Determination of delivery routes using the k-means and Elbow algorithms

Juan José Paniagua Medina[1]
University of Guanajuato, Multidisciplinary of Studies Department
Yuriria, Guanajuato, México.
jj.paniaguamedina@ugto.mx

Dr. Roberto Baeza Serrato[2]
University of Guanajuato, Multidisciplinary Studies Department
Yuriria, Guanajuato, México.
r.baeza@ugto.mx

***Abstract*— The purpose of this work is to define the delivery routes of product from a hen farm dedicated to egg production, to perform this task were implemented the k-means and Elbow algorithms in MATLAB software getting the number of delivery routes and the customers that form each of the routes. The results of this work show that five routes guarantee the delivery of the product with the shortest distance traveled. In the same way with help of the Google tools the area formed by the delivery routes path was analyzed reiterating what was illustrated by the results obtained.**

***Keywords*— Routes, k-means algorithm; Elbow algorithm; MATLAB; path.**

I. Introduction

The poultry industry is the sector within the agricultural activities with the highest rate of growth and industrialization in most parts of the world [1]. Currently, the main egg producers in the world are China, the United States, India and Mexico, which together produce approximately 53.7% of the world production [2]. Egg production ranges from small family businesses, which help to ensure vital needs and supply local or specialized markets, to large industrialized companies [3]. Egg marketing depends on logistic, which is essential for the product to be moved to the location.

Logistics comprises several basic activities within a company such as sourcing, production and distribution [4]. Within these activities, transportation is a vital element that allows the movement of finished products, raw materials and even people through transport chains designed to deliver on time and manner [5]. Transportation is one of the key points in customer satisfaction by adding value to the products when they are delivered on time, undamaged and in the required quantities [6]. However, it is one of the highest logistic costs and constitutes a representative proportion of products prices. The expenses generated by transportation is quite significant in logistics activities [7]. That is why it is necessary to implement techniques to establish routes with the shortest travel distances.

There are a variety of techniques used for clustering, which is defined as an unsupervised learning where data is clustered according to the similarity between them [8], the k-means algorithm solves the clustering problem by minimizing the sum of squared errors (SSE), uses distances as metrics and with the given number of k classes assigns the data to a set of points called centroids iteratively [9]. To know the optimal number of k classes, the Elbow algorithm is used which uses the sum of SSE as a performance indicator [10].

The development of this work aims to help a small hen farm dedicated to the commercialization of eggs to define the delivery routes of its 50 customers, which prior to this work did not have a structured logistic process for the delivery of its product.

The k-means clustering algorithm was implemented together with the Elbow algorithm to find the sets of customers with the shortest distances between them, which define each of the delivery routes, thus ensuring the shortest possible travel distances to transport the product from the farm to the customers.

II. Literature review

Clustering techniques are very important for knowledge acquisition. Due to the great advances in digital platforms, the generation and the way in which data is stored has increased exponentially, which is why it is necessary to apply techniques to improve the use of this information. Among these techniques is the k-means algorithm which is used in several sectors.

In recent years applications of the k-means algorithm are applied in fields such as image processing to segment objects in images [11]. In data mining the algorithm is used to cluster the grades of students in a school according to the performance of each student in each of their subjects [12]. In the energy sector specifically in wind energy production, the k-means algorithm is used to improve the efficiency of clustering wind farms according to parameters such as wind speed [13]. In medicine it is used to know the constancy of blood donors, also to manage blood transfusion centers in a proper way [14]. Environmental protection is another sector in which the k-means algorithm is used, grouping statistical data on the growth of wild animals [15]. In finance it is used in conjunction with the Elbow algorithm to cluster risks in smart product investments, in order to perform an analysis on the efficiency of investments [16]. In logistics area it is used to segment customers according to different variables such as destination cities, interests, products and the price to manage the marketing of the company [17]. The k-means algorithm was used to analyze the data of cases and deaths caused by COVID-19 in Southeast Asia and thereby solving the transmission problem [18]. In energy sector k-means and Elbow algorithms help to analyze the energy consumption structure of different areas, with the results obtained it was verified that the operation is carried out according to the energy laws [19]. Text classification is another task where this clustering algorithm is used to classify text documents from commonly used preprocessed databases [20]. In agriculture the study of climate change helps farmers to make decisions about sowing dates, the k-means algorithm along with multiple linear regression is used for building a predictive model to help with making that decision [21]. Finally, to streamline package delivery by trucks and drones the k-means algorithm is implemented to minimize the time and energy in package delivery by drones [22].

The aforementioned research shows the application of the k-means algorithm in different areas of knowledge with the aim of analyzing the behavior of various data sets. This research adds to the literature the application of the k-means algorithm helps to define the delivery routes of a farm dedicated to the commercialization of eggs, the study can be extrapolated to any business that needs to minimize the distances traveled in their deliveries.

III. METHODOLOGY

The methodology for the study consists of four stages, ranging from the collection of customer coordinates to the definition of delivery routes with the shortest distances. *Fig. 1* shows a diagram with the process performed in each stage.

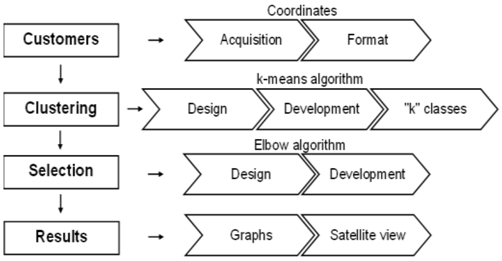

Fig. 1. *General methodology diagram.*

The flowchart shown in *Fig. 2* was used to design and implement the k-means algorithm.

