International Journal of Business and Management Invention (IJBMI)
ISSN (Online): 2319-8028, ISSN (Print):2319-801X
www.ijbmi.org || Volume 9 Issue 11 Ser. II || November 2020 || PP 20-25

Mining Future Readiness for Asian Countries by Using World Digital Competitiveness Data

Dian-Fu Chang[1], and Chia-Chi Chen[2]
Corresponding Author:Chia-Chi Chen.
[1]Graduate Institute of Educational Policy and Leadership, Tamkang University, Taiwan
[2]Doctoral Program of Educational Leadership and Technology management, Tamkang University, Taiwan

ABSTRACT: *This study selected 20 Asia countries' with future readiness ranking as targets to mining their patterns and differences. The future readiness ranking consists of two sub-factors: adaptive attitudes and IT integration. This study applied hierarchical clustering with Minitab to determine the optimal clusters for these countries. The dendrogram with three clusters drew by Ward linkage and Euclidean distance has a relatively high similarity level and a relatively low distance level in this study. The result reveals the selected Asia countries in cluster1 and cluster2 are considered the well performance of countries' future readiness. The countries in cluster3 might reveal their possible weaknesses of the digital competitiveness for future readiness. This study suggests cluster analysis can be used to explore the data with multiple dimensions and perspectives by way of their similarity and distance estimation.*

KEY WORD:*Cluster analysis, Data mining, World Digital Competitiveness Ranking (WDCR), Future Readiness*

Date of Submission: 07-11-2020 | Date of Acceptance: 20-11-2020

I. INTRODUCTION

Data mining is an essential process where intelligent methods are applied to extract data patterns. Data mining can also be applied to different types of data such as data streams, ordered or sequence data, graph or networked data, spatial data, text data, multimedia data, etc. (Brusco et al., 2017;Han et al., 2011) The data mining process is becoming more widespread nowadays and be more practical such as companies applying data mining to gaining a competitive edge. There are a variety of terms used to describe the data mining process, including analytics, predictive analytics, big data, machine learning, and knowledge discovery in databases. These process terms have the objective of mining actionable golds of knowledge from large data sets (Brusco et al.,2017; Feger, 2015; Han et al., 2011; Larose and Larose; 2014). Data mining has been attracting a significant amount of research; for instance, in the business field, the goal of data mining allows a corporation to improve its marketing, sales and customer support operations through a better understanding of its customers (Alsultanny, 2013; Rygielski et al., 2002) The algorithms of data mining require multiple passes over huge quantities of data. Meanwhile, data warehousing brings together data from many different sources in a common format with consistent definitions for keys and fields. On the other hand, the data warehouse can be designed exclusively for decision support. There are various data warehousing in the digital age, while with deepening mining processes are persisted issues in the field. In this study, we explore the notion of data mining to explore a specific research target by using a building data warehousing.

This study focuses on the data set collected by the International Institute for Management Development (IMD). The IMD World Digital Competitiveness Ranking (WDCR) measures the capacity and readiness of 63 economies in 2019 to explore digital technologies as a key driver for economic transformation in business, government and wider society (Brits and Cabolis, 2019). The digital competitiveness is designed to exam technology not only affects how businesses perform but also how economies function and prepare for the future. The ranking data has been accepted by public generally; however, the data is hard to tell how wide the discrepancy existed among the countries. An initial study is to exam 20 selected Asian countries as the research target and applied the notions of data mining to explore their future readiness issues. Facing with uncertainty future, the estimation of future readiness bas become a critical factor to achieve success. In the sense, this study takes the IMD's future readiness ranking data as an example 20 Asian countries as an example to tackle the meanings of their discrepancy. This study also applies the concept to know the issues future readiness of the Asian countries to know the meaning of correlations patterns and trends for further interpretation.

