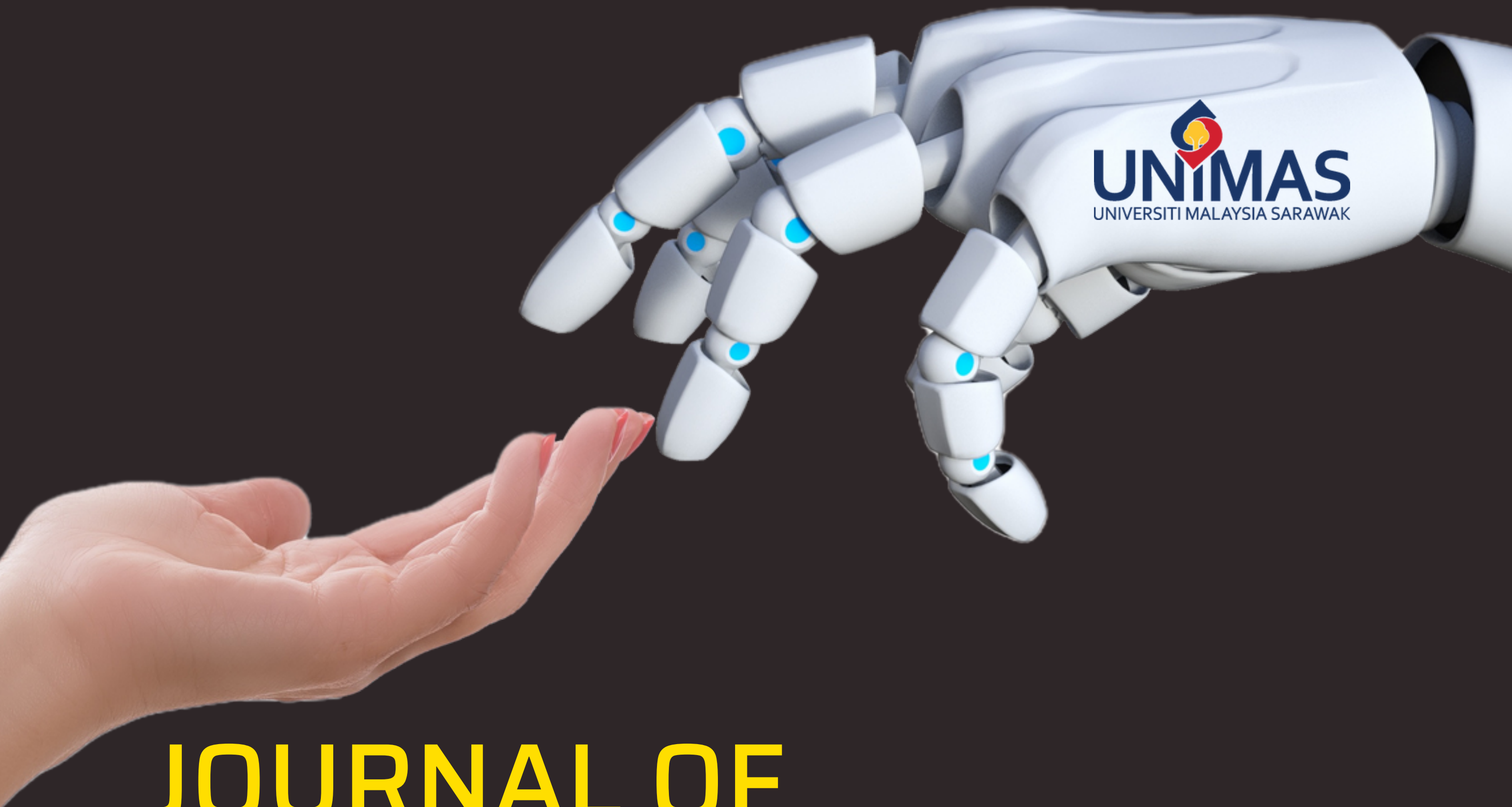

JOURNAL OF COMPUTING AND SOCIAL INFORMATICS

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITI MALAYSIA SARAWAK

eISSN 2821-3777

Journal of Computing and Social Informatics

The Journal of Computing and Social Informatics (JCSI) is an international peer-reviewed publication that focuses on the emerging areas of Computer Science and the overarching impact of technologies on all aspects of our life at societal level. This journal serves as a platform to promote the exchange of ideas with researchers around the world.

Articles can be submitted via *www.jcsi.unimas.my*

Assoc Prof Dr Chiew Kang Leng

Chief Editor
Journal of Computing and Social Informatics
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak, Malaysia

Contents

Open Government Data from the Perspective of SMEs: A Case Study in Indonesia

[1*,2]Amirudin Syarif, [2]Mohamad Aizi bin Salamat, [2]Aida binti Mustapha, [3]Supli Effendy Rahim

[1]Department of Management and Information System, Universitas Bina Darma, Jln. Ahmad Yani 3, Palembang 30000, Indonesia.
[2]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia.
[3]Universitas Kader Bangsa, Post Graduate Program, Jl. Mayjend. H.M Ryacudu No. 88, Palembang 30000, Indonesia.

email: [1*]amirudinsyarif@binadarma.ac.id, [2]aizi@uthm.edu.my, aidam@uthm.edu.my, [3]suplirahim31760@gmail.com

*Corresponding author

Received: 26 July 2021 | Accepted: 01 October 2021 | Early access: 15 November 2021

Abstract - *The Indonesian government conducts Open Government Data (OGD) through the development of a data portal (data.go.id) called Satu Data Indonesia (SDI) as part of an open government initiative. Numerous studies on OGD and its effect on SMEs business in a variety of countries have demonstrated that it has a positive effect on SME business growth, implying that OGD is critical and can benefit implementing countries such as Indonesia. The government must address several issues, including the use and benefits of data made available to stakeholders. Micro, Small, and Medium-Sized Enterprises (MSMEs) employ approximately 97 percent of the total workforce in Indonesia, and account for 99.9 percent of all businesses. MSMEs contribute roughly 60% of Indonesia's total GDP. As a pillar of the Indonesian economy, MSMEs must be considered when it comes to data availability that meets their needs. That is why it is critical to conduct this research to gain their perspective. The purpose of this article is to examine only the perspectives of Indonesian SMEs on the open data made available by the Indonesian government. Based on the data analysis findings, it is possible to conclude that the SMEs society has a demand for open data in terms of the existence of agency mechanisms for receiving and responding to data requests. Additionally, the Open Data Ecosystem is critical for SMEs in terms of government promotion of data reuse.*

Keywords: OGD, Satu Data Indonesia, SMEs business, MSMEs, Open Data Ecosystem.

1 Introduction

Open government data (hereinafter referred to as OGD) is a critical component of government organizations in almost every country in the world, including monarchies and communist countries. The OGD began with President Obama's 2009 inaugural address (Hossain et al., 2015; Okamoto, 2016; Srimuang et al., 2017; Kubler et al., 2018; Zuiderwijk et al., 2019; Syarif et al., 2020; Enriquez-Reyes et al., 2021), implying that the OGD is Despite this, its development is accelerated and results in a variety of benefits. As evidenced by the following research articles: 1) Discussing the benefits of OGD on innovation (Bedini et al., 2014); 2) Discussing the benefits of OGD in Sustainable Development (UNITED NATIONS, 2017); and 3) Discussing the benefits of OGD in a variety of ways (Safarov et al., 2017).

Gomes and Soares (2014) conducted a study on the implementation of OGD in Europe with the goal of analysing and identifying differences in how northern and southern European countries adopted, accepted, and promoted open government data portals. The results of the direct content analysis observation indicate that there are some current differences between the countries of the two regions, most notably in terms of their ability to reuse open data made available by public entities. At the moment, OGD is considered a desirable resource not only by citizens, but also by small, medium, and large businesses and other public and private organizations that view public data as a source of innovation and entrepreneurship, and thus the benefits of open data are increasingly recognized by a broad range of national and international organizations. According to research conducted by (Styrin et al., 2016), there are differences in the application and practice of OGD in Mexico, Russia, and the United States, with a particular emphasis on their respective policies. These distinctions are due to historical problems, policies, and politics that are context specific. Another article from (Saxena, 2017) discusses the enormous potential of OGD in facilitating the Gulf Cooperation Council's economic diversification initiatives (GCC). According to the author, open government data (OGD) is critical in facilitating the GCC region's economic turnaround because it fosters innovation and economic growth while also facilitating collaboration and participation among stakeholders.

The government of Indonesia carries out OGD by developing a data portal (data.go.id) under the name Satu Data Indonesia (SDI) as part of an open government initiative (Rahmatika et al., 2019), which is a significant starting point for the era of open Indonesian government. According to (Jacob et al., 2019), Indonesia, like other countries with obstacles to OGD practice, has problems related to obstacles to OGD implementation, namely structural barriers and obstacles to knowledge quality. According to (Syarif et al., 2020), the barriers to the practice of OGD in Indonesia are similar to those in countries that are members of the Association of Southeast Asian Nations because of their homogeneity, namely developing countries with a culture of openness, a degree of openness, and a similar level of confidence. The practice of OGD in Indonesia is at maturity stage 3, with a ranking ranging from 1 to 4. This means that while policies and procedures for implementing OGD are in place, their implementation is inefficient, and formalization of activities is still minimal (Rahmatika et al., 2019).

According to Huber et al. (2020), governments have decided to push for the use of open data due to its potential to catalyze digital innovation. Despite this, little research has been conducted on the role of open data in SME, in contrast to the growing literature on the collection and sharing of open data by the public sector. As a result, our research contributes to the field of open innovation by examining the critical capabilities required to successfully manage open data in SMEs. According to our findings, several critical factors influence the acquisition, assimilation, transformation, and exploitation of open data by SMEs. The findings indicate that without specific open data capabilities, SME adoption of open data will be limited, which may explain why open data adoption by SMEs in general has been limited thus far. If this 'raw material' for the digital economy is to be fully exploited, government must improve its development of these unique open data capabilities.

Several studies by researchers on OGD and its effect on SMEs business in various countries have shown that it has a positive influence on the progress of SMEs business, so it can be said that OGD is very important and can bring goodness to implementing countries such as Indonesia. This paper investigates the perspectives of Indonesian SMEs on the open data provided by the Indonesian government on the Indonesian One Data Portal (SDI), www.data.go.id.

2 Literature Review

Discussing OGD is akin to debating a country's future. Data has become a very important material in this fast-paced era of information technology. This is due to the growing volume of data collected, which must be published by the government as part of its responsibility and as a means of providing useful information to the public and businesses. OGD ushers in an era of open government based on information and communication technology, which is thought to provide flexibility and speed in making effective and efficient decisions. Stakeholders can use the available data for free and reuse or distribute it.

2.1 Understanding and Vision of Open Government Data

The following are some expert opinions that can be cited as OGD definitions: 1) The underlying philosophy of Open Government Data is "to make data freely accessible to all, without restriction" (Kalampokis et al., 2011), 2) According to Saxena and Janssen (2017), governments throughout the world took the initiative to "open" their administrative data to the general public, making it freely accessible and re-usable by all. 3) Open government data is defined as publicly available public sector data that individuals or organizations can use for personal or organizational purposes (Talukder et al., 2018). On the basis of these three definitions, OGD is digital data that has been made available with the technical and legal characteristics necessary to enable it to be freely used, reused, and redistributed by anyone, at any time and from any location.

Users can sort open data definitions according to their involvement in the business processes of each stakeholder. For instance, not all stakeholders desire the same type of data. The ice trader's information requirements will differ from those of the police. This is consistent with the view that open government is a "multilateral, political, and social process characterized by transparent, collaborative, and participatory action by government and administration," as stated by Wirtz et al. (2017). Moreover, Geiger and Von Lucke (2012) define Open Data as "stored data that can be made freely accessible in the public interest."

OGD is a global phenomenon aimed at making government data publicly and freely available in digital formats for use, reuse, and redistribution, according to Zhu (2017). OGD and open data are terms that are frequently used interchangeably. Even though it is significantly different. The term OGD refers to official data issued by the government, whereas open data refers to business data made available to the public. As stated by Gonzalez-Zapata and Heeks (2015), OGD can have several meanings. Because of the influence of stakeholders' interests on data, it has a lot of meaning.

The government's provision of data is an example of a public service. This is directly related to the government's responsibility in the public sector. According to Veljkovi et al. (2014), one of the factors impeding the development of Open Government (hereinafter referred to as OG) is a lack of conceptual clarity for OG. Thus, developing a conceptual model of Open Government will aid in better understanding the concept and guide the process of establishing benchmark indicators for OG evaluation. Governments should make data available in a reusable format that enables data indexing and retrieval without imposing restrictions on data retrieval. Given the widespread use and significance of government data, it must be comprehensive, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and license-free.

OGD is a component of electronic government, which is a form of honest, accountable, and dignified state administration. According to Purwanto et al. (2018), the policy of OGD practices in Austria, Greece, and England needs to take into account societal issues. According to Neuroni et al. (2013) research on OGD in Switzerland, OGD emphasizes encouraging transparency, collaboration, and participation. Their findings also show that executive authorities in Switzerland place a premium on the expansion of their core business and the potential for economic development. Another study Wirtz and Birkmeyer (2015) shows that the perception of government employees who are concerned about risks is a constraint to implementing OGD. This was identified as a significant constraint. Other significant barriers include legal barriers, authority hierarchies, a bureaucratic decision-making culture, and organizational transparency.

2.2 Open Government Data Indonesia

At the start of his tenure as Indonesia's seventh president, President Joko Widodo's vision and mission were Nawa Cita. One of the steps in Nawa Cita is to continue efforts to build a more open, participatory, and innovative government through the Open Government of Indonesia Commitments (OGI). This demonstrates the Indonesian government's commitment to reforming the country's public sector. OGI is expected to generate ideas, initiatives, and practices of government openness to all levels of society to foster collaboration and accelerate progress toward national goals and priorities. This will significantly improve the quality of public policy innovation and implementation to meet the needs of society.

