

Classification Model of Municipal Management in Local Governments of Peru based on K-means Clustering Algorithms

José Morales¹, Nakaday Vargas², Mario Coyla³
Gestión Pública y Desarrollo Social
Universidad Nacional de Moquegua, Moquegua, Perú

José Huanca⁴
Gestión Pública y Desarrollo Social
Universidad Nacional de Juliaca, Juliaca, Perú

Abstract—The K-means algorithm groups datasets into different groups, defines a fixed number of clusters, iteratively assigning data to the clusters formed by adjusting the centers in each cluster. K-means algorithm uses an unsupervised learning method to discover patterns in an input data set. The purpose of the research is to propose a municipal management classification model in the municipalities of Peru using a K-means clustering algorithm based in 58 variables obtained from the areas of human resources, heavy machinery and operating vehicles, information and communication technologies, municipal planning, municipal finances, local economic development, social services, solid waste management, cultural, recreational and sports facilities, public security, disaster risk management, environmental protection and conservation of all the municipalities of the 24 departments of Peru and the constitutional province of Callao. The results of the application of the K-means algorithm show that 32% of the municipalities made up of the municipal governments of Amazonas, Apurímac, Huancavelica, Huánuco, Ica, Lambayeque, Loreto and San Martín; are in Cluster 1; the 8% in Cluster 2 with the municipal governments of Ancash and Cusco; in the third Cluster the 28% with the municipal governments of the constitutional Province of Callao, Madre de Dios, Moquegua, Pasco, Tacna, Tumbes and Ucayali and in Cluster 4, 32% composed of the municipal governments of Arequipa, Ayacucho, Cajamarca, Junín, La Libertad, Lima, Piura and Puno Region.

Keywords—K-means; cluster; municipality; model; municipal management

I. INTRODUCTION

The clustering problem is one of the most studied topics in the data mining and machine learning communities. There is a wide variety of applications in social networks problems, multimedia, social science, health, education and other fields of knowledge. Grouping is a diverse topic, and the underlying algorithms are highly dependent on the data domain and the scenario in which the problems occur. The objective of clustering is to classify a set of elements into groups that are very similar among them, but different with elements from other groups. Author in [1] consider the k-Means grouping algorithm, to be one of the most efficient grouping algorithms for large-scale data sets.

The K-means algorithm allows clustering by grouping objects into k groups, this is why it becomes very important for researchers and so its results, which will be used in the

municipalities of Peru that promote local development, but also adapting themselves to current organizations' real situation; mainly with the objective to improve the provision of local public services; consequently it will allow a continuous improvement in the offered services.

The current investigation states the development of a model based on the K-means algorithm. It proposes an adjustment of the grouping algorithm, focusing on the detection of behavior patterns, for that it needs to read Peruvian municipalities databases with 58 variables, allowing the decision-making process to improve in all the levels.

Provincial and district municipalities are the governing bodies that promote local development, with legal status under public law and full capacity to fulfill their purposes (Law No. 27972 - Organic Law of Municipalities). Local governments are classified according to their jurisdiction: provincial municipality, located on the territory of the respective province and the provincial capital district. District municipality, located on the territory of the district and town center municipality, whose jurisdiction is determined by the respective provincial council.

The National Institute of Statistics and Informatics – INEI has statistical information of the provincial, district, and population centers at the national level, in order to generate municipal indicators that support regional and local management for planning and proper decision-making. Information that has been used in the investigation of the 196 provincial municipalities, 1,678 district municipalities and 2,656 municipality of populated centers of the country, compiled in the National Register of Municipalities 2019 [2]. The factors considered in the study are related to human resources, heavy machinery and operational vehicles, information and communication technologies, municipal planning, municipal finances, local economic development, social services, solid waste management, cultural, recreational and sports facilities, public security, disaster risk management and environmental protection and conservation.

The purpose of the research is to propose a classification model for municipal management of local governments in Peru based on K-means clustering algorithms, at the same time evaluating the characteristics of local governments in every cluster.

This article is organized as follows: The second section is a review of some studies on the k-means algorithm and municipalities features. The next section focuses on the theoretical background. Section fourth shows the results of the research, while the fifth part presents the argumentation of the results. Finally, the conclusion of the research, which proposes some ideas for further investigation, are shown in section sixth.

II. RELATED WORKS

This section presents the references of different investigations related to K-means and municipal management.

In [3] two grouping algorithms are compared, centroid K-Means algorithms and Fuzzy C-Means, based on their grouping efficiency, the conclusion states the K-Means algorithm is better than the Fuzzy C-Means algorithm, and considers they can be used to discover association rules and functional dependencies.

Customer segmentation is the subdivision of a business customer database into groups called customer segments, so that each customer segment consists of clients who share similar market characteristics. In [4] the k-Means grouping algorithm has been applied in customer segmentation in a retail business, identifying four steady groups or customer segments.

