

K-Means Clustering Based Model for Fair Water Distribution of Urban Regions Depending on Consumption

Mohammed Zaid-Alkelani¹, Mohammed Awad^{1*}

¹Department of Computer Science, Arab American University, Jenin, Palestine,

²Department of Computer Systems Engineering, Arab American University, Jenin, Palestine,

ABSTRACT:

The Middle East countries are suffering from problems of the increasing demand for water in light of the scarcity of resources to obtain sufficient quantities and satisfy the needs of citizens of different needs in different fields. These issues drive the water supply companies and authorities to seek other ways to deal with customers and to provide useful policies suitable for the available resource. This paper proposes a smart clustering method to distribute the water in urban regions with the aim to develop a new local classification of neighbourhoods water sustainability according to justice ways to automatically classify them in different categories suitable to their consumption; this methodology uses intelligent clustering techniques that depend on the historical water consumption in each region, rather than the classical methods that distribute the water in a stationary way regardless the quantity of the water used or needs in this region. The data of each region is processed and clustered using the K-Means clustering algorithm to identify the fair distribution of water as a function of water supply days for each region. Our study offered a look at available water resources and the quantities in order to help the water authority to evaluate the challenges and find the alternatives to satisfy the citizens. The K means clustering algorithm achieved superior results in adjusting, rearranging and clarifying the characteristics of water consumption by regrouping similar objects according to quantities and pattern of consumption within clear and organized clusters. Our results showed that this technique will be awesome for the self-classification for every neighbourhood water consumer based on historical data related to water demand belongs to these neighbourhoods.

Keywords: Fair Water Distribution, K-means Clustering Algorithm, Urban Water Demand.

Date of Submission: 18-10-2021

Date of Acceptance: 02-11-2021

I. INTRODUCTION

Throughout the ages, freshwater and its availability in quantity and quality have been a major concern for all humans' civilizations for its association and its reflection on the health of society. Demand, supply, service and community health are therefore based on the infrastructure of supply, distribution systems, and future strategic plans that have the capacity to meet the needs and sustain the success of all elements of human, industrial and agricultural development [1]. Thus, the water sector is an important sector of sustainable development at the national level. Potable water in a generalized way reaches consumers through a sanitary system that consists of engineering work with the function of distributing and purifying water from natural sources. This supply network is composed of a series of stages that process the raw water and distribute it to each of the supply points. The distribution network corresponds to the

plumbing system that reaches each of the points of supply for consumers. This network starts in the regulation pond and consists of pumping stations, pipes, valves that divides the water supply in case of ruptures and emergencies due to water scarcity, and, finally, of water volume measuring devices at the points of supply. The deliberate neglect of the development of the water sector in Palestine by the Israeli occupation over the decades has contributed directly to reduce the chances of real development. The development of the water sector remained constrained by the obstacles imposed by the occupation authorities even within the agreements signed between the Palestine Liberation Organization (PLO) and Israel. The high demand for water and the significant gap between demand and supply in the water sector is one of the major challenges facing the sector over the next few years. The water demand is increasing because of natural

population growth and national development requirements, Water and its use due to Israeli obstacles, add to this the large difference and the gap between the per capita Palestinian water compared with the Israeli, also the complex licensing requirements from the Israeli side, this is a great challenge and it is necessary to find creative solutions to supply the necessary quantities of water to different sectors and achieve balance for supply optimal water in Palestine.

If we look at recent years, we will find different methodologies that have been used in the field of water distribution in urban regions that have been proposed. The major systems used in Palestine are dependent on a stationary method that supplies each region with the water on a specified day, regardless of the consumption needs in the determined region. In recent years, different projects aimed at achieving more efficient and sustainable cities. The management of fair water distribution is a challenge that can be solved using smart technology. Intelligent Water Management uses intelligent technologies, such as Big Data, the Internet of Things and the detection and monitoring of real-time systems that can help save the water supply services. It is also important to analyze in detail how the distribution of water in the network works to anticipate the risks of incidents (leaks, pressure drops, etc.) and solve them. Currently, there are several tools that aim to optimize the performance of water networks. These systems work in optimizes all the performance indicators of the drinking water network (flow, pressure, and quality) through sensors that monitor the hydraulic behaviour of the network in real-time. This makes it possible to detect water leaks as well as to continuously monitor the quality of the water. But in the countries where the water supply organized by days for each region, it's important to produce a smart system that manages this process fairly.

