Clustering the E-Government Services Awareness of the Local Government Unit through K-Means Algorithm

Charies L. Malicay[a], Hidear Talirongan[b],

Florence Jean B. Talirongan[c], Janeth G. Saren[d]

[a] labanza.c@gmail.com, [b] hidear@mu.edu.ph, [c] badilles.fj@gmail.com, [d] Janethsaren822@gmail.com
[abcd] *College of Computer Studies, Misamis University, Ozamiz City, 7200. Philippines*

Abstract

The Department of Interior and Local Government (DILG) of the Republic of the Philippines utilized performance tools to measure the local government units (LGUs) to its community. This undertaking is an initiative of the DILG to ensure that the citizens feel the service of the LGUs and to ensure the level of awareness. The study utilized the dataset of gender and employment status queries from respondents who utilize the performance tool and use K-Means clustering to provide the same cluster of people's awareness by gender and employment status. As a result, the males are more aware of services like governance and response, public works and infrastructure, environmental management, and economic and investment promotion. On the other hand, females are more aware of services like health, support to education, and social welfare. In terms of employment status, employed are aware of services like governance and response, public works and infrastructure, and economic and investment promotion. Contrarily, the unemployed are aware of services like health, education, social welfare, and environmental management.

Keywords: LGU; clustering; awareness; gender; employment

1. Introduction

Government plays a vital role in the implementation of the different programs for the welfare of the country. In the Philippines Government, a major devolution of national government services, which included the first wave of health sector reform, through the introduction of the Local Government Code, which was first introduced in 1991 (Abrigo & Ortiz, 2018). Government service like environmental service is integral to basic social services (Hadley & Hatch, 2018).

The citizens are deemed in a better position to determine whether these services are delivered according to their needs, at the right time, and to the extent that they fulfill their service requirements. The importance of getting the citizens' true sentiments, opinions, and insights based on their perception and evaluation of the extent and quality of services that local government units (LGUs) are readily available and accessible to the constituents. The Citizen Satisfaction Index System (CSIS) development serves as a tool for drawing applicable information for gauging citizen satisfaction (Guades et. al, 2016).

People are longer unaware of their surroundings. The residents are more aware of their problems, are more open, and speak freely of their needs, as perceived by the local officials. Even the residents are aware of the changes in themselves, but not all are aware of the government's service.

The government provides some services that can be avail by the people. In the case of the citizen in Plaridel, Misamis Occidental, where the researcher will be using the data from the survey questionnaire filled

up by the person, the services that are aware of include health, support education, social welfare, government and response, public works and infrastructure, management, economic and investment. The data sets collected will be kept private and not available to anyone's authority but will be available for the researcher (Saldaña, 2015). This paper's main objective is to use a mining technique using the K-Means algorithm to determine the number of persons aware of the local government's basic services.

One of the most popular and simple clustering algorithms is K-means (Hamerly & Drake, 2015), a partitioning method applied to analyze data and treat the data as objects based on location distance between various input data points (Schabenberger & Gotway, 2017). Data mining is a process used to turn raw data into useful information (Vijayarani et al., 2015). In this research, the clustering task is used to group basic services of local government like governance and response, public works and infrastructure, economic and investment promotion, health, support to education, social welfare, and environmental management as to gender (male and female) and as to employment status (employed and unemployed). The study is only focused on determining the number of awareness through clustering the basic government services.

2. Theoretical Framework

2.1. Review of Related Literature

Data mining has been considered a useful means for identifying huge data patterns and trends (Marjani et al., 2017). It analyzes datasets that aimed to find the relationship among datasets and summarized them to be useful to the users (Lourenco, 2015).

The study of Sabado and colleagues (2016) on the delivery of environmental services of Metro Manila LGU revealed that the delivery of services was rated as moderately efficient and effective by both the constituents and local officials. Guades et al. (2016) presented the investigation results on tourism services of the local government of Calbayog using the Citizen Satisfaction Index System (CSIS). Among the five tourism promotion services offered, respondents are most aware of ecotourism or natural conservation programs.