Cluster is often considered as a preliminary step in a data mining process. Cluster analysis can be used as an independent data mining task to disclose intrinsic characteristics of data, or as a preprocessing step with the clustering results used further in other data mining tasks(Brusco et al.,2017; Feger, 2015; Han et al., 2011;

Larose and Larose; 2014). The clustering has been used in numerous applications in many disciplines and seeks to segment the whole data set into relatively homogeneous subgroups or clusters (Alsultanny, 2013; Rygielski et al., 2002). This study selected 20 Asian countries to tackle the cluster result to interpret the WDCR's future readiness. Specifically, the purposes of this study are as follows:

a) to restructure the IMD World Digital Competitiveness among for Asia countries;

b) to examine fitted cluster pattern with future readiness for Asian countries and interpret their future readiness.

Given these purposes, the structure of this paper will present as follows: First, the method section provides a brief description of the data structure and research method. Second, the result of cluster analysis will be displayed. Finally, conclusions are drawn.

II. METHOD

This study conducted cluster analysis with the World Digital Competitiveness data set to deepen the meanings of future readiness among 20 Asian countries. First, we prepare the data and examine the patterns of data or with specific trends. Then, we check the main impact factors which could reflect the meaning of future readiness among 20 Asian countries. For transforming the data's reason, we specify the future readiness with two sub-factors, namely adaptive attitudes and IT integration.

2.1 Definition of World Digital Competitiveness

World Digital Competitiveness Ranking (WDCR) has presented the 2019 overall rankings for 63 economies covered by the World Competitiveness Yearbook 2019 (WCY). The rankings are calculated on the basis of the 51 ranking criteria (31 Hard and 20 Survey data) (Brits and Cabolis, 2019). To evaluate an economy aptitude to adapt and explore digital technologies, WDCR examines three the following domains: knowledge, technology and future readiness. In WDCR framework, the first domain is "knowledge" which measures the capacity to understand and learn the new technologies. The second domain is "technology" which evaluates the competence of an economy to develop new digital innovations. The third domain is future readiness that assesses the preparedness for the coming developments. In 2019, The WDCR indicated several Asian economies advanced significantly in the ranking with respect to 2018. For example, Hong Kong and South Korea enter overall ranking of the top ten, while Taiwan and China move up to the 13th and 22nd position. Moreover, each domain of rankings can also provide a more detailed examination of specific aspects of the digital transformation (see Figure 1). This study selects 20 Asian countries as research target. The 2019 World Digital Competitiveness' future readiness ranking has been presented in Table 1.

Figure 1: World digital competitiveness' framework with future readiness and sub-factors