The water circumstance on the northern side of Palestine, such as the city of Jenin, is similar to the rest of Palestine regions. But in Jenin city, there are more difficulties depending on the amount of water leaking in the ground due to weak of networks and interruptions in supply and consumption during the time, in addition to other factors, such as the scarcity of water resources and they are linked with the Israeli side which controls the valves [16]. The water department in the municipality of Jenin has no programs or applications in terms of the ability to distribute the water in a fair manner, even in the other governorates also there are no modern methods, the methods just depend on supply the water for each region in determined days, thus the regions with big

number of citizens will be affected that the water quantity is not fair enough. Studies related to water conditions for supply and demand issues in the West Bank for the years 2000 to 2020 according to the applied research institute Jerusalem indicate that the city of Jenin will require large amounts of water to meet the expected progress in the field of agriculture and industry [2]. From this point of view, smart water distribution systems can help supply companies to address the challenge through efficiently managing this scarce resource. the importance of computerized information technology come and its benefit in clarifying the image and raising awareness of the designated authorities on the importance of sustainability of the water sector and exploit new techniques in improving and controlling the behaviour and attitudes of consumers [3]. Thus, water demand and predictability represent the core of modern technology for more efficient and sustainable use of water. Having accurate data on the state of the water consumption in each region allows supply companies not only to understand the supply and demand of water but also discrimination all that it entails.

In addition, having the ability to determine when and how much time of water supply needs this region will increase the life of the pipeline or even a pump. These capital investment decisions are almost always based on time, the amount of time an asset has been used, or whether it is an intelligent asset, among other factors. In addition, a reduction in water demand means that the water supply infrastructure carries fewer loads and, therefore, has a longer useful life. On the other hand, the important data provided by a clustering intelligent water solution allows the automation of tasks associated with routine maintenance and operation of the water distribution system, which results in greater general efficiency. Recently, several and different researches have been proposed especially that offer the appropriate methodologies, such as Kohonen Self-Organized Maps [12], fuzzy cognitive maps [4], k-means clustering algorithm and others which prove the worthiness of classification and predicting the demand of water. In this paper we used an algorithm called K-means clustering algorithm [13] to analyse water consumption for neighbourhoods and made clustering in Jenin city based on data was collected which depends on the water consumption in each region (Jenin city divided into 39 regions), where the similar regions were clustered to organize the water distribution with the suitable number of days depending on the consumption and develop a new local classification of neighbourhoods water sustainability in a justice ways.

II. LITERATURE REVIEW

The problem of fair distribution of water is not an urgent problem in the developed countries, but because of the situation of developing and poor countries, the problem of the fair distribution of water by region is one of the important problems that the researchers seek to solve, to ensure the fair distribution of water and the continuity of supplying the consumer with water within a specific time. Consumers in these countries use water tanks, which are usually placed on the roofs of the buildings for continuity of the water reaching the house. This subject has not been addressed by researchers extensively because of its relationship to this issue in the distribution of water in developing and poor countries. In this section, we review a number of the related work that was proposed as shown in the next papers which used various manners. In [5], the authors analyzed a dataset of 142 cities that related to annual per capita water use and population. They made hierarchical cluster analysis to determine the similarities between the 142 cities based on climatically similarity and also not identical, the authors applied statistical data mining style to develop primary urban water classify according to the city size, per capita water consumption, and net annual water balance. They deduced that the statistical clustering is a valuable approach to

Based on the preliminary results of population, housing and establishment's census 2019, the population of the Jenin city and its camp is about 60325 people distributed in many different neighbourhoods within the borders of the municipality [7], these neighbourhoods vary in degree of overcrowding with respect to the number of buildings, population and monthly water

improving a quantitative basis for (small/large) urban water management case study research. In [4], the authors analysed water consumption data from a set of consumers at the Greek island, the authors obtained extra information about facts related to their water consumption behaviour. These facts are adopted such as input vectors for the build the Kohonen Self-Organized Maps that are exploited as classification ways to cluster consumers based on their consumption. The authors deduced that such analysis can be useful for the automatic classification of water consumption in urban water demand data. The authors in [6], used two various clustering algorithms, called fuzzy c-means and K-means together with a genetic algorithm in order to identify the homogeneous zones in terms of groundwater water goodness. There are obtained sampled contain 14 hydrochemical parameters from 108 wells northeast of Iran. The authors confirm that optimal clusters of the K-means and fuzzy c-means were Varying Regarding (total dissolved solids) and (chlorine) parameters. they deduced, when compared to the results of clustering algorithms, they found the fuzzy c-means algorithm has achieved better results than the K-means clustering algorithm, and this return to uncertainty conditions in locating the class boundary.