Satu Data Indonesia (SDI), with the address https://data.go.id is Indonesia's OGD portal. The goal of this government data management policy is to produce high-quality data that is easily accessible and shared among Central and Regional Agencies (OECD Public Governance Reviews, 2016). This policy is outlined in Presidential Regulation No. 39 of 2019 on One Indonesian Data. The SDI Portal contains all government data as well as data from other relevant agencies. The Ministry of National Development Planning / Bappenas manages the Central Secretariat of One Data Indonesia to achieve government transparency and accountability while also supporting national development. As stipulated in Law 14 of 2008 on Freedom of Information, the entire data set in the One Data Indonesia Portal can be accessed freely and classified as public data, as long as it does not contain information containing state secrets, personal secrets, or other sensitive information. According to the OECD report 2016, in order to improve accountability and citizen engagement, Indonesia needs to: 1) foster a greater awareness of the relevance and usefulness of open governance reforms in public administration, 2) officials with the requisite capacity to enact reforms at the national and local levels of government, and a greater reliance on society, 3) Understanding civilians encourages the rise of more non-government actors capable of playing a constructive role in the open government agenda.

Sayogo and Yuli (2018) conducted research on the complexities of open government and the implementation of open data from the perspective of local government in Indonesia, discussing challenges, success factors, learning, and success indicators. This study discovered five major challenges in the practice of OGD: 1) data abuse and misuse, 2) limited technological capabilities, 3) data credibility, 4) availability of information policies to regulate transparency, and 5) maintaining public involvement and enthusiasm. There are four factors that will determine the success of OGD: 1) collaboration between the government, academia, the private sector, and the general public, 2) openness of government offices to accept criticism and suggestions, 3) accommodating leaders, and 4) commitment of government agencies to be involved in open government and open data.

According to research conducted by Syaripul and Bachtiar (2016) on the relationship between UKM and OGD in Jakarta, which is the capital of the Republic of Indonesia, shows that 78.05% of respondents cannot take insights from the data, and only 21.95% are able. The results of this study also indicate that respondents 100% agree that the OGD of the Jakarta City Government is still difficult to read and understand.

Referring to the OGD of the Indonesian government from the SDI portal, it can be said that OGD is still far from ideal expectations. The data.go.id portal (see figure 1a, 1b, and 1c) accessed on March 25, 2021 is still in Beta and only in the Indonesian version and there is no English version yet. The available dataset is 94,588. The number of datasets is still very minimal considering that this program is part of President Joko Widodo's Nawa Cita, which was launched in 2014. From several available datasets, a search using the keyword "UMKM" means that “MSMEs” only produces 165 datasets (see figure 1b), whereas if the keyword is used "Business", only produced 27 datasets (see figure 1c).

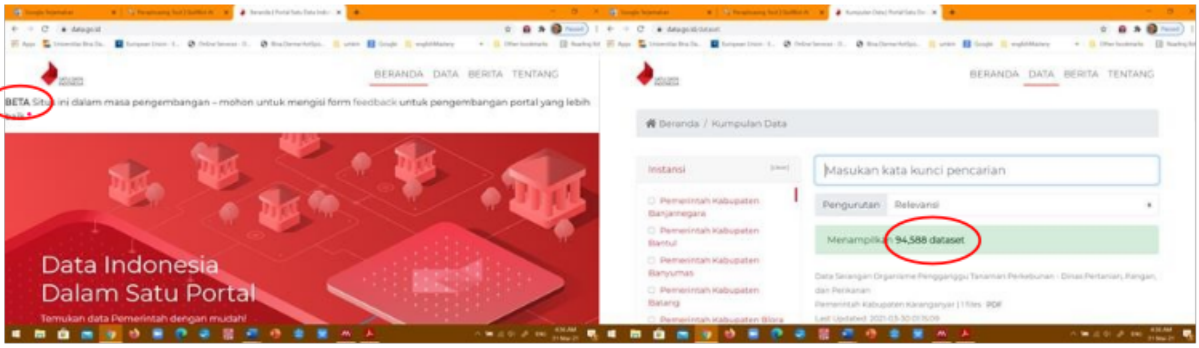

Figure 1a: The view from the portal https://data.go.id

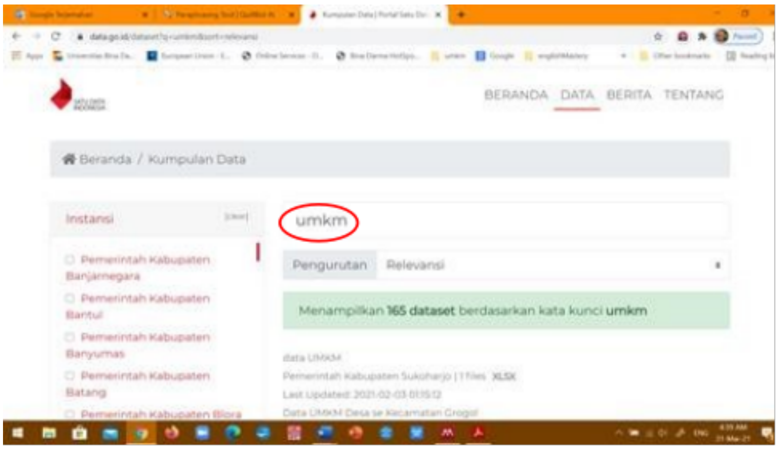

Figure 1b: Searching using MSMEs keyword

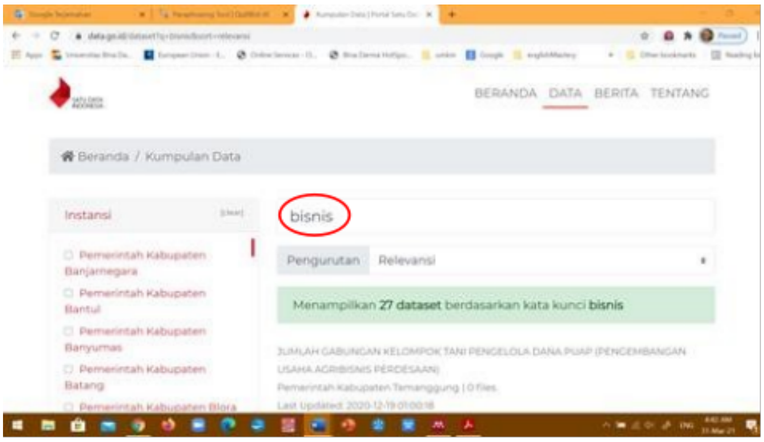

Figure 1c: Searching using Business keyword for MSMEs and Business respectively

OGD is a component of the country's investment in information and communication technology (ICT), which is believed to have the potential to grow the economy. According to Rachman (2017), determining the future potential for ICT investment is contingent upon the relationship between the adoption of ICT services by small to medium-sized enterprises (SMEs) and national economic growth. ICT services have a significant impact on SME productivity growth, which benefits the Indonesian economy. OGD is critical for society and business in Indonesia, particularly SMEs.

2.3 Chanel for Data and Information Sharing

OGD is a method of delivering open data to stakeholders. All countries that are members of open government initiatives (OGI) have joined the open data movement. According to (Parung et al., 2018), there are five types of barriers in order of importance: 1) law and privacy; 2) government culture; 3) social; 4) technical; and 5) economy. There are five priorities: 1) Priority 1 is to involve stakeholders in OGD preparation and form an OGD competency center, 2) Priority 2 is to establish a legal enforcement structure, 3) Priority 3 is to implement OGD in phases, 4) Priority 4 is to build collaboration features on the platform for communication and raising public awareness of OGD stakeholders, and 5) Priority 5 is to conduct training.

Indonesia is a prime example of a country that is only now embracing the open data movement. Nugroho (2013) reports that data are published for a variety of reasons, including public reuse, more efficient governance, and increased transparency. The most recent development in this area is the expectation that published data will be in a machine-readable format. There is, in general, a dearth of guidelines to assist in organizing and facilitating the data opening process. The comparison of open data policies was conducted to serve as a foundation for drawing conclusions and making recommendations about Indonesia's open data policy. The report makes several recommendations, including strengthening the legal framework, establishing an ecosystem between data publishers and users, enhancing information technology and organizational support for open data, and launching government-level initiatives that leverage open data.

The steps taken to implement the OGD, based on Guidelines on Open Government Data for Citizen Engagement published by United Nations, (Data, 2018) are as follows:

1. Make government data and information available online. All published data and information must be available online, the information uploaded must be in a format that the public can easily access and reuse based on their needs, the government must provide the public with the opportunity to provide feedback on the data and information provided, and the government is required to respond to any requests and notes made by the community.
2. Build and establish government institutions with an open culture. Openness is a culture that encourages the government to be more transparent and accountable. Things that must be done include: a) the publication of plans made by the government that can provide an overview to the public of the government's efforts to improve and develop a culture of transparency, collaboration, and public participation; and b) following the implementation stage, the evaluation stage to the application of the principles and the government's response. The results must also be published. c) Promoting transparency, collaboration, and public participation through innovative ideas.
3. Developing a policy framework. This policy framework is used to provide clear direction and guidance

for the long-term sustainability of open government implementation by allowing for collaboration with various parties. The goal is to increase government professionalism so that the implementation of open government can be properly developed in the future.

2.4 The Open Data Readiness Assessment (ODRA)

The World Bank has established a methodology to assist governments around the world in assessing and developing their Open Data programs. The Open Data Readiness Assessment, or ODRA, of the World Bank. It is an action-oriented assessment based on desk research and stakeholder consultation, according to (Government & Working, 2013) ODRA is intended to identify the steps required to launch an open data initiative. There are eight dimensions used, which are as follows: 1) At the highest level of government, leadership that evaluates the vision, understanding, and championing of Open Data; 2) A policy and legal framework that investigates how the state can be a legal supportive framework for the development of the Open Data Initiative. 3) government institutional structures, responsibilities, and capabilities that see how the government works horizontally and what the capacities of various institutions are usually very important in the implementation of the Open Data Initiative; 4) Data management policies, procedures, and data availability within the government that map existing data assets and data procedures; 5) Open data requests that assess open data awareness and initiatives among non-government actors, particularly NGOs, the private sector, academia, media and journalists, and startups and innovators; 6) Citizen engagement and open data capabilities that assess the state of interaction between government and nongovernment actors, the state of the country's information society, and overall societal capacity in ICT. 7) Fund open data programs that investigate whether budgets are available for the development of Open Data initiatives; 8) National technology and infrastructure skills that assess the state of the country's IT infrastructure.

This is driving the implementation of new Open Data initiatives, such as those initiated by the countries listed below: ODRA has been implemented in Moldova (Rahemtulla et al., 2012), Indonesia (Alonso et al., 2013) Kyrgyzstan (Zijlstra, 2015), Croatia (Vracic et al., 2016), Malaysia (World Bank Group, 2017), Uganda (Chrzanowski et al., 2017), Ethiopia (Boyera et al., 2017), Bahrain (Katbi & Al-ammary, 2019), Sierra Leone (World Bank Group, 2020b), Bangladesh (World Bank Group, 2020a), and other countries.

2.5 Definition of Small and Medium-sized Enterprises (SMEs)

According to the Organization for Economic Cooperation and Development (OECD, 2005), small and medium sized enterprises (SMEs) are self-contained, non-subsidiary businesses with fewer than a specified number of employees. This figure varies considerably by country. As is the case in the European Union, the most frequently used upper limit for SME designation is 250 employees. Certain countries, on the other hand, impose a limit of 200 employees, whereas the United States defines SMEs as businesses with fewer than 500 employees. Small businesses typically employ fewer than 50 people, while microbusinesses employ no more than ten, and in some cases, five. Additionally, SMEs are defined by their financial assets. In the European Union, a new definition took effect on January 1, 2005. It applies to all Community acts and funding programs, as well as State aid, and allows SMEs to receive a greater intensity of national and regional aid than large corporations. The new definition raises the financial thresholds, requiring medium-sized enterprises (50-249 employees) to have a maximum revenue of EUR 50 million, small enterprises (10-49 employees) to have a maximum revenue of EUR 10 million, and micro enterprises (less than 10 employees) to have a maximum revenue of EUR 2 million. On the other hand, the balance sheets of medium-sized, small, and micro enterprises should not exceed EUR 43 million, EUR 10 million, and EUR 2 million, respectively.

According to Capri's snapshot of Indonesian SMEs, approximately 57 million micros, small, and medium-sized enterprises (MSMEs) operated in 2017. MSMEs employ approximately 97 percent of the total workforce in Indonesia and account for 99.9 percent of all businesses. MSMEs contribute roughly 60% of Indonesia's total GDP. The following table summarizes data on Indonesia's small and medium-sized enterprises (SMEs).

Table 1: Indonesian SMEs Asset Classification

Business Type	Asset (IDR)
Micro	Max 50 million
Small	50-500 million
Medium	500 million

Notes: 1 USD equal to 14,500 Indonesian Rupiah (IDR) as of March 2021.

Micro, small, and medium-sized businesses have been forced to transform into digital businesses as a result of the Covid 19 pandemic, with available data indicating that 4.8 million MSMEs have gone digital as of March 2021 (https://www.liputan6.com/bisnis/read/4544531/48million-umkm-have-gone-digital-in-march-2021).

The evolution of business culture from traditional to digital has had an effect. It is unknown whether it has had a positive or negative effect on the growth of MSME businesses in Indonesia. According to the author, what may occur is a hybrid business in which two business platforms operate concurrently.

According to (Albats et al., 2019) who noticed that the majority of the cases examined corroborated an assumption about the triggers that preceded SME innovation processes. Simply put, all of the cases studied demonstrated an innovation process that was triggered by either internal factors (founders' ideas, intelligence, and entrepreneurial mindset), external factors (market demand and emerging opportunities, market turbulence and crises), or a combination of the two. While entrepreneurs initially sensed and scoped market opportunities and initial business strategies (internally), external knowledge sources were also utilized during these early stages. So, it can be said that this Covid pandemic may also be a business trigger.

