

Received September 13, 2021, accepted October 18, 2021, date of publication November 2, 2021, date of current version November 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3124817

Dual Water Choices: The Assessment of the Influential Factors on Water Sources Choices Using Unsupervised Machine Learning Market Basket Analysis

TIYASHA TIYASHA^{1,2}, SURAJ KUMAR BHAGAT^{1,2}, FIRAOL FITUMA^{1,2},
TRAN MINH TUNG¹, SHAMSUDDIN SHAHID³, AND ZAHER MUNDHER YASEEN^{4,5}

¹Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City 700001, Vietnam

²Department of Civil Engineering, Institute of Technology, Ambo University, Ambo 00251, Ethiopia

³Department of Water and Environmental Engineering, Faculty of Engineering, School of Civil Engineering, Universiti Teknologi Malaysia (UTM), Skudai, Johor 81310, Malaysia

⁴Department of Urban Planning, Engineering Networks and Systems, Institute of Architecture and Construction, South Ural State University, 454080 Chelyabinsk, Russia

⁵New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar 64001, Iraq

Corresponding author: Firaol Fituma (goftafituma@gmail.com)

ABSTRACT An unsupervised machine learning model of association rule known as market basket analysis is proposed in this study to analyze the influence of various socio-economic factors on the choice of the water source. Data of 51 socio-economic factors collected from 295 individuals living in 65 households in Ambo city in the Oromia region of Ethiopians were used for this purpose. The results revealed (i) 64% of the family preferred multiple water sources (i.e., public tap and river water), (ii) the water was collected females in 92% of the households, and (iii) majority of people preferred bathing and laundering in the river (support = 32% and confidence = 87%). Direct utilization of river water is not a preferable choice for the user since it may lead to severe health issues and cause water pollution from bathing and laundering. Education and monthly income have a significant impact on the choices of water sources. Local management authorities can improve sanitation and public health management using the results obtained in the study. The paper only gives a glimpse of the important factors that should be considered for improving the way of life for the underdeveloped areas of the world using advanced machine learning techniques.

INDEX TERMS Developing countries, socio-economic factors, unsupervised machine learning, water policy.

I. INTRODUCTION

Unsustainable human activities have caused water pollution and degradation of safe water sources across the globe [1], [2]. Decreasing freshwater resources and increasing water demand to meet the growing population and economic demand have caused water scarcity even in many water-rich regions. About 2.4 billion of the global population is under water stress, which would increase to 9.6 billion in 2050 [3], [4]. The scarcity will be more in developing countries due to inefficiency in water resources management.

Despite water scarcity, the availability of low-cost water purification technologies and changes in government policies to fulfil the Sustainable Development Goal of ensuring clean

water and sanitation for all have made the availability of safe water to more population in recent years [5]. A study by [6], estimated access to improved water sources to 16% more population in 2015 compared to 1990. However, improved provision of safe water alone cannot ensure access to clean water. There are many social and economic factors, such as education and behavior, that influence the choice of water sources and the use of safe water. For example, people with better economic ability have more access to freshwater. Education makes people more aware of safe water and enhances their willingness to get access to safe water. Reference [7] showed that the choice of water source is significantly affected by place of residence, geopolitical zone, education, wealth index, ethnicity, access to electricity and gender. Therefore, it is very important to consider these factors in designing and implementation of water services programs. Quantification

The associate editor coordinating the review of this manuscript and approving it for publication was Xujie Li¹.

of the relative influence of different social and economic factors on safe water use is also important for sustainable water resources management, ensuring social equity is water access, protection of public health, and improvement of the quality of life.

Association of safe water use with different socio-economic factors like health, education, gender, distance to the water source, size of family and so on has been established by many researchers [8], [9]. The previous studies have used multinomial logistic regression [10], generalized linear models [11], conditional logit model [12] and univariate analysis [13]. Table 1 presents a brief survey of the previous research studies done to identify the factors affecting household water access in developing countries using various techniques. All the reported literature studies were conducted using classical statistical models.

The statistical models provide a numerical measurement of association by testing the hypothesis of an association between two variables [14]. Such statistical approaches are applicable when the two variables of interest are known, or a hypothesis is already defined. However, such a condition is not valid when prior knowledge of the relationship among the variables is not known. Besides, such methods are limited to numerical and ordinal data. Therefore, a reliable method is needed to unwrap the correlation in a dataset with a wide range of data types [15], [16].

This study aims to use an unsupervised machine learning model which uses association rule, also known as market basket analysis. It is an efficient analytical methodology to analyze and optimize the choice and behavior of the customer [17]. The main advantage of unsupervised machine learning is the potential to solve complex problems using the capacity of intelligent mimicking [18], [20]. It can observe frequently occurring pattern, correlation and association from the dataset with the help of the three thresholds support, confidence and lift [21]–[23].

The goal of the current research is to use machine learning method for establishing the relationship between the daily water requirement for households, available water supply facilities, choices of water supply and how these choices are affecting the consumer regarding several elements, such as education, job, daily time invested, household income, responsible member of the house to do the chores, awareness, the willingness of maintaining the sanitation and hygiene. The paper desires to highlight the condition of developing countries that lack awareness and affordability of proper water supply. The methodology used for the assessment of the influence of socio-economic factors on water use can be replicated for a reliable analysis of socio-economic interactions with natural resources.

II. FACTOR INFLUENCING WATER SOURCE SELECTION

A. SOCIO-ECONOMIC CHARACTERISTIC

Water sources and their utilization are holistically affected by socio-economic factors. The choice of water source depends on the distance from the source, way of access, data

availability, ethnicity group of the area, status of the family, education-water access relationship and so on [24], [25]. Besides inequalities in water access, water policies for predominantly poor and economically disadvantaged rural settlements are greatly affected by socio-economic variables in developing countries. Hence, it can be considered the key indicator of water sources utilization [26], [27].

B. PRICING

Water consumption and water price are related to each other. The increase in water price significantly affects water consumption, specifically in low-income households where water bill surges affect the monthly budget [28]. The developing and underdeveloped countries commonly consist of high population density, and raised tariffs might undesirably affect the financial health of households [29]. In specific cases, the pricing can change as per the seasonal variation to minimize the consumption and enhance the water accessibility as per WHO guidelines [30].

C. COLLECTION TIME

Research studies have found that the choice of water source is likely influenced by household characteristics and distance to the water source [38], [39]. As the women spent most of the time in water collection and change in distance, the water sources affect more to the female and children [40]. The collection time is also affected by the water activity since water for drinking and cooking need better quality, even though the travel time increases irrespective of whether the user chooses the source. In addition to that, the water infrastructure and water policies significantly change the household behavior towards the efforts given to water collection, which may include proximity, pricing, quality, accessibility to the source and geographical structure of the area [12].

D. MULTIPLE SOURCES

Choice of multiple water sources is affected by the distance from the water source, quality of the water depending on activities, such as bathing, washing, cooking and conflict among the people. Reference [41] the study reports that the factors influencing the multiple uses of water sources are water services, water supply scheme, technology-system design, water quality–quantity, collection distance and time. Another study reported by [42] showed that users preferred river and public tap depending on the quality, improved access to the water source and efforts on water collection (time taken, the volume of water transported each time, frequency of filling).

E. IMPACT OF COMMUNITY WELLBEING

Water availability and community health are essential parameters to measure community wellbeing. Reference [43] reported in the study that small-scale community water supply significantly affects hygiene behaviors and daily life.

TABLE 1. Collection of studies done in developing and low-income countries in the area of household water access.