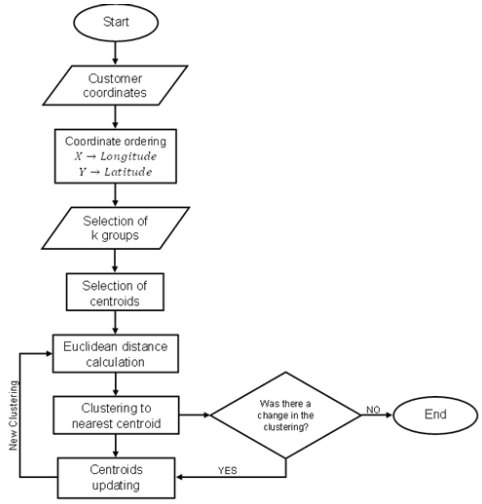

Fig. 2. *Flowchart of the k-means algorithm.*

IV. RESULTS AND DISCUSSION

For the implementation of this work were used the following coordinates of the 50 customers of the farm, the coordinates were obtained with the information of the owner of the farm and with the help of the Google Maps, in *Tab. 1* shows the names of the customers along with their coordinates in decimal form. Where: X = Longitude, Y = Latitude.

TABLE I. CUSTOMER COORDINATES.

Customers	X	Y
Fruit shop “El Árbol De Cristo”	-101.12998	20.21269
Packaging company “Yuriria”	-101.1382	20.20524
Grocery Store “López”	-101.13844	20.2024
Private House	-101.13221	20.20199
Fruit shop “El mercado”	-101.12925	20.21355
Bistro “San Antonio”	-101.13729	20.21295
Burguers Shop	-101.13858	20.20482
Butcher's Shop “Pérez”	-101.13496	20.20958
Grocery Store “García”	-101.17137	20.11949
Food Warehouse	-100.8827	20.22842
Supermarket “Paraísos Rangel”	-101.19241	20.12715
Grocery Store “Maru”	-101.2011	20.37918
Grocery Store "Sonido Negro"	-101.18663	20.13157
Grocery Store "Álvarez"	-101.17875	20.13731
Grocery Store "Auditorio"	-101.18727	20.11786
Grocery Store "López"	-101.19153	20.12459
Grocery Store "Lemus"	-101.2047	20.12974
Grocery Store "La Esperanza"	-101.20717	20.11982
Grocery Store "de Uriangato"	-101.17416	20.14916
Grocery Store "Romero"	-101.19046	20.1422
Grocery Store "Téllez"	-101.1746	20.12742
Grocery Store "Diego"	-101.16037	20.14988
Grocery Store "Parangarico"	-101.16706	20.19007
Supermarket "Casacuaran"	-101.0418	20.19958
Supermarket "Nayeli"	-100.88063	20.21175
Grocery Store "TONATIUH"	-100.87831	20.20631
Grocery Store "tomas o no tomas"	-100.89018	20.2021
Fruit shop "Rivera"	-100.72185	20.02878
Fruit shop "Ramírez"	-100.78993	20.34677
Fruit shop "Herrera"	-101.18444	20.38699
Fruit shop "Yépez"	-101.19579	20.39336
Fruit shop "Hernández"	-101.18413	20.39429
Grocery Store "Velia"	-101.04493	20.20055
Grocery Store "Silva"	-100.81585	20.51157
Grocery Store "Guanajuato"	-100.81611	20.50022
Fruit shop "El Mandado"	-100.80818	20.52907
Fruit shop "Doña Jose"	-100.83044	20.52317
Restaurant "Charly"	-101.06569	20.37283
Supermarket "Básicos"	-101.06299	20.37636
Snack Shop "La Glotona"	-101.06479	20.37742
Fruit shop "Tere"	-100.9516	20.48761
Fruit shop "Campos IV"	-100.96701	20.47725
Grocery Store "3B"	-100.96285	20.48447
Supermarket "Acsa"	-101.12492	20.21179
Bakery Shop "Mil Delicias"	-101.18052	20.3109
Grocery Store "González"	-101.17538	20.30965
Restaurant "Lupita"	-101.18223	20.31243
Grocery Store "La Cañada"	-100.94743	20.37381
Bakery Shop "La Providencia"	-100.75861	20.29206
Fruit shop “Los Pinos”	-100.75412	20.29387

Once all the coordinates were collected, they were inserted into the Google Maps tool to visualizethe distribution of the clients, also the location of thefarm was added to observe the distance between the farm and every costumer. *Fig. 3* shows the satellite view with the location of the farm and the costumer.

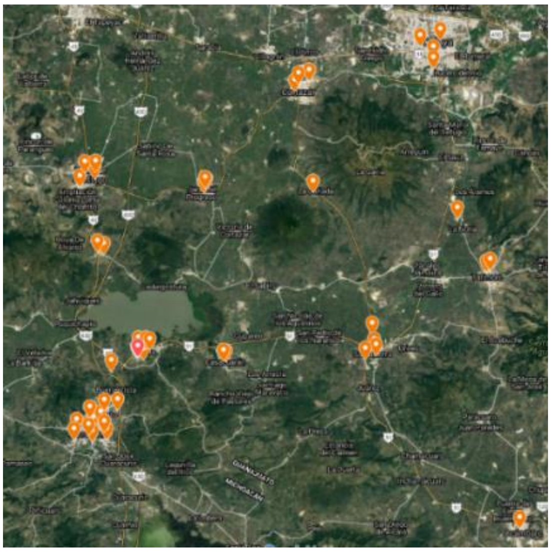

Fig. 3. *Satellite view of the location of the clients (orange dots) and the farm (red dot).*

A. K-MEANS ALGORITHM

The K-means algorithm allows the discovery of clusters in data sets. The K-means clustering algorithm is one of the most used techniques due to its simplicity. It has the objective of dividing a data set into k clusters in which each data belongs to the cluster whose distance is the closest.