The research done by Kurfali et al. (2017) investigated both trust and effort expectancy as underlying factors in citizen's decision to use e-government services in Turkey. Performance expectancy, social influence, facilitating conditions, and trust in the internet positively affect behavioral intention to use e-government services.

The study conducted by Machova and Lnenicka (2016) examined and evaluated the specific challenges, trends, and the progress of e-government in the European Union Member States. Using cluster analysis, the areas identified which are affected by crisis need improvement on services offered. Choi et al. (2016) researched an e-government evaluation model based on the principal component analysis. K-means clustering is applied to evaluate the local e-government stages of growth of the different cities, providing several important implications for e-government research and practice. Further, Wu et al. (2017) studied clustering web services based on function similarities and cluster them into functionality similar web service groups.

Another study by Morris et al. (2016) used cluster analysis in determining the factor scores on cognitive measures. It yielded three groupings and revealed that people with schizophrenia have different community integration barriers and could potentially benefit from different treatments.
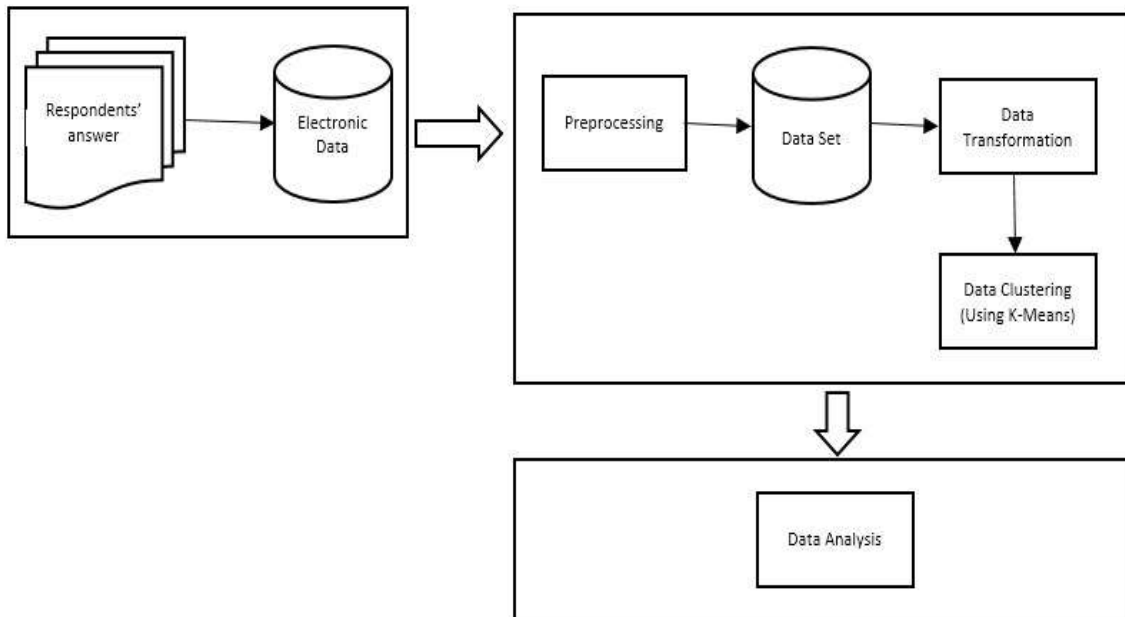
Manfren (2017) studied a systematic procedure on the development of new technology for examining occupant behavior used cluster analysis showed improving the behavior of building occupants and provides multifaceted insights into building energy end-use pattern. It revealed prioritizing efforts to modify occupant behavior to reduce building energy consumption and improve modeling of occupant behavior in numerical simulation.