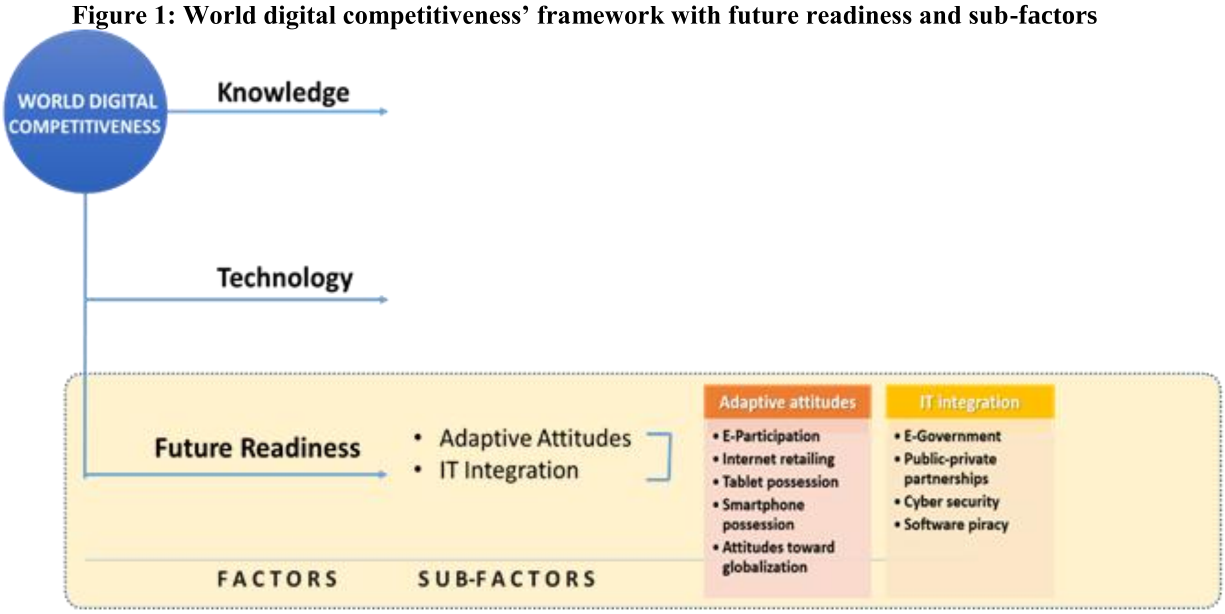


Table 1: World digital competitiveness of future readiness 2019 scores and ranking among 20 Asian Countries

	Country	Future Readiness %	OverallRanking	Adaptive attitudesRanking	IT IntegrationRanking
1	Korea Rep.	89.662	4	4	21
2	UAE	87.626	9	20	8
3	Singapore	86.407	11	19	4
4	Taiwan	85.556	12	14	24
5	Hong Kong SAR	84.23	15	12	22
6	Israel	81.816	19	21	16
7	China	80.743	21	24	41
8	Qatar	78.612	22	18	27
9	Japan	77.347	24	15	18
10	Malaysia	71.509	28	30	33
11	Kazakhstan	63.595	35	39	46
12	Saudi Arabia	61.95	38	50	30
13	Turkey	57.567	41	38	48
14	Russia	56.539	42	40	43
15	India	54.946	46	54	56
16	Thailand	52.864	50	58	51
17	Jordan	52.352	52	61	54
18	Philippines	52.093	54	53	58
19	Indonesia	48.166	58	60	60
20	Mongolia	42.936	61	31	62

Source: World Digital Competitiveness Ranking, 2019

2.2 Logic of cluster analysis

Cluster analysis is available for classifying objects on the basis of their similarities or dissimilarities (Favero and Belfiore, 2019; Rodrigues et al., 2019). Typically, cluster analysis includes hierarchical methods, partitioning methods and methods that allow overlapping clusters. Cluster analysis represents a set of very useful exploratory techniques also called segmentation analysis or taxonomy analysis for partitions sample data into groups or clusters (Favero and Belfiore, 2019; Pavel, 2006). There are a lot of literature on electric load profile segmentation using cluster analysis and mostly focuses on the empirical findings of papers that have used this approach (Fegar, 2015; Favero and Belfiore, 2019; Pavel, 2006; Tryon, 1939). The related analyses can be found in a variety of fields, such as business, education, psychology and social science (Alsultanny, 2013; Pavel, 2006; Rodrigues, 2019; Rygielski et al., 2002). The most frequently applied classification procedures are hierarchical was mentioned in social sciences application. Clusters are calculated by similarity or distance of objects with different characteristics. K-means is the number of clusters K observations from dataset randomly and assign it as each observation (Favero, 2019; Pavel, 2006). Conducting cluster analysis, we follow the following steps: First, selecting the data; Then, hierarchical clustering with Minitab statistic package to determine the clusters. Basic cluster algorithms are as follows (Berkhin, 2006; Chang and Chen, 2018; Kogan et al., 2006; Rygielski et al. 2002; Sinharay, 2010; Wittek, 2014):

- Select *k* point as initial centroids,
- Repeat,
- From *k* clusters by assigning each point to its closest centroids,
- Re-compute the centroids of each cluster,
- Until centroids do not change.

