International Journal of Educational Sciences and Arts (IJESA)

Vol. (3), No. (12)

IJESA

December 2024

المجلة الدولية للعلوم التربوية والآداب

الإصدار (3)، العدد (12)

Analysis of geographical and economic factors affecting water consumption in Sabha city-Libya using (WEKA method)

Mabroka Mohamed Daw
Sebha University, Libya
mabr.mohamed6@sebhau.edu.ly

Elhadi Ramadan Ali Abdulhadi
Wadi Alsahtti University, Libya
e.abdulhadi@wau.edu.ly

Abstract

In the middle and southern part of Libya, groundwater is the main water source for all uses due to the extremely low rainfall distribution. The excessively rapid population growth in the urban areas, coupled with very high temperatures – especially in summer - have adversely impacted the groundwater quantity. To study the situation, this study aimed to determine key factors affecting the water consumption. Specifically, the study assesses the factors that impact on consumption for urban residential water in the study area, using the relationship equation (liner regression equation) between the variables of the study. The Waikato Environment for Knowledge Analysis (WEKA) software workbench was selected as a predictive modeling alternative to linear regression model. The strategy used to determine the best operational procedure included (ARFF) to define the type of data, discerning the optimal dataset size, iteration setting, predictor subsets, target attributes. The WEKA models and settings were tested using multiple data strategies on all available data and in situ data since 1973 to 2023. The optimal settings and the linear regression models were used to determining the effect of some factors on water consumption in study area, also to

compare between the actual and predict data and get the error value and then compare the error value with the Kappa statistic value. The result of the linear regression model showed a relationship and positive effect between the water demand (WD) and (temperature T), (summer temperature ST), (population P) and (urban population UP). At the same time no relationship existed between (WD) and (income IN), (water price WP).

Keywords: WEKA Method, Water Consumption, Sabha City-Libya, Water Resources, Water Situation.

Introduction

In human history, over 50% of the global human population lives in the urban area. This population growth in urban areas in the 21st century has a lot of challenges and implications for water supply management. Forecasting and managing water consumption is a major challenge due to the complexity of the relation between human and natural systems(Martine et al. 2007). In the study area, underground water is the major source for drinking and other consumptions. However, groundwater's availability and quality are greatly influenced by climate change and over-abstraction, especially in regions with a low water table (Algnariani et al. 1955).The shortage of water creases the ability to maintain quality, especially if there are various sources of pollution. Also, the increased population in the urban regions have led to increasing water consumption and, unfortunately, leading to more contamination, which ultimately affects the quality of the water. The scarcity and contaminated water issues, among many others, serve as major drawbacks and challenges for the Libyan government. The rapid growth of population, urbanization, agriculture, and industrialization have increased great pressure on land, water and the environment, and their quality. The study area had witnessed rapid population growth during the period 1973-2023. The population was nearly 32164 in 1973, 75519 in 1984, and in 1995, the population reached 102352. During the period 1995-2006, the population increased to 121840 and in 2010, 13743, The population continues rising to reached 211,641 in 2020 and 223,636 in 2023.

However, it is not being matched with an increase in water supply, and consequently, the situation is deteriorating (Abughlelesha, S. M. & Lateh, H. B. 2013). The concentration of the inhabitant's density in urban areas has led to the increases of buildings, roads, and factories .These human activities result in several effects, including water waste and storm drain, smoke, dust, and solid waste, which lead to severe water and air pollution .Evaluating the factors that affect the water consumption are an important aspect because the growth of population, urbanization and the other factors play an important role to determine the quality and quantity of water demand. Today, the study area is facing one of the severest water shortage problems: Rainfall scarcity. In the coastal regions, the groundwater sources are renewed by yearly rain, but the vast groundwater reserves underlying the desert are not refilled. Over a long period, the country's demand for water has far exceeded supply, leading to heavy over-drafting and mining of aquifers, which is linked to the growing problems of aquifer diminution, quality worsening, and saltwater intrusion (Elgali 2013). A WEKA model was identified as being the most suitable to find the relationships between X and Y Linear and logistic regression models are the basic predictive modeling approaches used to establish relationships between a set of independent variables (predictor, features, X) and one or more dependent variables (target, Y). A logistic regression is usually used as a classifier in deep learning models. Another important research area for regression analysis is the estimation of independent variables’ importance with respect to the dependent variable. The development cycle of a regression analysis may include data collection and wrangling, standardizing the variables, optimizing model hyper parameters, validation of the model outputs, and sometimes deployment of the model to real-world applications (Kosorus et al. 2011).

Water Resources in Study Area (Sabha city)

There is only one source of water in Sabha city: groundwater (which is comprised of approximately 95% of the country's water). Libya's groundwater is reserved in five basins as follow: Jabal El-Akhdar, Kufra/as-Sarir, AlJafara, Nafusah/al-Hamada and Murzek. The groundwater is further classified into two, based on their renewability; renewable groundwater, which is found in shallow aquifers, and the non-renewable groundwater, usually referred to as fossil water source, which is found in deep aquifers. The characteristics of Libya's groundwater reservoir are summarized in (Table 1).

Table (1) Groundwater reservoirs characteristics - Source: Edawi Wheida & Verhoeven, 2007

Basin	Area	Renewable	Nonrenewable	dissolved
Catachrestic	(km)2	106 m3	106 m3	Solids mg/1
Al-jabal Alakhdar	145,000	200	50	1000-5000
Alkufra, Al-sarir	700,000	-	1800	200-1500
Al Jafara Plain	18,000	200	50	1000-5000
Al Hamada	215,000	250	150	1000-5000
Murzek	350,000	-	1800	200-1500

The Southern Region (Basin of Murzuk)

Murzuk basin is considered as the second largest water basin in Libya. it is located in the south-western part in the country. This region contains a large number of wells divided into five areas with varied depths. The estimated annual consumption of water in this basin for all purposes is 1454 million m3 per year. There is a significant decline in the water level, due to the huge withdrawal, as a large part of this water has been transferred to Al-Jafara plain in the north-western part of Libya (GWA 2006).