A clustering-based approach on visual speech image data was studied by Arifin et al. (2017), where the best quality clustering occurs at k=9 consisting of 10 classes (9 classes of clustering result and one neutral class).

3. Operational Framework

The study utilized an architectural design (Figure 1) that includes *data extraction, data pre-processing, data transformation, data clustering, and data analysis.*

Figure 1. Architectural Design of the Study
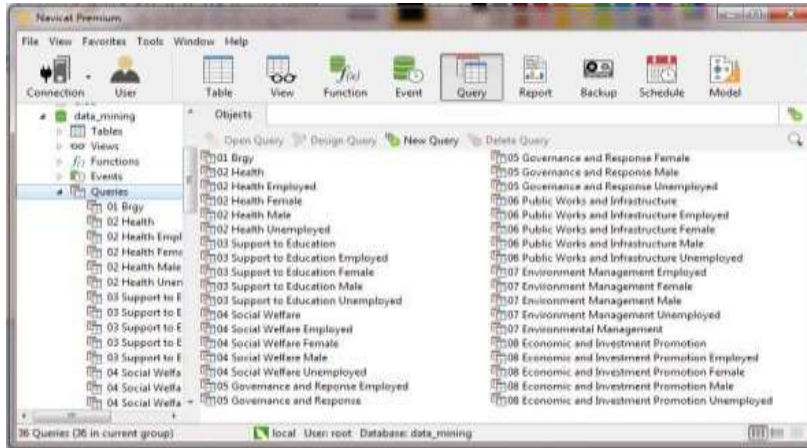

4. Methods

4.1. Data Extraction

The raw data, which is coming from the respondents' answers to the municipality of Plaridel, Misamis Occidental using the Citizen Satisfaction Index Survey (CSIS) questionnaire, was converted to electronic data. Querying data using MySQL was done to extract the desired data, which is to count the number of respondents who are aware of the different LGU basic services. The data taken is grouped by gender and employment status.

4.2. Data Preprocessing

Several queries are undertaken in this step to get the desired data sets. The following figure illustrates the queries:

Figure 2. Queries on Data Sets

4.3. Data Transformation

This step presents the new tables created in the database to be used in the k-means clustering. Figures 3 and 4 present the inputs for gender and employment status.

Figure 3. Datasets on Gender

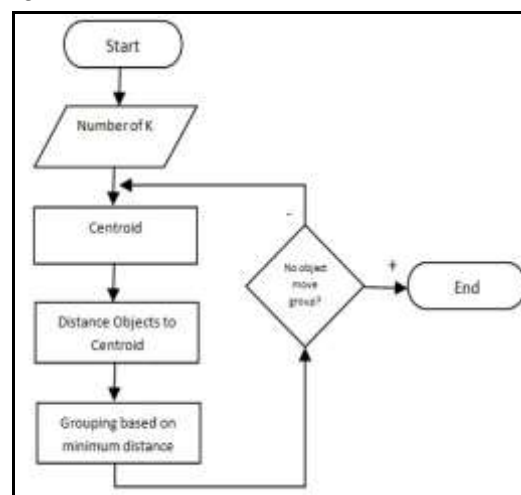*Figure 4. Datasets on Employment Status*

4.4. Data Clustering

Clustering algorithms (Gan & Ng, 2017) aim at dividing the set of objects into groups (clusters), where objects in each cluster are similar to each other (and as dissimilar as possible to objects from other clusters). Gender Dataset was the group as to male and female. Employment Dataset was the group as to employed and unemployed. The K-Means Nearest Neighbor algorithm is a machine-learning instance-based technique. This method does not construct models but stores the training instances. For each new instance, the algorithm compares the distance feature-vectors to the training set. The nearest neighbors are selected based on the distance of the new instance's features, i.e., the similarity between the new instance and the training set vectors. "K" in the algorithm defines the number of nearest neighbors. The classification of the new object is based on the distance between K-clusters and the object. The object is assigned to the cluster with the minimum distance.