3 Methodology

Information systems is a multidisciplinary field of study. Mathematics, management, natural science, engineering, linguistics, and behavioral sciences all contribute to its success (Myers, 1997). As a result, determining and selecting appropriate research methodology is difficult.

This study is based on mixed methods research (Ojeda & Juárez-Cerrillo, 1996; Yin, 2009; Flick, 2013; Antwi & Kasim, 2015; Alismaili et al., 2015; Creswell, 2018), specifically quantitative and qualitative research, as well as a case study. The researchers used exploratory methods for qualitative research and explanatory methods for quantitative research. Qualitative research is used to elucidate attitudes, underlying causes, and motivations (Oktaba & Piattini, 2008). It sheds light on the problem or aids in the formulation of concepts or hypotheses. According to (Kaplan & Duchon, 1988), Quantitative research elucidates a problem through generalizable findings; it does not prioritize data depth, whereas analysis prioritizes data breadth. Quantitative research employs the Likert scale as a measure of the indicator variable under study for qualitative method interpretation. The study examined 30 SMEs in Palembang, Indonesia. The questionnaire distributed to respondents is in the form of closed questions with response options ranging from 1 to 5, indicating a level of agreement or disagreement. A Likert scale is a numerical scale ranging from 1 to 5. The question item is a translation of the ODRA dimensions and indicators into the languages of the countries that have adopted them (World Bank Group, 2020a). SmartPLS software is used to process data. The data processing results are presented descriptively, utilizing respondent data tabulation and PLS analysis to illustrate the importance ranking of the variables. The purpose of this paper will elaborate on SME actors' perspectives on the OGD.

4 Results and Discussion

The data was gathered by handing out questionnaires to 30 SMEs actors in Palembang. The criteria for respondents are small businesses with 0 to 20 employees and assets ranging from IDR 50 to 500 million, and medium businesses with 21 to 100 employees and assets ranging from IDR 500 million to IDR 500 million. The questionnaire contains approximately eight dimensions and indicators, as shown in the table below, which are tailored to the needs of SMEs.

Table 2: Table of ODRA Dimensions and Indicators

Dimensions	Indicators
Leadership	(1) Leaders have expressed publicly visible support for OGD. (2) Support for OGD among key data-owning agencies. (3) The broader political context and top national priorities/plans help or hinder OGD.
Policy & Legal Framework	(1) Existence and effectiveness of an access to information law. (2) Privacy protection. (3) Systems security and archiving preservation. (4) Ownership and licensing of government data.

Government Institutional Structures & Skills for Data	(1) Expressed readiness of an agency with sufficient political weight and competency to lead on OGD for SMEs. (2) Track record of inter-agency mechanisms coordinating major ICT for OGD initiatives. (3) Existence and effectiveness of positions comparable to a CIO/CTO within agencies responsible for strategic ICT decisions and management.
Data of the ministry responsible for SMEs policies	(1) How and where data is held by the ministry. (2) The visibility of agencies into their data holdings. (3) The existence of key data-owning agencies with demonstrable capabilities in data management.
Demand for Open Data from SMEs Society	(1) Evidence of data demand by SMEs society, and the private sector. (2) Existence of agency mechanisms in place to intake and respond to requests for data.
Open Data Ecosystem	(1) Government record on SMEs engagement. (2) Existence of Business Apps. (3) Government promotion of reuse of the data.
Funding an Open Data Program	(1) Existence of resources and personnel for an Open Data Program. (2) Availability of government funding for necessary ICT infrastructure and training. (3) the government's track record for investing in innovation.
National Technology Infrastructure and Skills	(1) Overall ICT ecosystem and skills. (2) Access to high-speed Internet and mobile phones. (3) Maturity of the government's ICT infrastructure and use of technology, especially the use of shared infrastructure and services. (4) ICT literacy among the population of SMEs.

Three indicators are used for the intervening variable, which is based on the United Nations' Guidelines on Open Government Data for Citizen Engagement. 1) Make government data and information available on the internet. 2) Establishing government institutions with an open culture; and 3) Creating a policy framework. The dimensions of 5 priorities are used as an endogenous variable. There are five main objectives: 1) Priority 1 is to involve stakeholders in OGD preparation and establish an OGD competency center, 2) Priority 2 is to establish a legal enforcement structure, 3) Priority 3 is to phase OGD implementation, 4) Priority 4 is to build collaboration features on the platform for communication and raising public awareness of OGD stakeholders, and 5) Priority 5 is to conduct training. The obtained data was analyzed using the SmartPLS software. Procedure 1 involved dividing the three groups of exogenous variables from the eight dimensions of ODRA. Group 1 consists of Leadership, Policy & Legal Framework, Government Institutional Structures, and Data Skills (X1, X2, X3), Group 2, namely X4, X5, contains data from the ministry in charge of SMEs policies, as well as a demand for open data from the SMEs society. Open Data Ecosystem, Funding and Open Data Program, National Technology Infrastructure and Skills are part of Group 3, namely X6, X7, and X8. Procedure 2 involves creating an intervening variable Y1 with three indicators and a priority endogenous variable Y2 with five indicators. Respondents to the questionnaire were Palembang-based SMEs with varying levels of education and business knowledge. This is evident when asked to complete a questionnaire. Many people did not fill it out completely.

The following data processing results show that many respondents are inconsistent.

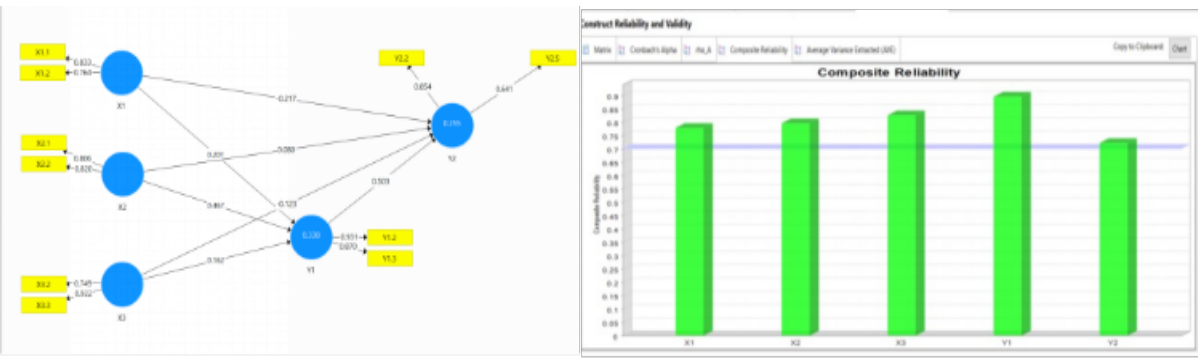

Figure 2: Group 1. The results of the validity test and reliability test

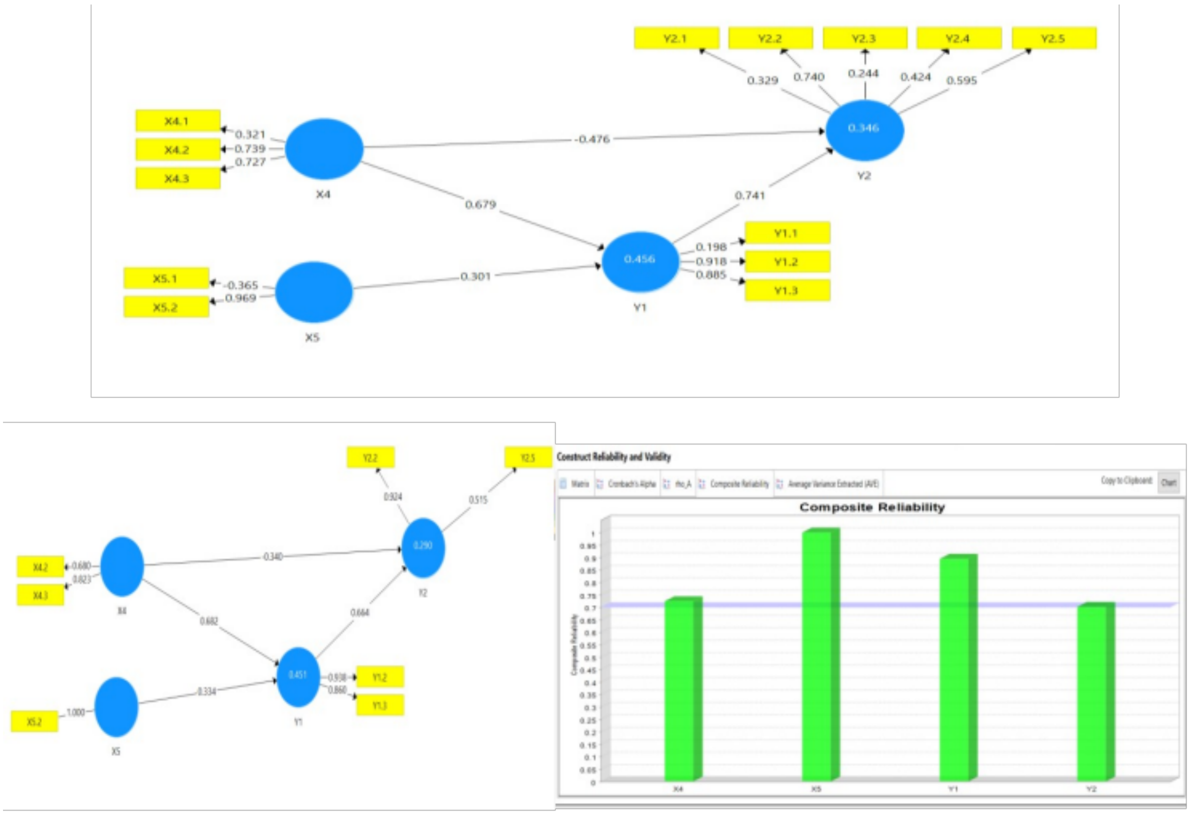

Figure 3: Group 2. The results of the validity test and reliability test

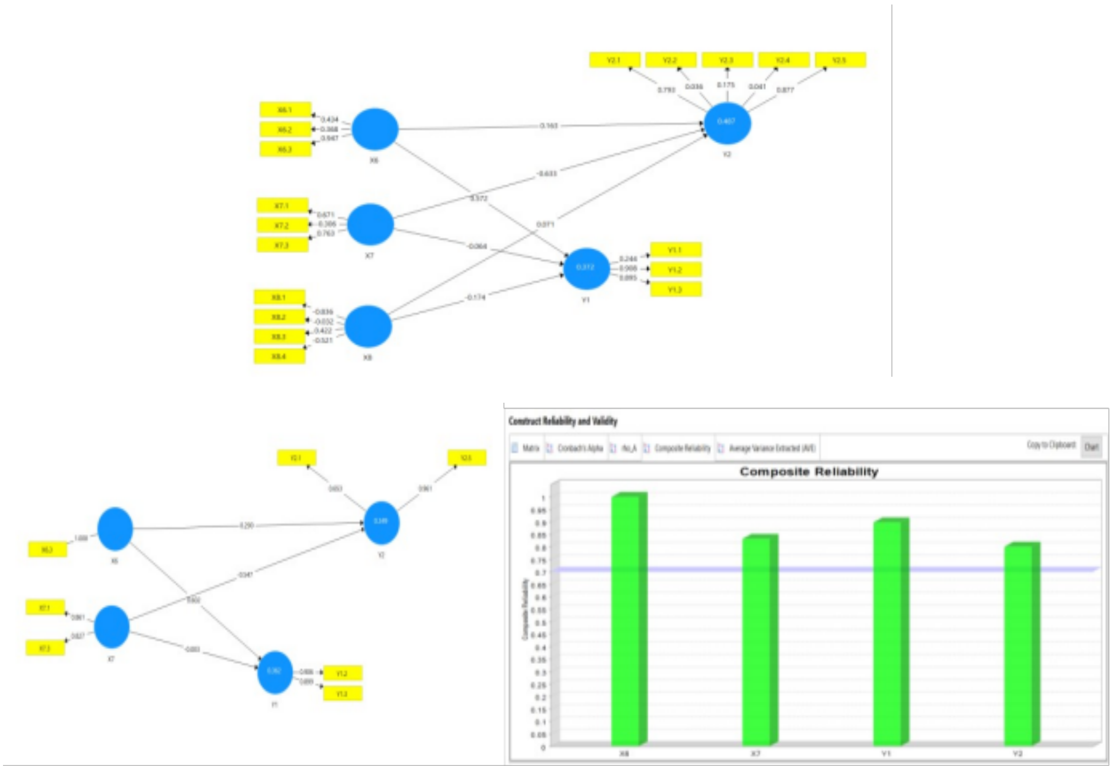

Figure 4: Group 3. The results of the validity test and reliability test Resumes

The results of data processing are as follows:

Table 3: Summary of The Perspectives of The SMEs' Actors

Variabel	Validity	Explanation
X1.1	0.833	Leadership/ Leaders have expressed publicly visible support for OGD.
X1.2	0.764	Leadership/ Support for OGD among key data-owning agencies.
X2.1	0.806	Policy & Legal Framework/ Existence and effectiveness of an access to information law.
X2.2	0.820	Policy & Legal Framework/ Privacy protection.
X3.2	0.749	Government Institutional Structures & Skills for Data/ Track record of inter-agency mechanisms coordinating major ICT for OGD initiatives.
X3.3	0.922	Government Institutional Structures & Skills for Data/ Existence and effectiveness of positions comparable to a CIO/CTO within agencies responsible for strategic ICT decisions and management.
X4.2	0.680	Data of the ministry responsible for SMEs policies/ The visibility of agencies into their data holdings.
X4.3	0.823	Data of the ministry responsible for SMEs policies/ The existence of key data-owning agencies with demonstrable capabilities in data management.
X5.2	1.00	Demand for Open Data from SMEs Society/ Existence of agency mechanisms in place to intake and respond to requests for data.
X6.3	1.00	Open Data Ecosystem/ Government promotion of reuse of the data.
X7.1	0.861	Funding an Open Data Program/ Existence of resources and personnel for an Open Data Program.
X7.3	0.827	Funding an Open Data Program/ the government's track record for investing in innovation.