For the implementation of the k-means algorithm, the process shown in *Fig. 2* was used, the algorithm was implemented 5 times with values of k = 3, 4, 5, 6 y 7, for each of the implementations, the corresponding number of centroids were selected for each application by randomly using the customer coordinates to assign them to the centroids, once the centroids were assigned, the distances between customers and centroids were calculated using the *Eq. 1*.

$$D(Coor, Cent) = \sqrt{\sum_{i=1}^{n}(Coor_x - Cent_x)^2 + (Coor_y - Cent_y)^2} \quad (1)$$

Where: $D =$ is the Euclidean distance, $Coor =$ customers coordinates, $Cent =$ centroid coordinates, $n =$ number of customers, $Coor_{x,y} =$ longitude and latitude of the customers, $Cent_{x,y} =$ longitude and latitude of the centroids.

With the calculation of the distances, the clustering process is started, customer is clustered with the centroid with the smallest distance calculated. At the end of the clustering, the new centroid location is calculated using the coordinates of all the customers clustered to this one, using *Eq. 2*.

$$Update\ Cent_{x,y} = \frac{1}{m}\sum_{i=1}^{m} Coor_{x,y} \quad (2)$$

Where: $m =$ number of customers clustered at the centroid, $Coor_{x,y} =$ coordinates of the customer assigned to the centroid.

The next step is to cluster each customer again with its nearest centroid and again relocate the centroids over and over again.

The algorithm ends once it is no longer necessary to relocate the centroids and therefore the customers are not clustered. To do this, a comparison of the new clusters of customers with the previous clusters is made to verify if they have changed, if not, the algorithm ends, otherwise, the centroids are updated again, the customers are clustered again until the customers and centroids no longer change their position.

The pseudocode for the implementation of the k-means algorithm is as follows:

k-Means algorithm

Input: $Coor_{x,y}, Cent_{x,y}, n$
Output: d, k
1: **While** (the centroids do not change) **do**
2: **For** n=1 to n=50
3: $D = \sum_{i=1}^{n} \sqrt{\sum (Coor_{x,y} - Cent_{x,y})^2}$
4: ***End For***
5: **If** D is the smallest then
6: The customer is assigned to the centroid
7: **Plot** D
8: **End If**
9: $Update\ Cent_{x,y} = \frac{1}{m} \sum_{i=1}^{m} Coor_{x,y}$
10: **End While**

The pseudocode was developed using the flowchart shown in *Fig. 2*.

The results obtained from the implementation of the k-means algorithm developed in MATLAB with a value of k=3 are shown below, *Fig. 4* shows each of the iterations performed to complete the clusters with the smallest distances between the customer and the corresponding centroid. The first graph shows the randomly selected centroids, the locations of the customers and the clusters formed by the customers to the nearest centroid. The subsequent graphs show the updates of both the centroid location and the change in the clustering formed by the customers to each of the centroids. 3 iterations were necessary to cluster each customer of the farm in its respective route.

Fig. 5 shows a satellite map with the clusters formed for the customers, and also shows the polygons for each of the customers belonging to each of the 3 routes.

