Data Mining Methods in Municipality Social Data System (DMSDS)

AIMAN DAW GHANIM [1], ZAKARIA SULIMAN ZUBI[2],

[1]Department of Computer Science, Faculty of Exact Sciences and Computer Science
, Hassiba Bin Bou Ali University, Chlef, ALGERIA
[2]Department of Computer Science, Faculty of Arts and Science, Sirte University, Zamzam ,Sirte,
LIBYA

Abstract: - The wide technical evolution of the use of data mining technology in dealing with huge electronic datasets is increasing day by day. This great and growing progress motivates us to employ this technology in dealing with the municipality's social data. The municipality is an organizational structure that includes a number of departments, offices and service localities that deal with sectors, citizens and businesses in a specific geographical area. Within the municipality there are areas affiliated to the municipality, which we called in Libya as "Al-Mahalla." This "Mahalla" represents an administrative division belonging to the municipality. There are social data represented in a dataset about the citizens residing in Al-Mahalla, which belongs to the municipality. The dataset holds a general statistical data on the citizens residing in the municipality, as well as the "Al-Mahalla" inhabited by the citizens. In this study, we will address the municipality of Al-Khums in Libya, as one of the municipalities in Libya as a case study.

 In this paper, we will use classification and clustering methods to extract knowledge from social data and predict marriage rate, death rate, poverty rate and rate of increase or decrease of citizens in the municipality, which in turn helps leaders and decision makers in the municipality to make the appropriate decision. As well as it aims to discuss the impact of the municipality's social data on the current Libyan lifestyle. We used also Weka software as a data mining solution to apply classification and clustering methods on the datasets we obtained from the municipality.

Key-Words: - Data mining, Municipality dataset, Social dataset, Classification, Clustering, KNN, K-means, Dataset.

1 Introduction

A municipality is usually a single administrative division having corporate status and powers of self-government or jurisdiction as granted by national and regional laws to which it is subordinate. Sometimes a municipality can be any political jurisdiction of a sovereign state. There are currently three kinds of municipalities:

- Metropolitan municipalities are enormous cities.
- Local municipalities are towns and their surrounding rural areas.
- District municipalities coordinate several local municipalities in a region.

The municipality provides services to the citizens referring to the basic social services like registration of population statistics in the municipality, registration of the number of deaths in the municipality, registration of the number of births in the municipality, registration of the number of workers, and non-workers in the municipality. It also provides other types of services as an exchange for the taxes which citizens pay, operates or contract for additional utilities like electricity, gas, water supply, sewage collection and disposal, municipal health services, street lighting, municipal parks and recreation, and cable television.

Aiman Daw Ghanim, Zakaria Suliman Zubi

These services provided by the municipality to the citizen result in very important data. These data are classified as social data that can be utilized to study some social phenomena same as measuring the rate of population increase, measuring the marriage rate, measuring the death rate, or measuring the unemployment rate in the municipality. It also includes information collected during tax rate processing, year-end financials, valuations, and other financial and socioeconomic data as well.

The social datasets collected from the municipality will be utilized in this study. This data will be applied to some data mining methods and algorithms. The utilization of data mining methods and algorithms on the social data of the municipality will produce valuable information to provide decision-makers in the municipality to take the appropriate decision based on the facts and data collected from the social data in the municipality. Since data mining extracts the hidden information in the datasets and provides useful information to support the decision of the leaders. These decisions will allow improving the investments in the civilian service/life quality through periodical assessment /evaluation of the public in the municipal. This paper aims to discuss the impact of the municipality's social data on the current Libyan lifestyle.

On the other hand, there are parts sited inside a municipality as an administrational division called 'Mahalla' represented by a specific geographical area within the municipality's divisions that includes several residents, and helps the municipality to organize the administrative work in terms of helping eople and providing services for them. As well as we will address the municipality of Al-Khums in Libya, as one of the municipalities in Libya as a case study.

2 Social Dataset
The social dataset indicated in this study are divided into three parts as follows:
2.1 Income Social Dataset

The income social dataset represents all

population data of the citizen's income in the Khoums Municipality since the year 2019; it indicates the real population data of the whole municipality.

Table 1. Illustration the income data built on salaries

The table includes the following data:
- Number of Mahallas inside the municipality.
- Population in every Mahalla.
- Percentage of males and females from the total population.
- Numbers of citizens have salaries and a percentage of male and female who had salaries.
- A number of workers in public and private sectors and their genders.
- Estimated percentage of poverty rate in every Mahala.
- A number of births from the whole population data.

2.2 Marriage Social Dataset
The marriage social dataset demonstrates all the attributes related to the state of marriage which also indicate the number of marriages and the their gender in a population data as well as the lowest and highest age among males and females, the percentage of males to female from the total number of the married couples and the percentage of male to female from the total number of singles in the dataset.