In [5] proposes a better K-Means algorithm to improve the classification precision, when K-Means cannot adequately classify data under certain data distribution conditions. The proposal considers the effect of variance on the classification so that the data can be classified with greater exactness.

In [6] presents a data model capable of extracting, classifying and then mapping data in order to generate new, more structured data that meets the organization's needs. This arrangement is based on the K-means grouping algorithm.

Data recording on the Internet is a way of Big Data to use the K-Means technique as a solution to the analysis of user's behavior. [7] In this research, a grouping process has been carried out using the K-means algorithm, an algorithm that classifies users into three groups, high, medium and low. The result of the research shows that each of these groups visits frequently some websites, through search engines, social networks, and news and information.

In [8], a balanced K-means grouping algorithm is proposed to classify apples automatically. The results show the precision of the multiple characteristics classification method is more than 96%. [9] Grouping is a data analysis technique that is used to investigate the underlying structure of the data. It is described as the technique that groups objects which have similar characteristics.

It suggests a methodology to investigate the inherent patterns in the relationships between air traffic and macroeconomic development. To do so, it uses data mining techniques, including the grouping of K-means. The most important contribution in the methodology is the ability to select variables objectively and quickly [10].

In [11] four groups of municipalities were identified based on socioeconomic indicators. The purpose was to examine the socioeconomic differences among Slovenian municipalities

and classify them into relatively homogeneous groups. The classified groups based on socioeconomic indicators reflect their development features, the results confirm the fact that the eastern part is less developed meanwhile the western part is the most developed in Slovenia. There is a small group of municipalities where the socioeconomic situation is grave.

In [12] the purpose was to identify municipalities where aspirations for energy autonomy could make technical and economic sense, consequently replicate successful projects in other municipalities within the same group; a cluster analysis to establish a municipal typology is used, in order to analyze the techno-economic municipalities appropriateness, for the autonomous energy systems. The results identify municipalities which successful measures from other municipalities can be applied to and provide a basis for future energy studies around the country.

III. THEORETICAL BACKGROUND

The K-Means technique is an unsupervised clustering algorithm, used with large amounts of data. The objective of the K-Means algorithm is to find "K" groups (clusters) among the data set. This algorithm is a grouping technique that is used in different machine learning applications [13] [14].

The K-means clustering algorithm [9] has been discovered more than 50 years ago by Steinhaus (1956), since then it has been applied in various fields of knowledge such as marketing, psychology, medicine, social sciences and biology, becoming one of the most widely used methods for its simplicity, easy implementation and efficiency [15].

The K-Means algorithm works iteratively by assigning each row of the input data set to one of the "K" groups based on their characteristics. The columns (variables) are grouped considering the similarity. The result of running the algorithm is:

- Each group centroids are coordinates of each K set that will be used to label new data sets.
- Labels for the training dataset. Each tag belongs to one of the K clusters formed.

The clusters adjust to a new position in each iteration of the process, until the algorithm converges. Once the centroids have been found, the data must be analyzed to observe which characteristics are unique, regarding the other groups. These groups are the labels that the algorithm generates.

A. K-Means Algorithm

The K-means algorithm begins by specifying a set of initial cluster centers that are derived from data. Then assigns the data to the most similar cluster, based on the input variables values. After all cases have been assigned, the cluster centers are updated to reflect the new dataset assigned to each cluster. Then the records are checked once more to see if they need to be reassigned to a different group, and the data allocation and iteration process continues until the maximum number of iterations is reached or the change between one iteration and the next does not exceed a specified threshold. Depending on the similarity or dissimilarity characteristics, data sets are

grouped into several different groups, for similar data within the same group and for different data among groups [16].

The K-means algorithm is one of the most important unsupervised clustering algorithms [17] that produces high quality results in less computation time [1]. The K-means machine learning algorithm [18] is used to group a known, assumed, or indicated in advance dataset. In [5] K-means is a classic prototype-based partitioning grouping technique that attempts to group data into k groupings that have been specified by the user.

Steps for executing K-means clustering algorithms [4] [5]

- 1) Define the cluster number, k.
- 2) Initialize the k centroids of the group.
- 3) Assign the n data points to the closest clusters.
- 4) Update the centroids of each group.
- 5) Repeat steps 3 and 4 until there are no more changes in the positions of the centroids.

[19] The k-means algorithm is one of the most widely used grouping algorithms, it is designed to group numerical data, in which each grouping contains a center called centroid. The algorithm works with cases in which all the variables are the quantitative type, and the Euclidean quadratic distance [20].

$$d(x_i, x'_i) = \sum_{j=1}^p (x_{ij} - x'_{ij})^2 = \|x_{ij} - x'_{ij}\|^2 \quad (1)$$

it is chosen as a difference measurement. Take into account the weights in the Euclidean distance can be used by redefining the x_{ij} values.