Authors	Country	Methods	Factors	Remarks
[31]	Sri Lanka	House-hold survey data (n=40) / 1994-1999/ log-log and t-test	Water consumption, water price, household size	The study considered the influence of domestic water consumption and correlated it with income and price.
[32]	South Africa	House-hold survey data (n=275) / 1998-2003/ regression, generalised least square and panel corrected standard error model	Climate, household characteristics, water consumption	The paper deals with understanding the factor that influences water consumption with a correlation with income level. The results show that the low-income group hardly changes the consumption capacity. However, price policy should consider multiple objectives.
[12]	Madagascar	House-hold survey data (n= 2179)/ 2004-2005/ conditional logit model	Water sources, household characteristics, distance to water sources, gender, age	The authors addressed the issue of water supply choice and time spent for collection. It concludes that reducing the distance to the water source will have a huge impact on adults and children.
[10]	Nigeria	House-hold survey data (n=5000)/ 2009 / multinomial logistic regression	Water access, income, age, marital status, employment type, working hours, financial information	The objective was the study was to identify the factor influencing water access and conclude that infrastructure development can help increase the water access issues
[33]	Cameroon	House-hold survey data (n= 11391)/ 2007/ multinomial logistic regressions	Water sources, household characteristics, price, gender, demographic and socio-economic factors, distance to water sources	
[11]	Ghana	House-hold survey data (n=11778) /2008 / generalized linear models	Water access, sanitation sources, income, education, age, household size, region, financial information	The authors suggest that nearby access to improved water sources and sanitation is important for sustainable water utilization. The study aims to aid the management in better water and sanitation policy preparation and execution.
[34]	Indonesia	Household survey data (n= 13535)/ 2007/ Wealth index, bivariate and multivariate regression analysis	Place of residence, sanitation ladder, type of dwelling, household members, sex, health, financial information, Household head information	The author aim was to assess the geographical and socio-economic inconsistencies in household's drinking water sources. The study concludes that irregularities lead to other problems like diseases, sanitation, internal migration, etc.
[35]	Vietnam	House-hold survey data (n= 9117, 9473, 11614)/ 2000, 2006,2011/ Binary and multivariate logistic regressions	Access to water, location, demographic and socio-economic factors, sanitation, geography.	The study concluded that poor access to improved sanitation facilities are the cause of various infectious diseases, increased mortality rate, poor hygiene
[36]	Mali	House-hold survey data (n=100) / 2016/ multiple regression model	House members, travel time, total travel time, water consumption, water quality	The study applied multiple regression models coupled with spatial distribution for building improved sources. Water quality and time collection were considered.
[13]	Zambia	Household survey data (n= 15920)/ 2013-2014 / Univariate analysis, Erreygers concentration, logistic regression, Wealth index	Wealth index, gender of household head, region, residence information, access to improved water, ethnicity	The study deals with inequalities and determination of access to water sanitation considering rural and urban households. The study conducted a univariate analysis to assess the relationship between various factors and help to mitigate the issues
[37]	Nigeria	House-hold survey data (n= 40,680) /2013/ logistic regression	Place of residence, geopolitical zone, educational level, household wealth, gender, ethnicity	The research examines the factors that influence Household access due to inequalities in regions and populations.

The factors considered were water source from the river to groundwater, collection time, household water consumption and income.

F. HEALTH

Water impact on health is a well-known fact but as per the studies yet neglected. Poor water supply or unprotected water source can cause acute infectious diarrhea, Trachoma,

Ascariasis, Hook work infection, Dracunculiasis, Schistosomiasis and other diseases from heavy metal exposure. Without access to piped water, the household is 4.8% more prone to infant mortality from diarrhea and other water-related diseases [44]–[46]. It was also reported that an awareness program alongside water supply improvement is also necessary to improve health, sanitation and decrease infant mortality [47]–[49].

G. EDUCATION

There is no well-established relation between education and water yet; however, most research supported the relation between water and education [50]. A good impact on the educational campaign in line with water savings and conservation has been noticed in water-scarce countries [51]. Consequently, a study in China has reported that children (especially girls) get to attain more school days in rural areas if they get treated water in their households [52]. Educated parents showed more interest in the easy access of water and hygienic maintenance for their children's health [16], [53].

H. INCOME

Income supports more increasing water demand than the hygienic way of using it; however, the child involvement for water fetching remains the same [16], [54]. The research reported less or no relation between the duration of water access and income [55]. Another study depicts that income is affected by the education level of the head of the household, assets and the preference to the home lifestyle, including water-related decisions. The income structure significantly affects the willingness to pay the water bill, inter-sectoral water transfer, and user-specific water demand and consumption response [56].

III. CASE STUDY AND DATA EXPLANATION

A. STUDY AREA

Ethiopia experienced a severe water shortage due to inadequate rainfall in the 2017 rainy season, which led to catastrophic agricultural and socio-economic losses [57]. Recently, 10.5 million Ethiopians need humanitarian assistance due to clean water shortage, as per the report published in August 2017 by the Government of Ethiopia [58].

Ambo is a town in the Oromia region and is 120 km from the capital, Addis Ababa. Ambo is one of the important geographical locations where water insecurity is at its peak, affecting the social and healthy life among 180 Woreda of West Shoa Zone, Oromia Regional State of Ethiopia [59]. Ambo woreda has six kebeles (the smallest administrative unit of Ethiopia similar to a ward) such as Ambo01, Ambo02, Ambo03, Senkale Foris, Kisose odo and Awaro. Kerchelle masa of Ambo 01 has insufficient water supply due to their human behavior, literacy rate, and insufficient infrastructure of the water supply system. The selected study area is well known for Afan-Oromo people in terms of their mother tongue. Ethiopia is running a severe water shortage due to inadequate rainfall during the 2017 rainy season, leading to insufficient crop production and catastrophic socio-economic losses [57]. The study area in Ambo is displayed in Figure 1.

The study area comes under the economically water-stressed zone and only rely on the surface water sources, especially river water. Besides, the groundwater has high fluoride, which makes it difficult for direct use [60]. The study considered the groundwater source, but none of the

users reported any use. Therefore, it was not discussed in the result.

B. DATA COLLECTION

A structured questionnaire was developed considering the local language, was used by the interviewees. The questionnaire was translated into Oromo then back-translated to English. The household was selected randomly within a 1 km distance from the river water source. The sample size consists of 295 individuals living in 65 households. The selected area comes under the rural area, Kerchelle masa village situated near Taltelle river of Ambo region. Ethiopia is one of such under developed country which is continuously need humanitarian assistance due to low economic condition [57]. Moreover, severe shortage of water due to inadequate rainfall in rainy season has led to catastrophic agricultural and socio-economic losses which added more stress to the present condition [58]. The study is a village which is far from the city due to which awareness and education is very less. This is one of the major reason that gathering data was one of the most difficult issue in this study. The selected area was one of the nearest village but issue focused in the study is the same for many such villages. Thus, this study aims to highlight the problems related to the water supply system of such villages all over Ethiopia and tried get attention required to improve the situation. This study found that with this number of data can be used for such case study and highlight the water supply problems and it's the influencing factors. The study also found that other research studies efficiently achieved their goal with minimum number of sample size such as $n = 100$ [36] and $n = 40$ [31].

The dependent variable used is the water sources, and independent variables for the study area were chosen based on the public water-related awareness and significant factors influencing wellbeing. Among which to was a major focus that factors such as education, job and awareness factors a lot when choosing the water source. The aim of considering many parameters was to evaluate and point out the issues which can elevate the life condition of the local people. Even though water was our primary concern but many key factors influences it and if at least one of the key point can be improved then major change can be observed. The independent variables were household member (number of male adults, number of female adults, number of male child and number of female child), education level (Uneducated, below 10th standard, below 12th standard and graduate), number of member employed, monthly income in Ethiopian Birr (ETB), daily water requirement of the household in litre, water collector information (sex and age), water collected each trip in litre, number of times water collected, total time taken for water collection in minutes, water collection method (manual and transportation), water bill paid per month in ETB, water sources selected for different household activity (bathing, washing and cooking), water collection time (Morning and evening), seasonal water quality variation (summer and rainy), seasonal water collection efficiency (difficult and