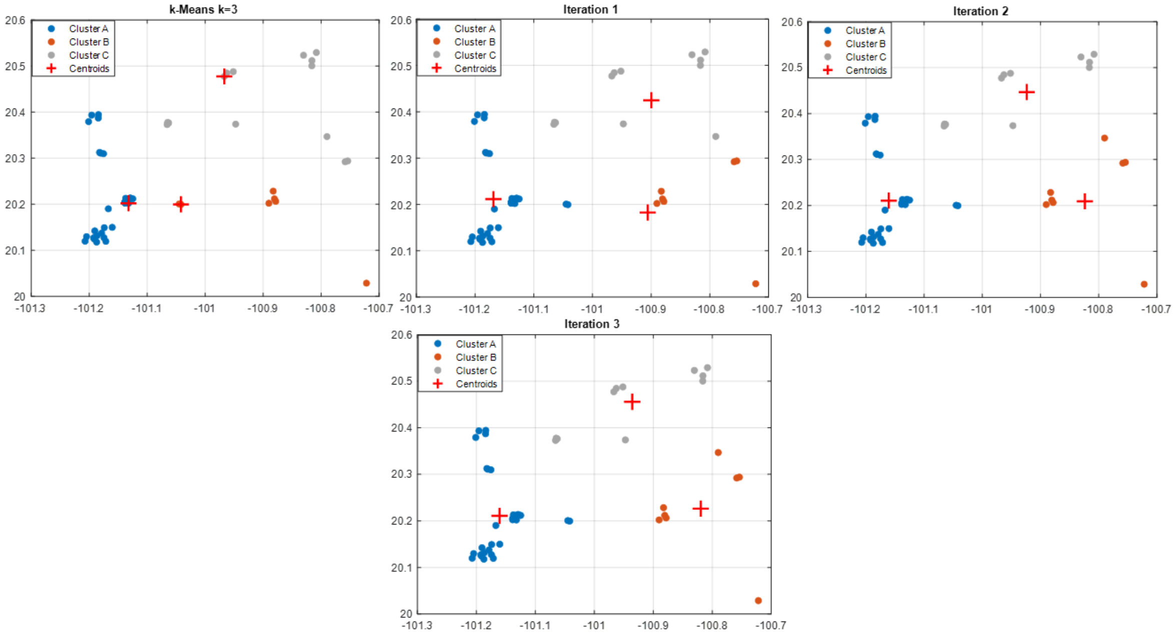

Fig. 4. *Results of the k-means with k=3.*

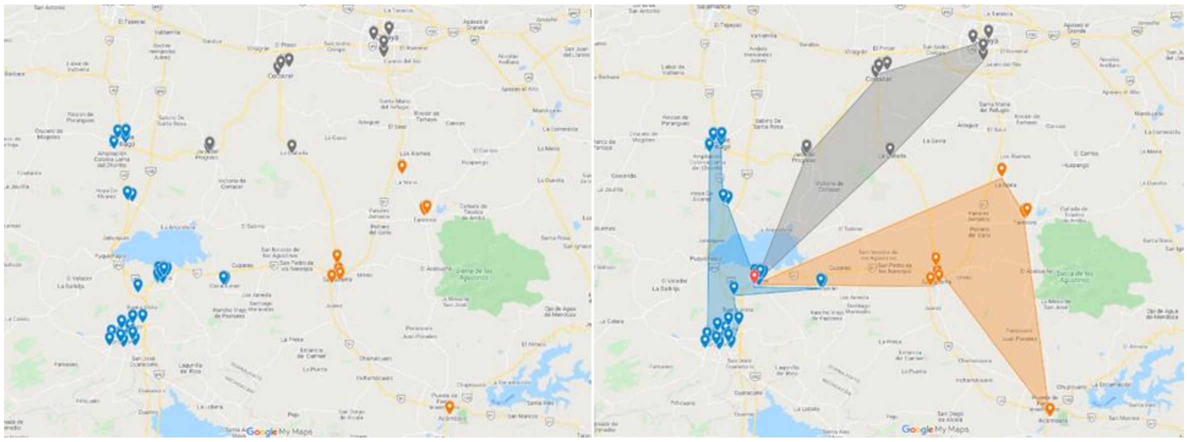

Fig. 5. *Satellite view of the clusters and the area formed in each route (k=3).*

Fig. 6 shows the results obtained from the k-means algorithm implementation with a value of k=4. The figure shows each of the iterations performed until completing the clusters with the smallest distances between the client and its corresponding centroid. The first graph shows the randomly selected centroids, the locations of the clients and the clusters formed by the clients to the nearest centroid and the subsequent graphs show the updates of both the centroid location and the change in the clustering formed by the customers at each of the centroids. 2 iterations were necessary to cluster each customer of the farm in its respective route.

Fig. 7 shows a satellite map with the clusters formed for the customers, and also shows the polygons for each of the customers belonging to each of the 4 routes.

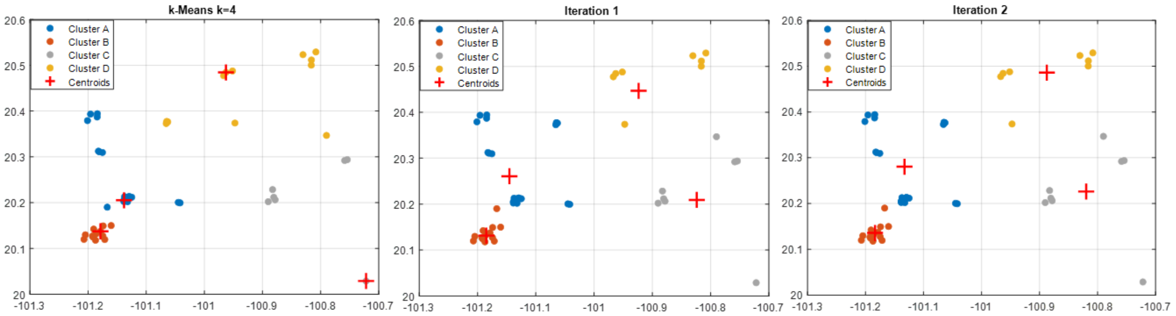

Fig. 6. *Results of the k-means with k=4.*

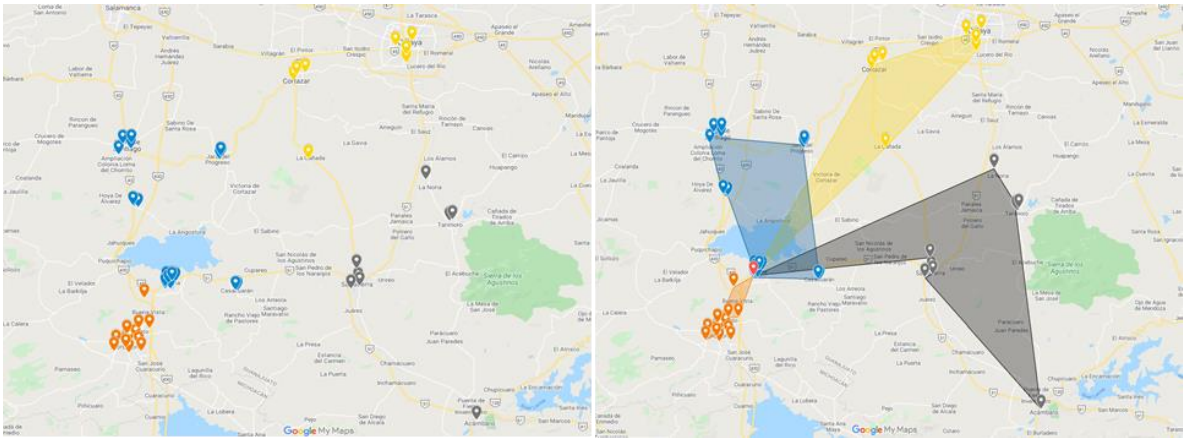

Fig. 7. *Satellite view of the clusters and the area formed in each route (k=4).*

Fig. 8 shows the results obtained from the k-means algorithm implementation with a value of k=5. The figure shows each of the iterations performed until completing the clusters with the smallest distances between the client and its corresponding centroid. The first graph shows the randomly selected centroids, the locations of the clients and the clusters formed by the clients to the nearest centroid and the subsequent graphs show the updates of both the centroid location and the change in the clustering formed by the customers at each of the centroids. 2 iterations were necessary to cluster each customer of the farm in its respective route.