III. STUDY AREA

consumption extracted from the database of municipal. During the period of preparation of this study, engineers from the Municipality of Jenin cooperated to help clarify and explain these neighbourhoods and details regarding the demarcation of the boundaries and locations on the city map based on a logical and real distribution of the supply zones.

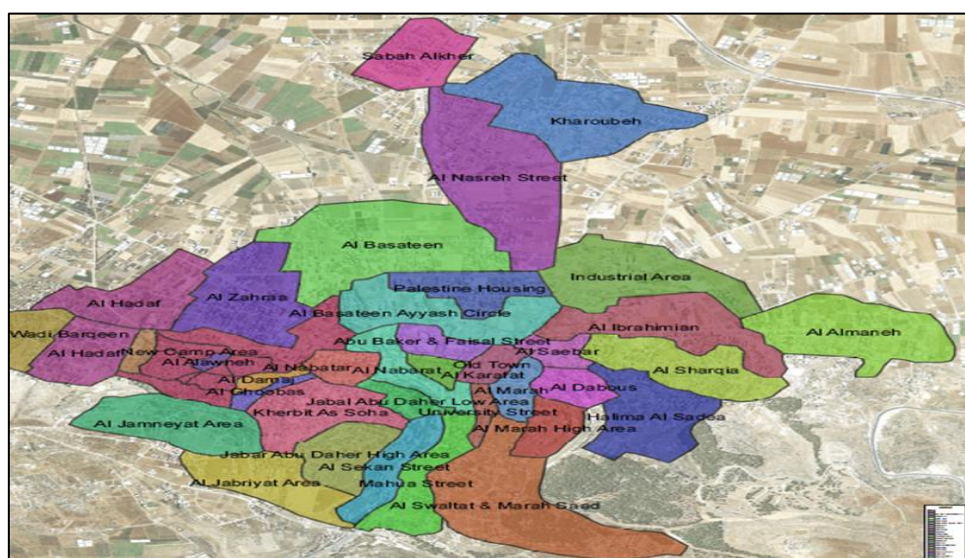


Figure. 1: Jenin city supply zones

According to the map in Figure 1, we note that the city of Jenin is divided into 39 neighbourhoods where the municipality of Jenin city responsible for providing and pumping the required quantities of water. These 39 neighbourhoods, as shown in the following figure, consume about 1160927 cubic meters according to the statistics of 2019. The quantities of water supplied to these sectors are one day a week and the equivalent of four days per month for each sector, regardless of the area of each neighbourhood, regardless of the number of subscribers and meters active, and

therefore every neighbourhood, whether the size of users little or large, he has four days' supply per month.

In light of the method adopted in the water supply, we find that it is unfair and lacks an optimal distribution, Therefore, there are small areas that provided with a number of days of pumping water more than the real needs and there are large areas have pump water days far below the need according to the numbers and quantities of water consumed and located in our hands for all regions.

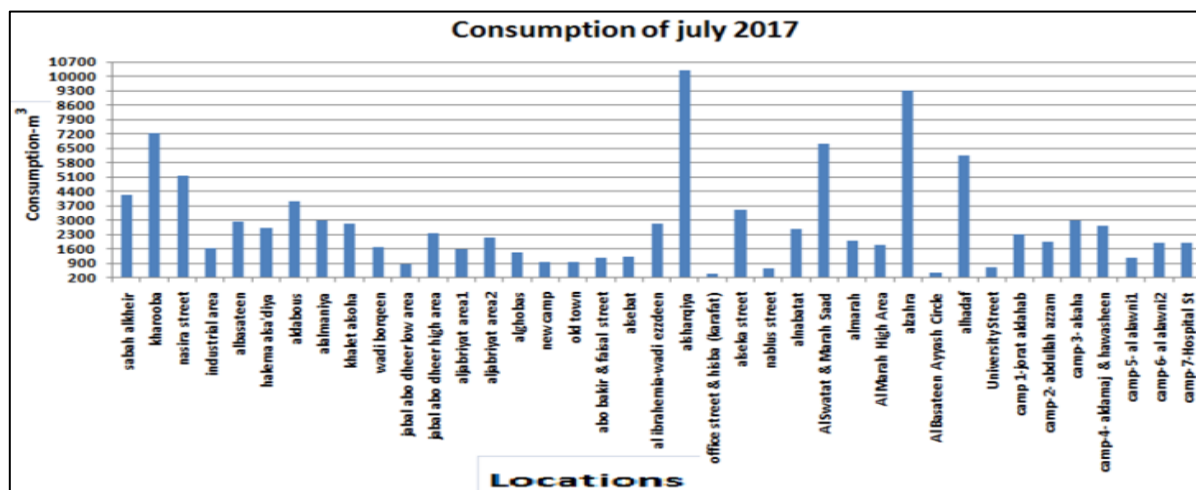


Figure. 2: distribution of quantities of water demand

In figure 2 we can note the distribution of quantities of water demand for the regions and how much it differs and however they get an equal number of supply days. Indeed, we face a real

problem and we seek effective and fair solutions in terms of equitable and optimal distribution, which achieves higher levels of consumer satisfaction rather than the traditional method used in Jenin city.