Water Resource Situation

There is a high water consumption in urban areas due to their domestic use, in gardens, hospitals, school and universities, public buildings, hotels, and businesses such as cafés and markets (E. Wheida & Verhoeven 2006). This high consumption is a corollary to the population boom. Domestic water consumption account for 90% of all water consumed in the study area. The remaining 10% is distributed between industrial uses and the service sector (Abdudayem & Scott 2014). The volume of water exploited for various uses has increased drastically in recent times, and is expected to reach a prohibitive level in a near future if the present population growth continues with a corresponding increased water usage. Taking the population boom trend into consideration, the per capita renewable water and surface water rate is expected to decline steadily from 170m3 in 1995 to 70m3 in 2025 (Figure1). The population of Sabha expanded from less than 33,000 in 1973 to 223,636 in 2023, and the figure is expected to rise to over twelve million by 2025 (Abdudayem & Scott 2014). (Table 2) shows the available sustainable freshwater reserves to be 2279.5 million cubic meters per year (GWA 2006). Based on these figures, the mean available water per person has reduced from 2280 m3 in 1995 to 380 m3 in 2005 and the decline is expected to continue to 190 m3 by 2025. The implication of this is that Libya is already experiencing water shortage, which will get sever with time (Table 3).

Table (2) Sustainable water supply from all available sources in the major water basins of Libya in million cubic meters per year (MCM/Y) - Source :GWA 2006

Water basin	Groundwater Resource	Surface Water Resources	Unconventional Water Resources	Total
Jafara Plain	200	52	27.5	279.5
Jabal Alakhdar	200	92	45.5	337.5
Alhamra	230	48	50.5	328.5
Kufra and Sarir	563	-	-	563
Murzuk	771	-	-	771
Total	1964	192	123.5	2279.5

Table (3) Historical trend and future predictions of water use by all socio-economic sectors in Libya in million cubic meters per year (MCM/Y) - Source: GWA 2006

Water use	Year				
	1984	1995	2000	2010	2020
Agriculture	1978	3376	3860	4825	5790
Industrial	90	145	176	261	386
Municipal	247	364	457	708	1060
Total	2315	3885	4493	5794	7236

National average, however, mask the spatial and temporal variability of the severity of water shortage on a water basin by water basin basis. Even within the same water basin, water reserves vary widely based on location or region. This implication for this is that strategies aimed at managing water should be unique for each basin and designed with a complete understanding of the hydro-climatic components and their inter-relationship with each reserve. This situation influenced the national authorities' decision to seek for more effective measures to bridge this gap through the development of new water resources to augment the presently available water supplies, in addition to the selection and implementation of specific technical and institutional arrangement related to water demand management .The Great Man-Made River project supplies and distributes over 2000 million cubic meters of water annually through probably the largest and most complicated man-made water distribution network of its kind in the dry areas of North Africa and the Middle East. A project as enormous as this will make provision for significant augmentation of the national water budget, which do not provide a complete solution to the problem of water shortage in Libya as revealed in (Figure 1). Meanwhile, decisions on developing other large-scale non-conventional water resources such as seawater desalination and wastewater treatment and reuse were ongoing.

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

International Journal of Educational Sciences and Arts (IJESA)
Vol. (3), No. (12)
IJESA
December 2024
المجلة الدولية للعلوم التربوية والآداب
الإصدار (3)، العدد (12)

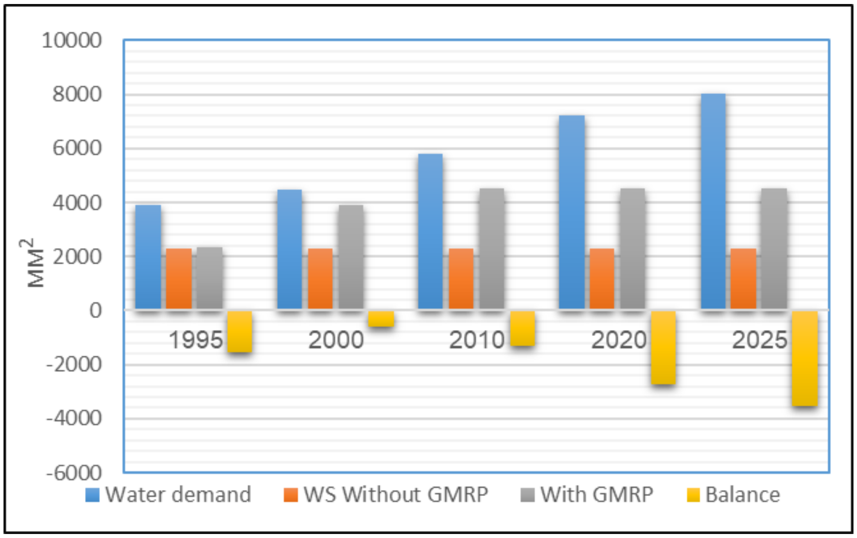

Figure (1) Past, present, and expected future water balance at the national level including the contribution of the Great Man-Made River Project in million cubic meters per year - Source: (GWA 2006).