The variables summarized in the tables are significant from the perspective of SMEs. These findings were also corroborated directly by three experts, who concurred with the conclusions. Additional variables not included in the table will be investigated further with a larger sample size and a broader distribution of questionnaires.

5 Conclusion

OGD is an essential component of today's digital government. One of the concerns that the government must address is the use and benefits of data made available to stakeholders. As one of the pillars of the Indonesian economy, SMEs must be considered in terms of data availability that suits their needs. That is why this research is important in gaining their perspective.

Based on the data analysis findings, it is possible to conclude that the SMEs society has a demand for open data in terms of the existence of agency mechanisms for receiving and responding to data requests. Additionally, the Open Data Ecosystem is critical for SMEs in terms of government promotion of data reuse. The resume in Table 3 contains a list of additional variables that are critical to SMEs.

The findings of this study are expected to serve as a framework for subsequent research in several other countries. While significant variables may share some similarities, they may also be distinct. This is fascinating, given the cultural and national differences in each country's business practices. We recommend that researchers from other countries conduct research in their home countries on open government data and SMEs as well.

Acknowledgments

The authors express appreciation to the Malaysia Ministry of Higher Education (MoHE). This research was supported by the Fundamental Research Grant Scheme (FRGS/1/2019/ICT01/UTHM/02/1) grant vot number K210.

References

Alismaili, S., Li, M., Shen, J., & He, Q. (2015). *A multi perspective approach for understanding the determinants of cloud computing adoption among Australian SMEs*. ACIS 2015 Proceedings - 26th Australasian Conference on Information Systems.

Alonso, J. M., Boyera, S., Grewal, A., Iglesias, C., & Pawelke, A. (2013). *Open Government Data Readiness Assesment Indonesia*. World Wide Web Foundation. http://labs.webfoundation.org/wp-content/uploads/2014/12/OGD_Indonesia.pdf

Antwi, S. K., & Kasim, H. (2015). Qualitative and Quantitative Research Paradigms in Business Research: A Philosophical Reflection. *European Journal of Business and Management*, *7*(3), 217–225. https://www.researchgate.net/publication/295087782, www.iiste.org

Bedini, I., Farazi, F., Leoni, D., Pane, J., Tankoyeu, I., & Leucci, S. (2014). Open Government Data: Fostering Innovation. *JeDEM - EJournal of EDemocracy and Open Government*, *6*(1), 69–79. https://doi.org/10.29379/jedem.v6i1.329

Boyera, S., Schalkwyk, F. van, & Grewal, A. (2017). *Open Data in Ethiopia*. https://www.academia.edu/38416197/Open_Data_in_Ethiopia

Capri, A. (2017). Micro and Small Businesses in Indonesia's Digital Economy: Keys To Developing New Skills and Human Capital Table of Contents. *The Asia Pacific Foundation of Canada*, 56. https://apfcanada-msme.ca/sites/default/files/2019-03/Micro and Small Businesses in Indonesia's Digital Economy.pdf

Chrzanowski, P., Herzog, T., & Holm, J. (2017). *Open Data Readiness Assessment Prepared for the Government of Uganda* (Issue May).

Creswell, J. W., & Creswell, D. J. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th ed.). SAGE Publications, Inc.

Enriquez-Reyes, R., Cadena-Vela, S., Fuster-Guillo, A., Mazon, J. N., Ibanez, L. D., & Simperl, E. (2021). Systematic Mapping of Open Data Studies: Classification and Trends From a Technological Perspective. *IEEE Access*, *9*, 12968–12988. https://doi.org/10.1109/access.2021.3052025

Flick, U. (2013). *The SAGE Handbook of Qualitative Data Analysis* (1st ed.). SAGE Publications Ltd. https://doi.org/http://dx.doi.org/10.4135/9781446282243

Geiger, C. P., & von Lucke, J. (2012). Open Government and (Linked) (Open) (Government) (Data). *JeDEM - EJournal of EDemocracy and Open Government*, *4*(2), 265–278. https://doi.org/10.29379/jedem.v4i2.143

Gomes, L., & Soares, D. (2014). Open government data initiatives in Europe. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*. Published. https://doi.org/10.1145/2691195.2691246

Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, *32*(4), 441–452. https://doi.org/10.1016/j.giq.2015.09.001

Hossain, M. A., Dwivedi, Y. K., & Rana, N. P. (2015). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 14–40. https://doi.org/10.1080/10919392.2015.1124007

Huber, F., Wainwright, T., & Rentocchini, F. (2018). Open data for open innovation: managing absorptive capacity in SMEs. *R&D Management*, *50*(1), 31–46. https://doi.org/10.1111/radm.12347

Jacob, D. W., Fudzee, M. F. M., Salamat, M. A., & Rahman, N. H. A. (2019). Analyzing the Barrier to Open Government Data (OGD) in Indonesia. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(1), 136–139. https://doi.org/10.30534/ijatcse/2019/2681.32019

Kalampokis, E., Tambouris, E., & Tarabanis, K. (2011). Open Government Data: A Stage Model. *Lecture Notes in Computer Science*, 235–246. https://doi.org/10.1007/978-3-642-22878-0_20

Kaplan, B., & Duchon, D. (1988). Combining Qualitative and Quantitative Methods in Information Systems Research: A Case Study. *MIS Quarterly*, *12*(4), 571. https://doi.org/10.2307/249133

Katbi, A. K., & Al-Ammary, J. (2019). Open Government Data in Kingdom of Bahrain: Towards an Effective Implementation Framework. *Advances in Intelligent Systems and Computing*, 699–715. https://doi.org/10.1007/978-3-030-16181-1_66

Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, *35*(1), 13–29. https://doi.org/10.1016/j.giq.2017.11.003

Myers, M. D. (1997). Qualitative Research in Information Systems. *MIS Quarterly*, *21*(2), 241. https://doi.org/10.2307/249422

Neuroni, A. C., Riedl, R., & Brugger, J. (2013). Swiss Executive Authorities on Open Government Data -- Policy Making beyond Transparency and Participation. *2013 46th Hawaii International Conference on System Sciences*. Published. https://doi.org/10.1109/hicss.2013.19

Nugroho, R. (2013). *A comparison of open data policies in different countries: Lessons learned for an open data policy in Indonesia | TU Delft Repositories*. Http://Resolver.Tudelft.Nl/Uuid:Ae4e0a64-579d-40c4-Bed0-D51614ddea9c.https://repository.tudelft.nl/islandora/object/uuid:ae4e0a64-579d-40c4-bed0-d51614ddea9c?collection=education

OECD. (2005). OECD SME and Entrepreneurship. https://stats.oecd.org/glossary/detail.asp?ID=3123

Open Government in Indonesia. (2016). OECD Publishing. http://dx.doi.org/10.1787/

Ojeda, M., & Juárez-Cerrillo, S. (1996). Biplot display for diagnostic in a two-level regression model for growth curve analysis. *Computational Statistics & Data Analysis*, *22*(6), 583–597. https://doi.org/10.1016/0167-9473(96)00013-8

Okamoto, K. (2016). Introducing Open Government Data. *The Reference Librarian*, *58*(2), 111–123. https://doi.org/10.1080/02763877.2016.1199005

Oktaba, H., & Piattini, M. (2008). Software Process Improvement for Small and Medium Enterprises : Techniques and Case Studies (H. Oktaba & M. Piattini (eds.)). *Information Science Reference*. www.igi-global.com

Parung, G. A., Hidayanto, A. N., Sandhyaduhita, P. I., Ulo, K. L. M., & Phusavat, K. (2018). Barriers and strategies of open government data adoption using fuzzy AHP-TOPSIS. *Transforming Government: People, Process and Policy*, *12*(3/4), 210–243. https://doi.org/10.1108/tg-09-2017-0055

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2018). Group Development Stages in Open Government Data Engagement Initiatives: A Comparative Case Studies Analysis. In P. Parycek et al. (Ed.), *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*): Vol. 11020 LNCS. Springer International Publishing. https://doi.org/10.1007/978-3-319-98690-6

Rachman, S. (2017). *The role of information and communication technology services within small and medium enterprise as a growth factor affecting Indonesia's economy* (Issue September) [RMIT University]. http://researchbank.rmit.edu.au/view/rmit:162274

Rahemtulla, H., Custer, S., Tisacova, I., Jhalla, K., Gigler, S., & Brigham, C. (2012). *The Journey of Open Government and Open Data Moldova*.

Rahmatika, M., Krismawati, D., Rahmawati, S. D., Arief, A., Sensuse, D. I., & Fadhil Dzulfikar, M. (2019). An Open Government Data Maturity Model : A Case Study in BPS-Statistics Indonesia. *2019 7th International Conference on Information and Communication Technology (ICoICT)*. Published. https://doi.org/10.1109/icoict.2019.8835352

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity*, *22*(1), 1–24. https://doi.org/10.3233/ip-160012

Saxena, S. (2017). Prospects of open government data (OGD) in facilitating the economic diversification of GCC region. *Information and Learning Science*, *118*(5/6), 214–234. https://doi.org/10.1108/ils-04-2017-0023

Saxena, S., & Janssen, M. (2017). Examining open government data (OGD) usage in India through UTAUT framework. *Foresight*, *19*(4), 421–436. https://doi.org/10.1108/fs-02-2017-0003

Sayogo, D. S., & Yuli, S. B. C. (2018). Critical Success Factors of Open Government and Open Data at Local Government Level in Indonesia. *International Journal of Electronic Government Research*, *14*(2), 28–43. https://doi.org/10.4018/ijegr.2018040103

Srimuang, C., Cooharojananone, N., Tanlamai, U., & Chandrachai, A. (2017). Open government data assessment model: An indicator development in Thailand. *2017 19th International Conference on Advanced Communication Technology (ICACT)*. Published. https://doi.org/10.23919/icact.2017.7890110

Styrin, E., Luna-Reyes, L. F., & Harrison, T. M. (2016). Open Data and Open Government. *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*. Published. https://doi.org/10.1145/2912160.2912161

Syarif, A., Salamat, M. A. B., & Syafari, R. (2020). A Comparative Analysis of Open Government Data in Several Countries: The Practices and Problems. *Advances in Intelligent Systems and Computing*, 346–357. https://doi.org/10.1007/978-3-030-36056-6_33

Syaripul, N. A., & Bachtiar, A. M. (2016). Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi DKI Jakarta: topik ekonomi dan keuangan daerah. *Jurnal Sistem Informasi*, *12*(2), 82. https://doi.org/10.21609/jsi.v12i2.481

Talukder, M. S., Shen, L., Hossain Talukder, M. F., & Bao, Y. (2019). Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh. *Technology in Society*, *56*, 147–156. https://doi.org/10.1016/j.techsoc.2018.09.013

United Nations. (2017). Guide on Lessons for Open Government Data Action Planning for Sustainable Development. https://publicadministration.un.org/en/ogd

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, *31*(2), 278–290. https://doi.org/10.1016/j.giq.2013.10.011

Vracic, T., Varga, M., & Curko, K. (2016). Effects and evaluation of open government data initiative in Croatia. *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Published. https://doi.org/10.1109/mipro.2016.7522380

Wirtz, B. W., & Birkmeyer, S. (2015). Open Government : Origin , Development , and Conceptual Perspectives. International Journal of Public Administration, 1–16(January), 37–41. https://doi.org/10.1080/01900692.2014.942735

Wirtz, B. W., & Birkmeyer, S. (2015). Open Government: Origin, Development, and Conceptual Perspectives. *International Journal of Public Administration*, *38*(5), 381–396. https://doi.org/10.1080/01900692.2014.942735

World Bank Group. (2017). *The Malaysia Development Experience Series - Open Data Readiness Assessment (ODRA)* Report (Issue May). wbg.org/Malaysia

World Bank Group. (2020a). *Bangladesh Open Data Readiness Assessment Report Open Data for Economic and Social Development and Improving Public Services*.

World Bank Group. (2020b). *Open Data Readiness Assessment Prepared for the Government of Sierra Leone*.

Yin, R. K. (2009). Case Study Research: Design and Methods. In L. Bickman & D. J. Rog (Eds.), *Essential guide to qualitative methods in organizational research* (Vol. 5, Issue 5). Sage Publications.

Zhu, X. (2017). The failure of an early episode in the open government data movement: A historical case study. *Government Information Quarterly*, *34*(2), 256–269. https://doi.org/10.1016/j.giq.2017.03.004

Zijlstra, A. (2015). *Open Data Readiness Assessment the Kyrgyz Republic.*

Zuiderwijk, A., Romer, M., & Kroesen, M. (2019). Open data policy development. *Proceedings of the 15th International Symposium on Open Collaboration*. Published. https://doi.org/10.1145/3306446.3340819

Modified PRDG Model for Caregiver Segmentation Using Zarit Burden Interview Instrument

[1*]Nabillah Annisa Rahmayanti, [1]Retno Aulia Vinarti, [1]Arif Djunaidy, [2]Anna Tjin, [3]Jeng Liu

[1]Department of Information Systems, Faculty of Intelligent Electrical and Information Technology, Institut Teknologi Sepuluh Nopember, Indonesia.
[2]Trinity College Dublin, The University of Dublin, Trinity Centre for Global Health, 7-9 Leinster Street, South, Dublin 2, Ireland.
[3]Tunghai University Department of Sociology No.1727 Section 4, Taiwan Blvd, Xitun District, Taichung City, Taiwan.

email: [1*]bellasnabillah@gmail.com, [1]zahra_17@is.its.ac.id, [1]adjunaidy@its.ac.id, [2]annat@tcd.ie, [3]jliu@thu.edu.tw

Corresponding author

Received: 09 December 2021 | Accepted: 10 March 2022 | Early access: 22 March 2022

Abstract - *The increasing demand for Indonesian workers in Taiwan has an impact on caregiver problems which can be triggered by the burden of caring for the elderly. Therefore, the aim of this study is to identify the characteristics of caregivers who are resilient to burdens based on Indonesian female caregivers who work in Taiwan data to be a guide for selecting prospective caregivers. The process includes analyzing the personal characteristics that have the most influence on the burden using multiple regression and then clustering caregiver data using K-Means with the Elbow Method and Silhouette Index. Then, segmentation in each cluster based on a comparison of the average values. The results of clustering accuracy on dimensions (PRDG) and modified dimensions (S+PRDG) were compared and the smallest error cluster was in case 4 in the S+PRDG dimension with the Elbow Method of 3.6%. Based on segmentation on that dimension, cluster 2 is a resilient caregiver cluster. Then the results of the multiple regression analysis (Number of Children, Education and Work Location) were studied further for each caregiver in cluster 2 and the conclusions are, their average number of children is 1, final education is in junior high school and their work location is in the capital of Taiwan.*

Keywords: Caregivers, S+PRDG Model, ZBI Instruments, K-Means Algorithm, Elbow Method, Silhouette Index.