IV. METHODS AND DATA CLUSTERING WITH K-MEANS CLUSTERING ALGORITHM

Data clustering is the operation of putting data points in similar clusters; it is a section of data mining. The clustering algorithm separates a data set into varied clusters since the similarity between points within a particular cluster is greater than the similarity between two points within two different clusters [8]. The idea, in general, is simple in it is natural and very close to humans in its way of thinking, so whenever dealing with a large amount of data we tend to summarize the vast amount of data into a few groups or categories in order to facilitate the analysis process [14].

Clustering algorithms are widely used in data classifications, data compression, and data model construction, since if we can find clusters of data, a model of the problem can be built on the basis of those clusters. There are various algorithms applied in the data clustering process, and we will

examine the simplest algorithms called the K-means clustering algorithm. K-means clustering is a kind of class of unsupervised learning, which is applied with unorganized or unlabelled data with defined categories or groups [9]. The purpose of this algorithm is to find groups or clusters of data, with the number of sets, declare by the variable K. The algorithm runs frequently to allocate each data node to one of the K sets based on characteristic similarity [15]. In other words, this algorithm is used to collect several data points depending on their properties to the K clusters, and the clustering process is done by reducing the distances between the data and the cluster center [10]. The general steps of the K-means clustering algorithm are shown in figure 3. The performance and effectiveness of this algorithm depend on the initial positions of the centers of the cluster, and it is recommended that

this algorithm is run several times with different positions each time from previous times. So the (K-means) aims to minimize total cluster variance or the squared error function and this given by:

$$J(c) = \sum_{i=1}^k \sum_{j=1}^n (\|X_j - C_i\|)^2 \quad 1)$$

Where k is the number of clusters, n is the number of data points in the i^{th} cluster, X_j : data point, C_i : centroid for cluster i , $\|X_j - C_i\|$: is the Euclidean distance between X_j and C_i . The general process for K-means clustering is illustrated as shown in figure 4.

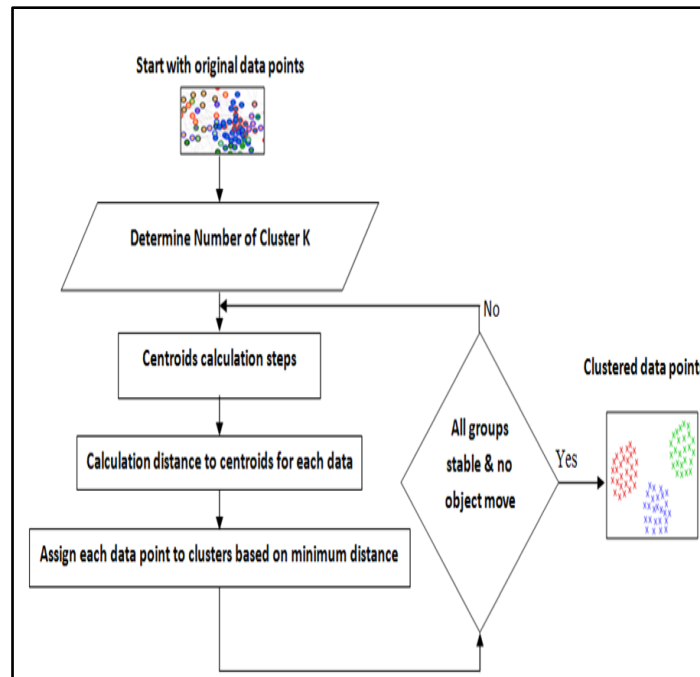


Figure. 3: General steps of K-means clustering algorithm

K-Means is comparatively an efficient method; therefore, we need to designate the number of clusters, and hence the final results often end at a local optimum. Unluckily there is no uniform notional method to find the best number of clusters, but the alternative approach is to compare the results of many rounds with various 'k' and adopt the best

one according to a predefined condition. In the next section, we present the practical and scientific results produced by the model proposed in this paper. There will be a presentation of the results by graphical methods, charts, and tables which will be closer to understanding the idea of the practical experiments of the model.