Mahalla	population	PoP-Young	PoP-Girls	lowest-Age	Highest-Age
Baladya	20000	7,044	5,550	16	32
Bin juha	17067	7056	6011	17	29
Mergeb	7200	1406	1194	18	37
Sileen	8022	2012	2010	17	36
Ganema	6200	1769	1431	16	35
Ejhawat	3434	1034	900	16	33
Libda	9500	2300	2200	18	39
Ras hamam	4800	2000	1100	17	31
Sed kalefa	4600	1300	1003	18	32
Suq khamis	20000	9,000	4000	16	36
Sahel	6105	2688	1417	17	34
Gogas	2300	900	800	19	29
Keam	7700	2200	2050	18	35

NumMarried-Male	NumMarried-Famel	Total Married	Rate reluctance	NumSingle-Male	NumSingle-Famel
5000	6177	11177	20%	7947	1430
3323	4567	7890	17%	6789	1101
1073	1107	2180	7%	2504	1046
1545	1667	1545	8%	1956	1756
1111	1188	2299	6%	1422	877
725	698	1423	3%	800	623
1215	1389	2604	10%	3156	3448
438	420	858	5%	750	108
674	757	1431	4%	1965	146
534	654	1188	20%	988	200
1082	1121	2203	6%	1648	655
409	355	764	2%	380	384
1767	1955	3722	7%	2222	1500

Table 2. illustrates the marriage dataset.

According to the previous table of marriage, we will create a derivatives table that contains all the relevant attributes to measure the ratios in case of marriage based on the percent of marriage among genders (males, female), the marriage age inside the municipality, the percentage of singles among genders (males, female) as well as matching it with the married rates from genders..

2.3 Deaths Social Dataset

The deaths social dataset was collected from the social data in the municipality's based on real population. The dataset holds important attributes such as the number of deaths of males and females in every Mahalla, the percentage of the total deaths in both genders, and the percentage of the new births and death among the population of the municipality [1].

Mahalla	population	PoP-Male	PoP-Famel	Deat count-Male	Deat count-Famel	Total Deat	num-New birthy
Baladya	20000	12,450	7,550	2700	1876	4576	824
Bin juha	17067	9056	8011	2043	1108	3151	935
Mergeb	7200	4506	2694	1922	991	2913	931
Sileen	8022	4012	4010	1005	844	1849	161
Ganema	6200	3769	2431	655	877	1532	222
Ejhawat	3434	2034	1400	867	642	1509	225
Libda	9500	4300	4200	1017	1012	2029	366
Ras hamam	4800	3000	1800	907	483	1390	424
Sed kalefa	4600	2300	2300	1095	1104	2199	401
Suq khamis	20000	14,000	6000	2207	1880	4087	327
Sahel	6105	3688	2417	1873	1304	3177	569
Gogas	2300	1300	1000	460	277	737	183
Keam	7700	4200	3500	902	800	1702	102

Table 3. Illustrates the deaths dataset.

3 The Statement of the Problem

In this paper, we will endeavor to address the following problems: -

1-Collecting the social data in the municipality of AL Khoums and creating datasets that hold the population census, the number of deaths, the number of births, the number of citizens who have a job, the number of citizens who do not have a job, the number of marriages and the number of divorce cases.

2-Using data mining techniques and algorithms to classifying the data and extracting the important rules and facts from these datasets.

3-Assisting the shareholders and decision-makers in the municipality in obtaining the necessary information to develop plans and ideas to make the correct decision.

4-Obtaining results about the population and the citizens' states in the municipality and creating a data warehouse that can be referenced when it is requested.

4 Methods and Model

After preparing the datasets in the cleaning process in the preprocessing stage, we are going to apply data mining methods and algorithms such as classification and clustering [2].

4.1 Classification Method and Algorithm

We are going to train the datasets using one of the classification algorithms called K- Nearest Neighbor (KNN). The KNN algorithm is a supervised learning algorithm used usually for classification problems. KNN algorithm hugely depends upon the number of K where it should be chosen carefully. The algorithm will improve the boundary conditions, which could be utilized to determine each target class by calculating the distance measures between the attributes. Once the boundary conditions are determined, the next task is to predict the target class. Using KNN as a classification algorithm will ensure the accuracy of interpreting the results in the training datasets [3].

4.2 Clustering Method and Algorithm

Clustering is a common data mining approach based on cluster analysis of the data, the aim is to obtain an intuitive understanding of the problem data space. Such an approach is especially useful where the data records are unlabeled. We are going to utilize one of the clustering algorithms called the K-means algorithm to create cluster partitions. The partition clusters in the training process of the dataset will be created based on the K-means algorithm. In another word, it is aimed at parting n observations into K clusters in which each observation belongs to the cluster with the nearest mean. It is similar to the anticipating maximization algorithm for mixtures of Gaussians in that they both endeavor to find the centers of natural clusters in the data as well as in the iterative lucidity approach utilized by both algorithms. K-means algorithm utilizes also Euclidean distance to measure the distance between points in the parting clusters [4].

4.3 WEKA Software

The Weka software is one of the well-known programs used as a solution for data mining, through which it is possible to apply or use methods like classification and clustering as well as algorithms such as KNN and K-means. We will use the Weka program on the social datasets that we have collected from the municipality of Al-Khoums. These data include social dataset on marriage, social dataset on income, and social data on deaths. When using the Weka program on social dataset for income, we will be able to measure the poverty rate. We can also measure the marriage rate if we use social dataset of marriage. As well as we can measure deaths rate if we use social dataset of deaths. These measurements and rates can greatly help leaders and decision makers to take appropriate decisions based on these rates.