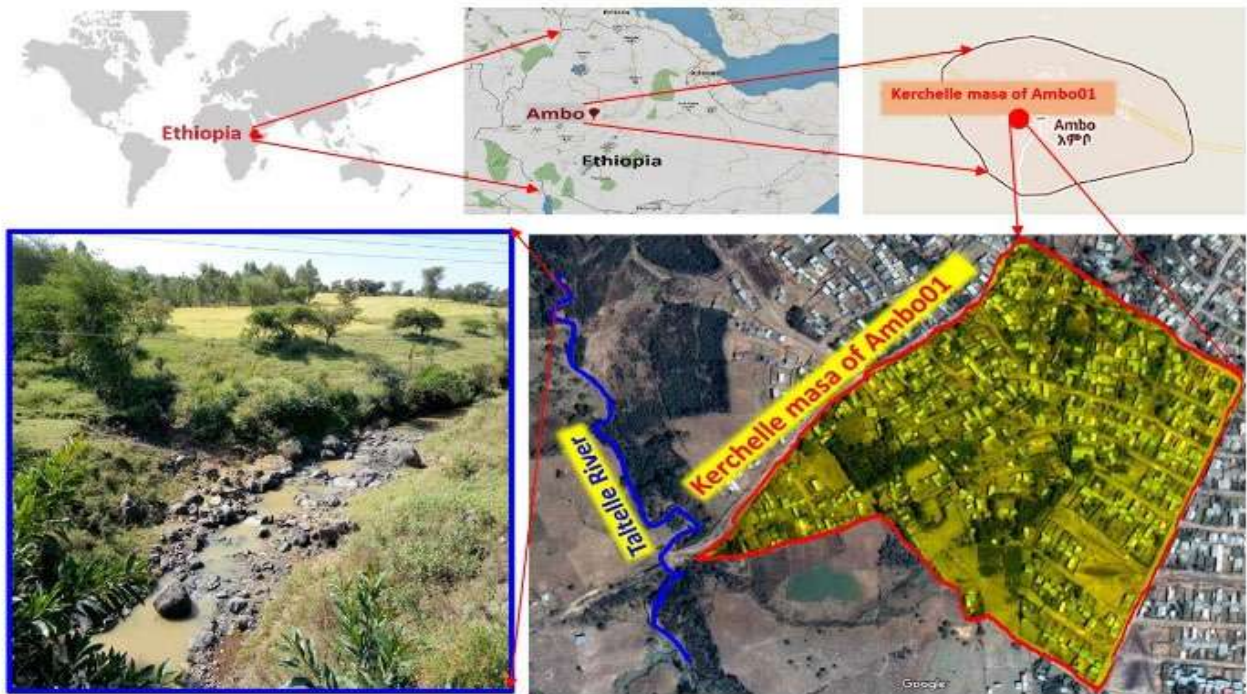


FIGURE 1. The selected case study area.

easy), water interruption in days, household water treatment, cleaning frequency of water storage container (daily, never, weekly), hygiene information (use of soap before handling water, during bath, after defecation), defecation in river, toilet at home, health information (diarrhea, common cold and other diseases).

IV. APPLIED MACHINE LEARNING AND STATISTICAL APPROACHES

A. ASSOCIATION RULE/ZZZZ/WWWWW/MARKET BASKET ANALYSIS

Association rule is the rule-based machine learning methodology where highly confident associations among multiple variables are calculated. This rule is a centered technique which has exhibited higher accuracy [61]. This tool has been used in different fields of science and engineering but have been applied in a limited number of researches with respect to analysis of water uses [62]–[64]. The best association is selected based on the various statistical analysis outcome such as the high number of counts, higher value of confidence, lift and support. *Apriori* algorithm is suited to the character variable and leads to better performance [65].

In a given transaction, where each transaction is a set of items, the association rule reveals that the transaction in the dataset containing X also contains Y expressed as equation (1). Support metric parameter is applied to measure the occurrence of the transactions of the item set and classifies the best rule for auxiliary analysis. Thus, support is the fraction of the total number of transactions for sets of items

and can be expressed as equation (2). Another parameter to measure the occurrence of consequent and antecedents is known as confidence. The presence of the probability of occurrence of the sets of items and expressed as equation (3). The confidence value is sometimes high even though the association between the items is most likely unrelated. To overcome such a situation, Lift is introduced as the third metric parameter, which controls the frequency of consequent while measuring the conditional occurrence probability of {Y} given {X} and can be expressed as equation (4) [66]. Support and confidence reflect the usefulness and certainty of the identified rules. R version 3.6.1. Packages “arules” and “arulesViz” were used for association estimation and visualization, respectively.

$$\text{Association rule } X \rightarrow Y \tag{1}$$

where, X and Y are sets of items

$$\begin{aligned} \text{Support} (\{X\} \rightarrow \{Y\}) &= \frac{\text{Transaction containing both X and Y}}{\text{Total number of transactions}} \end{aligned} \tag{2}$$

$$\text{Confidence} (\{X\} \rightarrow \{Y\}) = \frac{\text{Transaction containing both X and Y}}{\text{Transactions containing X}} \tag{3}$$

$$\begin{aligned} \text{Lift} (\{X\} \rightarrow \{Y\}) &= \frac{(\text{Transaction containing both X and Y})}{(\text{Transactions containing X})} \\ &= \frac{\text{Fraction of transactions containing Y}}{\text{Fraction of transactions containing X}} \end{aligned} \tag{4}$$

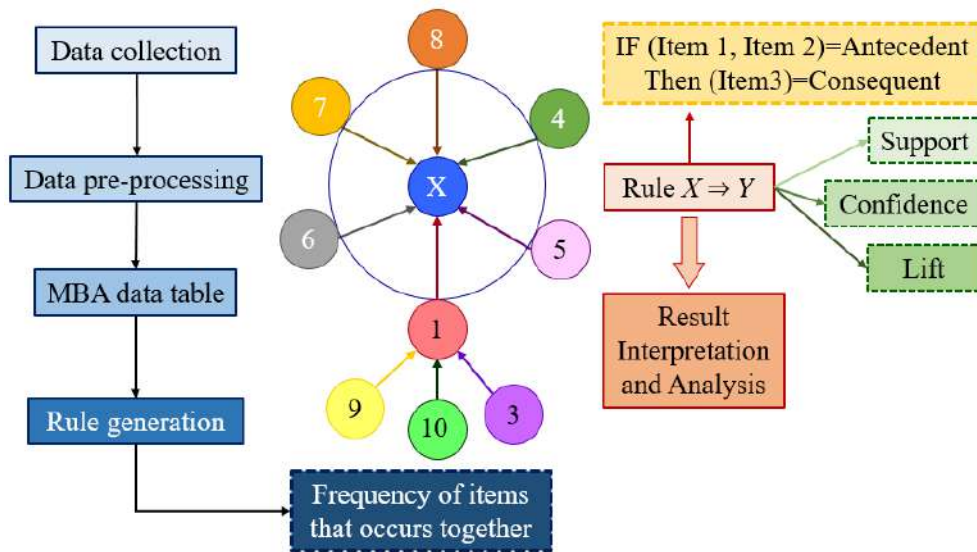


FIGURE 2. The general market basket analysis/association rule process chart.

1) BASIC CONCEPT

The process occurs in two steps (i) finding frequent items and (ii) generation of strong association rule based on support and confidence. Let consider $J = (i_1, i_2 \dots i_n)$ to be a set of items (*itemset*). Let D , the task-relevant data, be a set of database transactions where each transaction T is a set of items such as $T \subseteq J$. Each transaction is associated with an identifier, known as Transaction ID (TID). Let consider a set of items, where a transaction $\cup \cap _(\subseteq) \rightarrow |T$ is a set to contain X if and only if $X \subseteq T$. An association rule is an inference of the form $X \Rightarrow Y$, where $X \subset J, Y \supset J$, and $X \cap Y = \phi$. The rule $X \Rightarrow Y$ holds in the transaction set D with the support. That contains $X \cup Y$ (that is X and Y) is taken as the probability, $P(X \cup Y)$ the rule $X \Rightarrow Y$ has confidence C in the transaction set D , if C is the percentage of the transactions in set D which contains X that also contains Y . This is taken as the conditional probability as $P(Y|X)$ that is, Support = $(X \Rightarrow Y) = P(X \cup Y)$ and Confidence == $(X \Rightarrow Y) = P(Y|X)$ [67]. Figure 2 presents the general flowchart of the Association rule process.

Pearson’s correlation analysis is widely used in multiple sectors of science due to its accuracy in terms of the relationship between the variables [68], [69]. It measures a linear dependence between two variables (x and y)

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}} \quad (5)$$

where, x and y are the variables; m_x and m_y are the mean of x and y , respectively.

B. MODELING DEVELOPMENT

To transform the dataset into the transaction database so that it can be subjected to the selected analytical method. The data frame was divided according to information from the questioner prepared for the study. The data frame consists

of 51 variables and 65 observations. Out of 51 variables, 28 variables were assessed using association rule, 6 variables using Pearson’s correlation analysis and others were analyzed in general in which interviewer’s no. and interviewer’s general information was irrelevant for the analysis where the 3rd variable was the distance from the river which found to be average of 385m (min-max: 300-950m). The remaining variables were interpreted using general data analysis like the mean and min-max analysis.

The selected 28 variables were converted in logical type or binary type (i.e., *TRUE* or *FALSE*) data set as part of data pre-processing. The main transaction database (i.e., 28 selected variables and 65 observations) was divided further to better understand their relationship with each other and handle the numerous amounts of rules generated by the program. The application was separately established considering two different water sources, including Public tap and (ii) River. The water sources were coupled with four scenarios considering the dual water sources, i.e., public water source and river water sources, along with other variables presented in Figures 3 and 4. Total 65 observations are plotted for two water source choices, public tap water in Figure 3 and river water in Figure 4.