<p>General procedure: k-means clustering Algorithm</p> <p>Input: data points (X_i) into 'k' sets, number of 'k' (cluster) are formally determined.</p> <p>Step -1: Pick out 'k' points at randomly as cluster centers.</p> <p>Step -2: Allocate data points or nodes (X_i) to their closest and nearest cluster center according to the Euclidean distance function.</p> <p>Step -3: Recalculate the centroid or new cluster center (position) of all objects in each cluster by using $C_i = (1/n_i) \sum_{j=1}^n X_j$, '$n_i$' represents the number of data points in i^{th} cluster.</p> <p>Step -4: Recalculate the distance between each data point and new obtained cluster centers as the same steps in (2 and 3) until no data point reassigned and all data points in clusters are stable.</p> <p>Output: data point with cluster memberships.</p>
--

Figure .IV: General procedure of k-means clustering Model

V. EXPERIMENTAL RESULTS AND DISCUSSION

In all experiments procedures and applied model which we have designed, we tested it by employing the MATLAB R2013a under Windows 7 with Core i3-M 380 CPU 2.53GHz, 4GB RAM memory. In order to carry out an optimal and equitable distribution of water supply, we initially extracted and calculated the total consumption of each area separately during a given month by selecting individual months from 1 to 11 of 2019 for 39 districts representing all districts of Jenin City. After all, months which extracted and calculated as

we obtained three tables contain all districts and their consumption of water in certain months. To start our work by using the K-means clustering algorithm we already adopted a number of clusters (K=15) represent the maximum number of water supply days in the month, a number of data points (Xi=39) represent the total consumption of water for all districts in the city. We adopted the K=15 because we obtained the best results after we tried different values according to try and error, another reason we thought that this value represents adequate days and satisfy the customers in the large and overcrowded area

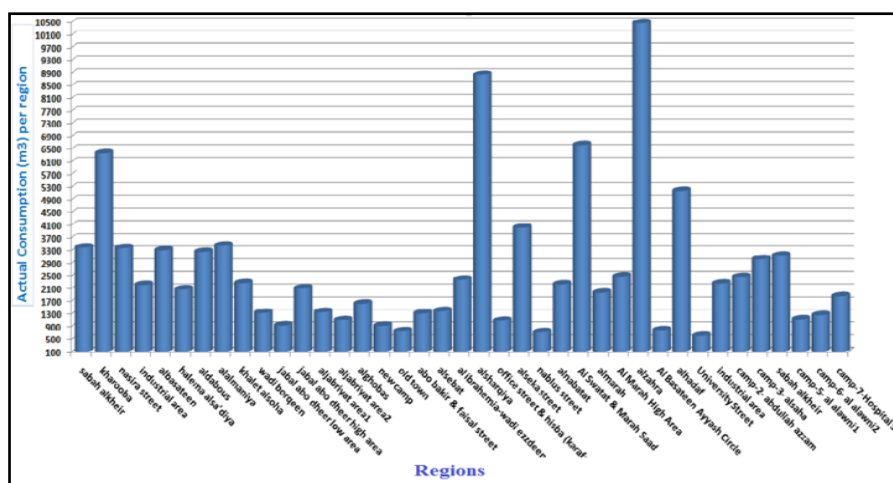


Figure. 5: January-2019- actual water consumption according to regional distribution

To perform our clustering process according to the monthly style, we calculated the total monthly consumption per neighborhood separately until we finally get a 39 value to be input vector to the algorithm. Figure (5, and 6) shows the

local distribution that represents the actual water consumption in all Jenin city neighborhoods for January, and July in 2019, respectively, pointing to the large differences in consumption between the neighborhoods.