This algorithm is iterative and repeats for each object. It converges until the objects are stable (i.e., no object changes in the group) (Kumar & Jarial, 207). K-Means clustering is simple, and the basic steps it follows are 1. Several clusters, K, is determined. 2. Assume a centroid or center of the K clusters. Any object can be randomly chosen and initialized as an initial centroid. The first K objects can also serve as the initial centroids. 3. The distance of each object from each of the centroids is calculated. 4. Group the objects based on minimum distance (find the closest centroid for each object). Figure 5 presents the k-means Nearest Neighbor algorithm's flowchart.

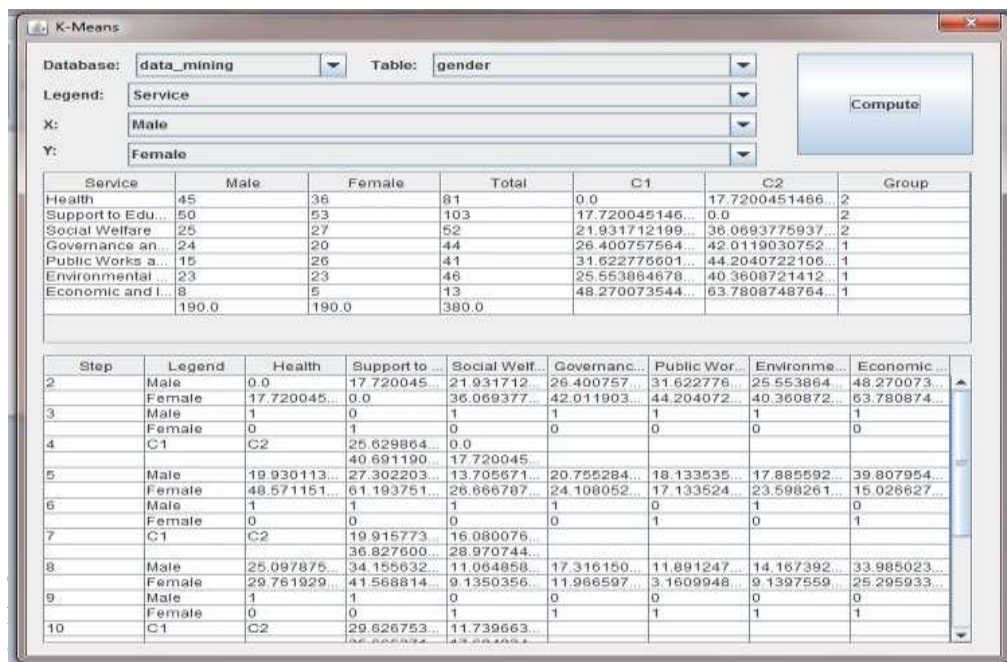*Figure 5. K-Means Nearest Neighborhood Flowchart*

4.5. Data Analysis

This step presented the results of computing the value of centroids and their distances following several iterations to attain similar groups. It also illustrated the results in tables and graphs.

5. Results and Discussions

Government services are integral in enhancing the level of awareness among the people. This study showed that in terms of gender, males are more aware of services like governance and response, public works and infrastructure, environmental management, and economic and investment promotion. On the other hand, females are more aware on services like health, support to education, and social welfare. In terms of employment status, employed are aware of services like governance and response, public works and infrastructure, and economic and investment promotion. On the other hand, unemployed are aware of services like health, education, social welfare, and environmental management. The researchers created an application that will automatically generate the results following the k-means algorithm shown in Figure 6.

Figure 6. K-Means Computation by Gender

answer for services row (male) is derived from the formula

=SQRT(((raw data of services(male)- C1) ^ 2) + ((raw data of services(female)-C1) ^2).