Scenario A: Eight variables: (i) Water Source: Public tap/River, (ii) Uneducated, (iii) BELOW 10, (iv) BELOW 12, (v) Graduate, (vi) WCI: Female, (vii) Activity: bathing, (viii) Activity: washing.

Scenario B: Seven variables: (i) Water Source: Public tap/River, (ii) Water collection time: Morning, (iii) Water collection time: Evening, (iv) water quality in rainy season: Bad, (v) water quality in rainy season: Difficult, (vi) water quality in rainy season: Good, (vii) water quality in the dry season: Easy, Water interruption.

Scenario C: Eight variables: Water Source: (i) Public tap/River, (ii) Any water treatment, (iii) use of storage container, (iv) Cleaning frequency Weekly, (v) Cleaning

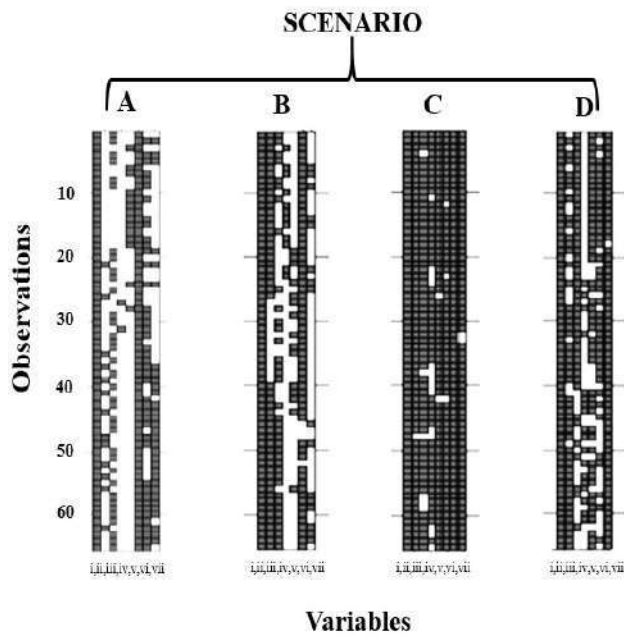


FIGURE 3. Scenario A: Eight variables: (i) Water Source: Public tap, (ii) Uneducated, (iii) BELOW 10, (iv) BELOW 12, (v) Graduate, (vi) WCI: Female, (vii) Activity: bathing, (viii) Activity: washing. Scenario B: Seven variables: (i) Water Source: Public tap, (ii) Water collection time: Morning, (iii) Water collection time: Evening, (iv) water quality in rainy season: Bad, (v) water quality in rainy season: Difficult, (vi) water quality in rainy season: Good, (vii) water quality in the dry season: Easy, Water interruption. Scenario C: Eight variables: (i) Public tap, (ii) Any water treatment, (iii) use of storage container, (iv) Cleaning frequency Weekly, (v) Cleaning frequency Daily, (vi) Cleaning frequency: Never, (vii) clean hand before handling water, (viii) using soap. Scenario D: Seven variables: (i) Water Source: Public tap, (ii) bathing in river, (iii) Use soap during bathing, (iv) defecation at river, (v) after defecation use of soap, (vi) toilet at home, (vii) common cold. The variables are arranged as mentioned and plotted, respectively.

frequency Daily, (vi) Cleaning frequency: Never, (vii) clean hand before handling water, (viii) use of soap.

Scenario D: Seven variables: (i) Water Source: Public tap/River, (ii) bathing in river, (iii) Use soap during bathing, (iv) defecation at river, (v) after defecation use of soap, (vi) toilet at home, (vii) common cold.

The implementation part consists of the first step of finding an association rule is searching for frequent item sets. *Apriori* algorithm available in the R studio “arules” package was used to find frequent item sets [70]. The transaction database (a) was specified with minimum support of 20%, minimum confidence of 70% (b) with minimum support of 90%, minimum confidence of 95% (c) specified with minimum support of 40%, minimum confidence of 70% (d) specified with minimum support of 50%, minimum confidence of 70%. Different minimum support and confidence were used for the association rule formation to decrease the number of rules and computational time. The application was able to find the association rules for (a, c and d) around 20-50 seconds since the rules were less. In dataset (b), there were more than 1000 rules that took several minutes, which was later reduced using high minimum support and confidence.

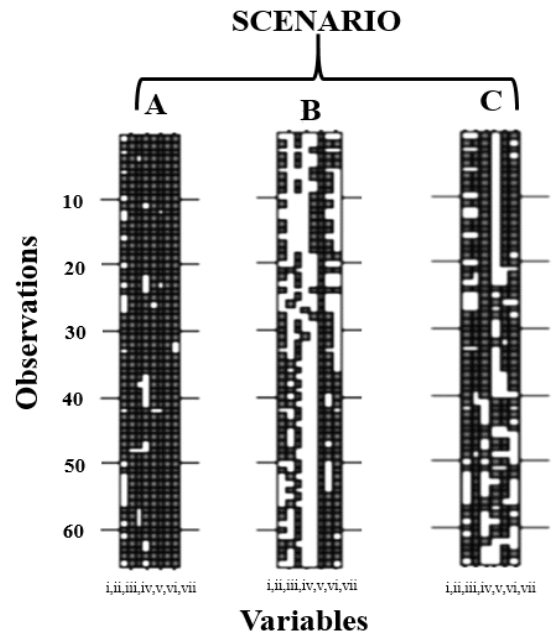


FIGURE 4. Scenario A: (a) 8 variables: Water Source: river, Uneducated, BELOW 10, BELOW 12, Graduate, WCI: Female, Activity: bathing, Activity: washing; (b) Scenario B: 7 variables: Water Source: river, Water collection time: Morning, Water collection time: Evening, water quality in rainy season: Bad, water quality in rainy season: Difficult, water quality in rainy season: Good, water quality in dry season: Easy, Interruption; and (c) Scenario D: 7 variables: Water Source: river, bathing in river, Use soap during bathing, defecation at river, after defecation use of soap, Do you have toilet, the common cold. The variables are arranged as mentioned and plotted, respectively. The scenario C result are the same for both water sources.

Further, for better management and interpretation “Head” function of R studio was used with specified 50 top rules sorted as per the descending order of the “Lift”. The results produced was presented in table 2 where the LHS (left-hand side) is called antecedent, and the RHS (right-hand side) are called consequent along with specified minimum support and minimum confidence, lift and count. The principle between the item set is “IF” the antecedent, item {X} is there “THEN” the consequent, item {Y} will be there, and this is supported by the three thresholds minimum support and minimum confidence and lift. The results produced from 8 simulations were quite high; thus, the most relevant relations were sorted out. Table 2 only presents 44 of the most pertinent association of the variables.

V. RESULT AND DISCUSSION

The results show that washing, bathing, and choice of public water sources are highly correlated (support = 36%, confidence = 75%). Villagers preferred river water for bathing and washing (support = 32%, confidence = 87%). The survey shows that single users opted 100% for on public tap, and 64% used both water sources.

Female water collectors with an education level less than the 10th standard used the river for bathing (support = 23%, confidence = 75%), which depicts that public tap is used

TABLE 2. The list of selected association rules the between the various antecedent and the consequent with the amount of support, confidence, lift and number of count.