1 Introduction

Global migration has increased rapidly since 1960 to date. This increase is an important part of the globalization process. In addition, the global labor market has emerged, and laborers can now be transferred across national borders based on capital demands. In the Southeast Asia region, Indonesia and Taiwan are a pair of countries that has established a relationship between sender and recipient in terms of labor (Kennedy, 2012). Therefore, families in Taiwan have had the option of employing foreign caregivers since 1991; the Taiwan government established a procedure and quota system for importing workers into the country (Cheng, 2003). The increasing demand for female workers, especially those caring for the elderly in developed countries such as Taiwan, is proportional to the increase in the education level of women in importer countries. As their education level is generally at the university level and most of them are career women, taking care of their elderly is burdensome to some extent. Therefore, to care for their parents, they employ female workers from other countries (e.g., Indonesia) (Yeoh & Huang, 2010).

Due to the large demand from families who care for the elderly, the Taiwanese government has given the option of employing foreign caregivers. However, this does not mean that it is an easy thing for foreign workers to work in Taiwan. They have to adapt to varying job demands, which affect their stress levels (Loveband, 2004). Some of these workers are able to adapt to the demands of their work as caregivers, but most of them struggle with the difficulties (Law et al., 2021). Workers who take care of the elderly have a higher level of burden because some of the elderly have comorbidities, one of which is memory problems (e.g., dementia) (Tam et al., 2018).

Basnyat and Chang (2017) stated that the level of burden experienced by caregivers was not only caused by the varying demands of their work but also due to the individual differences of each caregiver themselves. The different characteristics of each individual caregiver greatly affect their resilience in adapting to their workload. Adelman et al. (2014) stated that caregivers who care for patients for a long period of time have a considerable burden. Furthermore, something that is quite serious and is becoming a concern is when the caregiver is not paid and has a long period of time caregiving, it all causes financial difficulties, and sometimes the caregiver will feel unable to adjust to the work environment, and to deal with social factors such as undue pressure and failure to perform tasks, difficulty in maintaining relationships, harassment, and neglect. All of them can be several possibilities for the caregiver to close her own life history (Thapa et al., 2021). Additionally, some of the caregivers who are migrant workers would try to escape and end up becoming illegal workers (Loveband, 2004). Caregivers spend about 20.5 hours per week and 20% of those caregivers work >40 hours per week (Adelman et al., 2014). Therefore, the burden of caregivers is of further concern because it has a very serious impact on mental and physical health (Tsai et al., 2021) which affects the quality of the care provided. Recruiter agencies and government should be aware of the characteristics of female foreign workers who wish to work in Taiwan. For example, they should know about the workers' work experience, foreign language skills, and family background. The number of Indonesian Migrant Workers or better known as Tenaga Kerja Indonesia (TKI) in Taiwan, according to data from BNP2TKI (National Agency for the Placement and Protection of Indonesian Workers) (BNP2TKI, 2020) in 2019, is 79,574 legal workers from more than 500 TKI companies, which is an increase of 10% from the previous year (Lidwina, 2020). It does not include the number of TKI who migrated to Taiwan illegally. Utami (2020) said that there are six problems commonly faced by Indonesian migrant workers in Taiwan, namely, unpaid salaries, work accidents, repatriation of Indonesian migrant workers, sexual harassment, abuse, and the act of running away from the employer. For example, in the case of the problem where migrant workers run away from their employers' houses, it is suspected that the possible causes are unclear understanding of the employment contract and violation of rights committed by the employers during the implementation of the employment contract (Rahman et al., 2021). Therefore, caregivers who work overtime and have no days off because of the dependence of the elderly in their care are triggered by these causes. As a result, the caregivers feel stressed and run away from their employers' houses. Apart from the dependence of the elderly on caregivers, this problem can also be caused by other burdens related to personal strain, role strain, social life, and feelings of guilt felt by caregivers while caring for the elderly. Therefore, to reduce the problems faced by caregivers in Taiwan, we need to find out information about caregivers who are resilient to the burden.

Such information is useful for Indonesian female foreign worker recruiters and the government in determining what kind of actions should be implemented to reduce the burden of the caregivers before sending them abroad. Therefore, information about caregivers who can withstand the burden needs to be investigated further. A study on identifying personal characteristics to predict the burden experienced by Indonesian female foreign workers in Taiwan has been carried out by Troy et al. (2020) who investigated the relationship between burden and personal characteristics of Indonesian caregivers living in Taiwan using the data of 299 caregivers who care for the elderly. The burden and severity of the caregiver's burden were measured using Zarit Burden Interview (ZBI), and the results of the Exploratory Factor Analysis (EFA) test showed that the burden of the caregivers was multidimensional in nature and related to Personal strain, Role strain, Dependency, and Guilt (PRDG), which were in line with previous studies (Springate & Tremont, 2014; Lau et al., 2019). However, the study did not show any results in the form of information about caregivers who were able to withstand the burden based on available data.

Information on caregivers who are able to withstand the burden can be seen based on their load score. A high load score indicates that the caregiver is depressed or has a high burden, while a low load score indicates otherwise. To be able to calculate the load score of the caregiver, the ZBI instrument is used. The ZBI itself is a common instrument in studies conducted to assess the workload of caregivers. The ZBI was originally developed to assess the burden of caregivers caring for the elderly with dementia. In the ZBI instrument, there are dimensions determined by several researchers using the EFA test. The dimensions are not limited to caregivers who treat elderly people with dementia, but they also include caregivers who treat patients with a particular disease. Therefore, the dimensions of the ZBI instrument can be adjusted to the case study but cannot be separated from the basic dimensions of personal strain, role strain, dependency, and guilt, which can be abbreviated as PRDG in this study. The personal strain dimension has ten instruments based on the rotated factor matrix of the EFA test results. These instruments are related to the personal burden of the caregivers. After further research on nine items (6, 12, 9, 11, 3, 5, 4, 13, 17) of the personal strain instruments, it was found that three items (6, 12, 13) are more related to problems around the social life of the caregivers. Therefore, this study proposes adding the dimension of Social life, so that the dimensions become S+PRDG, with the hope that caregivers who are able to withstand the burden can be examined in more detail.

The purpose of this study is to experiment and estimate caregiver segmentation using the clustering method of K-Means algorithm. Since the K-Means algorithm has difficulties in determining the number of initial clusters, then two methods are used: Elbow Method and Silhouette Index. To evaluate which model is better (PRDG or S+PRDG), cluster accuracy is used. After finding out which cluster is better, an analysis of caregiver characteristics is carried out to explain the cluster results. The results of this study can be used as guidelines by the government and recruiter agencies that recruit Indonesian female foreign workers in selecting prospective caregivers, so that they can make appropriate strategies and decisions before sending them to work abroad.

2 Literature Review

2.1 Caregiver

A caregiver is a family member or a helper who intends to care for children or older adults, some of whom may have certain diseases or physical or mental disabilities (e.g., recipient) (Windham, 2015). The caregivers discussed in this paper are specific to Indonesian female caregivers who work in Taiwan (e.g., TKI). These caregivers care for the elderly for approximately 20 hours per week with 20% of them working overtime, meaning that their total working time exceeds the average total working time per week, but is less than 40 hours per week. Somehow, this condition puts a burden on the caregivers. In addition to the working loads, the burden can also be determined from the caregiver's personal characteristics. For example, caregivers with lower education levels (e.g., primary schools) feel more pressure on the balance of work and family than those who graduate from junior or senior high school. Furthermore, unmarried caregivers tend to have difficulties in controlling their emotions when caring for the elderly (Chang et al., 2021).

Hoenig and Hamilton (1966) proposed a concept regarding the meaning of the caregiver burden, which is divided into two aspects: subjective and objective. Subjective burden is a burden related to the feelings of caregivers when caring for patients. Objective burden is an activity related to the negative experience felt by the caregiver. Therefore, understanding the caregivers' attributes associated with burden is important. According to Zhu Liu et al. (2020), the attributes related to caregiver burden are self-perception, multifaceted determination, and overtime. Self-perception is how caregivers reflect their personal experiences while working with their patients. Multifaceted strain is a multidimensional caregiver burden such as emotional and psychological stress that is often experienced by caregivers.

Furthermore, the routine and social activities of caregivers in providing care for a long period or working overtime can be disrupted (Yoon et al., 2014; Arian et al., 2017). The higher the number of hours worked or overtime put in per day to provide care to elderly care recipients, the more impact it has on the high burden. The study by Juvang et al. (cited in Li et al., 2007) showed that there was a positive relationship between the number of hours caregiver worked to care for the elderly and the objective burden faced by them. The longer the time spent caring for the elderly, the more objective burden is felt by the caregiver (Rafiyah et al., 2011). When caregivers spend more time working on caring for the elderly, they may have less time to themselves. In fact, they also need time for their social life, such as chatting and visiting their friends and family. Finally, such condition has an impact on the caregivers' daily activities. Based on the attributes related to the caregiver burden, these attributes can be grouped into several dimensions similar to those in the proprietary instrument, namely, Personal strain, Role strain, Dependency, and Guilt (PRDG) dimensions (Troy et al., 2020). Another instrument added in this research is Social life related to overtime. Therefore, the dimensions become S+PRDG.

2.2 Zarit Burden Interview (ZBI)

Zarit Burden Interview (ZBI) is an instrument developed by Professor Steven H. Zarit of the University of Pennsylvania (Zarit et al., 1980), which is often used to assess the burden of care. This instrument has been translated into several languages and used in various countries such as America and Europe. In addition, the validity and reliability of this instrument have also been tested by researchers in China, Korea, and Japan. Caregivers were asked 22 questions about the impact of caring for the elderly on their lives. Then, they will be assessed to find out how burdened they are. In each question, caregivers were asked to mark how often they felt a certain way. Each item is given a score of 0 to 4 (0 = never, 1 = rarely, 2 = sometimes, 3 = quite often and 4 = almost always). The total score is calculated by adding the score of each item, with a value varying from 0 to 88. There is no cutoff score, but the higher score means the higher caregiver burden. There are several dimensions to the ZBI instrument, and these dimensions can be adjusted based on the studies of each researcher.

Table 1: ZBI Instrument Question Indicators

ZBI Items	Indicators
Item 1	The elderly patient asking for more help than needed
Item 2	Caregiver doesn’t have enough time for herself because of the elderly patient
Item 3	Caregiver feels stressed between caring care recipient and her family
Item 4	Caregiver feels embarrassed over the elderly’s behavior
Item 5	Caregiver feels angry when she’s around her elderly patient
Item 6	Caregiver feels that the current elderly affects relationship with family or friends in a negative way
Item 7	Caregiver feels afraid of the future of elderly care patient
Item 8	Caregiver feels the elderly is too dependent on her
Item 9	Caregiver feels strained when she is around the elderly patient
Item 10	Caregiver feels her health has suffered because of involvement with the elderly patient
Item 11	Caregiver feels that she doesn’t have much privacy as she want, because of the elderly patient
Item 12	Caregiver feels that social life has suffered because of the elderly patient
Item 13	Caregiver feels uncomfortable about having friends over, because of the elderly patient
Item 14	Caregiver feels that the elderly expect her to be the only one who can be depend on
Item 15	Caregiver doesn't have enough money to support the elderly patient other than the rest of the expenses
Item 16	Caregiver feels unable to take care of the elderly much longer
Item 17	Caregiver feels lost control of her life
Item 18	Caregiver want to leave the care of the elderly patient to someone else
Item 19	Caregiver is confused about what to do for elderly patients
Item 20	Caregiver feels she has to do more for the elderly patient
Item 21	Caregiver feels she can do a better job for caring the elderly patient
Item 22	Caregiver feels that caring for the elderly is a burden

Table 2: Comparison of Caregiver Burden Dimensions

ZBI Items	Previous Study				Current Study
	(Parpa et al., 2017)	(Unson et al., 2016)	(Ankri et al., 2005)	(Troy et al., 2020)	
Item 1	Personal Strain	Anger	-	Dependency	Dependency
Item 2	Personal Strain	Personal Strain	-	Personal Strain	Dependency
Item 3	Personal Strain	Personal Strain	-	Personal Strain	Personal Strain
Item 4	Social Strain	Social Strain	Emotional	Personal Strain	Personal Strain
Item 5	Social Strain	Anger	Emotional	Personal Strain	Personal Strain
Item 6	Social Strain	Social Strain	Social Strain	Personal Strain	Social Strain
Item 7	-	Inadequacy	-	Role Strain	Role Strain
Item 8	Personal Strain	Dependency	-	Dependency	Dependency
Item 9	-	Anger	Emotional	Personal Strain	Personal Strain

Item 10	Personal Strain	Personal Strain	Social Strain	Role Strain	Role Strain
Item 11	Social Strain	Social Strain	Social Strain	Personal Strain	Personal Strain
Item 12	Social Strain	Social Strain	Social Strain	Personal Strain	Social Strain
Item 13	Social Strain	Social Strain	Social Strain	Personal Strain	Social Strain
Item 14	Personal Strain	Dependency	-	Dependency	Dependency
Item 15	Personal Strain	Personal Strain	Guilt	Role Strain	Role Strain
Item 16	Uncertainty	Inadequacy	Guilt	Role Strain	Role Strain
Item 17	Personal Strain	Personal Strain	Social Strain	Personal Strain	Personal Strain
Item 18	Uncertainty	Inadequacy	Guilt	Role Strain	Role Strain
Item 19	Uncertainty	Inadequacy	Guilt	Role Strain	Role Strain
Item 20	Guilt	Inadequacy	Guilt	Guilt	Guilt
Item 21	Guilt	Inadequacy	Guilt	Guilt	Guilt
Item 22	-	-	Emotional	Role Strain	Role Strain

3 Methodology

The data used for this current study is based on Indonesian female caregivers working in Taiwan. This data is the same as the one used by Troy et al. (2020). Data was obtained using the questionnaire instrument. The questionnaire contains 22 ZBI (Zarit Burden Interview) questions. There were 299 respondents spread across the territory of Taiwan.