The scatter points can be written as:

$$W(C) = \sum_{k=1}^K \sum_{C(i)=k} \sum_{C(i')=k} \|x_i - x_{i'}\|^2 \quad (2)$$

$$W(C) = \sum_{k=1}^K N_k \sum_{C(i)=k} \|x_i - \bar{x}_{i'}\|^2 \quad (3)$$

In the k-means algorithm, the sum of the squares of the Euclidean distances of data points to their closest representatives, is used to quantify the objective function of the clustering [20]. Therefore, we have:

$$Dist(\bar{X}_i, \bar{Y}_j) = \|\bar{X}_i - \bar{Y}_j\|^2 \quad (4)$$

where $\bar{x}_k = (\bar{x}_{1k}, \dots, \bar{x}_{pk})$, is the vector of means associated with the k-th cluster, and $N_k = \sum_{i=1}^N I(C(i) = k)$. Thus, the criterion is to give the N observations to the K clusters so within each cluster the average of the differences of each observation to the mean of the cluster, defined by the cluster points, might be minimal.

IV. RESULTS

This section shows the classification of municipal management in Peru local governments, after applying the K-Means Clustering algorithm.

A. Research Variables

58 variables have been used that have been gotten from the areas of human resources, heavy machinery and operational vehicles, information and communication technologies, municipal planning, municipal finances, local economic

development, social services, solid waste management, cultural, recreational and sports facilities, public security, disaster risk management, protection and conservation of the environment of all local governments in the 24 departments of Peru and the constitutional province of Callao. The variables were the following:

- Municipal staff
- Female municipal staff
- Male municipal staff
- Technical assistance requirements
- Training requirements
- Heavy Operating Machinery
- Vehicles and Equipment
- Operating Computers
- Internet access
- Standard Transparency Portal
- Telephone lines installed and in service
- Urban and / or rural management and development instruments
- Sole Code of Administrative Procedures
- Method of payment of taxes and / or services
- Coercive Execution office
- Municipal revenue collected at departmental level
- Current income
- Capital income
- Transfers
- Financing
- Municipal expenses executed
- Capital expenditures
- Donations and capital transfers
- Debt Service
- Operating licenses for business
- Building permits
- Activities to promote Micro and Small Businesses
- Activities to promote crafts
- Social Organizations
- Care Center for the Elderly (CIAM)
- Municipal Office of Attention to People with Disabilities (OMAPED)
- Municipal Ombudsman offices for Children and Adolescents (DEMUNA)

- Municipalities that collected solid waste (garbage)
- Daily average amount of solid waste collection
- Solid waste management instruments
- Number of users served at the Municipal Library
- Number of visitors / users to culture centers
- Number of visitors / users to theaters
- Number of visitors / users to museums
- Municipal police force
- Female Municipal police force
- Male Municipal police force
- Interventions registered by the Municipal police force
- Equipment and infrastructure for Municipal police force activities
- Dividing by sectors for patrolling activities
- Integrated patrol of municipal police force with the Peruvian Police force
- District Citizen Security Committee (CODISEC)
- Citizen Security Plan
- Disaster Risk Management Working Groups (GTGRD)
- National Civil Defense System Platform (PDC)
- Municipal Emergency Operations Center (COEL)
- Results-based budgeting Program N ° 068
- Natural and man-made hazards
- Municipalities with an Environmental Office or Unit
- Registered environmental complaints
- Elements that cause environmental pollution
- Environmental management instruments available in the municipality
- Actions carried out by the municipality to encourage environmental conservation

B. K-Means Clustering Algorithm

The K-Means Clustering algorithm used has the following characteristics:

- Number of clusters: 4
- Variables: 58

Build Settings

- Generate distance field: true
- Cluster label: String
- Label prefix: cluster
- Optimize: Memory

- Mode: Expert
- Stop on: Default

Training Summary

- Algorithm: K-means
- Model type: Clustering

Fig. 1 shows the K-means algorithm used to establish the classification of municipal management in local governments in Peru.

C. K-Means Clustering Algorithm Results

Fig. 2 shows the classification of local governments by department. The K-Means algorithm considers that the 32% (8) of the local governments divided by departments in Peru is in cluster 1, the 8% (2) of them to cluster 2, the 28% (7) to cluster 3 and to cluster 4 another 32% (8).

Fig. 3 shows the intensity of membership of the departments in the groups formed by K-means clustering.

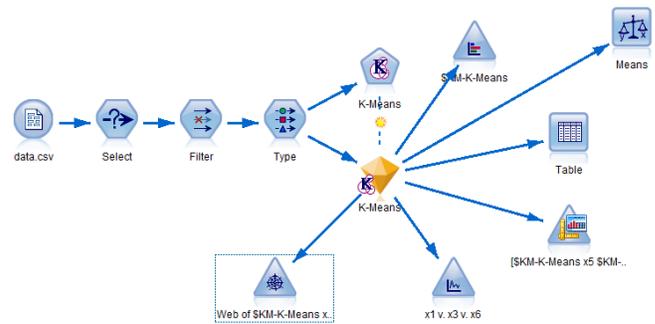


Fig. 1. K-Means Clustering Algorithm.