Our work aims to apply some data mining techniques, which elucidate how the data mining methods such as classification and clustering analysis are utilized in municipal social dataset. In classification methods we are going to apply K-Nearest Neighbor (KNN) algorithm to calculate the distance and the similarity measures. Beside in clustering we are going employ K – means algorithm to calculate the distance and the centroid point which belongs to the cluster with the nearest means.

5 Algorithms

The algorithms used in this paper are derived from the methods used in the data-mining environment. One of these methods is classification, in which we used the KNN algorithm; on the other hand, we used the K – means algorithm as one of the algorithms of the clustering method.

Those algorithms will be applied on the social datasets gathered from the municipality of Al-Khoums which holds datasets such as income dataset, marriage dataset and deaths dataset. .

5.1 K- Nearest Neighbor (KNN) Algorithm

The KNN algorithm will generate the social income dataset mentioned in tables 1,2 and 3, which holds datasets about salaries, marriages and deaths in the year of 2019 [5].

- **Poverty Rate Measurement**

Measuring the poverty rate, which are associated with the increasing rate of the population in the dataset? We will classify two categories that represent increase and the other represents decrease. In this case, the classification field contains "H" which represents the value of the increase and the value "L" will represent the value of the decrease indicated as follows:

High (H) = Increasing poverty rate.
Low (L) = Decreasing poverty rate.

The following table represents the poverty rate relied on the salaries of citizens. If the salary of the resident is less than 1000 dinners, it will be classified as low, but if the salary is greater than 1000 dinners, it will be classified as high. The following table shows this classification.

Mahalla	Main salary	Salary >=1000 X1	Salary <1000 X2	Class	POVERTY rate
Baladya	9377	1430	7947	L	7%
Bin juha	7890	1101	6789	L	6%
Mergeb	3550	1046	2504	L	6%
Sileen	3712	1756	1956	H	3%

Ganema	2299	877	1422	L	5%
Ejhawat	1423	623	800	H	2%
Libda	6604	3448	3156	H	4%
Ras hamam	858	108	750	L	5%
Sed kalefa	2101	146	1955	L	6%
Suq khamis	1188	200	988	L	4%
Sahel	2303	655	1648	L	5%
Gogas	764	384	764	H	2%
Keam	3722	1500	2222	L	5%

Table 4. Shows the poverty rate based on salaries.

According to table 4, applying the KNN algorithm, it requires us to calculate the distance. Initially, the distance between the new point in addition to each training point currently exists will be calculated by Euclidean Distance which currently exists as the square root of the sum of the squared differences between a new variance (x_1) in addition to an existing variance (x_2), also the class of the value of "Keam" currently, exists remaining unknown to all of us currently, based on that we are going to do the following:

First step we will calculates the distance dist. (D)

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

These distance measures will run as many times as there are rows in the dataset. The following table represents the distance calculation.

Mahalla	Main salary	Salary >=1000 X1	Salary <1000 X2	Distance	Class
Baladya	9377	1430	7947	9.17	L
Bin juha	7890	1101	6789	7.69	L
Mergeb	3550	1046	2504	3.35	L
Sileen	3712	1756	1956	3.51	H

Ganema	2299	877	1422	9.77	L
Ejhawat	1423	623	800	1.22	H
Libda	6604	3448	3156	6.40	H
Ras hamam	858	108	750	1.41	L
Sed kalefa	2101	146	1955	2.99	L
Suq khamis	1188	200	988	9.94	L
Sahel	2303	655	1648	7.77	L
Gogas	764	384	764	9.36	H
Keam	3722	1500	2222	3.52	L

Table 5. Illustrate the distance calculation

▪ **Marriage Rate Measurement**

The increase in the population growth rate in the municipality is accompanied by an increase in the marriage rate. Based on that, to measure the rate of increase in marriages, we will classify it with two variables, high (H) and low (L). The following table represents the increasing rate of the marriage among the citizens.

- **H** = Rate of marriage is increasing
- **L** = Rate of marriage is decreasing

Mahalla	Population	New-marriage X1	New-born X2	Rate of increasing %
Baladya	20000	1236	4234	4%
Bin juha	17067	1160	3786	4%
Mergeb	7200	1058	1167	3%
Sileen	8022	1072	2789	3%
Ganema	6200	988	2346	2%
Ejhawat	3434	660	987	2%
Libda	9500	712	3122	2%

Ras hamam	4800	444	1288	1%
Sed kalefa	4600	496	1201	2%
Suq khamis	20000	3000	5644	4%
Sahel	6105	1102	1550	4%
Gogas	2300	376	877	1%
Keam	7700	1080	2099	3%

Table 6. Demonstrate the increasing rate of the marriage among people.