After the data had been obtained, the next step was to preprocess the data by eliminating missing values, converting data types, and recoding some variables (Section A in Figure 1). Next, Confirmatory Factor Analysis (CFA) was conducted to see whether the number of factors and the grouping of ZBI items was in accordance with the theory in previous studies which produced 4 factors (Section B in Figure 1).

Based on the literature review, one dimension of S (Social life) was added to PRDG, turning it into S+PRDG. This addition would be validated using the clustering technique to see whether the number error was lower after the modification had been made or not. In the next step, multiple regression analysis was carried out on several factors such as education, number of children, marital status, work location in Taiwan, Chinese proficiency, and others. The purpose of using multiple regression analysis was to see the significance of the relationship between these factors on the 5 dimensions (S+PRDG). The most significant factors were called the key indicator (Section C in Figure 1).

The next step was performing Elbow Method and Silhouette Index to find the optimum initial cluster, followed by clustering using the K-Means algorithm (Section D in Figure 1). The clustering was performed using 6 scenarios on two dimensional models: PRDG and S+PRDG. After that, an evaluation of the clustering results was carried out to see the number of clustered errors from each scenario and determine which case was the best. The best case was the one with the smallest clustered error value (Section E in Figure 1). Next was the process of segmenting the results of the clustering based on the average value of each dimension in each cluster and comparing it with the total average value in each dimension.

The process was done to obtain the PRDG and S+PRDG segmentation results. After that, the clustering results were given a rating from the case scenario to indicate which cluster had the lowest burden and which one had the highest burden based on the results of the previous segmentation (Section F in Figure 1). Lastly, after going through all of those processes, the results of the cluster that had the lowest burden were obtained. Based on the results, it can be seen which respondents or caregivers are included in the cluster.

Then the background of each respondent related to the key indicators that had been obtained from the results of the multiple regression analysis in the previous stage was reviewed. Therefore, information about the background of respondents who were in the cluster with the lowest burden was obtained (Section G in Figure 1). Recruiter agencies and the government can use the conclusions from the results obtained in making strategies and decisions for the next wave of prospective Indonesian migrant workers. For the details, each stage of the process in this study can be seen in Figure 1, along with the marking of each process from A to G, which will be described in more detail according to the marking given in the following explanation.

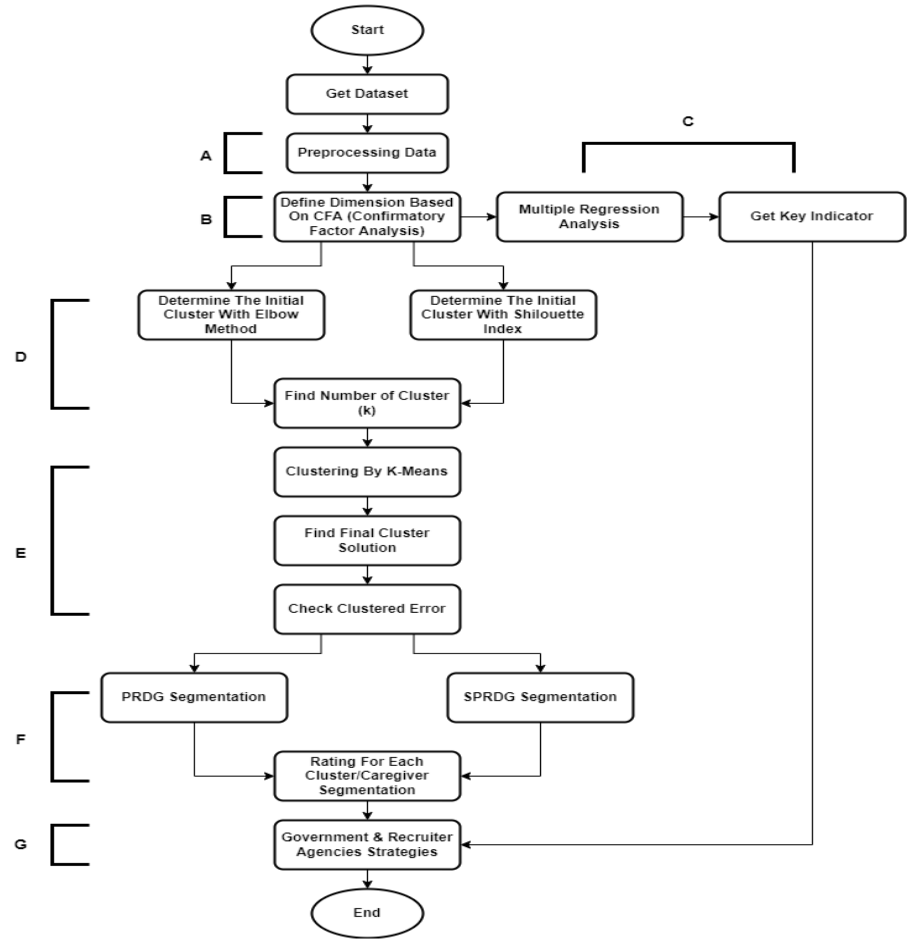

Figure 1: Stages of caregiver segmentation based on PRDG, S+PRDG model and clustering techniques.

3.1 Preprocessing Data (A)

After getting the dataset, the next step was the preprocessing stage. At this stage, the data was cleaned because there was a missing value in the existing data. In addition, some data were converted from string to numeric so that the analysis process can be carried out later. The recoding ranged from 0-4 to 1-5. All stages, from preprocessing to multiple regression analysis, were carried out using the SPSS 25.

3.2 Defining the Dimensions Based On CFA (Confirmatory Factor Analysis) & Literature Review (B)

The S+PRDG dimension was proposed after the Confirmatory Factor Analysis (CFA) test was carried out to see whether the number of factors and the grouping of the variables were in accordance with the theory in the previous study (Troy et al., 2020) which resulted in 4 dimensions (PRDG). The results from the CFA show that the Kaiser Meyer Oikin Measure of Sampling Adequacy (KMO MSA) value is 0.887 > 0.50, then factor analysis can be carried out and the Anti-Image Correlation value is > 0.50, thus the MSA assumption is fulfilled and the results of the Rotated Component Matrix show the distribution of items in the four factors (see Table 3). However, after careful consideration of the literature, it turns out that the personal strain dimension can be further developed into a new dimension by separating items that lead to caregiver social relationships into dimensions of Social life.

Thus, the dimensions become S+PRDG. This is reinforced by the statement of Ong et al., (2018) that the social support felt by the caregiver, such as relationships with family and friends, is able to mediate the relationship between resilience and the burden of the caregiver, and this is very important because it functions as a protective factor to protect the caregiver from the perceived burden while caring for the elderly. In addition, the Social life dimension was added based on a literature review of previous studies that also used the ZBI Instrument and the existence of a Social life dimension in that study (see Table 2).

Table 3: Confirmatory Factor Analysis Results

ZBI Items	Rotated Component Matrix			
	Factor 1 Personal Strain	**Factor 2 Role Strain**	**Factor 3 Dependency**	**Factor 4 Guilt**
Item 3	0.626			
Item 4	0.674			
Item 5	0.671			
Item 6	0.767			
Item 9	0.635			
Item 11	0.618			
Item 12	0.668			
Item 13	0.590			
Item 17	0.534			
Item 7		0.541		
Item 10		0.575		
Item 15		0.766		
Item 16		0.709		
Item 18		0.748		
Item 19		0.557		
Item 22		0.404		
Item 1			0.688	
Item 2			0.436	
Item 8			0.803	
Item 14			0.772	
Item 20				0.835
Item 21				0.832

3.3 Multiple Regression Analysis to Get the Key Indicators (C)

After finding the dimensions based on the CFA results, the next step was to determine what key indicators can be used to measure the relationship between several predictor variables (representing the caregiver's family structure, job qualifications, and job characteristics) and each response variable (S, P, R, D, G). The results of multiple regression analysis in previous studies concluded that there was a positive or negative relationship between each of the existing factors (e.g., education, Mandarin proficiency, work location, etc.) and the dimensions of the PRDG and also severity of the burden of each factor based on the results of multinomial analysis. Meanwhile, in the current study, multiple regression analysis was used to see the significance of the relationship between the response variable and two or more predictor variables. The difference between the current study and previous studies (Troy et al., 2020) is that this study does not discuss the unidirectional relationship of each variable with the dimensions of the load and the severity of the load. However, this paper discusses more about how the addition of a new dimension (S) can affect the caregiver clustering model. The cluster model is then used to better understand the personal characteristics, which consitute the background of caregivers who are tough against the burden. Thus, the pattern is obtained.

Each personal caregiver's characteristic represents several categories (see Table 4) such as family structure, work qualifications, and employee characteristics, where each category consists of several personal characteristics which are independent variables, while load dimensions such as Personal Strain and so on are dependent variables. Thus, this study uses multiple regression to analyzetheir relationship. The effect sizes, which are R Square from the multiple regression results, show how much variation in the predictor variables can be explained by the dependent variable. In the family structure category (see Table 4), the largest effect size is in the Guilt dimension which shows 0.027 or 2.7% of the variation of the Guilt dimension can be explained by the number of children and the marital status of the caregiver. The large effect sizes have an effect on the p-value (significance influence of the independent variable individually on the dependent variable) which in the family structure category, the significance of the number of children variable on the Guilt dimension has the smallest p-value among others with p = 0.005 or p <0.05 or it can be said that the number of children variable has a significant influence on the Guilt dimension. Likewise, in the category of work qualifications, 0.089 or 8.9% of the variation in the Personal Strain dimension can be explained by the education, Chinese proficiency and elderly work experience of the caregiver. Education has a significant influence on the dimensions of Personal Strain, with p = 0.000 or p <0.05. Furthermore, in the employment characteristic category, the effect size of 0.111 or 11.1% variation from the Role Strain dimension can be explained by recipient age and location of work. The location of work variable has a significant effect on the Role Strain dimension, with p = 0.000 or p <0.05. Based on the results of multiple regression analysis (see Table 4), it can be concluded that there are three key indicators that represent each category of caregiver personal characteristics, namely the number of children, education, and location of work.

Table 4: Multiple Regression Test Results

Variables	**Persona l Strain**	**Role Strain**	**Depen dency**	**Guilt**	**Social Life**
Family Structures					
Number of Children	0.168	0.019	0.028	*0.005*	0.725
Marital Status	0.338	0.829	0.422	0.379	0.225
R Square	0.018	0.026	0.016	0.027	0.009
Work Qualifications					
Education	*0.000*	0.203	0.024	0.076	0.007
Chinese Proficiency	0.209	0.009	0.215	0.028	0.216
Elderly Care experience	0.020	0.269	0.142	0.281	0.119
R Square	0.089	0.031	0.026	0.029	0.034
Employment Characteristics					
Recipient Age	0.067	0.967	0.012	0.363	0.242
Location of Work	0.333	*0.000*	0.542	0.003	0.574
R Square	0.015	0.111	0.022	0.034	0.005

3.4 Determining the Initial Cluster (D)

A common problem when using K-Means algorithm is that it is very sensitive to initial partitioning (Khan & Ahmad, 2004). Thus, it becomes a weakness of the algorithm as we must determine the number of the clusters ourselves. As this algorithm cannot define the appropriate number of clusters by itself, it relies on the user to guess the number of clusters in advance. Therefore, to determine the optimal number of initial clusters, several approaches using other algorithms are needed. A couple of algorithms that are commonly used to determine the optimum initial cluster are the Elbow Method and the Silhouette Index. The Elbow Method is used to determine the best number of clusters that can be used to produce the best cluster results and maximize the quality of cluster results (Dewi & Pramita, 2019) and the Elbow Method determines the best number of clusters from the percentage of comparison between the number of clusters that will form an elbow at a point (Madhulatha, 2007).