Analysis the factors affecting water consumption in study area using WEKA

The Waikato Environment for Knowledge Analysis (WEKA) (figure 2) is a software workbench for a suite of machine learning models that meets the selection criteria. It is widely used and highly recommended in the machine learning community and within various disciplines that use machine learning tools. It is also free, open-source software available under the GNU General Public License guidelines. Another benefit is that all the models are available within a uniform, Java-based graphical user interface that facilitates simplified comparisons among multiple models. The WEKA can also process data supplied in the form of a single relational table. WEKA is designed to achieve the following objectives :

a) Assist users in extracting useful information from data; and
b) Ensure ease of identification of a suitable algorithm for generating an accurate predictive model from the data (Hall et al. 2009)

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

Finally, the workbench includes tools for data exploration, data preprocessing, cluster analysis, classification, regression and visualization, which enables robust analysis of the data (Kosorus et al. 2011). WEKA is an open-source software which consists of a collection of machine learning algorithms for data mining tasks (Thomas Moh Shan Yau, Mohd Fauzi bin Othman 2006). This paper employs WEKA including tools for data exploration, data preprocessing, cluster analysis, classification, regression and visualization, which enables robust analysis of the data (Kosorus et al. 2011). A WEKA model was identified as being the most suitable to find the relationships between X and Y Linear, and logistic regression models are the basic predictive modeling approaches used to establish relationships between a set of independent variables (predictor, features, X) and one or more dependent variables (target, Y). A logistic regression is usually used as a classifier in deep learning models. Another important research area for regression analysis is the estimation of independent variables' importance with respect to the dependent variable.

Figure (2) The main user interface of Weka - Source: (Hall et al. 2009)

WEKA Data Processing

Data processing includes two main functions; preparing data so it can be imported into WEKA and summarizing WEKA model outputs for analysis. Before data can be imported into WEKA, it must be converted, organized, and formatted correctly. These processes were performed on all the variables (dependent variables and independent variables). To apply the Linear Regression in WEKA, there are some steps that must be taken. The first step is loading data into WEKA. The data must be in a format that will be understood. WEKA's preferred method for loading data is in the Attribute-Relation File Format (ARFF), which is able to define the type of data being loaded, then supply the data itself. In the file, one must define each column and what each column contains. In the second step, the model is created. The procedure involves clicking on the 'Classify' tab, then clicking the 'Choose' button, and expanding the functions branch. Thereafter, the Linear Regression leaf is selected. Figure (2) shows the steps that will be applied by the study tool (WEKA method) until the stage of reaching the results.

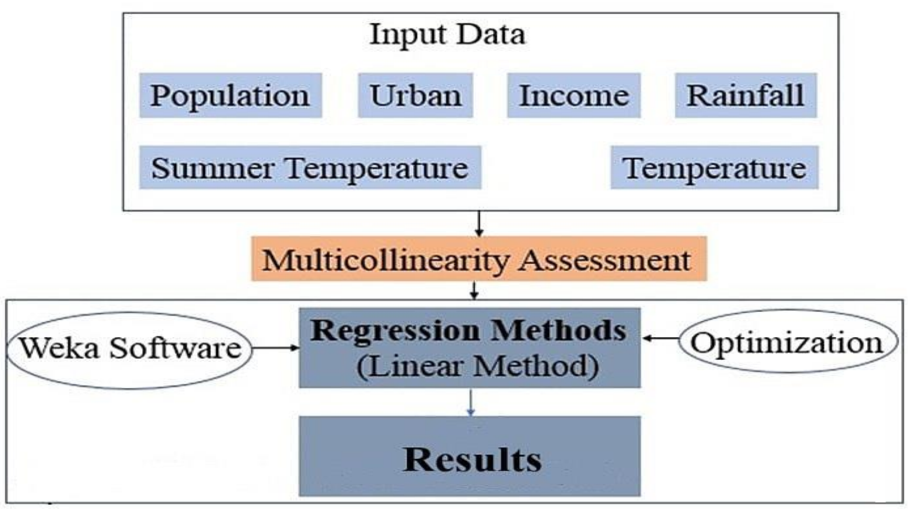

Figure (3) WEKA framework

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

From figure 3, the Class for employing linear regression for prediction uses the Akaike criterion for model selection and is able to deal with weighted instances. The Capabilities include Class (Date class, Numeric class, missing class values) and Attributes (Missing values, Numeric attributes, Date attributes, Binary attributes, Unary attributes, Nominal attributes, Empty nominal attributes). In the third and final step of applying the Linear Regression in WEKA, the dependent variable is selected (the column to be predicted) leading to creation of the model. In this study, the dependent variable is water demand which is the variable of focus. The result is compared with the Kappa value, which is a statistical measure of the agreement between the known results and the predicted results to evaluate a model's predictive capability.

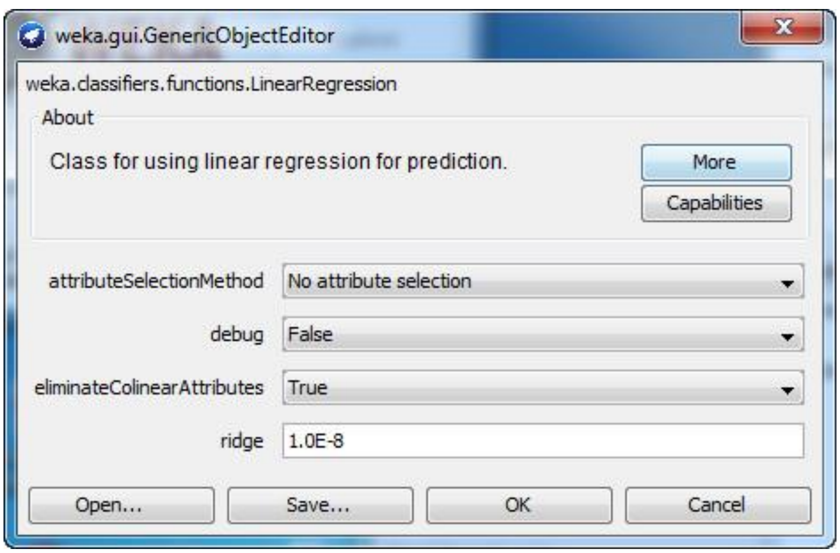

Figure (4) Linear regression model in WEKA

International Journal of Educational Sciences and Arts (IJESA)