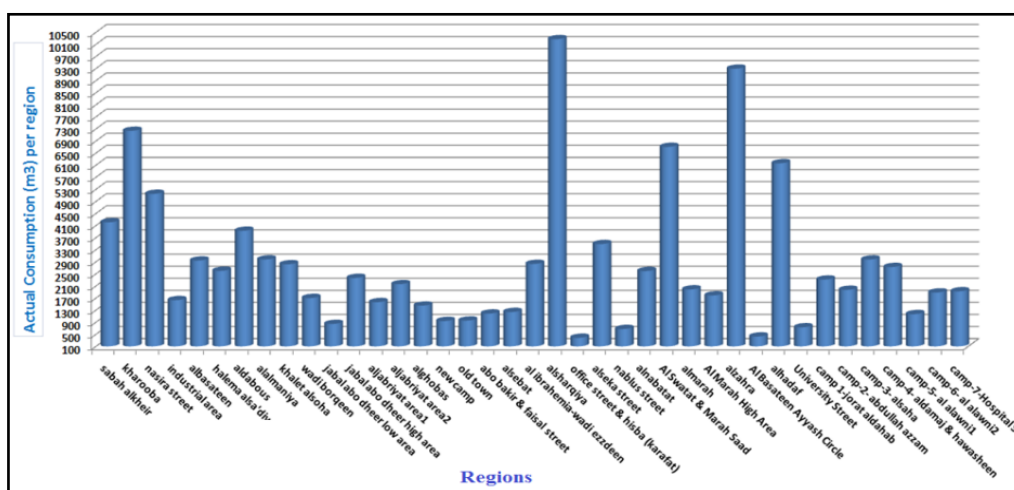


Figure. 6: July – 2019- actual water consumption according to regional distribution

We carried out our proposed way by using a 39 value as input to the algorithm to get 39 outputs to represent the index of clusters for each value. Table 1, and 2 how the final result which obtained according to the K-means clustering algorithm based on 39 neighbourhoods for January, July, and in 2019, respectively.

With regard to the proposed values that determine the number of days required for water supply operations for different neighbourhoods, they have been obtained in a carefully considered manner, so the consumption of cluster index for each region individually converted and normalized to the range

between the value 0.5 and 15 according to the following equation:

$$y_i = \frac{0.5 + (x_i - \min(x)) * (15 - 0.5)}{(\max(x) - \min(x))} \quad (2)$$

Where (x_i) is the actual consumption and (y_i) is the normalized value, min and max are the maximum and minimum values for actual consumption [11]. Graphically we illustrate the clustering process more precisely and closer to understanding, so the following figures 7, and 8 also show all clusters based on 39 neighbourhoods for January and July in 2019, respectively.

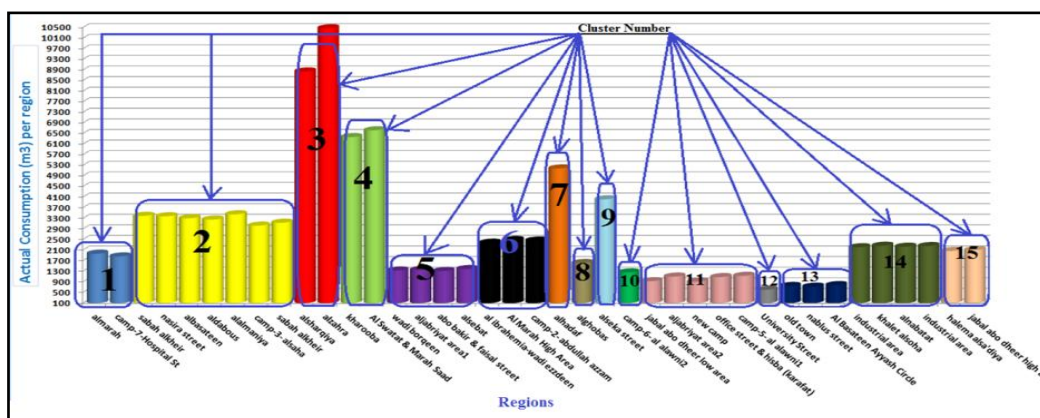


Figure. 7: Distribution clusters for all neighbourhoods January-2019

According to table (1 and 2) and figures (7 and 8), it is clear that the K-means clustering algorithm achieved an excellent result for aggregate the neighbourhoods in highly accurate form according to an appropriate quantity of clusters (K=15). There is no any irregular data point inside the clusters. These excellent clusters distribution has helped us to understand the quantities of consumption and know of the characteristics of different neighbourhoods and also represented the base stone of the proposal to distribute a fair and

ideal to end the situation and the method of unfair distribution.

The Proposed numbers of supply days represent how much days this neighbourhood need of supply, each number obtained according to converted the Consumption into days by converting all quantities to fixed normalized values between 0.5 and 15, this procedure depends on each index value after arranging all clusters from smallest values to the largest values.