The answer for services row (female) is derived from the formula

=SQRT(((raw data of services(male)- C2) ^ 2) + ((raw data of services(female)-C2) ^2).
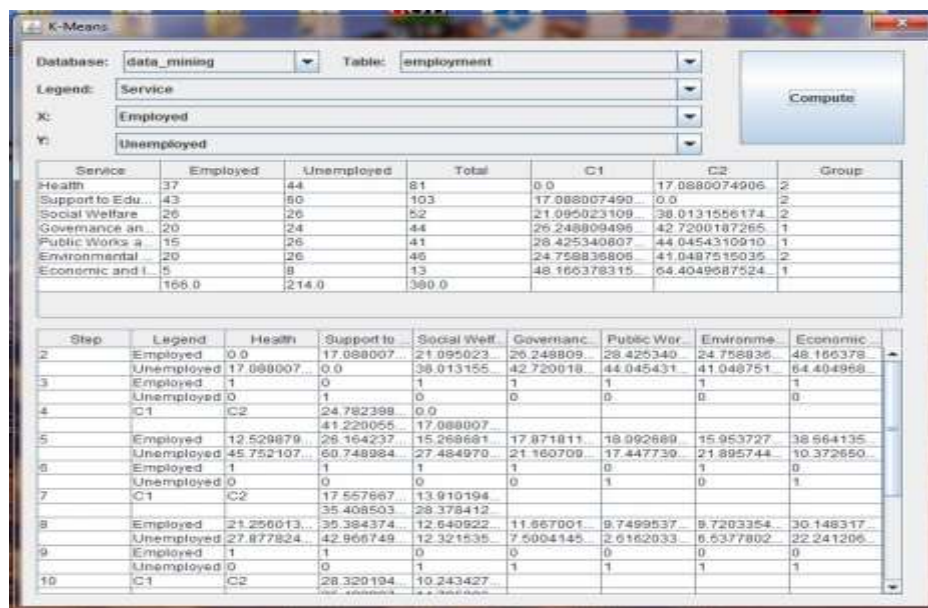
The results of the figure on employment status computation based on its final iteration is presented in the table below:

Table 1. Clustered LGU Basic Services by Gender

LGU Basic Services	Male	Female	Group
Health	45	36	2
Support to Education	50	53	2
Social Welfare	25	27	2
Governance and Response	24	20	1
Public Works and Infrastructure	15	26	1
Environmental Management	23	23	1
Economic and Investment Promotion	8	5	1

The table above presented four members for Group 1 and 3 members for Group 2. It showed that males are more aware of services like governance and response, public works and infrastructure, environmental management, and economic and investment promotion. On the other hand, females are more aware on services like health, support to education, and social welfare.

Figure 7. K-Means Computation by Employment Status

The answer for services row (employed) is derived from the formula
=SQRT(((raw data of services(employed)- C1) ^ 2) + ((raw data of services(unemployed)-C1) ^2).

The answer for services row (unemployed) is derived from the formula
=SQRT(((raw data of services(employed)- C2) ^ 2) + ((raw data of services(unemployed)-C2) ^2).

The results of the figure on employment status computation are based on its final iteration, which is presented in the table below:

Table 2. Clustered LGU Basic Services by Employment Status

LGU Basic Services	Employed	Unemployed	Group
Health	37	44	2
Support to Education	43	60	2
Social Welfare	26	26	2
Governance and Response	20	24	1
Public Works and Infrastructure	15	26	1
Environmental Management	20	26	1
Economic and Investment Promotion	5	8	1

The table above presented three members for Group 1 and 4 members for Group 2. It showed that employed are aware of services like governance and response, public works and infrastructure, and economic and investment promotion. On the other hand, unemployed are aware on services like health, support to education, social welfare, and environmental management.