Vol. (3), No. (12)

IJESA

December 2024

المجلة الدولية للعلوم التربوية والآداب

الإصدار (3)، العدد (12)

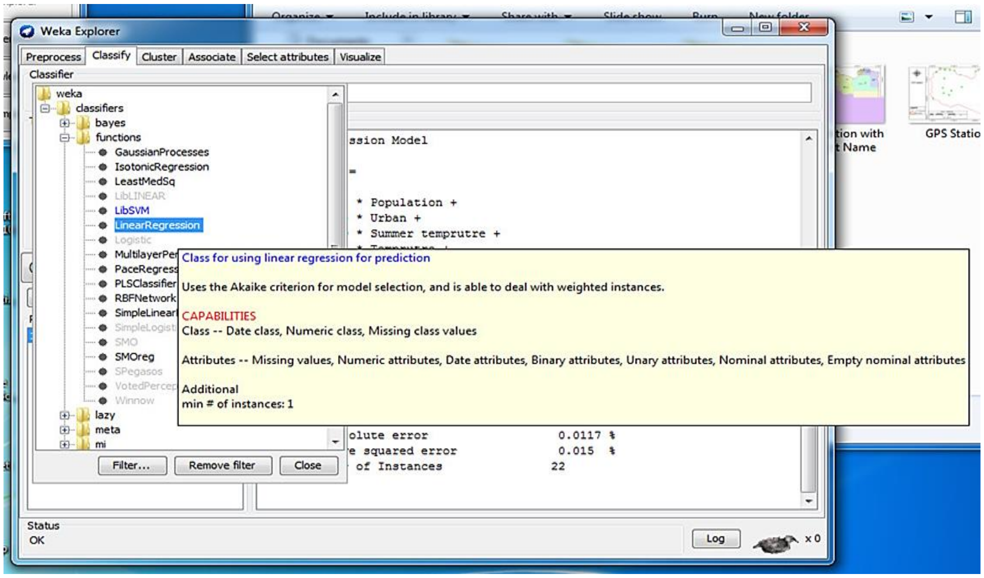

Figure (5) The advantages and capabilities of the use of the linear regression

(Table 4) presents an interpretation of the Kappa statistical values by using estimative language to indicate levels of model agreement for corresponding Kappa value ranges (Aten 2013).

Table (4) Kappa value ranges with estimative language indicating agreement - Source:(Aher n.d.)

Kappa Value	Agreement
< 0	Less than chance agreement
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Almost perfect agreement

By using the linear regression equation, the actual value and the predicted value and the error value in tables can be compared to verify the validity of the data.

Design of the Standard Model and Sources of Data

The Standard Model of the variables studied is formulated by converting the relationship between these variables to a standard formula, which is consistent with previous literature, studies and reality. The CW dependent variable can be expressed as independent variables in its traditional form as follows:

CW = f (ST, T, RF, POP, UP, IN, WP)

Where:

- CW refers to the Consumption of Water in Libya from the periods (1973 - 2023), the General Ministry of Water, the status of water in study area
- ST refers to Summer Temperature. Water consumption is expected to increase according to summer temperatures. The summer temperature series was obtained from the National Observatory of Meteorology, Libya and air reports for the period of 1973-2023.
- T refers to Temperature. Water consumption was expected to increase according to temperature and vice versa. The time series data for temperature was also obtained from the National Observatory of Meteorology, Libya and air reports for the period 1973-2023.
- RF refers to the rainfall rate. It is expected that an inverse relationship exists between rainfall and water consumption. Rainfall data was also obtained from the National Observatory of Meteorology, Libya, and Air reports for the period 1973-2023.
- POP refers to the population census in the study area. A positive relationship was expected to exist between the population census and the water consumption, i.e., the larger the population, the greater water consumption.

The population census data was supplied by the results of the general population census for the period (1973-2023).

- UP refers to the urban population. A positive relationship was expected to exist between the increase of the urban population and the water consumption. Data was obtained urban population census from both the General Ministry for Planning and Economy, the Statistics and Census Authority, the results of the general census of the population for the period (1973-2023).
- IN refers to per capita income. A positive relationship was expected to exist between per capita income and water consumption as the economics theory shows and vice versa. Data on income was obtained from the Central Bank of Libya and the World Bank.
- WP refers to the price of water. A reverse relationship was expected to exist between the price and water consumption as the economics theory confirmed. The researcher obtained time series data of the water price and the situation of water in Libya 2023 from the General Authority for Water, Tripoli.

Literature Review

A review of literature on consumption water reveals that consumption of water for study area dwellers exceeds the other surrounding areas. Individual WEKA method with different functional types, were used to determine the elasticity of consumption of water for income, price, household features, and structure among alternatives. These revisions use plate data, chronology, or cross-section data. The most important and modern studies carried out in this field are:

- **K. Sai Madhuri et al. (2023)** presented comparison between the performance of WEKA and Scikit-learn in predicting whether a person has diabetes or not based on different machine learning algorithms. The dataset contains information on the patient's age, insulin level, blood pressure, and other relevant features. They

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

compared the performance of algorithms such as KNN, decision trees, random forests, and logistic regression dataset, and evaluated their accuracy using various performance metrics. The results of this study can provide insights into which algorithm is most suitable for predictive analysis of the diabetes dataset and can help the user make informed decisions to improve their performance The findings of this study can aid in selecting the appropriate tool for similar data analysis tasks. Through this research, they hope to highlight the effectiveness of WEKA and Scikit-learn in data analysis and how these tools can be leveraged to uncover insightful information that may guide individual choices.