Table 1: K-means clustering algorithm index result for January-2019

Original Data Set			Outputs by K-Means	Proposed Num. of Supply
No	Location	Con Jan M ³	Index / Cluster No	Days
1	Sabah Alkheir	3394	2	5
2	Kharooba	6368	4	9
3	Nasira Street	3379	2	5
4	Industrial Area	2210	14	3
5	Albasateen	3305	2	5
6	HalemaAlsa'diya	2070	15	3
7	Aldabous	3255	2	5
8	Alalmaniya	3453	2	5
9	KhaletAlsoha	2266	14	3
10	WadiBorqeen	1330	5	2
11	Jabal A.D. Low Area	927	11	1
12	Jabal A.D.High Area	2102	15	3
13	Aljabriyat Area1	1356	5	2
14	Aljabriyat Area2	1102	11	2
15	Alghobas	1614	8	2
16	New Camp	917	11	1
17	Old Town	748	13	1
18	AboBakir Faisal. St	1313	5	2
19	Alsebat	1387	5	2
20	Al Ibrahim-W.Ezzd	2380	6	4
21	Alsharqiya	8846	3	13
22	Office Street &Hisba	1083	11	2
23	Alseka Street	4020	9	6
24	Nablus Street	721	13	1
25	Alnabatat	2229	14	3
26	Al Swatat&M.Saad	6624	4	10
27	Almarah	1966	1	3
28	Al Marah High Area	2480	6	4
29	Alzahra	10479	3	15
30	AlBasateenAya sh .C	780	13	1
31	Alhadaf	5184	7	7
32	University Street	605	12	1

Table 2: K-means clustering algorithm index result for July -2019

Original Data Set			Outputs by K-Means	Proposed Number of Supply
No	Location	Con Jan M ³	Index / Cluster No	Days
1	Sabah Alkheir	4208	8	6
2	Kharooba	7236	4	11
3	Nasira Street	5149	13	7
4	Industrial Area	1630	14	2
5	Albasateen	2942	7	4
6	HalemaAlsa'diya	2600	9	4
7	Aldabous	3926	8	6
8	Alalmaniya	2976	12	4
9	KhaletAlsoha	2817	9	4
10	WadiBorqeen	1701	14	2
11	Jabal A.D. Low Area	842	11	1
12	Jabal A.D.High Area	2366	15	3
13	Aljabriyat Area1	1564	14	2
14	Aljabriyat Area2	2160	15	3
15	Alghobas	1449	10	2
16	New Camp	938	5	1
17	Old Town	953	5	1
18	AboBakirFaisal. St	1186	5	2
19	Alsebat	1237	5	2
20	Al Ibrahim-W.Ezzd	2829	9	4
21	Alsharqiya	10271	6	15
22	Office Street &Hisba	383	11	1
23	Alseka Street	3487	3	5
24	Nablus Street	675	11	1
25	Alnabatat	2592	9	4
26	Al Swatat&M.Saad	6704	4	10
27	Almarah	1985	2	3
28	Al Marah High Area	1786	1	3
29	Alzahra	9294	6	14
30	AlBasateenAya sh .C	425	11	1
31	Alhadaf	6162	4	9
32	University	730	11	1

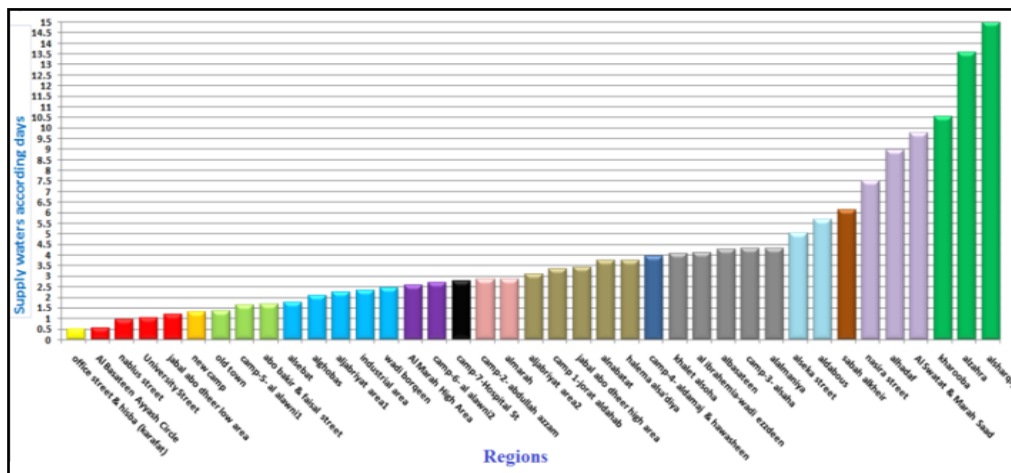


Figure. 10: supply water According to the number of day's allocated- July-2019

According to previous figures, you can note how much the clusters are accurate and it is a regularity, add to that the neighborhoods and data points values are near together in the same cluster and there are no large differences. Depending on the experimental result of clustering, the features computed on the water consumption data for each region in Jenin city as shown in the map actually