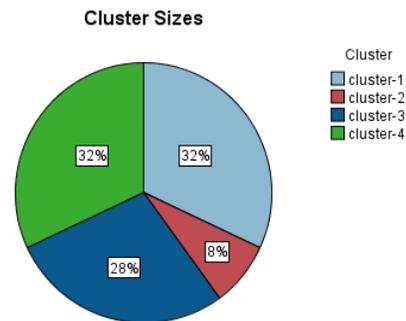


Fig. 2. Cluster Sizes using the K -Means Clustering Algorithm.

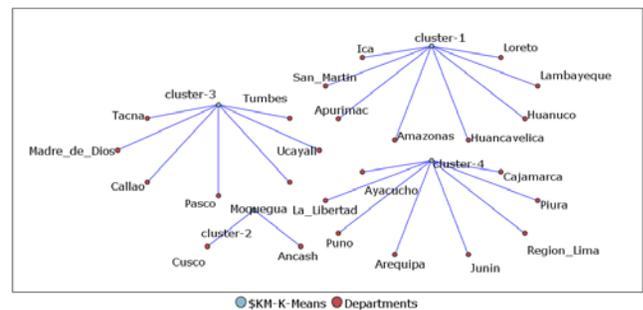


Fig. 3. Membership of Departments to Groups Formed by K -means Clustering.

D. Evaluation and Validation of the Algorithm

The K-means clustering algorithm shows the classification of municipal management in the local governments of Peru in 4 clusters:

Table I shows that in cluster 1 there is an average of 4898 workers per department, of which 1623 are female and 3265 are male. In cluster 2, on average there are 19796 workers per department, 5857 of which are female and 13939 are male. Cluster 3 has an average of 3,578 workers per department, divided in 1,423 female workers and 2,155 male employees. Cluster 4 averaged 9075 workers per municipality, of which 3077 were female and 5998 males. Likewise, it is observed that 62 municipalities require technical assistance in cluster 132 municipalities in cluster 2, 16 municipalities in cluster 3, and 99 municipalities in cluster 4. It is also shown that 67 municipalities in cluster 135 in cluster 2, 17 municipalities in cluster 3 and 104 in Cluster 4 requires training.

According to Table II, it can be seen that in cluster 1, 37 municipalities have at least heavy operating machinery and 65 municipalities have vehicles and equipment. In cluster 2 shows that 98 municipalities have at least heavy operating machinery and 133 municipalities with vehicles and equipment. Cluster 3 describes that 13 municipalities have at least heavy operating machinery and 18 have vehicles and equipment. In cluster 4 there are 76 municipalities that at least have one heavy operating machinery and 104 with vehicles and equipment.

Table III shows that in cluster 1 there are 2,117 computers on average per department, of which 1,812 have internet access and 23 municipalities have their transparency portal. Cluster 2 shows 7126 operating computers on average by department of which 5716 with internet access and 41 municipalities with its transparency portal. Cluster 3 presents an average of 1,513 operating computers per department with 1,219 computers with internet access and 10 municipalities with its transparency portal. Cluster 4 shows 4049 operating computers of which 3604 have internet access and with 37 municipalities that have their transparency portal.

According to Table IV it is observed in cluster 1 that on average 65 municipalities have at least one management instrument, among which the Organizations and Functions Regulation stands out, and 49 municipalities have the Sole Code of Administrative Procedures (TUPA). In cluster 2, an average of 127 municipalities have a management instrument and 100 municipalities have the Sole Code of Administrative Procedures (TUPA). Cluster 3 shows that on average 17 municipalities have a management instrument and 15 municipalities have the Sole Code of Administrative Procedures (TUPA). Cluster 4 shows that on average 94 municipalities have at least one management instrument and 76 municipalities have the Sole Code of Administrative Procedures (TUPA).

According to Table V, from the total municipal income collected it is observed that in cluster 1 the 46% is from Central Government transfers, financing 42%, current Income 12% and capital Income 0,03%. Cluster 2 shows that 67% corresponds to transfers from the Central Government, 25% to financing, 8% to current income and 0.07% to capital income. In cluster 3

it is seen that 51% comes from transfers from the Central Government, 29% to financing, current income 20% and capital income 0.42%. Cluster 4 shows that 50% corresponds to transfers, 36% to financing, 14% to current income and 0.12% to capital income.