Applying the KNN algorithm requires us to calculate the distance. Initially, the distance between the new point in addition to each training point currently exists will be calculated by using Euclidean Distance measure, which currently exists as the square root of the sum of the squared differences between a new variance (x_1) in addition to an existing variance (x_2), also the class of the value of "Libda" currently, exists remaining unknown to all of us. According to that we are going to apply Euclidean Distance measure as follows:

The Euclidean Distance masseuse dist. (D)

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

The distance measures calculated according to Euclidean Distance equation are shown in the below table.

Mahala	Population	New-marriage X1	New born X2	Rate of increasing %	Class	Distance
Baladya	20000	1236	4234	4%	H	5.27
Bin juha	17067	1160	3786	4%	H	4.74
Mergeb	7200	1058	1167	3%	H	**2.02**
Sileen	8022	1072	2789	3%	H	3.66

Ganema	6200	988	2346	2%	L	3.13
Ejhawat	3434	660	987	2%	L	**1.44**
Libda	9500	712	3122	2%	L	6.40
Ras hamam	4800	444	1288	1%	L	**1.52**
Sed kalefa	4600	496	1201	2%	L	**1.5**
Suq khamis	20000	3000	5644	4%	H	8.44
Sahel	6105	1102	1550	4%	H	2.45
Gogas	2300	376	877	1%	L	**1.05**
Keam	7700	1080	2099	3%	H	8.58

Table 7. Shows the distance measures of the marriage rate.

▪ **Death Rate Measurement**
The death rate is usually measured by the rate of decrease, and from this aspect, the measurement of the rate of decrease represents all decreases in the dataset. As mentioned earlier, we will determine two categories called (H, L) with their particular class provides. Whereas; "H" indicate the adult deaths and "L" indicates the young deaths among people at a specific level of age.
- H = Rate of adult deaths.
- L = Rate of young deaths.

Mahalla	Population	Adult deaths	Young deaths	Rate of decreasing %
Baladya	20000	4576	986	2%
Bin juha	17067	3151	970	2%
Mergeb	7200	2913	808	2%
Sileen	8022	1849	732	2%
Ganema	6200	1532	1498	5%
Ejhawat	3434	1509	1430	4%
Libda	9500	2029	1522	5%
Ras	4800	1390	1254	4%

hamam				
Sed kalefa	4600	2199	1196	3%
Suq khamis	20000	4087	913	2%
Sahel	6105	3177	806	2%
Gogas	2300	737	1046	3%
Keam	7700	1702	880	2%

Table 8. Demonstrate the deaths rate among people.

As a result of the previous table, we will apply the KNN algorithm, which requires calculating the distance using the previous law on the two variables (x1, x2) on the class of the value of "Ganema" which it was unknown previously. The following table represents the distance measures of the deaths.

Mahalla	Population	Adult deaths	Young deaths	Rate of decreasing %	Class	Distance
Baladya	20000	4576	986	2%	L	5.36
Bin juha	17067	3151	970	2%	L	3.92
Mergeb	7200	2913	808	2%	L	3.51
Sileen	8022	1849	732	2%	L	**2.38**
Ganema	6200	1532	1498	5%	H	2.82
Ejhawat	3434	1509	1430	4%	H	**2.74**
Libda	9500	2029	1522	5%	H	3.35
Ras hamam	4800	1390	1254	4%	H	**2.44**
Sed kalefa	4600	2199	1196	3%	H	3.19
Suq khamis	20000	4087	913	2%	L	4.8
Sahel	6105	3177	806	2%	L	3.78
Gogas	2300	737	1046	3%	H	**1.49**
Keam	7700	1702	880	2%	L	**2.38**

Table 9. Illustrations the distance calculate for all the data points to deaths.

5.2 Using K-means Algorithm Cluster

Using the k-means algorithm as a one of the clustering algorithms on our income salary social dataset indicated in the below table to measure the poverty rate among citizens inside a municipality. The poverty rate was calculated by two attributes (x1) and (x2) which represent the Mahala that had a greater number of people whose salaries are higher than the others. In this case, we set k=2, which will demonstrates the number of clusters that we are willing to create, the following table shows that dataset that we are going to apply the k-means algorithm [6] :-

Sample	Salary>= 1000 X1	Salary< 1000 X2	Class	POVERTY rate
1	1430	7947	L	7%
2	1101	6789	L	6%
3	1046	2504	L	6%
4	1756	1956	H	3%
5	877	1422	L	5%
6	623	800	H	2%
7	3448	3156	H	4%
8	108	750	L	5%
9	146	1955	L	6%
10	200	988	L	4%
11	655	1648	L	5%
12	384	764	H	2%
13	1500	2222	L	3%

Table 10. Shows that dataset of poverty rate for the k-means algorithm .

As a clustering algorithm, it required to calculate the distance. The distance will be calculated by

using the Euclidean distance with the following equation.

Distance $(x,y),(a,b) = \sqrt{(X-a)^2 + (x-b)^2} = 0$

The below table shows the results of the Euclidean distance measures which produced, the result of two clusters.