VI. CONCLUSIONS

In this paper, we have made an important analysis of water consumption data related to neighborhoods of Jenin city, with the aim to develop a new local classification of neighborhoods water sustainability in a justice way also to automatically classify them in different categories, according to its consumption. Our work outcomes offer a primarily answer about the similarity and differences between the neighborhoods in terms of their water consumption patterns. One of the main aims of our study was to exploit the common unsupervised machine learning algorithms (K-means algorithm) to cluster the groups of similar and different

REFERENCES

- [1]. HASSAN, Fekri A. Water history for our times. UNESCO International Hydrological Programme, 2010.
- [2]. ARI, Environmental profile for the West Bank, .Jerusalem: Applied Research Institute. Vol. 5 1996
- [3]. Laspidou, Chrysi. "ICT and stakeholder participation for improved urban water management in the cities of the future." *Water Util. J* 8, 79-85. 2014
- [4]. Laspidou, Chrysi, et al. "Exploring patterns in water consumption by clustering." *Procedia Engineering* 119. 1439-1446. 2015

represented consumers with similar properties to get "naturally" grouped together. The presented result illustrates in graphs in 3 months of January and July. As shown in the figures above the k-means clustering algorithm classifies each similar gropes of regions in one cluster depends on the consumption. This process aims to help in the control of the water distribution for each region.

neighborhoods according to their water consumption and demand. Our study also offers a look available water resources and the quantity in order to help the water authority to evaluate the challenges and find alternatives to satisfy the citizens. The K means clustering algorithm achieved superior results in adjusting, rearranging and clarifying the characteristics of water consumption by regrouping similar objects according to quantities and pattern of consumption within clear and organized clusters. Our results show that this technique will be awesome for the self-classification of neighborhood water consumers that we wanted to achieve based on historical data related to water demand belong to these neighborhoods.

- [5]. Noiva, Karen, John E. Fernández, and James L. Wescoat Jr. "Cluster analysis of urban water supply and demand: Toward large-scale comparative sustainability planning." *Sustainable cities and society* 27 . 484-496. 2016.
- [6]. Mohammadrezapour, Omolbani, Ozgur Kisi, and Fariba Pourahmad. "Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality." *Neural Computing and Applications*. 1-13. 2018
- [7]. Palestinian Central Bureau of Statistics (PCBS). Preliminary results of the population, housing and establishments census. 2017. Ramallah - Palestine: 2018.

- [8]. Pandey, Abhishek, and Ankita Malviya. "Optimizing Map Reduce Scheduling Using Parallel Processing Model On Data Nodes With K-Mean Clustering Algorithm In Hadoop Environment." (2018)..
- [9]. Ding, Chris, and Xiaofeng He. "K-means clustering via principal component analysis." Proceedings of the twenty-first international conference on Machine learning. ACM, 2004.
- [10]. Kanungo, Tapas, et al. "An efficient k-means clustering algorithm: Analysis and implementation." IEEE Transactions on Pattern Analysis & Machine Intelligence 7: 881-892. 2002
- [11]. Awad, Mohammed, and Mohammed Zaid-Alkelani. "Prediction of Water Demand Using Artificial Neural Networks Models and Statistical Model." Vol 11 N 9. 2019.
- [12]. GALUTIRA, Edwin F.; FAJARDO, Arnel C.; MEDINA, Ruji P. A Novel Kohonen Self-organizing Maps Using Exponential Decay Average Rate of Change for Color Clustering. In: Intelligent and Interactive Computing. Springer, Singapore, 2019. p. 23-33.
- [13]. RADLIYA, N. R.; FACHRIZAL, M. R.; RABBI, A. R. Monitoring Application for Clean Water Access and Clustering using K-Means Algorithm. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2019. p. 022096.
- [14]. YANG, Ao, et al. Enhancing Residential Water End Use Pattern Recognition Accuracy Using Self-Organizing Maps and K-Means Clustering Techniques: Autoflow v3. 1. Water, 2018, 10.9: 1221.
- [15]. Awad, M., &Abuhasan, A. (2016). A smart clustering based approach to dynamic bandwidth allocation in wireless networks. International Journal of Computer Networks & Communications (IJCNC) Vol, 8.
- [16]. TROTTIER, Julie. Palestinian Water Management–Policies and Pitfalls. 2019.

Mohammed Zaid-Alkelani, et. al. "K-Means Clustering Based Model for Fair Water Distribution of Urban Regions Depending on Consumption." *International Journal of Engineering Research and Applications (IJERA)*, vol.11 (10), 2021, pp 43-53.