Typically, hierarchical clustering starts with k = N clusters and proceeds by combining the two closest observation into one cluster. Each clustering groups data can be tell from a cluster tree or dendrogram (Wittek, 2014). The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. The dendrogram is represented like tree and by relatively short branches. Based on the dendrogram, it was constructed using the complete linkage approach suggests the presence of number of subgroups and displays both the cluster and sub-cluster relationships (Berkhin, 2006; Sinharay, 2010; Wittek, 2014) This study decides the level or scale of clustering that is most appropriate for the data application. The Ward method considers the minimum variability as the criterion for merging to form the within-cluster sum of squares is minimized. It indicates that the similarity within the group is high. The Ward method was used to transform the data according to the following format (Berkhin, 2006; Chang and Chen, 2018; Sinharay, 2010;):

$$d_{A.B} = n_A \|\bar{x}_A - \bar{\bar{x}}\|^2 + \; n_B \|\bar{x}_B - \bar{\bar{x}}\|^2$$

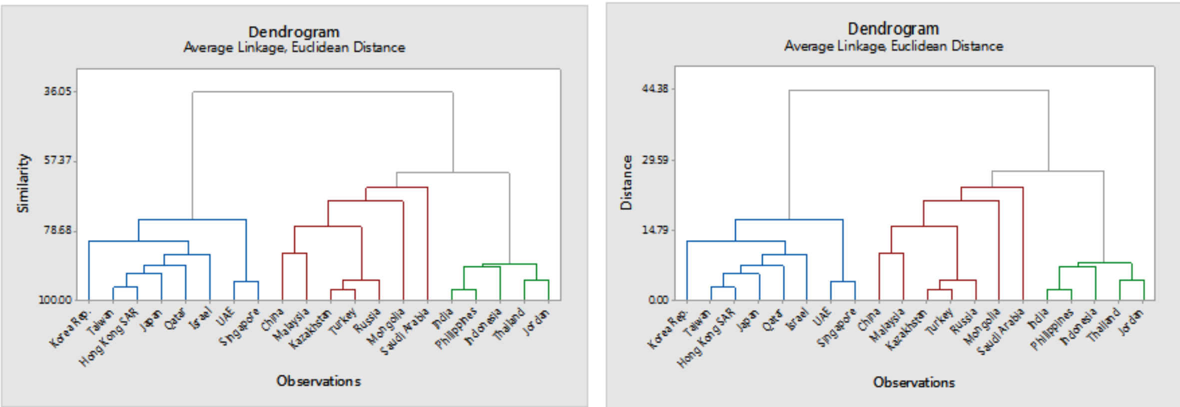
$d_{A.B}$indicates the calculated distance between A and $B.n_A$ and $n_B$represent number of variables in cluster A and B. Assuming that $\overline{x_A}$ and $\bar{x}_B$represent the indicators vector for talents, $\overline{x_A}$ and $\bar{x}_B$ in cluster A and B, and $\bar{\bar{x}}$ the centroid of cluster A or B, in other words, which will calculate the minimum distance squared of $\|\bar{x}_A - \bar{\bar{x}}\|^2$ and $\|\bar{x}_B - \bar{\bar{x}}\|^2$.

III. RESULTS

3.1 Demonstration of cluster analysis

Clustering methods aim to display the hierarchy of data samples using a so-called dendrogram. The key outputs of cluster observations analysis include the similarity and distance values, the dendrogram, and the final partition. The result reveals the higher the similarity level, the more similar the observations are in each cluster. The lower the distance level, the closer the observations are in each cluster. Additionally, the clusters should have a relatively high similarity level and a relatively low distance level. In this study result represent the dendrogram was constructed using the Ward linkage approach suggests the presence of three groups. The dendrogram with these three clusters drew by Ward linkage and Euclidean distance has a relatively high similarity level and a relatively low distance level, see Figure 2. The dendrogram displays with the groups in different colors by Minitab.