Fig. 9 shows a satellite map with the clusters formed for the customers, and also shows the polygons for each of the customers belonging to each of the 5 routes.

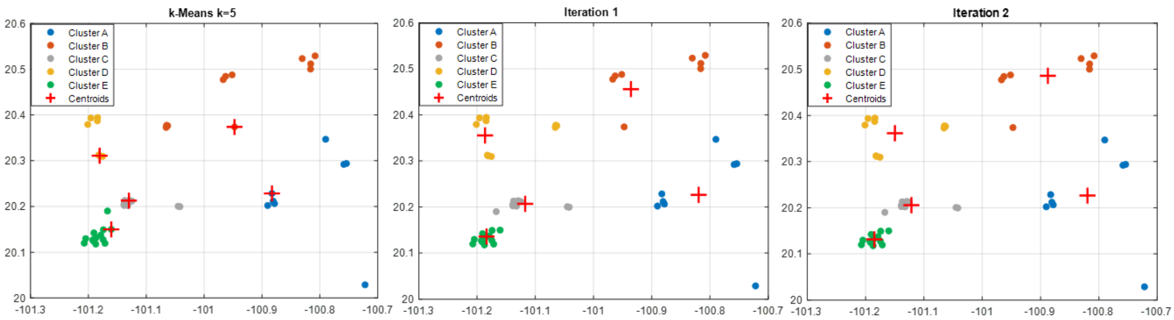

Fig. 8. *Results of the k-means with k=5.*

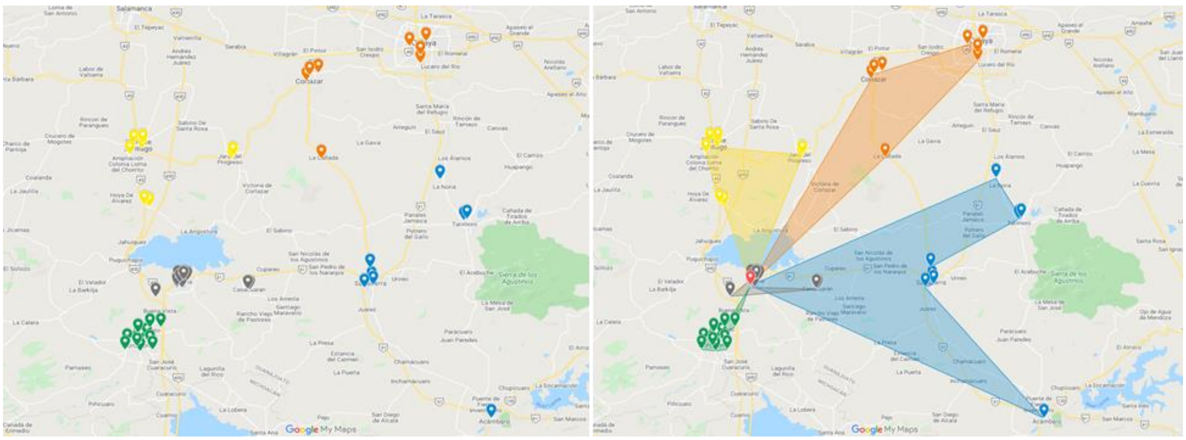

Fig. 9. *Satellite view of the clusters and the area formed in each route (k=5).*

Fig. 10 shows the results obtained from the k-means algorithm implementation with a value of k=6. The figure shows each of the iterations performed until completing the clusters with the smallest distances between the client and its corresponding centroid. The first graph shows the randomly selected centroids, the locations of the clients and the clusters formed by the clients to the nearest centroid and the subsequent graphs show the updates of both the centroid location and the change in the clustering formed by the customers at each of the centroids. 4 iterations were necessary to cluster each customer of the farm in its respective route.

Fig. 11 shows a satellite map with the clusters formed for the customers, and also shows the polygons for each of the customers belonging to each of the 6 routes.