Cluster	Euclidean Distance		
	Cluster1	Cluster2	Sample
Dataset (1430, 7947)	0	38.5	1
Dataset (1101, 6789)	38.5	0	2
Dataset (1046, 2504)	13.30	4.62	3
Dataset (1756, 1956)	11.92	21.26	4
Dataset (877, 1422)	41.94	32.52	5
Dataset (623, 800)	55.51	21.24	6
Dataset (3448, 3156)	22.38	98.06	7
Dataset (108, 758)	62.67	45.05	8
Dataset (146, 1955)	33.33	53.63	9
Dataset (200, 988)	54.89	62.05	10
Dataset (655, 1648)	29.75	25.35	11
Dataset (384, 764)	58.87	29.59	12
Dataset (1500, 2222)	16.90	58.76	13

Table 11. Shows k-means cluster on salaries dataset.

- **Marriage Rate Measurement**

We are going to measure the rate of the increasing population in terms of the marriage rate using the k-means algorithm according to the distance measures and the two clusters. The following table represents the marriage rate distance measures using cluster algorithm.

Cluster	Euclidean Distance		
	Cluster1	Cluster2	Sample
Dataset (1236, 4234)	0	77.30	1
Dataset (1160, 3786)	77.30	0	2
Dataset (1058, 1167)	97.15	82.88	3
Dataset (1072, 2789)	45.98	53.16	4
Dataset (988, 2346)	60.21	41.25	5
Dataset (660, 987)	18.50	44.20	6
Dataset (712, 3122)	38.87	69.56	7
Dataset (444, 1288)	96.46	58.61	8
Dataset (496, 1201)	98.72	3.205	9
Dataset (2000, 5644)	50.71	65.08	10
Dataset (1102,1550)	84.98	49.83	11
Dataset (376,877)	27.40	61.89	12
Dataset (1080 ,2099)	28.89	44.59	13

Table 12. Illustration marriage rate distance measures using cluster algorithm.

- **Deaths Rate Measurement**

As we mentioned before concerning the death rate is the rate of decreasing, and from this aspect, we are going to measure the decreasing population based on deaths using the k-means algorithm, to

measure the distance of two clusters. This calculation will be indicated in the following table.

Cluster	Euclidean Distance		
	Cluster1	Cluster2	Sample
Dataset (4576, 986)	0	47.81	1
Dataset (3151, 970)	47.81	0	2
Dataset (2913, 808)	53.14	52.88	3
Dataset (1849, 732)	66.41	83.11	4
Dataset (1532, 1498)	50.28	26.21	5
Dataset (1509, 1430)	97.99	31.19	6
Dataset (2029,1522)	55.52	93.45	7
Dataset (1390,1254)	16.67	26.00	8
Dataset (2199,1196)	51.90	63.39	9
Dataset (4087 ,913)	74.61	60.37	10
Dataset (3177,806)	44.60	28.97	11
Dataset (737, 1046)	60.12	77.53	12
Dataset (1702, 880)	91.03	30.96	13

Table 13. Shows the deaths rate distance measures using cluster algorithm.

6 Experiments and Results

The aims of this paper, is to investigate the main issues that impact e-Government implementation and suggest useful practice guidelines for a successful e-Government application within the context of developing countries such as Libya. The objectives of this paper that arise as a result of the above aim are: -

• Develop a strategy for creating a municipal data warehouse.

• Facilitate accessing, processing, and managing municipality data.

• Employees in the municipality can store easily and securely via intelligent devices.

• Providing services to citizens in the municipality through the municipality's websites.

• The ability to provide an external dataset to exchange data with the internal datasets in the data warehouse to provide fast data sharing in the time-series data environments.

6.1 Results of Using KNN Algorithm

The result that we obtained from applying the KNN algorithm to the statistical community of the referred data of the Mahala within the municipality, depending on the number of males and females who receive a salary from the public sector or the private sector, is as follows:
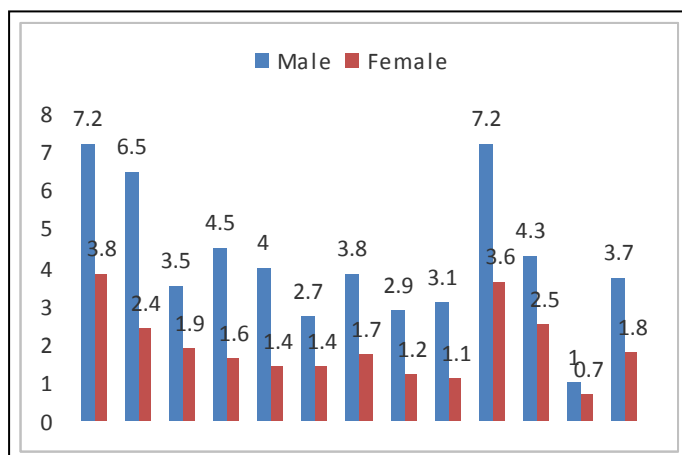

Figure 1. Shows the population in every Mahala inside a municipality based on the salary.

The above chart illustrates the measure of the poverty rate based on salaries that are received every year.