Figure 2:Three clusters among 20 Asian countries' future readiness



This dendrogram was created using a final partition of 3 clusters. The cluster1 (far left) is composed of eight observations (Korea Rep., Taiwan, Hong Kong SAR, Japan, Qatar, Israel, UAE, Singapore). The cluster2, shows in the middle, is composed of seven observations (China, Malaysia, Kazakhstan, Turkey, Russia, Mongolia, and Saudi Arabia). The cluster3 consists of five observations (India, Philippines, Indonesia, Thailand, and Jordan). After determining the final groupings, the study displays the final partition in Table 2, Table 3, and Table 4 which show the characteristics of each cluster's centroids and distances.

Table 2: Final partition of 20 Asian Countries' Future Readiness

Clusters	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	8	655.88	8.1271	13.9782
Cluster2	7	1097.43	10.8531	19.3707
Cluster3	5	99.60	4.4133	5.0478

Table 3:Cluster centroids for 20 Asian Countries' Future Readiness

Variables	Cluster1	Cluster2	Cluster3	Grand centroid
Adaptive attitudes	15.375	36.0000	57.2	33.05
IT integration	17.500	43.2857	55.8	36.10

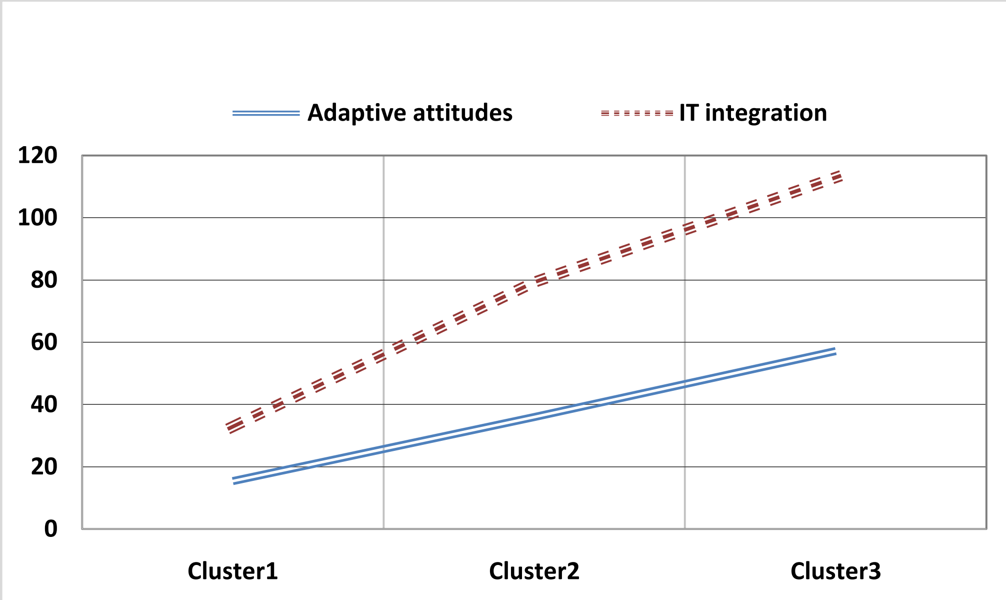
Table 4: Distances between cluster centroids

Clusters	Cluster1	Cluster2	Cluster3
Cluster1	0.0000	33.0196	56.7117
Cluster2	33.0196	0.0000	24.6180
Cluster3	56.7117	24.6180	0.0000