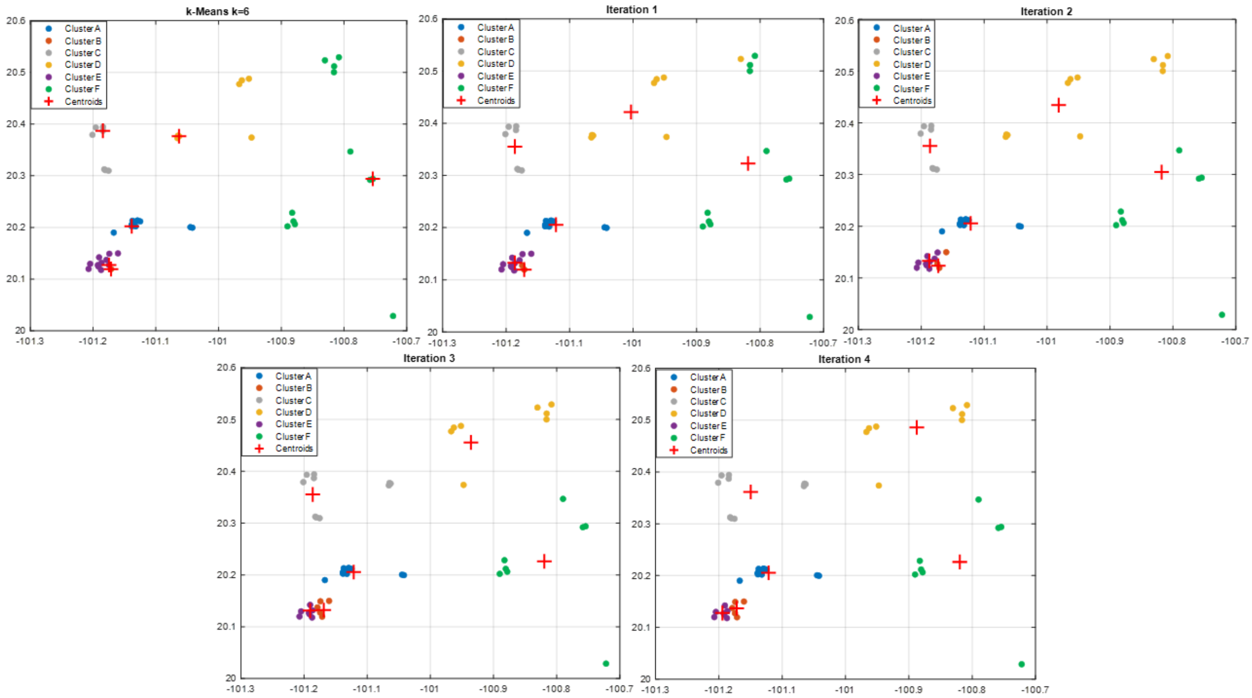

Fig. 10. *Results of the k-means with k=6.*

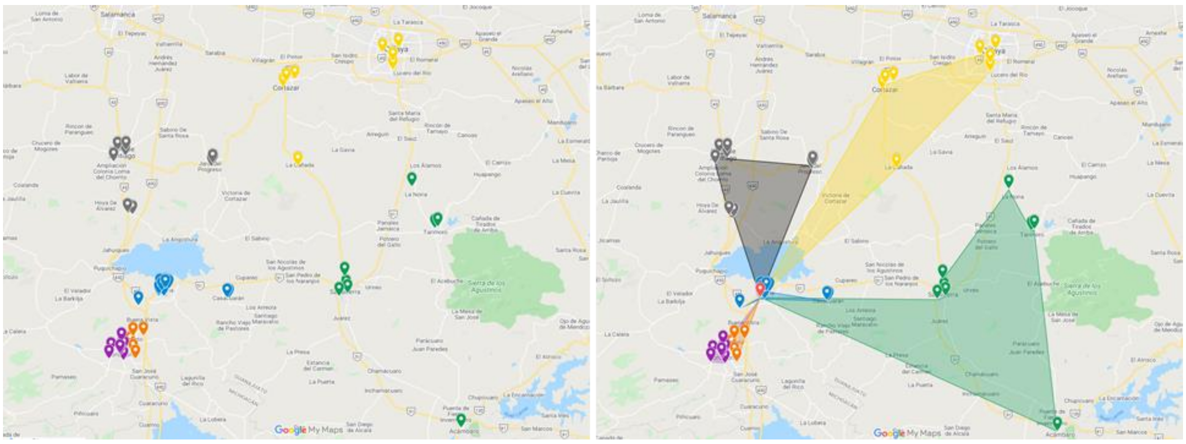

Fig. 11. *Satellite view of the clusters and the area formed in each route (k=6).*

Finally, *Fig. 12* shows the results obtained from the k-means algorithm implementation with a value of k=7. The figure shows each of the iterations performed until completing the clusters with the smallest distances between the client and its corresponding centroid. The first graph shows the randomly selected centroids, the locations of the clients and the clusters formed by the clients to the nearest centroid and the subsequent graphs show the updates of both the centroid location and the change in the clustering formed by the customers at each of the centroids. 4 iterations were necessary to cluster each customer of the farm in its respective route.

Fig. 13 shows a satellite map with the clusters formed for the customers, and also shows the polygons for each of the customers belonging to each of the 7 routes.