- **Gökhan AKSU and NuriDoğan** (2019) analyzed the one of the data mining methods which is very popular in recent years and commonly used in this area. The WEKA program and the decision trees, which is one of the methods used to estimate the dependent variable through independent variables, will be introduced. In this study, they discussed WEKA software, which is one of the programs in the field of data mining, how to run the program and the content of the analyzes and output files. The study also contains some suggestions for the practitioners who want to use this program about the superior aspects of the software and what kind of analysis can be done with it .
- **Asha Kiranmai and Jaya Laxmi (2018)** presented the classification of power quality problems such as voltage sag, swell, interruption and unbalance using data mining algorithms: J48, Random Tree and Random Forest decision trees. These algorithms are implemented on two sets of voltage data using WEKA software. The numeric attributes in first data set include 3-phase RMS voltages at the point of common coupling. In second data set, three more numeric attributed such as minimum, maximum and average voltages, are added along with 3-phase RMS voltages. The performance of the algorithms is evaluated in both the cases to determine the best classification algorithm, and the effect of

addition of the three attributes in the second case is studied, which depicts the advantages in terms of classification accuracy and training time of the decision trees .

- **Dr. Yousef Abuzir(2018),** presented an applied study using data mining to discover some factors affecting agricultural vegetable production and predicting the yield production in Palestine. to increase the amount of production to benefit the farmers in particular and individual, society in general. To achieve this goal, we used Waikato's Knowledge Analysis Environment (WEKA) tool and algorithms such as K-Means, Kohonen's Self Organizing Map (KSOM) and EM to identify the most influential factors that increase the production of agricultural vegetable. This research has proved that K-Means is worthwhile to increase the efficiency and reliability of the prediction process of determining the factors that affect the yield production and KSOM the most accurate to predict the yield production
- **Swasti Singhal and Monika Jena** (2013) described the steps of how to use WEKA tool for these technologies. The main objective for this study is to introduce the key principle of data pre-processing, classification, clustering and introduction of WEKA tool. In today's world data mining have increasingly become very interesting and popular in terms of all application. The need for data mining is that we have too much data, too much technology but don't have useful information. Data mining software allows user to analyze data. this study will provide the facility to classify the data through various algorithms .
- **Thomas Moh Shan Yau and Mohd Fauzi, bin Othman (2006)** presented the comparison of different classification techniques using Waikato Environment for Knowledge Analysis or in short, WEKA. The aim of this paper is to investigate the performance of different classification or clustering methods for a set of large data. The main aim of this study is to investigate the performance of different

classification or clustering methods for a set of large data. The algorithm or methods tested are Bayes Network, Radial Basis Function, Pruned Tree, Single Conjunctive Rule Learner and Nearest Neighbors Algorithm. The data breast cancer data with a total data of 6291 and a dimension of 699 rows and 9 columns will be used to test and justify the differences between the classification methods or algorithms. Subsequently, the classification technique that has the potential to significantly improve the common or conventional methods will be suggested for use in large scale data, bioinformatics or other general applications.

Results and Discussion

The Relationship Between the Variables

The Waikato Environment for Knowledge Analysis (WEKA) software workbench was selected as a predictive modeling alternative to linear regression model. The strategy used to determine the best operational procedure included (ARFF) to define the type of data, discerning the optimal dataset size, iteration setting, predictor subsets, target attributes. The WEKA models and settings were tested using multiple data strategies on all available data and in situ data since 1973. Data categories included all independent variables of population, Urban population, Temperature, summer temperature, rainfall, income, price and, in addition, the dependent variable of water demand. The optimal settings and the linear regression models were used to determine the residential urban water demand in the study area, also to compare between the actual and predict data and get the error value and then compare the error value with the Kappa statistic value.