We also utilized a function to calculate the average salaries of individuals of both genders in each Mahala within the municipality and the results are demonstrated in the following figure as:
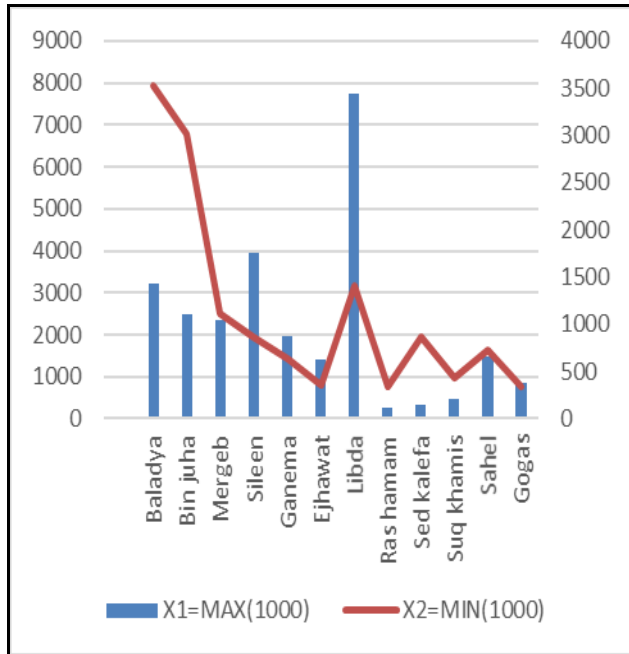

Figure 2. Illustrates the different salaries among every Mahala.

The above figure displayed the statistic calculation of the percentage for the income salaries in terms of the Max (>1000) and Min (<1000) for every Mahala and the results were promising and close to reality, where it is as follows:

In the Mahala of "Baladya", the number of people receiving salaries with Min (<1000) reached 8000 individuals representing 25%. Whereas; on the opposite side 3200 individuals receiving the highest salaries representing 12%. On the other hand, our consideration of the minimum population in the Mahala of "Gogas" is the number of citizens who received high salaries that reaches 900 individuals with an average of 8%, whereas the opposite of 700

individuals with an average of 5%, which illustrate the percentage are close to each other.

The Mahala of "Libda" appeared among the highest salaries in the municipality. Whereas; the Mahala of "Goga" were the lowest salary compared to the population of municipality social dataset.
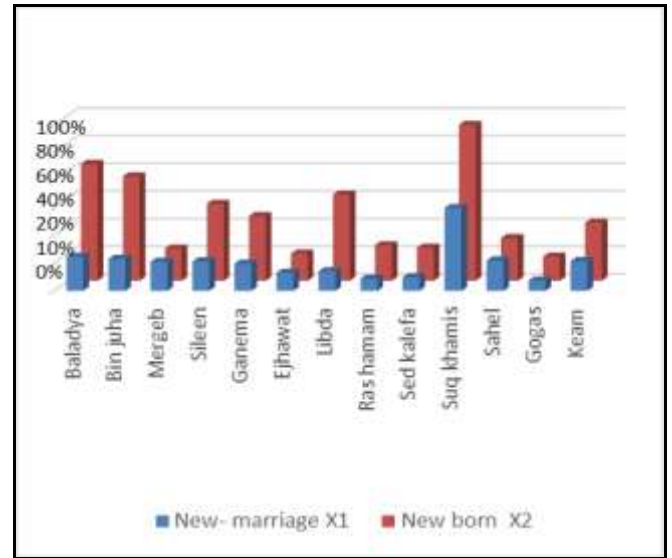

Figure 3. Shows the increasing number of people in case of new born and new marriages.

In figure 3, indicate the increasing rate of the population after measure the marriage rate which appeared to be a rising mark. According to that we calculate the marriage and new born people rates yearly to detect the increasing percentage of the general population inside the municipality.
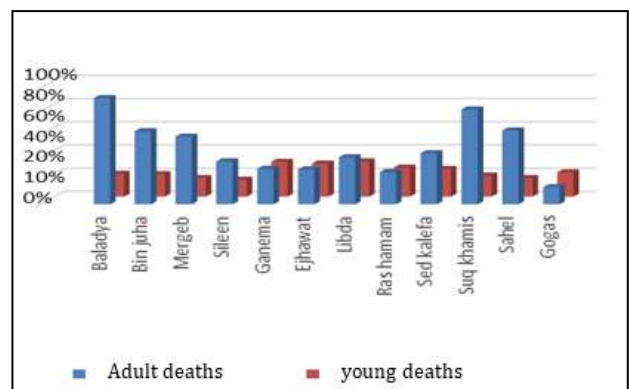

Figure 4. Displays the decreasing rate of the deaths.

The above figure is the statistical population decline from two decreasing components which are the deaths of the adults and the deaths of the young people among the citizens in the municipality. According to that the figure shows the adult's death

rate is estimated as 77%, whereas; the young people death rate is estimated as 33%.