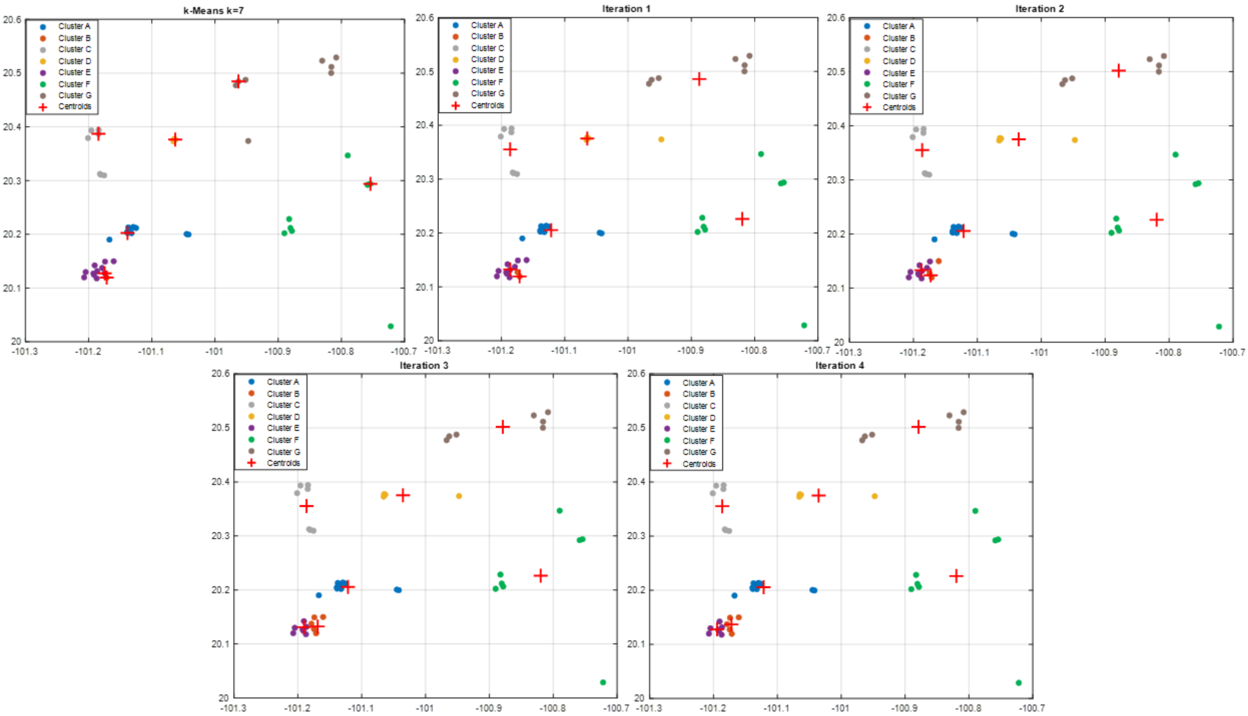

Fig. 12. *Results of the k-means with k=7.*

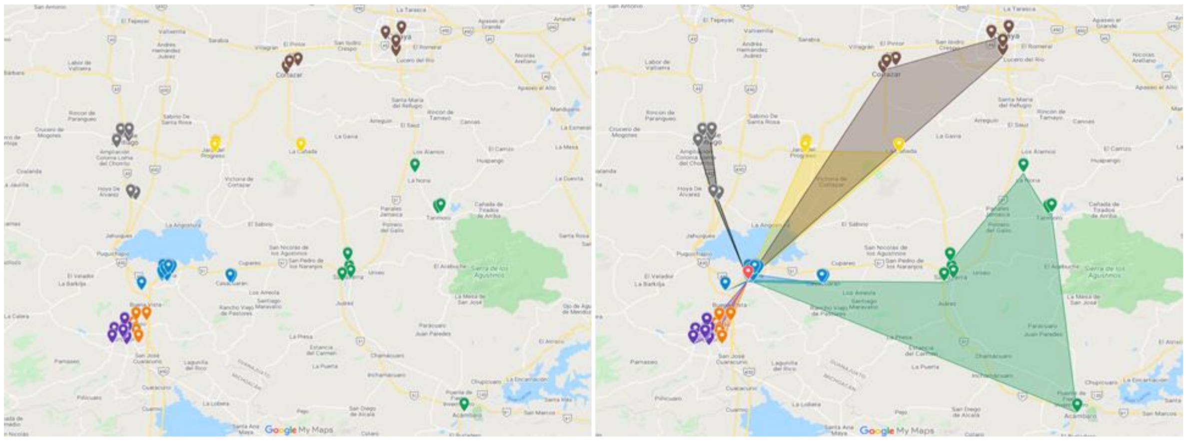

Fig. 13. *Satellite view of the clusters and the area formed in each route (k=7).*

B. Elbow Algorithm

Elbow is an unsupervised autonomous learning algorithm used to calculate the average distortion of the clusters, which is the average distance from the centroid to all points in the cluster and is obtained with the K-means algorithm as a function of number k. For the implementation of the Elbow algorithm, *Eq. 3* was used to calculate the value of the dispersion and when the dispersion decreases sharply, whereas, if the number of groups increases to the appropriate number, the value of the dispersion will decrease more slowly as k increases.

$$\sum_{i=3}^{k}\sqrt{\sum_{i=1}^{n}(Coor_x - Cent_x)^2 + (Coor_y - Cent_y)^2} \qquad (3)$$

Where: k = number of clusters used in the k-means algorithm, $Coor_x$ y $Coor_y$ = customer coordinates, $Cent_x\ y\ Cent_y$ = centroid coordinates.

The pseudocode for the implementation of the Elbow algorithm is as follows:

Elbow algorithm

Input: $Coor_{x,y}, Cent_{x,y}$

Output: d, k

For k=3 to k=7 **do**

2: $d = \sum_{i=3}^{k} \sqrt{\sum(Coor_{x,y} - Cent_{x,y})^2}$
3: **End For**
4: **Show** d
5: **Plot** d

Error! Reference source not found. shows the results obtained from the Elbow algorithm for each of the k values.

TABLE II. RESULTS OF THE APPLICATION OF THE ELBOW ALGORITHM.

K value	Dispersion results
3	4.5106795
4	3.5943049
5	2.5890165
6	2.5407106
7	2.2059442

To observe in a better way the abrupt change that indicates the optimum number of k, the results in *Tab. 2* were plotted, as shown in *Fig. 9*.

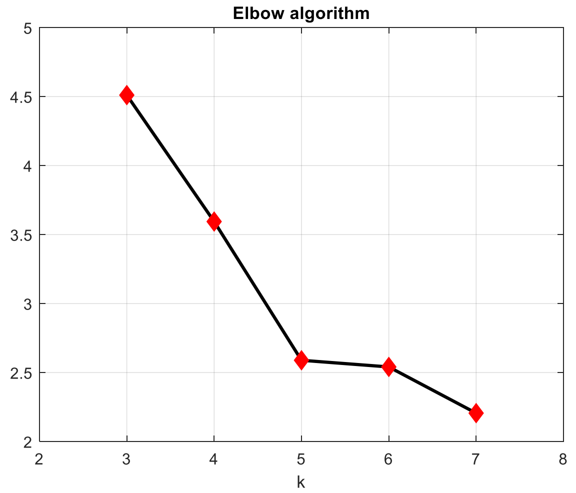

Fig. 14. *Results of the Elbow algorithm.*

V. CONCLUSIONS

The research developed offers a satisfactory solution to a real and unattended logistical problem of a farm dedicated to the sale of eggs.