TABLE I. HUMAN RESOURCES

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|-----------------------------------|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Municipal staff | | | | |
| Mean | 4898 | 19796 | 3578 | 9075 |
| Standard deviation | 1602 | 14750 | 2280 | 2327 |
| Technical assistance requirements | | | | |
| Mean | 62 | 132 | 16 | 99 |
| Standard deviation | 20 | 41 | 8 | 22 |
| Training requirements | | | | |
| Mean | 67 | 135 | 17 | 104 |
| Standard deviation | 22 | 35 | 9 | 23 |

TABLE II. HEAVY MACHINERY AND OPERATING VEHICLES

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---------------------------|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Heavy Operating Machinery | | | | |
| Mean | 37 | 98 | 13 | 76 |
| Standard deviation | 13 | 7 | 7 | 18 |
| Vehicles and Equipment | | | | |
| Mean | 65 | 133 | 18 | 104 |
| Standard deviation | 20 | 32 | 8 | 20 |

TABLE III. INFORMATION AND COMMUNICATION TECHNOLOGIES

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|------------------------------|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Operating Computers | | | | |
| Mean | 2117 | 7126 | 1513 | 4049 |
| Standard deviation | 412 | 3722 | 706 | 692 |
| Internet access | | | | |
| Mean | 1812 | 5716 | 1219 | 3604 |
| Standard deviation | 404 | 2563 | 543 | 765 |
| Standard Transparency Portal | | | | |
| Mean | 23 | 41 | 10 | 37 |
| Standard deviation | 7 | 0 | 4 | 8 |

TABLE IV. MUNICIPAL PLANNING

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|------------|-----------|-----------|-----------|
| | 14 Records | 2 Records | 8 Records | 1 Records |
| Urban and / or rural management and development instruments | | | | |
| Mean | 65 | 127 | 17 | 94 |
| Standard deviation | 21 | 33 | 8 | 20 |
| Single Text of Administrative Procedures | | | | |
| Mean | 49 | 100 | 15 | 76 |
| Standard deviation | 14 | 24 | 6 | 16 |

TABLE V. MUNICIPAL FINANCE

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Municipal revenue collected at the departmental level | | | | |
| Mean | 726562 | 2626041 | 405445 | 1359223 |
| Standard deviation | 151935 | 490451 | 190745 | 382405 |
| Municipal expenses executed | | | | |
| Mean | 674408 | 2051009 | 349766 | 1178082 |
| Standard deviation | 129651 | 751169 | 184323 | 294256 |

On the other hand, in cluster 1 it is shown that 63% of the expenses executed in the municipalities were spent capital expenses, cluster 2 shows 67%, in cluster 3 42% to capital expenses and in Cluster 4 the municipalities allocated 59% to capital expenses.

Table VI shows in cluster 1 the number of operating licenses for establishments granted by local governments was in average 1,564 licenses, in construction 1,551 licenses and 37 municipalities reported that they carried out activities to promote Micro and Small Businesses. In cluster 2, 2,821 operating licenses were granted, 1,230 for construction, and 76 municipalities reported that they carried out actions Activities to promote Micro and Small Businesses. In cluster 3, were granted 751 operating licenses, 407 building permits, and 11 municipalities reported that they carried out actions to promote Micro and Small Businesses. Cluster 4 shows 2967 operating licenses granted, 2444 building permits, and 62 local governments carried out activities to encourage Micro and Small Businesses.

According to Table VII, it is shown that in cluster 1 there are an average of 87,206 users in social organizations (glass of milk, soup kitchen, mothers' club and youth organizations), 4841 beneficiaries in the Care Center for the Elderly (CIAM), 6126 People with Disabilities and an average of 4387 cases attended each year in the Municipal Ombudsman for Children and Adolescents (DEMUNA). Cluster 2 shows an average of 118,281 users in social organizations, 9,133 beneficiaries in the Care Center for the Elderly, 9,190 people with disabilities and 17,257 cases on average attended each year by the Municipal Ombudsman for Children and Adolescents. Cluster 3 shows 34,890 users on average in social organizations, 3,307 CIAM beneficiaries, 3,153 people with disabilities and 5,626 cases attended on average by DEMUNA. Cluster 4 shows an average of 138087 users in social organizations, 11,768 beneficiaries of CIAM, 9,512 people with disabilities and 9,123 cases attended by DEMUNA.

Table VIII shows that 67 municipalities collect solid waste, on an average of 411,894 kilos per department in cluster 1. In cluster 2, there are 137 municipalities that collect solid waste, with an average amount of 742400 kilos per department. In cluster 3 there are 18 municipalities that collect solid waste, with an average amount of 297,729 kilos per department. Cluster 4 shows 104 municipalities that collect solid waste, with an average amount of 816506 kilos per department.

Table IX shows in cluster 1, the number of users served in the municipal library was 23075 users on average, 2741 users in cultural centers, in theaters 2471 and in museums 1795 users. In cluster 2, 103883 users were served in the municipal library, 19630 users in cultural centers, 79457 users in the theaters and in museums 45623 users. Cluster 3 served 5881 users in the municipal library, in cultural centers 2472 users, in the theaters 7649 and in museums 4768 users. In cluster 4 the municipal library served 73529 users, 13133 users in the House of Culture, in theaters 13917 users and in museums 14930 users.