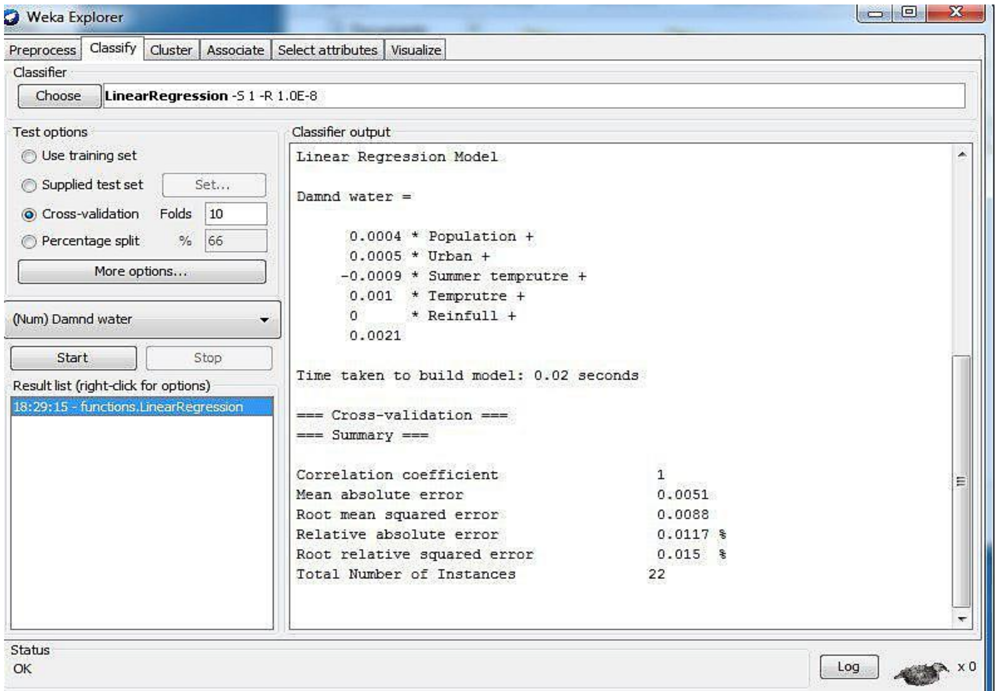

Figure (6) The linear regression model for the parameters

The Linear Regression for the Variables

a. The result of the linear regression model showed a relationship and positive effect between the water demand (WD) and (temperature T), (summer temperature ST), (population P) and (urban population UP). At the same time no relationship existed between (WD) and (income IN), (water price WP) as shown in (Figure 6.1) .

b. The relationship between (rainfall RF) and (water demand WD) is weak and the impact is weak compared to other factors as showed in the statistical analysis and linear regression model. This is contrary to the hypothesis of the study that there is a negative relationship between the rate of rainfall (rainfall RF) and demand for water (water demand WD). This result is attributed to the following reasons:

- study area population do not rely on rain water this is due to the nature of the desert climate of the region, where there is almost no rain.
- Low rainfall rates in Libya where a large area of the country has a semi-desert dry climate and low rainfall. Therefore, rainwater does not constitute a major source for water in Libya.

c. On the other hand, the relationship between (Water Demand WD) and (income IN) is negative, with a significant effect in the long term and short-term. This is also contrary to the hypothesis of the study and the economics theory which states that there is a positive relationship between (IN) and water demand (WD). The researcher explains the reasons for this result as follows:

- Water is a very necessary commodity that must meet the demand and stopping its consumption leads to death. Therefore, demand for water (DW) is usually inflexible to changes in income (IN). This is confirmed by the output of the model as the flexibility of income near zero in the short term (0.04) and was low in the long run (0.3).
- The low impact of income in Libya on demand in general and water in particular, where the main source of income in Libya has been exploited by the public sector for many years, which monopolized the economic activity represented by government salaries. These salaries have been in a standstill for more than two and a half decades with slight increases

d. Also, there is no relationship and negative effect between the (Water Demand WD) and (water price WP) in the short and long terms. This is consistent with the hypothesis of the study as well as the economic theory that state that there is a negative relationship between the price of water (WP) and the (water demand WD). For more in-depth analyses the results, the error value between the actual and predicted value for the water demand was very small, as shown

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

in the following tables below. In addition, the error value between the actual and predicted value for the water demand was very small, as shown in the table below. Also calculated were the Crepitation coefficient, mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error, Total Number of Instances

Table (7) The linear regression model values for the period 1973-2023 - Source: Prepared by the researcher based on the statistical program WEKA results

Parameters	1973- 1985	1986-1998	1999- 2011	2012-2023
Population	0.0004	0	0.0004	0.0006
Urban	0.0005	0.0008	0.0005	0.0008
Summer temperature	-0.0009	0.0009	0.0015	-0.0009
Temperature	0.001	-0.0015	-0.009	0.0005
Income	0	0	0	0
Rainfall	0	0	-0.0001	0
Intercept	0.0021	0.0033	0.1589	0.0129
Mean absolute error (MAE)	0.0037	0.0018	0.0347	0.002
Root mean squared error (RMSE)	0.0066	0.0022	0.0854	0.0024

(Table () contains the values of the linear equation for the period 1973 - 2023. These values were obtained after applying the linear regression model on each period. These values were extracted to get the predicted values for all the variables to compare the actual values with the predicted values and then get the error values. The results of the comparison between the actual values with the predicted values showed that the error values are very low after comparing them with the Kappa statistical measure where the results were as follows: For the census period 1973-1985, 0.9982 for the census period 1986 -1998, 0.9653 for the census period 1999 - 2011 and 0.9980 for the census period 2012 - 2023, thus, the computational capability of the linear regression model is high, with the lowest error rate of 0.9982, after comparing with Kappa statistical measure all results were almost perfect. The following tables show the actual, predicted, and error values for the four periods.

- The linear regression model (1973 - 1985) as illustrated in
- (

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

- Table Error! No text of specified style in document.)

Demand water = (0.0004 * Population + 0.0005 * Urban pop - 0.0009 * Summer temperature + 0.001 * Temperature + 0.0021)

Correlation coefficient	1
Mean absolute error	0.0037
Root mean squared error	0.0066
Relative absolute error	0.0088 %
Root relative squared error	0.0117 %
Total Number of Instances	22

Table **Error! No text of specified style in document.**The actual, predicted and error values for the census period 1973 -1985 - Source: Prepared by the researcher based on the WEKA statistical program results

No.	Actual	Predicted	error
1	36.4	36.427	0.027
2	37.61	37.61	0
3	47.62	47.614	-0.006
4	42.62	42.618	-0.002
5	155.88	155.878	-0.002
6	41.37	41.372	0.002
7	11.16	11.157	-0.003
8	32.68	32.676	-0.004
9	12.03	12.025	-0.005
10	125.43	125.434	0.004
11	98.82	98.819	-0.001
12	245.63	245.63	0
13	104.2	104.194	-0.006
14	67.12	67.121	0.001
15	66.36	66.363	0.003
16	69.91	69.911	0.001
17	21.68	21.683	0.003
18	30.76	30.76	0
19	18.14	18.14	0
20	18.19	18.189	-0.001
21	17.61	17.608	-0.002

22	5.36	5.353	-0.007

- The linear regression model for (1986 – 1998) as illustrated in (
- Table).