The clustering of customers for the assignation of routes will help the farm to make the delivery of products, optimizing the distance traveled and therefore the time to make deliveries, since these variables are not currently taken into account in the routes that are made to bring the products to the customers.

The motivation for the creation of this application has been to find the optimal number of routes for the delivery of products from a hen farm producing eggs, but in general, the work can be extrapolated to any business that needs to go through several points in an optimized way and that does not have any tool to perform this task. This has been achieved through the design of an application using MATLAB, in addition to the support of Google map services.

According to the results obtained by the Elbow algorithm, it was determined that the optimal number of routes is 5 routes with the respective customers for each of the routes.

As a final conclusion the k-means clustering algorithm developed in this work together with the application of the Elbow algorithm support the choice of the optimal number of routes to provide a solution for a farm's product delivery system.

REFERENCES

[1] Agricultura, O. d. l. N. U. p. l. A. y. l. (2021). Producción y productos avícolas. Retrieved from Obtenido de https://www.fao.org/poultry-production-products/production/es/

[2] Aguirre, R., & Pizarro, M. (2018). Panorama y mercado del huevo. *Departamento de Análisis de Mercado y Política Sectorial*. Chile. Obtenido de www.odepa.gob.cl.

[3] Agricultura, O. d. l. N. U. p. l. A. y. l. (2021). Producción y productos avícolas: Mercados y Comercio. Obtenido de https://www.fao.org/poultry-production-products/socio-economic-aspects/markets-trade/es/

[4] Cancelas, N. G. (2016). Presentación: transporte y logística. *Revista Transporte y Territorio*, (14), 1-4.

[5] Hurtado, B. A., Robles, J. M., Preciado, J. M., & Bañuelos, N. (2018). Logística de transporte y desarrollo local en organizaciones exportadoras de uva de mesa sonorense. 28(51), 1-25.

[6] Pasharibu, Y., Paramita, E. L., & Febrianto, S. (2018). Price, service quality and trust on online transportation towards customer satisfaction. Jurnal Ekonomi dan Bisnis, 21(2), 241-266.

[7] Fuentes, D. D. (2020). Transporte y logística en la economía mundial. (38), 19-22.

[8] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664-681.

[9] Blömer, J., Lammersen, C., Schmidt, M., & Sohler, C. (2016). Theoretical analysis of the k-means algorithm–a survey. *Algorithm Engineering*, 81-116. Springer, Cham.

[10] Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. J, 2(2), 226-235.

[11] Zheng, X., Lei, Q., Yao, R., Gong, Y., & Yin, Q. (2018). Image segmentation based on adaptive K-means algorithm. *EURASIP Journal on Image and Video Processing*, 2018, 1-10.

[12] Wang, X., Yu, H., Lin, Y., Zhang, Z., & Gong, X. (2020). Dynamic Equivalent Modeling for Wind Farms With DFIGs Using the Artificial Bee Colony With K-Means Algorithm. *IEEE Access*, 8, 173723-173731.

[13] Wang, X., Yu, H., Lin, Y., Zhang, Z., & Gong, X. (2020). Dynamic Equivalent Modeling for Wind Farms With DFIGs Using the Artificial Bee Colony With K-Means Algorithm. *IEEE Access*, 8, 173723-173731.

[14] Rajabi, J., Alizade, S., & Ashoori, M. (2018). Identifying behavioral patterns in blood donation using K-means algorithm based on recency, frequency and blood value. *Journal of Health Administration (JHA)*, 21(71), 66-77.

[15] Pan, L. (2020). Statistical Clustering Algorithm of Wild Animal Growth Environment. *Revista Científica de la Facultad de Ciencias Veterinarias*, 30(Sandoval Serrano & 11), 457-465.

[16] Li, L., Wang, J., & Li, X. (2020). Efficiency analysis of machine learning intelligent investment based on K-means algorithm. *IEEE Access*, 8, 147463-147470.

[17] Qabbaah, H., Sammour, G., & Vanhoof, K. (2019, October). Using K-Means Clustering and Data Visualization for Monetizing logistics Data. *2nd International Conference on new Trends in Computing Sciences (ICTCS)*, 1-6.

[18] Hutagalung, J., Ginantra, N. L. W. S. R., Bhawika, G. W., Parwita, W. G. S., Wanto, A., & Panjaitan, P. D. (2021). COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm. *Journal of Physics: Conference Series*, 1783(Sandoval Serrano & 11), 1-6.

[19] Kong, W., Wang, Y., Dai, H., Zhao, L., & Wang, C. (2021). Analysis of energy consumption structure based on K-means clustering algorithm. *E3S Web of Conferences*, 267 (p. 01054), 1-5.

[20] Mustafi, D., & Sahoo, G. (2019). A hybrid approach using genetic algorithm and the differential evolution heuristic for enhanced initialization of the k-means algorithm with applications in text clustering. *Soft* Computing, 23(15), 6361-6378.

[21] Babatunde, G., Emmanuel, A. A., Oluwaseun, O. R., Bunmi, O. B., Precious, A. E. J. I. J. o. E., & Engineering, M. (2019). Impact of climatic change on agricultural product yield using K-means and multiple linear regressions. *International Journal of Education and Management Engineering*, 5(8), 16-26.

[22] Chang, Y. S., & Lee, H. J. (2018). Optimal delivery routing with wider drone-delivery areas along a shorter truck-route. *Expert System with Applications*, 104, 307-317.