Antecedent for IF (LHS)	Consequent for THEN (RHS)	Support	Confidence Interval	Lift	Count
{Activity: washing}	{Activity: bathing}	0.3692308	0.75	1.133721	24
BELOW 10, Activity: washing}	{Activity: bathing}	0.2307692	0.75	1.133721	15
Water Source: Public tap, Activity: washing}	{Activity: bathing}	0.3692308	0.75	1.133721	24
{BELOW 10, WCI: Female, Activity: washing}	{Activity: bathing}	0.2307692	0.75	1.133721	15
{Graduate}	{WCI: Female}	0.2153846	1.00	1.000000	14
{Graduate}	{Water Source: Public tap}	0.2153846	1.00	1.000000	14
{Uneducated}	{Water Source: Public tap}	0.2000000	1.00	1.000000	13
{Activity: washing}	{Water Source: Public tap}	0.4923077	1.00	1.000000	32
{BELOW 10}	{WCI: Female}	0.5538462	1.00	1.000000	36
{Uneducated, WCI: Female}	{Water Source: Public tap}	0.2000000	1.00	1.000000	13
{Water collection time: Evening}	{Water collection time: Morning}	0.9076923	1.00	1.0156250	59
{water quality in rainy season: Good}	{Interruption}	0.9230769	0.9677419	0.9984639	60
{water quality in rainy season: Good}	{water quality in rainy season: Difficult}	0.9384615	0.9838710	1.0151050	61
{water quality in rainy season: Good}	{Water Source: Public tap}	0.9538462	1.0000000	1.0000000	62
{water collection in rainy season: Difficult}	{water quality in rainy season: Good}	0.9384615	0.9682540	1.0151050	61
{Interruption}	{water quality in rainy season: Difficult}	0.9384615	0.9682540	0.9989922	61
{Water Source: Public tap}	{Interruption}	0.9692308	0.9692308	1.0000000	63
{Cleaning frequency Weekly}	{Water Source: Public tap}	0.5846154	1.0000000	1.0000000	38
{clean hand before handling water}	{Water Source: Public tap}	0.9384615	1.0000000	1.0000000	61
{Cleaning frequency Weekly}	{Water Source: Public tap}	0.5846154	1.0000000	1.0000000	38
{bathing in river, common cold}	{after defecation use of Soap}	0.5692308	0.7872340	1.023404	37
{Water Source: Public tap, after defecation use of Soap}	{bathing in river}	0.5846154	0.7600000	1.029167	38
{Water Source: Public tap, Do you have toilet}	{after defecation use of Soap}	0.5076923	0.8048780	1.046341	33
{Do you have toilet}	{Water Source: Public tap}	0.6307692	1.0000000	1.000000	41
{bathing in river}	{Water Source: Public tap}	0.7384615	1.0000000	1.000000	48
{Activity: bathing ,Activity: washing}	{Water Source: River}	0.3230769	0.8750000	1.354167	21
{WCI: Female, Activity: bathing}	{Water Source: River}	0.5692308	0.8604651	1.331672	37
{Activity: washing}	{Water Source: River}	0.3846154	0.7812500	1.209077	25
{Uneducated}	{WCI: Female}	0.2000000	1.0000000	1.000000	13
{BELOW 10}	{WCI: Female}	0.5538462	1.0000000	1.000000	36
{Graduate}	{WCI: Female}	0.2153846	1.0000000	1.000000	14
{Water Source: River}	{water quality in rainy season: Bad}	0.4615385	0.7142857	0.9103641	30
{Water Source: River, water quality in rainy season: Difficult, Interruption}	{Water collection time: Evening}	0.5538462	0.8780488	0.9673419	36
{Water Source: River, Interruption}	{Water collection time: Evening}	0.5538462	0.8780488	0.9673419	36
{Water Source: River, water quality in rainy season: Difficult}	{Water collection time: Evening}	0.5692308	0.8809524	0.9705408	37
{using soap}	{Water Source: River }	0.2000000	0.8125000	1.2574405	13
{Water Source: River}	{clean hand before handling water}	0.6000000	0.9285714	0.9894614	39
{Cleaning frequency Weekly}	{Do u have storage container}	0.4615385	0.7894737	0.9868421	30
{Cleaning frequency Never}	{clean hand before handling water}	0.2307692	1.0000000	1.0655738	15
{Water Source: River}	{Do u have storage container}	0.4769231	0.7380952	0.9226190	31
{bathing in river, Use soap during bathing}	{Water Source: River}	0.4615385	0.9677419	1.497696	30
{Use soap during bathing, after defecation use of Soap}	{Do you have toilet}	0.4307692	0.8750000	1.387195	28
{bathing in river, Use soap during bathing, Do you have toilet}	{Water Source: River}	0.3538462	1.0000000	1.547619	23
{Water Source: River, Use soap during bathing}	{bathing in river}	0.4615385	0.9090909	1.231061	30
{Water Source: River}	{bathing in river}	0.5846154	0.9047619	1.225198	38

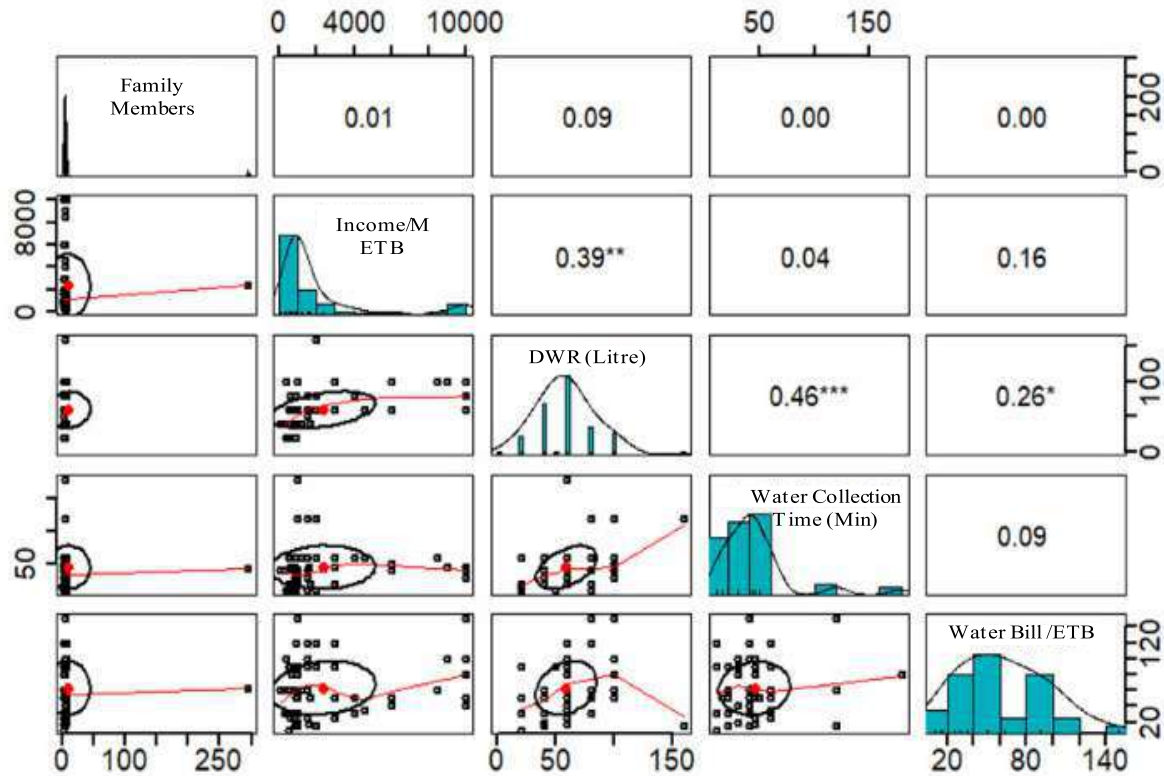


FIGURE 5. Scattered plot of matrices for the five variables below the diagonal, histogram on the diagonal and Pearson correlation above the diagonal. (DWR = Daily water requirement).

less for bathing; however public tap is preferred for washing (support = 49%, confidence = 100%). The uneducated and educated family opted for the female collector with support 20% and confidence 100%, whereas a family with basic education below 10th showed more aptitude towards the same (support = 55%, confidence = 100%, count = 36). The study finds that collection’s responsibility falls on females rather than males, where 92% of families have similar choices. Since most households collect water in the evening and morning (support = 90%, confidence = 100%), and thus, collectors have to devote more time.

The water quality during the rainy season is good, but collection becomes difficult, and interruption and evening collection is also affected (support = 55%, confidence = 87%). In this case, to manage the lack of water, a river source is chosen. The study also reveals that tap water quality remains good (support = 95%, confidence = 100%, count = 62%) but highly related to the regular interruption of water supply (support = 96%, confidence 96%). The average water supply interruption is 5.8 days with a minimum of 2 days and a maximum of 14 days long. The reason for frequent interruption can be due to (i) limited water supply (ii) rainy season affects the water treatment plant of the study area by increasing the sediment load, which leads to the temporary shutdown (iii) municipal corporation

poor maintenance leads to frequent leakage, broken pipes and blockages.