To give a graphical presentation of the two categories of clustering, Figure 8 presents the scattered plot of clustering by gender, and Figure 9 presents the scattered plot of clustering by employment status. It showed that the four highest services to which males are aware are governance and response, public works and infrastructure, environmental management, and economic and investment promotion, which is Group 1. For Group 2, females are more aware on services like health, support to education, and social welfare. Group 1 showed that the three highest services to which employed people are aware of our governance and response, public works and infrastructure, and economic and investment promotion in terms of the employment status. On the other hand, unemployed are aware on services like health, support to education, social welfare, and environmental management.
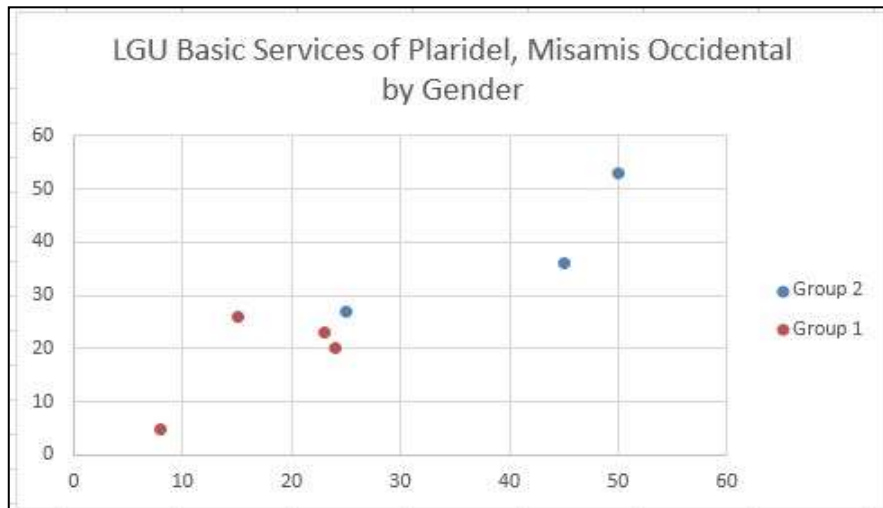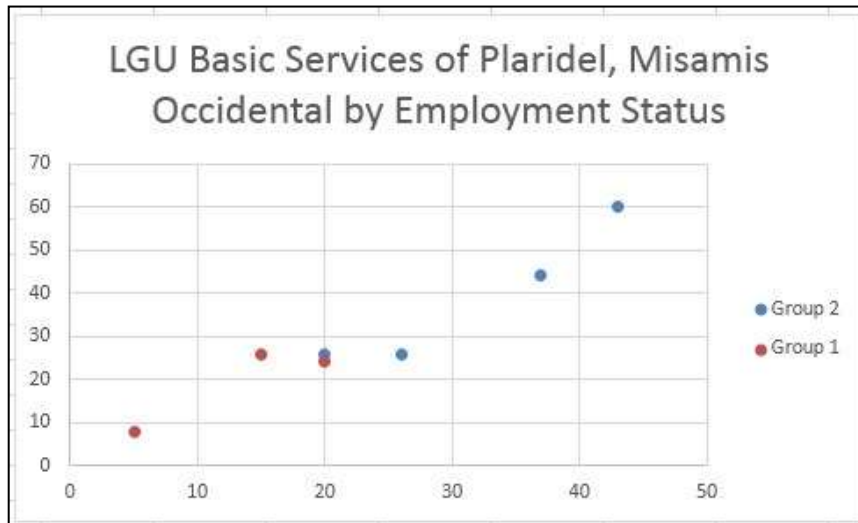
Figure 8. Scattered Plot by Gender

Figure 9. Scattered Plot by Employment Status

6. Conclusions and Recommendations

The study presented the different clusters of the LGU Basic services using K-means clustering based on the awareness of the citizens of Plaridel, Misamis Occidental into two categories that include gender and employment status. The results showed that in terms of gender, males are more aware of services like governance and response, public works and infrastructure, environmental management, and economic and investment promotion. On the other hand, females are more aware on services like health, support to education, and social welfare. In terms of employment status, employed are aware of services like governance and response, public works and infrastructure, and economic and investment promotion. On the other hand, unemployed are aware on services like health, support to education, social welfare, and environmental management.