3.2 Patterns of future readiness in Asia

Future readiness reflects an economy that shows the level of country preparedness to exploit digital transformation and integration of digital technologies in the economy. By using the ranking data to compute the cluster analysis shows three clusters grouping. Cluster1 is composed of eight observations countries include Korea Rep., Taiwan, Hong Kong SAR, Japan, Qatar, Israel, UAE and Singapore. The first group indicates the best performance for future readiness in Asian countries. Moreover, Cluster1 reflects the same performance in the original overall ranking and sub-factors of adaptive attitude and IT integration. The countries in first group are well performance on the top ranking among 20 Asian countries. The cluster2 shows the middle performance with seven countries: China, Malaysia, Kazakhstan, Turkey, Russia, Mongolia, and Saudi Arabia. The result reveals Mongolia is lower-ranked countries; however, the detailed adaptive attitudes have the middle-ranking among 20 Asian countries. As for the last group, the cluster consists of five observations: India, Philippines, Indonesia, Thailand, and Jordan. The third cluster is identified as the possible weaknesses for the digital competitiveness for future readiness. Overall, by observing the cluster centroids distances can also show the results evenly distributed as the same result in future readiness overall ranking in Figure 3.

Figure 3:Three clusters among 20 Asian countries' future read



IV. CONCLUSIONS

There are various data set in well-design data bank. WDCR is a good example providing by IMD to approaching future readiness purpose. As previous discussion, digital competitiveness ranking can the objective to assess and the extent to which a country adopts and explores digital technologies leading to transformation in government practices, business models and society in general, while cluster analysis can be extended their meanings in the target countries. This study focuses future readiness and transforms the ranking data set to cluster format. Cluster evaluation can be used to determine the optimal number of clusters for the data using different evaluation criteria in diverse settings. The result of mining future readiness reveals what is possible in digital competitiveness for Asian countries to reveals improving reference. For further studies, cluster analysis can provide different formats for well-prepare databank. Future readiness is a potential topic for mining practices for related digital competitiveness issues. We suggest set a research agenda for specific organizations to mining their future readiness for related policy makers.

BIBLIOGRAPHY

[1]. Alsultanny, Y. A. (2013). Labor market forecasting by using data mining. *Procedia Computer Science*, *18*, 1700-1709.

[2]. Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin, Heidelberg.

[3]. Brits, A., & Cabolis, C. (2019). IMD World Digital Competitiveness Ranking 2019. *IMD World Competitiveness Center*, *180*.

[4]. Brusco, M. J., Singh, R., Cradit, J. D., & Steinley, D. (2017). Cluster analysis in empirical OM research: survey and recommendations. *International Journal of Operations & Production Management*.

[5]. Chang, D. F., & Chen, C. C. (2018). Cluster analysis for student performance in PISA2015 among OECD economies. *ICIC express letters. Part B, Applications: an international journal of research and surveys*, *9*(11), 1139-1146.

[6]. Favero, L. P., & Belfiore, P. (2019). *Data science for business and decision making*. Academic Press.

[7]. Feger, H. (2015). Classification, Understandings of in the Social Sciences.

[8]. Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

[9]. Kogan, J., Nicholas, C., & Teboulle, M. (2006). *Grouping multidimensional data* (p. 5). Springer-Verlag Berlin Heidelberg.

[10]. Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.

[11]. Pavel, B. (2006). Grouping Multidimensional Data.

[12]. Rodrigues, A., Rodrigues, G. N., Knauss, A., Ali, R., & Andrade, H. (2019). Enhancing context specifications for dependable adaptive systems: A data mining approach. *Information and software technology*, *112*, 115-131.

[13]. Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in society*, *24*(4), 483-502.

[14]. Sinharay, S. (2010). An overview of statistics in education.
[15]. Tryon, R. C. (1939). Cluster analysis: correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality. Edwards brother, Incorporated. *Ann Arbor*.
[16]. Wittek, P. (2014). *Quantum machine learning: what quantum computing means to data mining*. Academic Press.

Dian-Fu Chang. "Mining Future Readiness for Asian Countries by Using World Digital Competitiveness Data." *International Journal of Business and Management Invention (IJBMI)*, vol. 09(11), 2020, pp. 20-25. Journal DOI- 10.35629/8028