The equation (1) below is used to calculate the SSE (Sum of Squared Error) mathematically. Also, the percentage of the calculation result becomes a comparison between the number of clusters (Kodinariya & Makwana, 2013). Elbow shape in this method is formed by comparing one's SSE value with another SSE value. If the SSE value of the first cluster is higher than the second cluster's SSE value, then it forms an elbow-like shape. The best initial cluster is the one with the sharpest decline of the SSE value (Purnima & Arvind, 2014). K is the number of clusters used in the K-Means algorithm, Xi is the number of data, and Ck is the number of clusters in the k cluster. The larger the number of K clusters, the smaller the SSE value (Irwanto et al., 2012). The smaller the SSE value, the better the cluster is.

$$SSE = \sum_{k=1}^{k} \sum_{x_i \in S_k} \| X_i - C_k \|^2 \tag{1}$$

Similar to the Elbow Method, Silhouette validity index is a statistical measure used to solve the problem of determining the optimal number of K clusters that can provide a brief graphical representation of how well each object is located in the cluster (Rousseeuw, 1987). This method is a combination of the method of separation and cohesion (Kodinariya & Makwana, 2013). The equation (2) below is used to calculate the Silhouette Index mathematically by calculating the average distance of each j object to all objects in cluster p, the same cluster as object j, denoted by ap,j. Then, calculate the average distance of each j object to all objects in cluster q, where p is not equal to q, called dq,j. Then, find ap,j from the minimum dq,j, which shows the objects' mean difference x(j) for the cluster closest to its neighbors. The higher the $S_{x(j)}$ value, the more precise the placement of $x_{(j)}$ to cluster p. Silhouette Index values are usually in the range of -1 to 1. The optimum number of clusters is obtained when the index value is close to 1.

$$S_{x(j)} = \frac{b_{p,j} - a_{p,j}}{max\{a_{p,j}, b_{p,j}\}} \tag{2}$$

In this stage, we determine the optimum number of the initial cluster (k) using the Elbow Method and Silhouette Index. WEKA software was used in this process. Table 5 contains a description of the results of k clusters for each different dimension after clustering using samples, namely the total value of each dimension (P, R, D, G and S, P, R, D, G).

Table 5: Initial Cluster Result

Initial Cluster Differences			
PRDG Dimension		**S+PRDG Dimension**	
Elbow	**Silhouette**	**Elbow**	**Silhouette**
k=3	k=5	k=3	k=4

3.5 Clustering Using K-Means (E)

To find caregivers who can withstand the burden, clustering is carried out according to the dimensions of caregiver burden that have been described previously (see Table 2). According to Yedla, Madhu, and Srinivasa Rao Pathakota (2010), clustering is the process of grouping data objects into the same group called a cluster. Clustering is an example of unsupervised classification. Unsupervised is a grouping that does not depend on class and training standards. The grouping can be done using various methods, one of which is K-Means. The K-Means algorithm was first proposed by Stuart Lloyd in 1957 (Lloyd, 1982), even though it was not published until 1982. K-means is a partition clustering method that is widely used in industries. K-means is considered a clustering algorithm since it is easy to implement and the most efficient in terms of execution time.

In this stage, we do clustering using the K-Means algorithm, which was conducted by trying out different scenarios. For example, in the first case scenario, one variable is chosen at random in each dimension of Personal strain, Role strain, Dependency, Guilt (PRDG) and also Social life when clustering is done in the S+PRDG dimension. Clustering was done by trying out 6 scenarios in each dimension. This clustering process adjusts the initial k from the results of both the Elbow Method and Silhouette Index.

The 6 case scenarios referred to in this study were selected randomly. One item from each dimension was then used 6 times with no repetitions, except for the Dependency, Guilt, and Social life dimensions which have a limited variety of variable items, meaning that it is possible to use the same item in the next case. For example, case 1 contained item 9 from the Personal strain dimension, item 16 from the Role strain dimension, item 8 from the Dependency dimension, item 20 from the Guilt dimension, and item 6 from the Social life dimension. Then, case 1 went through a clustering process using the K-Means algorithm and the optimum initial cluster that had been determined previously. The purposes of these scenarios were (1) to test the K-Means clustering algorithm and (2) to find the lowest error rate in one cluster.

In addition to the clustering process using K-Means, the thing that needed to be considered was how much clustering that had been done can group caregivers correctly according to the clusters. This can be seen from the number of clustered errors. The case selected for further analysis regarding which respondents or caregivers were grouped into the cluster as well as each respondent's educational background and work location in Taiwan was the case that had the least number of clustered errors.

3.6 Caregiver Segmentation (F)

After going through the clustering process with K-Means, the average results for each dimension (P, R, D, G and S, P, R, D, G) in each resulting cluster were obtained. As the average value of each dimension was known, the next step was to segment the results of the average value of each dimension in each cluster by comparing it with the average total value of each dimension (P, R, D, G and S, P, R, D, G). The total average value was obtained by adding up the overall average value of the dimensions in each cluster and dividing it by the number of clusters produced. After that, segmentation was done by comparing the results of the average value of each dimension (P, R, D, G and S, P, R, D, G) in each cluster with the average total value. If the average value of each dimension in each cluster (P, R, D, G and S, P, R, D, G) is higher than the average total value, then an over bar appears (e.g., $\overline{D}$: Higher dependency value). However, if the result of the average value of each dimension (P, R, D, G and S, P, R, D, G) in each cluster is lower than the average total value, then an under bar appears. Caregiver feels a dependency burden from care recipient (e.g., $\underline{D}$: Lower dependency value; caregiver does not feel a dependency burden from care recipient) (Ait et al., 2015).

3.7 Government & Recruiter Agencies' Strategies (G)

After rating each cluster based on its segmentation, information about which cluster had the lowest rating value in all aspects of the S+PRDG dimension was obtained. The lowest rating was chosen because the dimension in this research study is the burden of the caregiver, which is negative. Thus, if the rating is low, the caregivers in the cluster are resistant to the burden. Based on the cluster that had the lowest segmentation rating value for each aspect of the dimension, it was also known which caregivers or respondents were included in the cluster. Next, additional information about what factors affected caregiver burden based on key indicators that had been obtained from the multiple regression analysis process as well as each caregiver or respondent's education and work location in Taiwan were examined. Therefore, the information obtained from the results of a series of analysis processes and clustering can help recruiter agencies and the government in selecting, formulating strategies and decisions for prospective Indonesian Migrant Workers or TKI before they get hired abroad by looking at their educational background and work placement locations.

4 Results and Discussion

Two methods were used to determine the initial cluster, namely the Elbow Method and the Silhouette Index, which were followed by clustering using K-Means (see Figure 1). The difference in initial clusters affected the number of clustered errors that were significantly different in both the PRDG and S+PRDG dimensional models. Table 6 and Table 7 describe the trial results of 6 cases and also the total value of each dimension (P, R, D, G and S, P, R, D, G). It can be seen that the initial cluster generated from the Elbow method produced smaller clustered error values in both dimensional models. In addition, the S+PRDG dimensional model had smaller clustered error values using both the initial cluster from the Elbow Method and Silhouette Index when compared to the PRDG dimensional model. By comparing the two models, it was determined that cluster 2 had the lowest burden. The caregiver segmentation is described in Table 9 by taking one case scenario, namely case 4 in the S+PRDG dimensional model. Case 4 consists of 5 ZBI instruments.

Table 6: PRDG Model Clustering Results

Case	PRDG			
	Elbow Method K=3		Silhouette Index K=5	
	Lowest Cluster	Clustered Error	Lowest Cluster	Clustered Error
Total	Cluster 2	38.1 %	Cluster 2	53.17%
Case 1	Cluster 2	40.1%	Cluster 2	30.1%
Case 2	Cluster 2	21.4%	Cluster 2	45.4%
Case 3	Cluster 2	24.7%	Cluster 2	46.8%
Case 4	Cluster 2	11.3%	Cluster 2	33.1%
Case 5	Cluster 2	34.1%	Cluster 2	52.8%
Case 6	Cluster 2	27.7%	Cluster 2	40.8%

Table 7: S+PRDG Model Clustering Results

Case	S+PRDG			
	Elbow Method K=3		Silhouette Index K=5	
	Lowest Cluster	Clustered Error	Lowest Cluster	Clustered Error
Total	Cluster 2	5.3%	Cluster 2	10%
Case 1	Cluster 2	30.1%	Cluster 2	38.4%
Case 2	Cluster 2	8.6%	Cluster 2	19.3%
Case 3	Cluster 2	14%	Cluster 2	22.7%
Case 4	Cluster 2	*3.6%*	Cluster 2	19%
Case 5	Cluster 2	5%	Cluster 2	15.7%
Case 6	Cluster 2	11.3%	Cluster 2	36.7%

Item 3 represents the P dimension, item 7 represents the R dimension, item 8 represents the D dimension, item 21 represents the G dimension, and lastly, item 12 represents the S dimension. Case 4 was chosen because it is in the S+PRDG dimensional model which has a lower clustered error than the PRDG dimension model. In addition, it also has the lowest clustered error value. In Table 8, it can be seen that the average value of cluster 2 in each dimension (S, P, R, D, G) is smaller than the average total value (last row).

Table 8: Comparison of Clusters Based on The K-Means Method

Cluster	Size	P	R	D	G	S	Pattern
1	104	2.56	4.04	3.98	3.70	2.41	$\overline{P}\,\overline{R}\,\overline{D}\,\overline{G}\,\overline{S}$
2	125	1.56	2.12	1.99	2.30	1.47	$\underline{\boldsymbol{P}}\,\underline{\boldsymbol{R}}\,\underline{\boldsymbol{D}}\,\underline{\boldsymbol{G}}\,\underline{\boldsymbol{S}}$
3	70	2.21	1.15	4.4	3.45	2.17	$\overline{P}\,\underline{R}\,\overline{D}\,\overline{G}\,\overline{S}$
Total	299	2.11	2.44	3.45	2.94	2.01	

Thus, the under bar appears in each dimension of cluster 2. Cluster 1 has the highest rating on burden. The average value of each dimension in cluster 1 is higher than the average total value, resulting in the appearance of the upper bar. Based on the results from cluster 1, it is indicated that the caregiver group within the cluster experiences a burden on all dimensions. However, in cluster 3, only the average value of the role strain dimension is lower than the average total value. In cluster 3, the caregivers feel pressured because they have to take care of elderly people and fulfill other obligations for their families. The burden experienced by caregivers is related to personal strain.

In addition, caregivers also feel other burdens such as the dependence of the elderly who rely too much on their caregivers, feeling guilty about not taking care of the elderly optimally, and feeling stressed because they have to take care of the elderly continuously, resulting in them not having the time to socialize with friends and relatives. The next process was finding out which caregivers or respondents were included in cluster 2 and finding out the educational background and work location in Taiwan of each caregiver included in cluster 2. The purpose of the process is to provide information for recruiter agencies and also the government regarding key indicators that have an influence in determining the next prospective workers who will be sent to Taiwan, who are certainly more able to withstand the burden. Table 9 describes which respondents or caregivers are included in cluster 2 along with information about their number of children, education and work locations in Taiwan. Table 9 shows the caregivers (121 respondents) in cluster 2 grouped by their number of children, education and work location. Even though the number of caregivers in this cluster is high, it has a minimum cluster error (3.6%); approximately 4 caregivers (green in the red cluster) were misplaced (see Figure 2).

Table 9: Number of Children, Education, and Location of Work in Cluster 2

Respondent	**Total Respondent**	**The Average of Personal Characteristics**		
		Number of Children	**Education**	**Work Location**
2,8,10,11,16,17,20,21,23,24,99,100,104,105,107,108,109,110,111,113,30,32,35,36,38,40,41,42,44,45,268,269,273,275,279,281,283,287,288,290,293,49,53,54,57,61,63,66,67,70,71,72,74,76,77,78,80,85,86,93,94,145,146,148,153,155,156,159,161,165,167,120,122,130,134,135,136,137,138,141,143,170,174,175,178,179,181,183,184,186,193,196,198,200,203,210,211,213,214,215,221,222,225,229,235,236,237,240,244,245,247,248,249,254,255,256,257,258,262,263,266	121	2	Junior High School	Capital

Based on the results of the clustering analysis it can be said that cluster 2 in each case of the experimental scenario always has low segmentation results in each dimension based on both PRDG and S+PRDG comparisons. However, based on the results of the cluster error, where the cluster error itself indicates how well the cluster results are formed, we look for clustering results which have the smallest cluster error, and that is in case 4 in the S+PRDG dimension with a cluster error of 3.6%, which means that the cluster formed in case scenario 4 has a validity of 96.4%. The results of the selected clustering, which have been declared valid at 96.4%, are then re-analyzed regarding caregivers who are members of the cluster. The analysis carried out is related to the results of multiple regression, that is personal caregiver characteristics which have the most influence on the burden (number of children, education, and location of work). The results presented (see Table 9) are the average results of the overall personal characteristics of the cluster members.

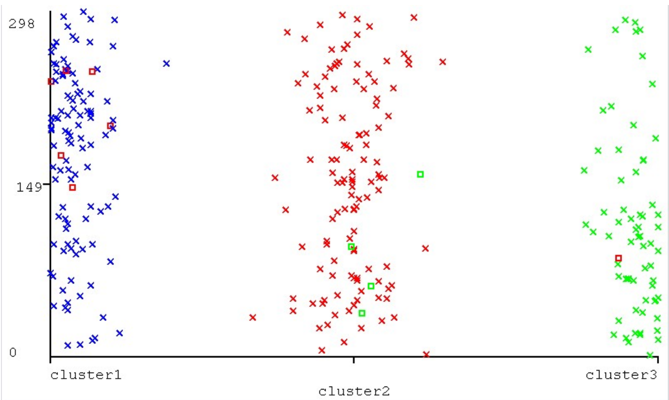

Figure 2: Result of Case 4 Clustering

The number of children, education, and work location of the caregivers in cluster 2 are dominated by those who on average, have 1 child and have a minimum education level of junior high school. Furthermore, most of them work in the capital city of Taiwan (see Table 9). Taipei (capital of Taiwan) is a base location for Indonesian female workers in Taiwan (Representative Office of Indonesia in Taipei, 2020). Taipei's position is in the northern region of Taiwan. The North and Central areas of Taiwan are the areas where the majority of the Indonesian population is located (Taichung City Government, 2010). The large population of Indonesians in both regions (North and Central) may affect the Social life of caregivers in those areas, as socializing can reduce their stress level. This could be one of the reasons why the caregivers in cluster 2 have a low burden. Another information gathered from cluster 2 is the education level of the caregivers. It can be seen that the majority of the caregivers in cluster 2 graduated from junior high school, followed by high school. A higher level of education has more impact on knowledge, maturity, and problem-solving ability in coping with their work as a caregiver (Dambi et al., 2016). Table 9 becomes a pattern that will serve as a guideline for female labor recruitment agencies and also the government in selecting prospective female workers who will work as caregivers for the elderly before working abroad, especially in Taiwan.