In hygiene and sanitation issues, the results show that cleaning the container weekly is the most frequent choice (support = 58%, confidence = 100%), whereas most cleaned their hands before handling water (support = 93%, confidence = 100%). However, a few of them never cleaned the container, which also used river water (support = 23%, confidence = 100%, count = 15). Furthermore, only 27% of the family cleaned hands after defecation, yet 72 % used soap while bathing. It was also noted that for bathing, 73% and for washing 49% choose river water source whereas 100% of users choose tap water for cooking. 27 % of doesn’t have a toilet at home and 27% defecates in the river. These preferences of different water source choices can be due to following reasons (i) Users prefer good quality water for consumption (ii) To reduce the water bill (iii) To reduce the time and amount of water collection from the river or public tap, and (iv) frequent cases of interruptions. The users also never do any kind of home treatment, whatever the sources may be.

The 5 integer variables were analyzed by correlation analysis. Figure 5 shows the scattered plot of the variables using Pearson’s correlation. If the number of family members is less, then the daily requirement is less with less time spent on

water collection and the bill paid per month is also decreased. The water bill increases with family income. The high positive correlation shows increased water requirement per day and increased collection of time. Water bill month is highly correlated to daily water requirement but negatively related to time spent on water collection. The correlation coefficient values of five variables are presented in Figure 5.

By the simple factor analysis, the age of the water collector has been divided into three: (i) 11-22: 45%, (ii) 22-33:40% and (iii) 33-44: 14%. The average time spent was found to be 44 minutes per day, in which 90% of users collect water two times a day, and all the users collect water manually. The survey also estimates that average income is 2327 ETB of the selected family in which the factor analysis reveals that (i) 56% user's monthly income is less than 1000 ETB (ii) 18% between 1000-2000 (iii) 9% between 2000-4000 (iv) 3% between 4000-6000 (v) 1.5% between 6000-8000 (v) 14% between 8000-10000. This shows that 74% of the users are from low-income backgrounds, which can explain that most have chosen dual water sources and paying water bills is a financial burden since the average water bill payment is 62 ETB per month. Considering the socio-economic structure of the study area, income is bound to affect the water choices. The education data discloses that 20% of the study population is uneducated, 55% of education level is below 10th standard, 3% below 12th standard, and only 21% are graduates. Moreover, the majority of the low-income family also belong to education level is below 10th standard, the as per the health issues very common diseases are diarrhea (75%), influenza (98%) and other seasonal water-related diseases like malaria, cholera (20%).

VI. CONCLUSION

The study aims to establish a relationship between the choice of water sources and the factors influencing the choice. The application of unsupervised machine learning techniques reveals that water is a key need of life, and safe choices of water sources are very important. The market basket analysis was able to associate the user's everyday life with the need for water. All the study users don't have a direct water supply, thus depending on the public and river water supply and the coinciding relationship between income, education, water awareness, water security, and health. Ethiopia has the majority of the young populace, and in the study area, 47% were adults, and 53% were children. However, the adult population of the study mostly consisted of youngsters below 25. Furthermore, females are responsible for the water collection, spending 44 minutes per day on travelling and spending 63 ETB per month. Considering the socio-economic background, the water expenditure is quite high considering 56% of family monthly earning is less than 1000 ETB. It is not surprising that many have chosen dual water sources to cope with expenditure and the monthly interruption, which may vary from 2-14 days. 73% of people preferred bathing with river water, and 49% preferred washing. 27% present of the users defecate in the river, which creates serious sanitation

and hygiene issues. This also reveals that river water gets polluted due to public use. The time spends on water collection also affects education and working hours. Education is the path for overall development, including the awareness of the right choices since none of the users was familiar with neither treated water nor knowledge of treatment techniques and their benefits. In addition to that, 23% of users don't use soap after defecation. This exposes the lack of awareness about WASH (water, sanitation and hygiene). This can be the reason for the health issues commonly reported by the respondents were diarrhea (75%), influenza (98%) and other seasonal water-related diseases like malaria, cholera (20%).

The study also concluded that there is no bore well or tube well in their area, which means that the government or local population does not exploit groundwater. The lack of water supply to each household shows a lack of water economic structures in the study area and government intervention [71]. It is safe to say that a reliable household water supply can be very helpful to increase the everyday quality of life of the villagers. However, the study could not get to that detailed information about the pollutant in the river, point sources of pollution, detailed study of the water-related diseases in the area and number of children missing school. In the future, in-depth study and analysis for a better understanding of the impact of water on human wellbeing should be done.

The study recommends the following recommendations (i) Piped water service should be extended to the households as it is the most reliable water source (ii) water-related water structure needed to be improved and expanded (iii) water reservoirs and tanks should be installed to overcome the regular interruption issue (iv) in case of contamination during water distribution, household water treatment and safe storage should be encouraged (v) water awareness should be more efficiently increase with the population to ensure water security and health. Furthermore, feature engineering and hyper-parameter tuning would be applying to minimize the noise as well as to uplift the estimation accuracy, as reported in several studies [20], [72]–[74].

ACKNOWLEDGMENT

The authors acknowledge Ambo University and the Municipality of Ambo, Ethiopia, to provide the proper platform and adequate information, respectively.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] P. Sullivan, J. J. Clark, F. J. Agardy, and P. E. Rosenfeld, *Toxic Legacy: Synthetic Toxins in the Food, Water and Air of American Cities*. Amsterdam, The Netherlands: Elsevier, 2010.
- [2] Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being*, vol. 5. Washington, DC, USA: Island Press, 2005.
- [3] K. Akramov and M. Malek, "Analyzing profitability of maize, rice, and soybean production in Ghana: Results of PAM and DEA analysis," Int. Food Policy Res. Inst., Ghana Strategy Support Program Working Paper 0028, 2012.