Demand Water = (0 * Population + 0.0008 * Urban + 0.0009 * Summer temperature -0.0015 * Temperature +0 * Rainfall +0.0033)

Correlation coefficient 1

Mean absolute error 0.0018

Root mean squared error 0.0022

Relative absolute error 0.0026 %

Root relative squared error 0.0023 %

Total Number of Instances 22

Table 7 The actual, predicted and error values for the census period 1986 – 1998 - Source: Prepared by the researcher based on the WEKA statistical program results

No.	Actual	Predicted	error
1	60.89	60.887	-0.003
2	62.86	62.856	-0.004
3	79.58	79.582	0.002
4	71.22	71.221	0.001
5	260.51	260.512	0.002
6	69.15	69.147	-0.003
7	18.65	18.651	0.001
8	54.61	54.611	0.001
9	20.11	20.106	-0.004
10	209.63	209.63	0
11	165.15	165.151	0.001
12	410.51	410.51	0
13	174.14	174.137	-0.003
14	112.18	112.176	-0.004
15	110.91	110.911	0.001
16	116.83	116.831	0.001
17	36.24	36.24	0
18	51.41	51.412	0.002
19	30.32	30.322	0.002

20	30.4	30.402	0.002
21	29.43	29.432	0.002
22	8.95	8.951	0.001

- The linear regression model for (**1999 – 2011)**, as illustrated in (Table).

Demand Water = (0.0004 * Population + 0.0005 * Urban + 0.0015 * Summer temperature -0.009 * Temperature -0.0001 * Rainfall +0.1589)

Correlation coefficient	1
Mean absolute error	0.0347
Root mean squared error	0.0854
Relative absolute error	0.0271 %
Root relative squared error	0.0495 %
Total Number of Instances	22

Table 8 actual, predicted and error values for the census period 1999-2011 - Source: Prepared by the researcher based on the WEKA statistical program results

No.	Actual	predicted	error
1	111.92	111.934	0.014

2	115.55	115.554	0.004
3	146.28	146.303	0.023
4	130.92	130.928	0.008
5	478.87	478.935	0.065
6	127.11	127.127	0.017
7	34.29	34.285	-0.005
8	100.39	100.401	0.011
9	36.96	36.972	0.012
10	385.77	385.394	-0.376
11	303.58	303.616	0.036
12	754.59	754.692	0.102
13	320.1	320.108	0.008
14	206.2	206.221	0.021
15	203.88	203.903	0.023
16	214.77	214.78	0.01
17	66.62	66.638	0.018
18	94.51	94.517	0.007
19	55.74	55.741	0.001
20	55.89	55.89	0
21	54.1	54.102	0.002
22	16.46	16.459	-0.001

- The linear regression model for **(2012 – 2023)**, as illustrated in (Table).

Demand Water = (0.0006 * Population + 0.0008 * Urban -0.0009 * Summer temperature + 0.0005 *Temperature + 0   * Rainfall + 0.0129)

Correlation coefficient	1
Mean absolute error	0.002
Root mean squared error	0.0024
Relative absolute error	0.0008 %
Root relative squared error	0.0008 %
Total Number of Instances	22

Table 9 The actual, predicted and error values for the census period 2012 – 2023 - Source: Prepared by the researcher based on the WEKA statistical program results

No.	Actual	predicted	error
1	204.42	204.424	0.004
2	211.05	211.048	-0.002

https://doi.org/10.59992/IJESA.2024.v3n12p15 E-ISSN 2976-7237

3	267.19	267.189	-0.001
4	239.14	239.138	-0.002
5	874.67	874.673	0.003
6	232.17	232.168	-0.002
7	62.63	62.628	-0.002
8	183.36	183.365	0.005
9	67.51	67.505	-0.005
10	703.84	703.843	0.003
11	554.5	554.5	0
12	1378.29	1378.287	-0.003
13	584.68	584.68	0
14	376.64	376.639	-0.001
15	372.39	372.391	0.001
16	392.28	392.279	-0.001
17	121.68	121.68	0
18	172.62	172.621	0.001
19	101.81	101.808	-0.002
20	102.08	102.08	0
21	98.82	98.822	0.002
22	30.06	30.062	0.002

Conclusion

This study addressed the issue of water consumption in Sabha city. The study focused on the key factors that led to increase the demand for water by studying sample of the Staff in water departments regarding their perceptions of the environmental situation in the study area and the quantity of water available. This study also evaluated the factors which affected the increased demand for water. Also, the (WEKA) method was used to determine the spatial distribution of groundwater quality. The study had successfully demonstrated that the application of WEKA technique is a powerful tool in evaluating and describing the study phenomenon. The main objective of this study is to determine the factors that affect water consumption in Sabha city. To reach this goal varieties of data were collected from questionnaires and time series data for the period 1973 - 2023 for the quantities of water consumed and the factors affecting it.

The result of the linear regression model (WEKA) showed a relationship and positive effect between water demand (WD) and temperature (T), summer temperature (ST), population (P) and urban population (UP). At the same time no relationship was found between WD and income (IN), water price (WP). The relationship between rainfall (RF) and water demand (WD) is weak and is less impactful than other factors.