6.2 Results of Using K- Means Algorithm

As we applied the k-means algorithm to the municipality's social dataset, were it is one of the algorithms that is used in the clustering method. We were able to deduce statistical readings of the poverty rate by increasing or decreasing salaries, the rate of increase in the population, deaths rate and population decrease. These statistics are illustrated in the following graphs:

Figure 5. Shows the rates of the two clusters of the salaries.

Figure 5, demonstrates the percentage of neighbouring high salaries as well as low salaries between the Mahalas in the municipality social dataset. Whereas; cluster1 represents the values of salaries greater than 1000 Libyan dinner, while cluster2 represents the salaries less than 1000 Libyan dinner. In the statistical analysis, the results appear to be fairly close.
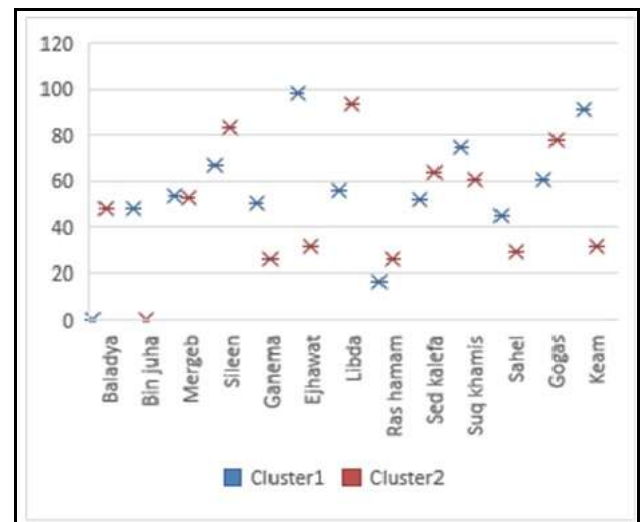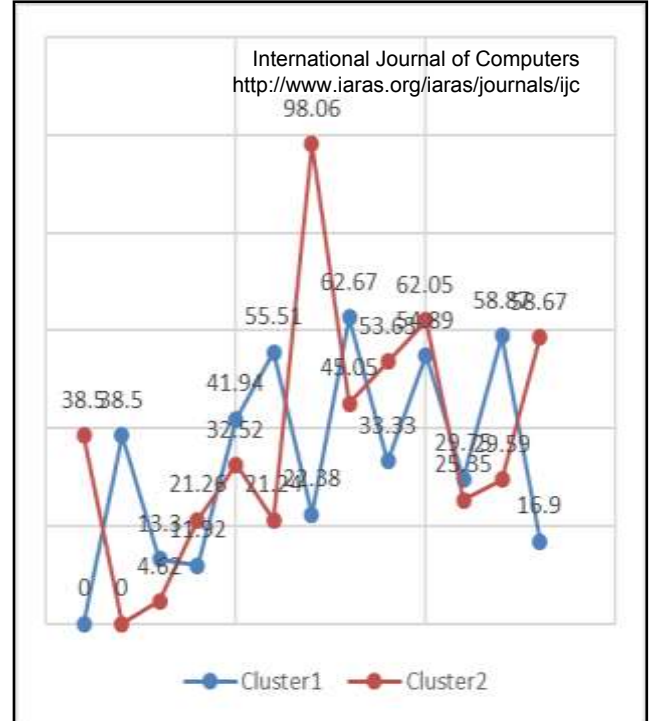

Figure 6. Illustrates the increasing rate of the population in each Mahala.

In this regard, and in the above figure, we have shown the rate of increase in population. This increase includes the rate of new marriages and new-born, as the average rate of marriage and

new born compared to the population within the municipality for each Mahala is somewhat close Based on the values of cluster 1 (new marriage) and cluster 2 (new-born).
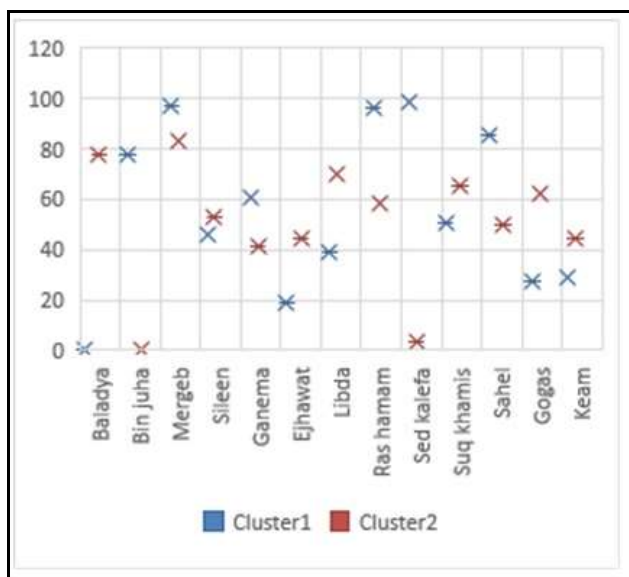

Figure 7. Displays the deaths rates in the population in each Mahala.

The graph indicated in the figure, shows the death rate between adults and children for each Mahala within the municipality social dataset. Whereas; cluster1 represents the values of adults death rate and cluster2 represents the values of children death rate.