- [4] J. A. Finegold, P. Asaria, and D. P. Francis, "Mortality from ischaemic heart disease by country, region, and age: Statistics from World Health Organisation and United Nations," *Int. J. Cardiol.*, vol. 168, no. 2, pp. 934–945, Sep. 2013.
- [5] S. M. Awadh, H. Al-Mimar, and Z. M. Yaseen, "Groundwater availability and water demand sustainability over the upper mega aquifers of Arabian Peninsula and west region of Iraq," *Environ., Develop. Sustain.*, vol. 23, no. 1, pp. 1–21, Jan. 2021.
- [6] H. Ritchie and M. Roser. (Jan. 2019). *Urbanization*. Our World Data. Accessed: Jan. 17, 2019. [Online]. Available: <https://ourworldindata.org/urbanization>
- [7] A. M. Abubakar, H. Elrehail, M. A. Alatailat, and A. Elçi, "Knowledge management, decision-making style and organizational performance," *J. Innov. Knowl.*, vol. 4, no. 2, pp. 104–114, Apr. 2019.
- [8] I. Hasan and Y. Saeed, "Trend analysis of hydrological drought for selected rivers in Iraq," *Tikrit J. Eng. Sci.*, vol. 27, no. 1, pp. 51–57, Mar. 2020.
- [9] S. Boniface, D. Lewer, S. L. Hatch, and L. Goodwin, "Associations between interrelated dimensions of socio-economic status, higher risk drinking and mental health in south east London: A cross-sectional study," *PLoS ONE*, vol. 15, no. 2, Feb. 2020, Art. no. e0229093.
- [10] E. S. Osabuohien, U. R. Efobi, and C. M. Gitau, "Environmental challenge and water access in Africa: Empirical evidences based on Nigeria's households survey," in *Proc. Berlin Conf. Hum. Dimensions Global Environ. Change*, Oct. 2012, pp. 5–6.
- [11] E. A. Adams, G. O. Boateng, and J. A. Amoyaw, "Socioeconomic and demographic predictors of potable water and sanitation access in Ghana," *Social Indicators Res.*, vol. 126, no. 2, pp. 673–687, Mar. 2016.
- [12] C. Boone, P. Glick, and D. E. Sahn, "Household water supply choice and time allocated to water collection: Evidence from madagascar," *J. Develop. Stud.*, vol. 47, no. 12, pp. 1826–1850, Dec. 2011.
- [13] J. N. Mulenga, B. B. Bwalya, and K. Kaliba-Chishimba, "Determinants and inequalities in access to improved water sources and sanitation among the Zambian households," *Int. J. Develop. Sustain.*, vol. 6, no. 8, pp. 746–762, 2017.
- [14] M. Shareef, "Assessment of Tigris river water quality using multivariate statistical techniques," *Tikrit J. Eng. Sci.*, vol. 26, no. 4, pp. 26–31, Dec. 2019.
- [15] K. Rieger, R. H. Holm, and H. Sheridan, "Access to groundwater and link to the impact on quality of life: A look at the past, present and future public health needs in Mzimba district, Malawi," *Groundwater Sustain. Develop.*, vol. 2–3, pp. 117–129, Aug. 2016.
- [16] E. Mangyo, "The effect of water accessibility on child health in China," *J. Health Econ.*, vol. 27, no. 5, pp. 1343–1356, Sep. 2008.
- [17] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient classification based on multiple class-association rules," in *Proc. IEEE Int. Conf. Data Mining*, Nov./Dec. 2001, pp. 369–376.
- [18] M. Kubat, *An Introduction to Machine Learning*. 2017.
- [19] T. Tiyasha, T. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000–2020," *J. Hydrol.*, vol. 585, Jun. 2020, Art. no. 124670.
- [20] S. K. Bhagat, K. Pyrgaki, S. Q. Salih, T. Tiyasha, U. Beyaztas, S. Shahid, and Z. M. Yaseen, "Prediction of copper ions adsorption by attapulgite adsorbent using tuned-artificial intelligence model," *Chemosphere*, vol. 276, Aug. 2021, Art. no. 130162.
- [21] Y. Boztuğ and T. Reutterer, "A combined approach for segment-specific market basket analysis," *Eur. J. Oper. Res.*, vol. 187, no. 1, pp. 294–312, May 2008.
- [22] Y. Boztuğ and T. Reutterer, "A combined approach for segment-specific market basket analysis," *Eur. J. Oper. Res.*, vol. 187, no. 1, pp. 294–312, 2008.
- [23] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2592–2602, 2009.
- [24] T. Jordanova, R. Cronk, W. Obando, O. Medina, R. Kinoshita, and J. Bartram, "Water, sanitation, and hygiene in schools in low socio-economic regions in Nicaragua: A cross-sectional survey," *Int. J. Environ. Res. Public Health*, vol. 12, no. 6, pp. 6197–6217, May 2015.
- [25] A. M. Bhatti and S. Nasu, "Domestic water demand forecasting and management under changing socio-economic scenario," in *Proc. Soc. Social Manag. Syst.*, 2010, pp. 1–8.
- [26] E. A. Adams, "Intra-urban inequalities in water access among households in Malawi's informal settlements: Toward pro-poor urban water policies in Africa," *Environ. Develop.*, vol. 26, pp. 34–42, Jun. 2018.
- [27] E. L. E. Angoua, K. Dongo, M. R. Templeton, J. Zinsstag, and B. Bonfoh, "Barriers to access improved water and sanitation in poor peri-urban settlements of Abidjan, Côte d'Ivoire," *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0202928.
- [28] R. Martínez-Espiñeira and C. Nauges, "Is all domestic water consumption sensitive to price control?" *Appl. Econ.*, vol. 36, no. 15, pp. 1697–1703, Sep. 2004.
- [29] F. Arbués, M. Á. García-Valiñas, and R. Martínez-Espiñeira, "Estimation of residential water demand: A state-of-the-art review," *J. Socio-Econ.*, vol. 32, no. 1, pp. 81–102, Mar. 2003.
- [30] *Progress on Sanitation and Drinking Water: 2015 Update and MDG Assessment*, UNICEF, World Health Org., Geneva, Switzerland, 2015.
- [31] H. M. Gunatilake, C. Gopalakrishnan, and I. Chandrasena, "The economics of household demand for water: The case of Kandy municipality, Sri Lanka," *Int. J. Water Resour. Develop.*, vol. 17, no. 3, pp. 277–288, Sep. 2001.
- [32] A. Jansen and C. E. Schulz, "Water demand and the urban poor: A study of the factors influencing water consumption among households in Cape Town, South Africa," *South Afr. J. Econ.*, vol. 74, no. 3, pp. 593–609, 2006.
- [33] L. A. T. Fotue, "Determinants of the households' choice of drinking water source in Cameroon," *J. Sustain. Develop. Afr.*, vol. 14, no. 3, pp. 86–97, 2012.
- [34] S. Irianti, P. Prasetyoputra, and T. P. Sasimartoyo, "Determinants of household drinking-water source in Indonesia: An analysis of the 2007 Indonesian family life survey," *Cogent Med.*, vol. 3, no. 1, pp. 1–13, 2016.
- [35] T. T. Tuyet-Hanh, T. Thi, J.-K. Lee, J. Oh, H. Van Minh, C. O. Lee, L. T. Hoan, Y.-S. Nam, and T. K. Long, "Household trends in access to improved water sources and sanitation facilities in Vietnam and associated factors: Findings from the multiple indicator cluster surveys, 2000–2011," *Global Health Action*, vol. 9, no. 1, 2016, Art. no. 29434.
- [36] P. Martínez-Santos, "Determinants for water consumption from improved sources in rural villages of southern Mali," *Appl. Geogr.*, vol. 85, pp. 113–125, Aug. 2017.
- [37] I. R. Abubakar, "Factors influencing household access to drinking water in Nigeria," *Utilities Policy*, vol. 58, pp. 40–51, Jun. 2019.
- [38] K. He, J. Zhang, Y. Zeng, and L. Zhang, "Households' willingness to accept compensation for agricultural waste recycling: Taking biogas production from livestock manure waste in Hubei, P.R. China as an example," *J. Clean. Prod.*, vol. 131, pp. 410–420, Sep. 2016.
- [39] M. Munasinghe, *Water Supply and Environmental Management*. Evanston, IL, USA: Routledge, 2019.
- [40] J. P. Graham, M. Hirai, and S.-S. Kim, "An analysis of water collection labor among women and children in 24 sub-Saharan African countries," *PLoS ONE*, vol. 11, no. 6, Jun. 2016, Art. no. e0155981.
- [41] L. Katsi, J. Siwadi, E. Guzha, F. S. Makoni, and S. Smits, "Assessment of factors which affect multiple uses of water sources at household level in rural Zimbabwe—A case study of Marondera, Murehwa and Uzumba Maramba Pfungwe districts," *Phys. Chem. Earth*, vol. 32, nos. 15–18, pp. 1157–1166, 2007.
- [42] P. Jagals, "Does improved access to water supply by rural households enhance the concept of safe water at the point of use? A case study from deep rural South Africa," *Water Sci. Technol.*, vol. 54, no. 3, pp. 9–16, Aug. 2006.
- [43] K. Sakisaka, E. A. Chadeka, S. Nagi, D. S. Mwandembo, and M. Jimba, "Introduction of a community water supply in rural western Kenya: Impact on community wellbeing and child health," *Int. Health*, vol. 7, no. 3, pp. 204–211, May 2015.
- [44] P. R. Hunter, A. M. MacDonald, and R. C. Carter, "Water supply and health," *PLoS Med.*, vol. 7, no. 11, 2010, Art. no. e1000361.
- [45] C. G. Victora, P. G. Smith, J. P. Vaughan, L. C. Nobre, C. Lombard, A. M. B. Teixeira, S. C. Fuchs, L. B. Moreira, L. P. Gigante, and F. C. Barros, "Water supply, sanitation and housing in relation to the risk of infant mortality from diarrhoea," *Int. J. Epidemiol.*, vol. 17, no. 3, pp. 651–654, 1988.
- [46] S. A. Esrey, J. B. Potash, L. Roberts, and C. Shiff, "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma," *Bull. World Health Org.*, vol. 69, no. 5, p. 609, 1991.
- [47] M. A. Hanjra and F. Gichuki, "Investments in agricultural water management for poverty reduction in Africa: Case studies of Limpopo, Nile, and Volta river basins," *Natural Resour. Forum*, vol. 32, no. 3, pp. 185–202, Aug. 2008.