References

1. Abdudayem, A. & Scott, A. H. 2014. Water infrastructure in Libya and the water situation in agriculture in the Jefara region of Libya. African J. Economic and Sustainable Development 3(1): 33–64. https://doi.org/10.1504/AJESD.2014.061634. (ISSN: P 2520-7431).
2. Abughlelesha, S. M. & Lateh, H. B. 2013. A Review and Analysis of the Impact of Population Growth on Water Resources in Libya. World Applied Sciences Journal 23(7): 965–971. doi:10.5829/idosi.wasj.2013.23.07.13102
3. Abuzir, Dr. Yousef, 2018. Predict the Main Factors that Affect the Vegetable Production in Palestine Using WEKA Data Mining Tool. Vol.1 – N0 1 – January 2018.Palestinian Journal of Technology and Applied Sciences (PJTAS)
4. Aher, S. B. (n.d.). data mining in educational institutes.pdf. International Journal of Computer Applications®. IJCA. Retrieved from https://pdfs.semanticscholar.org/0ad0/20d11b15dd3b07182151085f51e7c542b558.pdf.
5. Aksu, G. Ö. K. H. A. N., and N. U. R. İ. Doğan. "An analysis program used in data mining: WEKA." Journal of Measurement and Evaluation in Education and Psychology 10.1 (2019): 80-95.
6. Alghariani, S. a, Authority, G. M. M. & Jamahiriya, L. a. 1955. Water transfer versus desalination in North Africa: sustainability and cost comparison. Population (millions) 1990: 2025. https://doi.org/10.1017/S0263718900003460.
7. ATEN, M. L. 2013. TROPHIC STATE MONITORING OF LAKES AND RESERVOIRS USING REMOTE SENSING.
8. ATEN, M. L. 2013. TROPHIC STATE MONITORING OF LAKES AND RESERVOIRS USING REMOTE SENSING.
9. bin Othman, M.F., Yau, T.M.S. (2007). Comparison of Different Classification Techniques Using WEKA for Breast Cancer. In: Ibrahim, F., Osman, N.A.A., Usman, J., Kadri, N.A. (eds) 3rd

Kuala Lumpur International Conference on Biomedical Engineering 2006. IFMBE Proceedings, vol 15. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68017-8_131.
10. Eibe Frank, Mark Hall, Len Trigg, Geoffrey Holmes, Ian H. Witten, Data mining in bioinformatics using Weka, Bioinformatics, Volume 20, Issue 15, October 2004, Pages 2479–2481, https://doi.org/10.1093/bioinformatics/bth261.
11. Elgali, J. A. J. 2013. The impacts of groundwater pollution in Derna Region, Libya. Universiti Kebangsaan Malaysia.
12. GWA. 2006. Water Situation in Libya. Tripoli-Libya. Retrieved from https://scholar.googleusercontent.com/scholar.bib?q=info:AFnabFzMgcUJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWewlggR1mYUdcRwFbbXB9OOMB0Nj1IG4&scisf=4&ct=citation&cd=0&hl=en
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. H. 2009. The WEKA data mining software. ACM SIGKDD Explorations Newsletter 11(1): 10.https:// doi.org/10.1145/1656274.1656278.
14. K. Sai Madhuri, B. Prathusha, G. Manish Reddy, Gayatri Deshmukh, K. Bhuvan Sai Kumar, Aklilu Markos (2024). Performance Analysis of Machine Learning Algorithms by Using WEKA and Scikit-Learn. Conference paperpp 633–644.https://doi.org/10.1007/978-981-99-9179-2_47.
15. Kosorus, H., Honigl, J. & Kung, J. 2011. Using R, WEKA and RapidMiner in Time Series Analysis of Sensor Data for Structural Health Monitoring. 2011 22nd International Workshop on Database and Expert Systems Applications, hlm. 306–310. IEEE. doi:10.1109/DEXA.2011.88
16. Kosorus, H., Honigl, J. & Kung, J. 2011. Using R, WEKA and RapidMiner in Time Series Analysis of Sensor Data for Structural Health Monitoring. 2011 22nd International Workshop on Database and Expert Systems Applications, hlm. 306–310. IEEE. doi:10.1109/DEXA.2011.88
17. Martine, G., Deligiorgis, D., Fuersich, C., Leon, L. & Odelius, A. 2007. State of World Population 2007: Unleashing the Potential of Urban Growth. bases.bireme.br. Retrieved from http://bases.bireme.br/cgibin/wxislind.exe/iah/online/?IsisScript=iah/iah.xis&src= google&base=REPIDISCA&lang=p&nextAction=lnk&e xprSearch=171894&indexSearch=ID
18. Patel, Nikita, and Saurabh Upadhyay. "Study of various decision tree pruning methods with their empirical comparison in WEKA." International journal of computer applications 60.12 (2012).
19. S. A. Kiranmai and A. J. Laxmi, "Data Mining for Classification of Power Quality Problems Using WEKA and the Effect of Attributes on Classification Accuracy," in Protection and Control of Modern Power Systems, vol. 3, no. 3, pp. 1-12, July 2018,https://doi.org/10.1186/s41601-018-0103-3.

20. Singhal, Swasti, and Monika Jena. "A study on WEKA tool for data preprocessing, classification and clustering." International Journal of Innovative technology and exploring engineering (IJItee) 2.6 (2013): 250-253.
21. Wheida, E. & Verhoeven, K. 2006. Review and assessment of water resources in Libya. Water International, hlm. Vol. 31, 295–309. Taylor & Francis Group. doi:10.1080/02508060608691933
22. Wheida, E. & Verhoeven, R. 2007. An alternative solution of the water shortage problem in Libya. Water Resources Management 21(6): 961–982. doi:10.1007/s11269-006-9067-6.