For the wider dissemination of information, the researcher would like to recommend, based on the result, informing female persons on services like governance and response, public works and infrastructure, environmental management, and economic and investment promotion. On the other hand, the male should be informed on services like health, support to education, and social welfare. In terms of employment status, the unemployed should be informed on services like governance and response, public works and infrastructure, and economic and investment promotion. On the other hand, employed should be informed on services like health, education, social welfare, and environmental management. Besides, the government should create programs for information dissemination on the local government unit's basic services.

The researchers recommended this study be used as a baseline for future research about the awareness and use of other clustering methods.

References

Abrigo, M. R. M., & Ortiz, D. A. P. (2018). *Devolution of health services, fiscal decentralization, and antenatal care in the*

Philippines (No. 2018-42). PIDS Discussion Paper Series.

Arifin, S. S., Muljono, M., & Hariadi, M. (2017). A Model of Indonesian Dynamic Visemes From Facial Motion Capture Database Using A Clustering-Based Approach. *IAENG Int. J. Comput. Sci*, *44*(1), 41-51.

Choi, H., Park, M. J., Rho, J. J., & Zo, H. (2016). Rethinking the assessment of e-government implementation in developing countries from the perspective of the design–reality gap: Applications in the Indonesian e-procurement system. *Telecommunications Policy*, *40*(7), 644-660.

Gan, G., & Ng, M. K. P. (2017). K-means clustering with outlier removal. *Pattern Recognition Letters*, *90*, 8-14.

Guades, E. J., Lonzaga, E. A., Calumba, E. C., & Bacamante Jr, F. B. (2016). Citizens' Satisfaction on the Tourism Promotion Services of Calbayog City, Philippines.

Hadley, R., & Hatch, S. (2018). *Social welfare and the failure of the state: Centralised social services and participatory alternatives* (Vol. 6). Routledge.

Hamerly, G., & Drake, J. (2015). Accelerating Lloyd's algorithm for k-means clustering. In *Partitional clustering algorithms* (pp. 41-78). Springer, Cham.

Kumar, A., Kumar, D., & Jarial, S. K. (2017). A hybrid clustering method based on improved artificial bee colony and fuzzy C-Means algorithm. *Int. J. Artif. Intell*, *15*(2), 24-44.

Kurfalı, M., Arifoğlu, A., Tokdemir, G., & Paçin, Y. (2017). Adoption of e-government services in Turkey. *Computers in Human Behavior*, *66*, 168-178.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government information quarterly*, *32*(3), 323-332.

Machova, R., & Lnenicka, M. (2016). Modelling E-Government Development through the Years Using Cluster Analysis. *JeDEM-eJournal of eDemocracy and Open Government*, *8*(1), 62-83.

Manfren, M. (2017). Multi-Scale Computing for a Sustainable Built Environment. *Smart Cities: Foundations, Principles, and Applications*, 53-97.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: architecture, opportunities, and open research challenges. *IEEE Access*, *5*, 5247-5261.

Morris, S. E., Vaidyanathan, U., & Cuthbert, B. N. (2016). Changing the diagnostic concept of schizophrenia: the NIMH research domain criteria initiative. In *The Neuropsychopathology of Schizophrenia* (pp. 225-252). Springer, Cham.

Sabado, J. N., Narbarte, R. E., Borromeo, R. A., & Carpizo, R. (2016). Relationship of Internal Audit Functions to Governance of Local Government Units. In *Journal of International Scholars Conference-BUSINESS & GOVERNANCE* (Vol. 1, No. 3).

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Schabenberger, O., & Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. CRC press.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7-16.

Wu, C., Qiu, W., Zheng, Z., Wang, X., & Yang, X. (2015, June). QoS prediction of web services based on two-phase k-means clustering. In *2015 ieee international conference on web services* (pp. 161-168). IEEE.