- [48] L. Haller, G. Hutton, and J. Bartram, "Estimating the costs and health benefits of water and sanitation improvements at global level," *J. Water Health*, vol. 5, no. 4, pp. 467–480, Dec. 2007.
- [49] J. Newman, M. Pradhan, L. B. Rawlings, G. Ridder, R. Coa, and J. L. Evia, "An impact evaluation of education, health, and water supply investments by the Bolivian social investment fund," *World Bank Econ. Rev.*, vol. 16, no. 2, pp. 241–274, Aug. 2002.
- [50] M. de França Doria, "Factors influencing public perception of drinking water quality," *Water Policy*, vol. 12, no. 1, pp. 1–19, Feb. 2010.
- [51] C. Tortajada, F. González-Gómez, A. K. Biswas, and J. Buurman, "Water demand management strategies for water-scarce cities: The case of Spain," *Sustain. Cities Soc.*, vol. 45, pp. 649–656, Feb. 2019.
- [52] J. Zhang and L. C. Xu, "The long-run effects of treated water on education: The rural drinking water program in China," *J. Develop. Econ.*, vol. 122, pp. 1–15, Sep. 2016.
- [53] D. K. Lamichhane and E. Mangyo, "Water accessibility and child health: Use of the leave-out strategy of instruments," *J. Health Econ.*, vol. 30, no. 5, pp. 1000–1010, Sep. 2011.
- [54] A. Prüss-Ustün et al., "Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: A retrospective analysis of data from 145 countries," *Tropical Med. Int. Health*, vol. 19, no. 8, pp. 894–905, Aug. 2014.
- [55] S. B. Sorenson, C. Morssink, and P. A. Campos, "Safe access to safe water in low income countries: Water fetching in current times," *Social Sci. Med.*, vol. 72, no. 9, pp. 1522–1526, May 2011.
- [56] R. M. Saleth and A. Dinar, *Satisfying Urban Thirst: Water Supply Augmentation and Pricing Policy in Hyderabad City, India*. Washington, DC, USA: The World Bank, 1997.
- [57] *Ethiopia—Complex Emergency*. USAID, Washington, DC, USA, 2019.
- [58] *Ethiopia: Mid-Year Review of the Humanitarian Requirements Document (HRD) for 2017*. ReliefWeb, New York, NY, USA, 2017.
- [59] T. Alemu, E. Mulugeta, and M. Tadese, "Determination of physico-chemical parameters of 'Hora' natural mineral water and soil in Senkele Kebele, Oromia region, Ethiopia," *Cogent Chem.*, vol. 3, no. 1, Jan. 2017, Art. no. 1354800.
- [60] H. Kloos and R. T. Haimanot, "Distribution of fluoride and fluorosis in Ethiopia and prospects for control," *Tropical Med. Int. Health*, vol. 4, no. 5, pp. 355–364, May 1999.
- [61] J. Kaur and N. Madan, "Association rule mining: A survey," *Int. J. Hybrid Inf. Technol.*, vol. 8, no. 7, pp. 239–242, Jul. 2015.
- [62] C. Li, N. Ding, G. Zhang, and L. Li, "Association analysis of serial cases based on Apriori algorithm," in *Proc. 4th Int. Conf. Math. Artif. Intell.*, 2019, pp. 137–140.
- [63] S. Mishra, A. Pease, and J. R. Stribling, "Automated action performance using network data in association with machine learning and big data techniques," U.S. Patent Appl. 15 801 008, 2019.
- [64] F. Ciarapica, M. Bevilacqua, and S. Antomarioni, "An approach based on association rules and social network analysis for managing environmental risk: A case study from a process industry," *Process Saf. Environ. Protection*, vol. 128, pp. 50–64, Aug. 2019.
- [65] H. Toivonen, *Apriori Algorithm BT—Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2010, pp. 39–40.
- [66] R. Srikant and R. Agrawal, "Mining generalized association rules," in *Proc. 21st VLDB Conf.*, Zürich, Switzerland, 1995, pp. 1–13.
- [67] A. Raorane, R. Kulkarni, and B. Jitkar, "Association rule—extracting knowledge using market basket analysis," *Res. J. Recent Sci.*, vol. 1, no. 2, pp. 19–27, 2012.
- [68] A. K. Sharma, *Text Book of Correlations and Regression*. New Delhi, India: Discovery Publishing House, 2005.
- [69] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, vol. 5, no. 8. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
- [70] C. Borgelt, "Efficient implementations of Apriori and Eclat," in *Proc. IEEE ICDM Workshop Frequent Itemset Mining Implement. (FIMI)*, Nov. 2003, p. 90.
- [71] A. Al-Aboodi and Z. Hashim, "Assessment of groundwater vulnerability using LULC map and DRASTIC technique in Bahr AL-Najaf area, middle of Iraq," *Tikrit J. Eng. Sci.*, vol. 26, no. 3, pp. 1–9, Sep. 2019.
- [72] S. K. Bhagat, T. M. Tung, and Z. M. Yaseen, "Development of artificial intelligence for modeling wastewater heavy metal removal: State of the art, application assessment and possible future research," *J. Cleaner Prod.*, vol. 250, Mar. 2020, Art. no. 119473.
- [73] S. K. Bhagat, T. M. Tung, and Z. M. Yaseen, "Heavy metal contamination prediction using ensemble model: Case study of Bay sedimentation, Australia," *J. Hazardous Mater.*, vol. 403, Feb. 2021, Art. no. 123492.
- [74] T. Tiyasha, T. M. Tung, S. K. Bhagat, M. L. Tan, A. H. Jawad, W. H. M. W. Mohtar, and Z. M. Yaseen, "Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models," *Mar. Pollut. Bull.*, vol. 170, Sep. 2021, Art. no. 112639.



TIYASHA TIYASHA received the master's degree (Hons.) in environmental engineering from Rajasthan Technical University, India, in 2014. She is currently pursuing the Ph.D. degree with the Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. She successfully secured the National Eligibility Test conducted by the Government of India. She joined as a Lecturer at the Department of Civil Engineering, Institute of Technology, Ambo University, Ethiopia, where she also led a funded project which won the Energy Globe Award, issued by Wolfgang Neumann, an Austrian Energy Pioneer.



SURAJ KUMAR BHAGAT received the master's degree (Hons.) in environmental engineering from Rajasthan Technical University, India, in 2014. He is currently pursuing the Ph.D. degree with the Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. In 2014, he joined the Department of Civil Engineering, Institute of Technology, Ambo University, Ethiopia, as a Lecturer, where he also led a funded project which won the Energy Globe Award at national level, issued by Wolfgang Neumann, an Austrian Energy Pioneer. He also earned a Research Fellowship at the Indian Institute of Technology (IIT Delhi), India, in 2013. He received the Gold Medal from Rajasthan Technical University for his master's degree, in 2014.



FIRAOL FITUMA received the bachelor's degree from the Department of Civil Engineering, Institute of Technology, Ambo University, Ethiopia. He equipped with soft skilled while working several water management project to the local community in collaboration with international and national non-governmental organization. His hard-working and dedication sharpen the pencil of the management guidelines. His several recommendation has been peer-reviewed in the authority meetings and included some of them.



TRAN MINH TUNG received the master's degree in civil engineering from the Ho Chi Minh City University of Technology, Vietnam, and the Ph.D. degree in civil engineering from the University of Wollongong, Australia. He is currently a Lecturer and the Dean of the Faculty of Civil Engineering, Ton Duc Thang University, Vietnam. His majors include sustainability construction, structure engineering, environmental engineering, and climate. In addition, he has an excellent expertise in machine learning and advanced data analytics. He has published over 30 articles in international journals. He was awarded the Peter Schmidt Memorial Award for the best performance in postgraduate research, in 2013. In addition, he won the scientific projects at Ton Duc Thang University.



SHAMSUDDIN SHAHID is currently an Associate Professor with the Department of Hydraulic and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia (UTM), where he is also the Head of the Integrated Water Resources Management (IWRM) Research Group. His major research interests include water resources management, climate change impacts and adaptation to water resources, statistical hydrology, hydrological disasters, and groundwater hydrology. He uses statistical and mathematical tools for innovative solutions of hydrological problems for adaptation to global environmental changes. His major research achievements in recent years include forecasting water demand in a holistic way, downscaling climate using an alternative approach for better projection of climate change, modeling economical impacts of global warming induced changes in water resources, modeling droughts during crop-growing seasons, projection of drought severity-area-frequency curves for adaptation to climate change impacts on droughts and water stress, multi-criteria decision making for selection of adaptation and mitigation strategies, and determination of unidirectional trends to distinguish global climate change from natural variability. He has successfully completed about ten national and international research projects as a Principal Investigator. He published about 100 research articles in internationally reputed indexed journals and five academic books.



ZAHER MUNDHER YASEEN is currently a Lecturer and a Researcher in the field of civil engineering. The scope of his research is quite abroad, covering water resources engineering, environmental engineering, knowledge-based system development, and the implementation of data analytic and artificial intelligence models. He has published over 220 research articles within international journals and total number of citations over 6000 (Google Scholar H-Index: 45). He has collaborated with over 40 international countries and more than 430 researchers. He has served as a reviewer for more than 120 international journals.

• • •