7 Evaluations

According to the social dataset obtained from the municipality. As well as analyzing our experiments results from the use of the KNN algorithm as a method of classification, beside on our results of the use of k-means algorithm as a clustering method. The evaluation of this work will be evaluated into two variables, accordingly that the variable X1 represents high values in terms of salary >1000, new marriage and adult deaths. On the other hand, the variable X2 will represent low values such as salary<1000, new-born, and children deaths. Based on that we will be able to evaluate our work as follows:

X1: represent high values in all charts.

X2: represent low value in all charts.

Value	Poverty rate	Increasing rate	Decreasing rate
X1	25.11%	22.96%	35.55%
X2	12.02%	36.18%	28.41%

Table 14. The evaluation results of the municipality social dataset used.

In table 14, this displays the largest values, which also represent the increase, and the smallest, which also represent the decrease. These values are the results of using the aforementioned algorithms. We evaluated these results as follows:

- **Poverty rate**: The evaluation result based on the proposed system gave 25.11% for those with salaries less than 1000 LYD, compared to 12.02% for those with salaries more than 1000 LYD. We can say that the population in the municipality suffers from a percentage of 25% poverty.

-**Increasing the population rate**: Through evaluating the deteriorating results from the proposed system. The percentage of increase in the population was estimated at 22.96% based on new marriages, in addition to 36.18% for new born. As a net result of the rate of increase in population in the municipality will be 59.14 %.

-**Decreasing the population rate**: The high rate of population depreciation is considered unfortunate sometimes, but it is considered as results obtained from the proposed system that we cannot ignore. Whereas, the evaluation of the population decrease rate for the number of deaths was 35.55%, in addition to the 28.41% for those that reluctance to marriages. This indicator rises to 63.96%, which is slightly higher than the rate of increase.

8 Conclusions

We used the existing social dataset in the municipality of Al Khums in Libya. The dataset contained a lot of data on citizens within the municipality, but it was not organized. We used one of the data mining methods in classifying the data, which is classification, and in this

aspect we used the KNN algorithm. We also used another method of data mining in clustering the data through K-means algorithm. After applying the data mining methods, we were able to measure the following rates:

- Measurement of poverty rate (via income).
- Measurement of increasing the population rate (via marriage).
- Measurement of decreasing the population rate (via deaths).

To measure these rates using the above-mentioned data mining methods, it is necessary to use an application to extract these data and measure the rates mentioned above. In this regard, we have used the Weka software, which is one of the pioneering programs in the use of different data mining methods.

We applied Weka software to the municipality's social dataset, as well as implementing the KNN algorithm and the K-means algorithm. We were able to measure the poverty rate, the rate of increase and the rate of decrease in the municipality's population.

Besides, we were able also to conclude, some results, which are promising, as well as some unfortunate results on the situation of the population in general in the municipality.

Finally, this work that we have done on the social dataset in the municipality of Al-Khums in Libya is considered the first of its kind, which is based on helping decision makers in the municipality to make the right decision. We can also develop this proposed system in the future to include the evaluation of institutions within the municipality according to the quality standards used in the quality of international institutions.

References:

[1] Md. Rafiqul Islam, Morshed U. Chowdhury and Safwan Mahmood, Khan, *Medical Image Classification Using an Efficient Data Mining Technique,* School of Information Technology Deakin University, Burwood, Victoria 3125, Australia Complexity International.

[2] M. H. Dunham, *Data Mining, Introductory and Advanced Topics*, Prentice Hall, 2012.

[3] *KNN Algorithm - Finding Nearest Neighbors* (tutorialspoint.com).

[4] *Cluster Analysis - Definition, Types, Applications and Examples* (byjus.com) Login clock 03:55 am. date: 22-10-2021.

[5] *Datasets (Definition, Types, Properties & Example)* / Dataset in Mathematics. (byjus.com).

[6] *K-means clustering,* http://en.wikipedia.org/wiki. Login clock 01:28 am. date:14-11-

[7] Zakaria Suliman Zubi, Ayman Mahmmud, *Using Data Mining Techniques to Analyze Crime Pattern*, 3rd International Workshop of Innovation and New Trends in Information Systems) INTIS'2013), held in Tangier – Morocco from 29-30 November 2013.

[8] Zakaria Suliman Zubi, Ayman Altaher Mahmmud, *Crime Data Analysis Using Data Mining Techniques to Improve Crimes Prevention,* international journal of computers, ISSN: 1998-4308, Volume 8, 2014.

[9] Zakaria Suliman Zubi, Rema Asheibani Saad, *Improves Treatment Programs of Lung Cancer Using Data Mining Techniques*, Journal of Software Engineering and Applications, 2014, 7, 69-77, Published Online February 2014, (http://www.scirp.org/journal/jsea) , (http://dx.doi.org/10.4236/jsea.2014.72008).

[10] Zakaria Suliman Zubi, Ali A. Elrowayati, Ibrahim Saad Abu Fanas, "*A Movie Recommendation System Design Using Association Rules Mining and Classification Techniques*," WSEAS Transactions on Computers, vol.20, pp. 189-199, 2022.