | Central Java, West Nusa Tenggara | Tobacco, Coconut |
| Cluster 6 | West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, Jambi | Palm, Rubber |
| Cluster 7 | East Java | Tobacco, Sugarcane, Coconut, Coffee, Cacao |
| Cluster 8 | Lampung | Coffee, Sugarcane, Cacao, (Rubber, Coconut) |
| Cluster 9 | North Sumatera, Riau, South Sumatera | Palm, Rubber, Coffee, Coconut |
| **Minor Producers** | | |
| Cluster 3.1 | Bali, NTT, Moluccas | Coconut, Cacao, Coffee |
| Cluster 3.2 | Bengkulu | Palm, Rubber, Coffee, Tea |
| Cluster 3.3 | Aceh | Palm, Coconut, Rubber, Coffee, Cacao, Tobacco |
| Cluster 3.4 | Yogyakarta | Sugarcane, Tobacco, Coconut |
| Cluster 3.5 | Riau Archipelago, Banten, North Kalimantan | (Palm, Coconut, Rubber, Cacao) |
| Cluster 3.6 | Bangka Belitung | Palm, Rubber |

Cluster 3 is difficult to read because its products are too small compared to those at the national level. Thus, the Kohonen SOM is iterated again for all provinces in cluster 3. The results are interesting since the majority of the minor provinces are coconut producers, along with cacao and coffee (except for Bangka-Belitung).
Based on the Kohonen SOM, it is easier to determine what commodities should be developed in certain provinces such as Sulawesi for Cacao (except for Gorontalo and North Sulawesi), North Sulawesi and the Moluccas for Coconut, West Java for Tea, Lampung for Coffee and so on. This differs from LQ methods which difficult to interpreted since one province may have many base commodities which indicate by LQ > 1.

Based on the Kohonen SOM, Coconut becomes the main commodity in North Sulawesi and North Moluccas, followed by North Sumatera, Riau, South Sumatera, East Java, Lampung, Central Java, West Nusa Tenggara. Coconut also becomes the main commodity in a smaller province such as in Bali, NTT, Moluccas, Aceh, and Yogyakarta. Provinces, where Cacao and Coconut are co-existing, can be found in West Sumatera, Central Sulawesi, South Sulawesi, South-East Sulawesi, West Sulawesi. Yet, the share of coconuts in these provinces are small. In East Java, the share of coconut is large and Cacao is small. The opposite is true for Lampung. In minor provinces, cacao coexists with coconut in Bali, Moluccas and East Nusa Tenggara in which share of coconut is higher in these provinces. Aceh is the province where the share of both are quite similar. Coffee coexists with coconut in North Sumatera, Riau, South Sumatera, East Java, Lampung (share of coffee is greater than coconut), and Aceh. Empirical results indicated consistent results, such as in Sulawesi (mixed cropping), Inderagiri Hilir (processed nut), Banjarnegara and Yogyakarta (coconut sugar), where coconut become the main player.

Coconut, cacao, and coffee can be found in many provinces in Indonesia as shown by LQ, but Kohonen-SOM classified this coexistence more systematically. Cacao and Coconut coexist in many provinces such as in Sulawesi, Aceh, East Java, and in smaller provinces such as Bali, Moluccas, and East Nusa Tenggara, whilst coffee is coexisting with coconut in North Sumatera, Riau, South Sumatera, East Java, Lampung, and Aceh. Therefore, mixed cropping (in the form of inter-cropping) can be initiated from these provinces. By implementing mixed cropping, farmers have more resources for their income and all input can be shared more efficiently.

4. Conclusion
Clustering methods can be used as an alternative for a classical approach such as LQ and Klassen Typology. Kohonen SOM, as one of the machine learning models, has more flexibility since it can handle large data when doing classification. Data can be added with ease and incrementally. Raw data can be presented directly to Kohonen layers without too many efforts. For instance, when data are obtained from the regency level, more information and pattern can be obtained. Also, other commodities such as from the food sector, horticulture, husbandry, any sectors outside the agriculture and socio-economic attributes can be presented and calculated at once. Therefore, much information can be obtained. The dendrogram was more suitable for less complex classification whilst Biplot in PCA was more suitable for linear cases. Dendrogram and Biplot in PCA were reducing the dimension of the data and classified the database on their similarity. Kohonen SOM appropriate for nonlinear model and large data.