TABLE VI. LOCAL ECONOMIC DEVELOPMENT

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Operating licenses for establishments | | | | |
| Mean | 1564 | 2821 | 751 | 2967 |
| Standard deviation | 553 | 327 | 467 | 1045 |
| Building licenses | | | | |
| Mean | 1551 | 1230 | 407 | 2444 |
| Standard deviation | 1321 | 415 | 215 | 1776 |
| Actions to encourage Micro and Small Businesses | | | | |
| Mean | 37 | 76 | 11 | 62 |
| Standard deviation | 19 | 35 | 6 | 10 |

TABLE VII. SOCIAL SERVICES

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Social Organizations | | | | |
| Mean | 87206 | 118281 | 34890 | 138087 |
| Standard deviation | 30068 | 1237 | 27987 | 54310 |
| Comprehensive Care Center for the Elderly (CIAM) | | | | |
| Mean | 4841 | 9133 | 3307 | 11768 |
| Standard deviation | 1896 | 2819 | 3451 | 5174 |
| Municipal Office of Attention to Persons with Disabilities (OMAPED) | | | | |
| Mean | 6126 | 9190 | 3153 | 9512 |
| Standard deviation | 2894 | 4924 | 2019 | 5975 |
| Municipal Ombudsman for Children and Adolescents (DEMUNA) | | | | |
| Mean | 4387 | 17257 | 5626 | 9123 |
| Standard deviation | 1406 | 7640 | 9094 | 4880 |

TABLE VIII. PUBLIC CLEANING

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Municipalities that carried out solid waste collection (garbage) | | | | |
| Mean | 67 | 137 | 18 | 104 |
| Standard deviation | 21 | 35 | 8 | 22 |
| Average daily amount of solid waste collection | | | | |
| Mean | 411894 | 742400 | 295553 | 816506 |
| Standard deviation | 282801 | 32446 | 297729 | 394566 |

TABLE IX. CULTURAL, RECREATIONAL AND SPORTS INFRASTRUCTURE

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Number of users served in the municipal library | | | | |
| Mean | 23075 | 103883 | 5881 | 73529 |
| Standard deviation | 20759 | 8190 | 7185 | 83444 |
| Number of visitors / users to culture houses | | | | |
| Mean | 2741 | 19630 | 2472 | 13133 |
| Standard deviation | 2695 | 27675 | 3661 | 11642 |
| Number of visitors / users to theaters | | | | |
| Mean | 2471 | 79457 | 7649 | 13917 |
| Standard deviation | 2971 | 32311 | 15970 | 23538 |
| Number of visitors / users to museums | | | | |
| Mean | 1795 | 45623 | 4768 | 14930 |
| Standard deviation | 1869 | 47324 | 10916 | 13030 |

Table X shows in cluster 1 an average of 401 Municipal police force by department, 13% female, 87% male, and an average of 3,283 interventions by department were performed. Cluster 2 shows an average of 1313 Municipal police force troops, 14% female, 86% male, performed 11086 interventions. Cluster 3 shows an average of 359 serenade personnel by departments, 18% female, 82% male, and carried out 9486 interventions. Cluster 4 presents 1040 Female municipal police force troops on average by department, 12% female, 88% male, and performed 24146 interventions.

Table XI shows in cluster 1, that 52 municipalities have working groups for disaster risk management, an average of 52 municipalities have National Civil Defense System Platform and 21 with local emergency operations centers. Cluster 2 shows that on average 109 municipalities have working groups for disaster risk management, 111 municipalities have National Civil Defense System Platform and 56 have local emergency operations centers. Cluster 3 shows 14 municipalities that have working groups for disaster risk management, 15 municipalities National Civil Defense System Platform, and 8 that have local emergency operations centers. Cluster 4 shows 74 municipalities with disaster risk management working groups, 76 municipalities with Civil defense platforms and 35 with local emergency operations centers.

In Table XII it is observed that in cluster 1, 43 municipalities have an environmental office or unit, 23 municipalities received environmental complaints, 49 municipalities carried out actions to encourage environmental conservation. Cluster 2, shows 67 municipalities with an environmental unit, 38 of them have gotten environmental complaints and 88 municipalities carried out actions to conserve the environment. While in cluster 3, 14 municipalities have an environmental unit, 6 have received environmental complaints and 13 carried out actions to conserve the environment. Cluster 4 shows 66 municipalities with an environmental unit, 35 received complaints and 73 carried out actions to conserve the environment.

E. Classification Results

Fig. 4 shows the classification of municipal management in the local governments of Peru in the 24 departments and the

constitutional province of Callao, resulting from the application of the K-Means clustering algorithm.