5 Conclusion

The main objective of this study was to find out which caregivers could withstand the burden they face. It revealed that cluster 2 is the most important cluster as it is the cluster that has the lowest load in all dimensions based on the segmentation results. In addition, the caregivers who are included in cluster 2 have been identified based on the results of the clustering. Furthermore, information regarding each caregiver's number of children, educational background, and location of work in Taiwan have been obtained as those are the main indicators affecting caregiver burden.

In addition, this study adds Social life as a dimension based on the literature review of several previous studies that also used the Social life dimension. Other than that, there is a separation of items from the Personal strain dimension into dimensions related to social problems of the caregivers, which is called Social strain or Personal and Social strains in some studies. This study also found that the novel dimension (social life) other than PRDG, that is S+PRDG, has better cluster accuracy than the existing PRDG dimensions. This is a strong reason why Social life is added to the dimensions that already exist. Better cluster accuracy is obtained based on the results of clustering, where the dimensions of S+PRDG in all cases with the initial cluster using the Elbow Method k=3 or using the initial cluster Silhouette Index k=5 resulted in a lower percentage of clustered error than the PRDG dimension, both when using the initial cluster with the Elbow Method k=3 or with the Silhouette Index k=4. Besides, it is also known that in both the PRDG dimensions and the modified S+PRDG dimensions, the use of a smaller initial cluster k=3 results in a low cluster error. Based on Table 6 and Table 7, the lowest clustered error with a percentage of 3.6% is in case 4, which is in the S + PRDG dimension and the initial cluster using the Elbow Method k=3.

The findings of this study are expected to be used by recruiting agents and the Indonesian government as a guide for them in selecting prospective female workers, especially those who work as caregivers in Taiwan. This is explained specifically for selecting prospective female caregivers who will work in Taiwan because it relates to the data used in this study, which is collected from Indonesian female caregivers for the elderly who have worked in Taiwan. However, caregivers who work in Taiwan are not only female caregivers and from Indonesia, but there are also male caregivers and from other countries as migrant workers, and they do not only work specifically to care for the elderly, but also to care for patients who are chronically ill and patients who need special care. Therefore, we recommend that future research be done by collecting different datasets such as datasets for male caregivers or random (both male and female caregivers), based on their country of origin as in this study, and it would be better if, the case studies that are examined are also different, such as caregivers who treat patients with cancer, etc. Furthermore, the use of other analytical techniques is also required, such as the use of other clustering analysis techniques, namely Fuzzy Logic, K-Medoid, DBSCAN, etc., which of course will result in the number of clusters and different levels of accuracy, and will affect the cluster segmentation. In addition, future analysis can also combine optimization algorithms such as genetic algorithms, PSO (particle swarm optimization) algorithms, etc., for more optimal clustering results.

References

Adelman, R. D., Tmanova, L. L., Delgado, D., Dion, S., & Lachs, M. S. (2014). Caregiver Burden. JAMA, 311(10), 1052–1059. https://doi.org/10.1001/jama.2014.304

Ankri, J., Andrieu, S., Beaufils, B., Grand, A., & Henrard, J. C. (2005). Beyond the global score of the Zarit Burden Interview: useful dimensions for clinicians. International Journal of Geriatric Psychiatry, 20(3), 254–260. https://doi.org/10.1002/gps.1275

Arian, M., Younesi, S. J., & Khanjani, M. S. (2017). Explaining the Experiences and Consequences of Care Among Family Caregivers of Patients with Cancer in the Terminal Phase: A Qualitative Research. International Journal of Cancer Management, 10(10), 0–5. https://doi.org/10.5812/ijcm.10753

Basnyat, I., & Chang, L. (2016). Examining Live-In Foreign Domestic Helpers as a Coping Resource for Family Caregivers of People With Dementia in Singapore. Health Communication, 32(9), 1171–1179. https://doi.org/10.1080/10410236.2016.1220346

Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. International Journal of Computer Applications, 105(9), 17–24. https://doi.org/10.5120/18405-9674

Badan Nasional Penempatan dan Pelindungan Tenaga Kerja Indonesia. (2020, February 10). Data Penempatan dan Perlindungan TKI Periode Tahun 2019. Badan Pelindungan Pekerja Migran Indonesia. Retrieved November 30, 2021, from http://portal.bnp2tki.go.id/stat_penempatan/indeks

Chang, H. M., Lin, C. H., & Chuang, Y. W. (2021). Study of Caregivers, Job Stress and Burnout of Long Term Care Institution in Taichung City. Scientific and Social Research, 3(3), 1–6. https://doi.org/10.36922/ssr.v3i3.1081

Cheng, S. J. A. (2003). Rethinking The Globalization Of Domestic Service. Gender & Society, 17(2), 166–186. https://doi.org/10.1177/0891243202250717

Dambi, J. M., Mandizvidza, C., Chiwaridzo, M., Nhunzvi, C., & Tadyanemhandu, C. (2017). Does an educational workshop have an impact on caregivers' levels of knowledge about cerebral palsy? A comparative, descriptive cross-sectional survey of Zimbabwean caregivers. Malawi Medical Journal, 28(4), 167–173. https://doi.org/10.4314/mmj.v28i4.4

Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. Matrix : Jurnal Manajemen Teknologi Dan Informatika, 9(3), 102–109. https://doi.org/10.31940/matrix.v9i3.1662

Hoenig, J., & Hamilton, M. W. (1966). The Schizophrenic Patient in the Community and His Effect On the Household. International Journal of Social Psychiatry, 12(3), 165–176. https://doi.org/10.1177/002076406601200301

Irwanto, I., Purwananto, Y., & Soelaiman, R. (2012). Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra. Jurnal Teknik ITS, 1(1), A197–A202. https://doi.org/10.12962/j23373539.v1i1.631

KDEI Taipei. (2021). Infografik Ketenagakerjaan KDEI 2020. SIMPATI KDEI-Taipei. Retrieved November 10, 2021, from https://simpati.kdei-taipei.org/v2/index.php/infografik#/simpati

Kennedy, J. (2014, May 28). Female Indonesian Domestic Workers in Taiwan. Academia. Retrieved November 1, 2021, from https://www.academia.edu/1750408/Female_Indonesian_Domestic_Workers_in_Taiwan

Khan, S. S., & Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. Pattern Recognition Letters, 25(11), 1293–1302. https://doi.org/10.1016/j.patrec.2004.04.007

Kodinariya, T. M., & Makwana, P. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal of Advance Research in Computer Science and Management Studies, 1(6), 90–95.

Lau, S., Ali, N., Lim, J. P., & Lim, W. S. (2018). ZBI multidimensionality: The salience of worry about performance. International Psychogeriatrics, 31(06), 909–911. https://doi.org/10.1017/s1041610218001217

Law, S., Ormel, I., Babinski, S., Kuluski, K., & Quesnel-Vallée, A. (2021). "Caregiving is like on the job training but nobody has the manual": Canadian caregivers' perceptions of their roles within the healthcare system. BMC Geriatrics, 21(1), 1–11. https://doi.org/10.1186/s12877-021-02354-z

Li, J., Lambert, C. E., & Lambert, V. A. (2007). Predictors of family caregivers' burden and quality of life when providing care for a family member with schizophrenia in the People's Republic of China. Nursing & Health Sciences, 9(3), 192–198. https://doi.org/10.1111/j.1442-2018.2007.00327.x

Lidwina, A. (2020, February 12). Jumlah TKI Menurun di Malaysia, Meningkat di Taiwan. katadata. Retrieved November 1, 2021, from https://databoks.katadata.co.id/datapublish/2020/02/12/jumlah-tki-menurun-di-malaysia-meningkat-di-taiwan

Liu, Z., Heffernan, C., & Tan, J. (2020). Caregiver burden: A concept analysis. International Journal of Nursing Sciences, 7(4), 438–445. https://doi.org/10.1016/j.ijnss.2020.07.012

Lloyd, S. (1982). Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. https://doi.org/10.1109/tit.1982.1056489

Loveband, A. (2004). Positioning the product: Indonesian migrant women workers in Taiwan. Journal of Contemporary Asia, 34(3), 336–348. https://doi.org/10.1080/00472330480000141

Omran, M. G., Engelbrecht, A. P., & Salman, A. (2007). An overview of clustering methods. Intelligent Data Analysis, 11(6), 583–605. https://doi.org/10.3233/ida-2007-11602

Ong, H. L., Vaingankar, J. A., Abdin, E., Sambasivam, R., Fauziana, R., Tan, M. E., Chong, S. A., Goveas, R. R., Chiam, P. C., & Subramaniam, M. (2018). Resilience and burden in caregivers of older adults: moderating and mediating effects of perceived social support. BMC Psychiatry, 18(1), 1–9. https://doi.org/10.1186/s12888-018-1616-z

Parpa, E., NG, K., Tsilika, E., Galanos, A., Papazoglou, I., & Mystakidou, K. (2017). Validity and Reliability of the Greek Version of the ZBI in Informal Carers of Adults with Intellectual Disabilities. International Journal of Mental Health & Psychiatry, 03(03). https://doi.org/10.4172/2471-4372.1000145

Rachid, A. D., Abdellah, A., Belaid, B., & Rachid, L. (2015). Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco. International Journal of Computer and Information Engineering, 9(8), 1976–1986. https://doi.org/10.5281/zenodo.1110165

Rafiyah, I. (2011). Review: Burden on Family Caregivers Caring for Patients with Schizophrenia and Its Related Factors. Nurse Media Journal of Nursing, 1(1), 29–41. https://doi.org/10.14710/nmjn.v1i1.745

Rahman, R. A., Salim, A. A., & Anggriawan, R. (2021). Peran Pemerintah Indonesia dan Taiwan dalam Menanggulangi Kasus Pekerja Migran Indonesia "Kaburan." Jurnal Wawasan Yuridika, 5(1), 81. https://doi.org/10.25072/jwy.v5i1.414

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Springate, B. A., & Tremont, G. (2014). Dimensions of Caregiver Burden in Dementia: Impact of Demographic, Mood, and Care Recipient Variables. The American Journal of Geriatric Psychiatry, 22(3), 294–300. https://doi.org/10.1016/j.jagp.2012.09.006

Taichung City Government. (2010, June). Looking back at over 20 years in Taichung. Taiwanfun. Retrieved November 11, 2021, from https://www.taiwanfun.com/central/taichung/articles/1006/1006TCTM004.htm

Tam, W. J., Koh, G. C. H., Legido-Quigley, H., Ha, N. H. L., & Yap, P. L. K. (2017). "I Can't Do This Alone": a study on foreign domestic workers providing long-term care for frail seniors at home. International Psychogeriatrics, 30(9), 1269–1277. https://doi.org/10.1017/s1041610217002459

Thapa, P., Lama, S., Pradhan, N., Thapa, K., Kumar, R., & Basnet, M. (2021). Attitude Towards Suicide among Caregivers of Patients Admitted with Suicide Attempt to a Tertiary Care Hospital: A Descriptive Cross-sectional Study. Journal of Nepal Medical Association, 59(236). https://doi.org/10.31729/jnma.6246

Troy, C., Tjin, A., Perez C., J. J., & Liu, J. (2020). Personal Determinants of Burden Among Indonesian Female Caregivers of Older Adults in Taiwan. Journal of Applied Gerontology, 41(1), 217–226. https://doi.org/10.1177/0733464820972888

Tsai, C. F., Hwang, W. S., Lee, J. J., Wang, W. F., Huang, L. C., Huang, L. K., Lee, W. J., Sung, P. S., Liu, Y. C., Hsu, C. C., & Fuh, J. L. (2021). Predictors of caregiver burden in aged caregivers of demented older patients. BMC Geriatrics, 21(1), 1–9. https://doi.org/10.1186/s12877-021-02007-1

Unson, C., Flynn, D., Haymes, E., Sancho, D., & Glendon, M. A. (2015). Predictors of types of caregiver burden. Social Work in Mental Health, 14(1), 82–101. https://doi.org/10.1080/15332985.2015.1080206

Utami, D. N. (2020). Upaya Indonesia Dalam Menyelesaikan Permasalahan Pekerja Migran Indonesia di Taiwan 2018–2020. eJournal Hubungan Internasional Fisip Universitas Mulawarman, 8(4), 848–860.

Windham, W. (2015). A caregiver is a family member or a helper who intends to care for children or older adults. In Definition of Parent, Guardian and Caregiver. Nesmith Library Board of Trustees. https://nesmithlibrary.org/sites/nesmithlibrary.org/files/attachments/Definition%20of%20Parent%20Guardian%20and%20Caregiver_0.pdf

Yedla, M., Pathakota, S. R., & t M, S. (2010). Enhancing K-means Clustering Algorithm with Improved Initial Center. International Journal of Computer Science and Information Technologies, 1(2), 121–125. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.451.7988&rep=rep1&type=pdf

Yeoh, B. S. A., & Huang, S. (2009). Foreign Domestic Workers and Home-Based Care for Elders in Singapore. Journal of Aging & Social Policy, 22(1), 69–88. https://doi.org/10.1080/08959420903385635

Yoon, S. J., Kim, J. S., Jung, J. G., Kim, S. S., & Kim, S. (2014). Modifiable factors associated with caregiver burden among family caregivers of terminally ill Korean cancer patients. Supportive Care in Cancer, 22(5), 1243–1250. https://doi.org/10.1007/s00520-013-2077-z

Zarit, S. H., Reever, K. E., & Bach-Peterson, J. (1980). Relatives of the Impaired Elderly: Correlates of Feelings of Burden. The Gerontologist, 20(6), 649–655. https://doi.org/10.1093/geront/20.6.649