This paper shows that the clustering method provides information in a more systematic way than Klassen Typology and LQ. Therefore, relational between region and commodity are easier to be mapped. This paper also shows that palm and coconut are not closely related, palm and cacao exist oppositely. When cacao becomes the dominant commodity, the palm is not the main commodity in many provinces (except in Aceh, North Sumatera, Riau, South Sumatera). Mixed cropping also can be proposed when two or more compatible commodities coexist.

References
[1] Terry M 1994 The tree of life Agribus. Worldwide vol. 16, no. 3, p. 18+
[2] Ningsih R 2015 Analisis Nilai Sektor Basis Perkebunan Kelapa-Dalam (Cocos nucifera L) di Kabupaten Tanjung Jabung Timur Provinsi Jambi,” Pros. Semin. Nas. Lahan Suboptimal
[3] Gro Intelligence 2016 Cuckoo for Coconuts: Demand Is Soaring, but Production Isn’t Keeping Up Available: https://gro-intelligence.com/insights/articles/coconuts-growing-demand-stagnant-production. [Accessed: 16-Aug-2019]
[4] Bourdeix R and Prades A 2018 A Global Strategy for the Conservation and Use of Coconut Genetic Resources 2018-2028. Montpellier, France: Cogent.

[5] Pirmansah A 2017 Coconut Statistical Yearbook 2016 Jakarta, 2017

[6] Crone L K, Haynes R W, and Reyna N. E. 1999 Different perspectives on economic base USDA For. Serv. - Res. Note PNW-RN, no-538, pp. 1–21

[7] Nainggolan H L and Aritonang J 2006 Pengembangan pertanian berbasis komoditi unggulan dalam rangka pembangunan berkelanjutan (Studi Kasus Kabupaten Humbang Hasundutan) Seminar Nasional Tantangan Pembangunan Berkelanjutan dan Perubahan Iklim di Indonesia Program Studi Magister dan Doktor Pengelolaan Sumberdaya Alam USU pp. 1–10

[8] Rupasingha A and Patrick J M 2009 Tools for Understanding Economic Change in Communities: Economic Base Analysis and Shift-Share Analysis Coop. Ext. Serv. Coll. Agric. Consum. Environ. Sci., Circular 6, pp. 1–8.

[9] New York State Department of Labor 2017 Location Quotients: A Statewide and Regional Analysis June, 2017

[10] Hendayana R 2003 Aplikasi Metode Location Quotient (LQ) dalam Penentuan Komoditas Unggulan Nasional,” J. Inform. Pertan., vol. 12, no. March, pp. 1–21

[11] Sainani K L 2014 Introduction to principal components and factor analysis Journal of Injury, Function, and Rehabilitation 2014, vol. 6, no. 3, pp. 275–8.

[12] Shlens J 2014 A Tutorial on Principal Component Analysis 2014

[13] Tharwat A 2016 Principal component analysis - a tutorial Int. J. Appl. Pattern Recognit. vol. 3, no. 3, p. 197

[14] Bartholomew D J 2010 Principal components analysis Int. Encycl. Educ., vol. 2, pp. 374–377, 2010.

[15] Phan T 2016 An Introduction to Principal Component Analysis with Examples in R http://www.octoberraindrops.com/publications/Phan_2016_IntroductionToPCA.pdf

[16] Eshghi J, A, Haughton D, Sayek S, Teebagy N, and Topi H 2003 Geography Matters: Kohonen Classification of Determinants of Foreign Direct Investment in Transition Economies J. Bus. Strateg., vol. 20, no. 1, p. 23

[17] Melody Y K and Ajith K 2001 An Evaluation of Self-Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications Inf. Syst. Res., vol. 12, no. 2, pp. 177–194

[18] Cottrell A M, Olteanu M, Rossi F, and Villa-vialaneix N 2018 Self-organizing maps , theory and applications Investig. Operacional, vol. 39, no. 1

[19] Curry B, Davies F, Evans M, Moutinho L, and Phillips P 2018 The kohonen self-organising map as an alternative to cluster analysis: an application to direct marketing Int. J. Mark. Res., vol. 45, no. 2, pp. 1–20