TABLE X. CITIZEN SECURITY

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Serenade cash | | | | |
| Mean | 401 | 1313 | 359 | 1040 |
| Standard deviation | 253 | 293 | 392 | 427 |
| Female serenade cash | | | | |
| Mean | 51 | 181 | 66 | 126 |
| Standard deviation | 47 | 71 | 93 | 54 |
| Male serenade cash | | | | |
| Mean | 350 | 1133 | 293 | 914 |
| Standard deviation | 214 | 221 | 301 | 377 |
| Interventions registered by the serenade | | | | |
| Mean | 3283 | 11086 | 9486 | 24146 |
| Standard deviation | 2015 | 2879 | 10300 | 19387 |

TABLE XI. DISASTER RISK MANAGEMENT

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Disaster Risk Management Working Groups (GTGRD) | | | | |
| Mean | 52 | 109 | 14 | 74 |
| Standard deviation | 14 | 1 | 7 | 11 |
| Civil Defense Platform (PDC) | | | | |
| Mean | 52 | 111 | 15 | 76 |
| Standard deviation | 14 | 8 | 7 | 10 |
| Local Emergency Operations Center (COEL) | | | | |
| Mean | 21 | 56 | 8 | 35 |
| Standard deviation | 8 | 6 | 3 | 17 |

TABLE XII. PROTECTION AND CONSERVATION OF THE ENVIRONMENT

| Variable | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|-----------|-----------|-----------|-----------|
| | 8 Records | 2 Records | 7 Records | 8 Records |
| Municipalities that have an Environmental Office or Unit | | | | |
| Mean | 43 | 67 | 14 | 66 |
| Standard deviation | 15 | 13 | 6 | 14 |
| Registered environmental complaints | | | | |
| Mean | 23 | 38 | 6 | 35 |
| Standard deviation | 10 | 13 | 2 | 7 |
| Actions carried out by the municipality to encourage environmental conservation | | | | |
| Mean | 49 | 88 | 13 | 73 |
| Standard deviation | 13 | 15 | 6 | 13 |

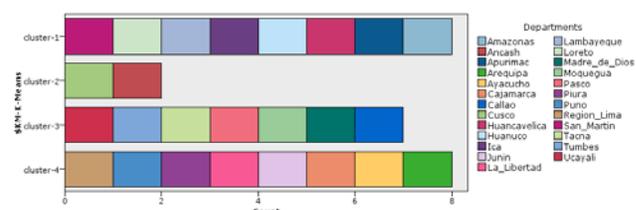


Fig. 4. Classification of Municipal Management in Local Governments of Peru.

V. DISCUSSION

The local governments of the departments of Peru have been distributed as follows:

1) *Cluster 1*: Formed by municipal governments of 8 departments: Amazonas, Apurímac, Huancavelica, Huánuco, Ica, Lambayeque, Loreto and San Martín.

2) *Cluster 2*: Formed by municipal governments of Ancash and Cusco.

3) *Cluster 3*: Formed by municipal governments of 7 departments: Constitutional Province of Callao, Madre de Dios, Moquegua, Pasco, Tacna, Tumbes and Ucayali.

4) *Cluster 4*: Formed by municipal governments of 8 departments: Arequipa, Ayacucho, Cajamarca, Junín, La Libertad, Region Lima, Piura and Puno.

From 100% of the local governments of the departments of Peru, it is observed that 32% of them are in Group 1, 8% in Group 2, 28% in the Third Group and 32% are found in Group 4. In [11] the classification of four groups of municipalities was the focus, because it offers a more detailed view of the socio-economic differences that exist between Slovenian municipalities and for that reason the results are more informative too. The variables that were used were socioeconomic development indicators. The K-Means algorithm is used to improve the results of Ward's method.

VI. CONCLUSIONS

This work uses the K-means algorithm to form a classification model for municipal management of local governments in Peru based on their indicators. The results have been presented in four groups, which shows a more detailed and informative picture of municipal management. The factors considered are related to human resources, heavy machinery and operational vehicles, information and communication technologies, municipal planning, municipal finances, local economic development, social services, solid waste management, cultural, recreational and sports facilities, public security, disaster risk management and environmental protection and conservation factor.

Cluster 1, formed by the municipal governments of Amazonas, Apurímac, Huancavelica, Huánuco, Ica, Lambayeque, Loreto and San Martín. Cluster 2 by municipal governments of Ancash and Cusco. Cluster 3 the municipal governments of the constitutional province of Callao, Madre de Dios, Moquegua, Pasco, Tacna, Tumbes and Ucayali. Cluster 4 the municipal governments of Arequipa, Ayacucho, Cajamarca, Junín, La Libertad, Region Lima, Piura and Puno. According to the results, it can be seen that the municipal governments of the constitutional province of Callao, Madre de Dios, Moquegua, Pasco, Tacna, Tumbes and Ucayali show poor municipal management and; on the other hand, the municipal governments of the Ancash and Cusco with efficient municipal management.

The application of the K-means algorithm is a contribution to the municipalities, in order to improve the municipal

management by making right decisions for the benefit of their inhabitants.

VII. FUTURE WORK

Research efforts in this work, have been focused on some specific questions, the application of K-means algorithms in the management of municipalities; the analysis of municipal management by regions has been reserved for future work.

REFERENCES

- [1] A. Sewisy, R. M. Abd El-Aziz, M. H. Marghny, and A. I. Taloba, "Fast Efficient Clustering Algorithm for Balanced Data," *SSRN Electron. J.*, vol. 5, no. 6, pp. 123–129, 2015, doi: 10.2139/ssrn.2545138.
- [2] G. C. Medina, J. C. Chahuayo, A. C. Huamanhorqqe, Y. R. Dulanto, F. C. Fonseca, and A. Q. Saavedra, "Peru: Municipal Management Indicators 2019," 2019.
- [3] S. Ghosh and S. Kumar Dubey, "Comparative Analysis of K-Means and Fuzzy C-Means Algorithms," *Int. J. Comput. Eng. Res.*, vol. 4, no. 4, pp. 35–39, 2013.
- [4] C. Pascal, S. Ozuomba, and C. Kalu, "Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 10, pp. 40–44, 2015, doi: 10.14569/ijarai.2015.041007.
- [5] T. Tamer, A. Haydar, and I. Ersan, "Data Distribution Aware Classification Algorithm based on K-Means," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, 2017, doi: 10.14569/ijacsa.2017.080946.
- [6] J. Kachaoui and A. Belangour, "Enhanced Data Lake Clustering Design based on K-means Algorithm," vol. 11, no. 4, pp. 547–554, 2020.
- [7] M. Zulfadhilah, Y. Prayudi, and I. Riadi, "Cyber Profiling Using Log Analysis And K-Means Clustering," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 7, pp. 430–435, 2016, doi: 10.14569/ijacsa.2016.070759.
- [8] Y. Yu, S. A. Velastin, and F. Yin, "Automatic grading of apples based on multi-features and weighted K-means clustering algorithm," *Inf. Process. Agric.*, no. xxxx, 2019, doi: 10.1016/j.inpa.2019.11.003.
- [9] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020, doi: 10.1016/j.apr.2019.09.009.
- [10] J. H. Chen, H. H. Wei, C. L. Chen, H. Y. Wei, Y. P. Chen, and Z. Ye, "A practical approach to determining critical macroeconomic factors in air-traffic volume based on K-means clustering and decision-tree classification," *J. Air Transp. Manag.*, vol. 82, no. February 2019, p. 101743, 2020, doi: 10.1016/j.jairtraman.2019.101743.
- [11] J. Rován and J. Sambt, "Socio-economic Differences Among Slovenian Municipalities: A Cluster Analysis Approach," *Dev. Appl. Stat.*, 2003, [Online]. Available: <http://www.stat-d.si/mz/mz19/rovan.pdf>.
- [12] J. M. Weinand, R. McKenna, and W. Fichtner, "Developing a municipality typology for modelling decentralised energy systems," *Util. Policy*, vol. 57, no. March, pp. 75–96, 2019, doi: 10.1016/j.jup.2019.02.003.
- [13] L. Wei, R. Zhou, X. Wang, C. Zhu, J. Yin, and X. Zhang, "Subspace segmentation via self-regularized latent K-means," *Expert Syst. Appl.*, vol. 136, pp. 316–326, 2019, doi: 10.1016/j.eswa.2019.06.047.
- [14] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, "Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm," *Comput. Electron. Agric.*, vol. 165, no. March, p. 104962, 2019, doi: 10.1016/j.compag.2019.104962.
- [15] L. Bai, J. Liang, and F. Cao, "A multiple k-means clustering ensemble algorithm to find nonlinearly separable clusters," *Inf. Fusion*, vol. 61, no. January, pp. 36–47, 2020, doi: 10.1016/j.inffus.2020.03.009.
- [16] E. Zhu, Y. Zhang, P. Wen, and F. Liu, "Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index," *Neurocomputing*, vol. 363, pp. 149–170, 2019, doi: 10.1016/j.neucom.2019.07.048.

- [17] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using K-Means algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, 2020, doi: 10.1016/j.jksuci.2019.12.011.
- [18] P. J. Jones et al., "FilterK: A new outlier detection method for k-means clustering of physical activity," *J. Biomed. Inform.*, vol. 104, p. 103397, 2020, doi: 10.1016/j.jbi.2020.103397.
- [19] Y. Gimenez, "Unsupervised Classification: The K – Means Method," Buenos Aires' University, 2010.
- [20] S. Mita, Y. Yamazoe, T. Kamataki, and R. Kato, Metabolic activation of a tryptophan pyrolysis product, 3-amino-1-methyl-5H-pyrido[4,3-b]indole(Trp-P-2) by isolated rat liver nuclei, vol. 14, no. 3. 1981.