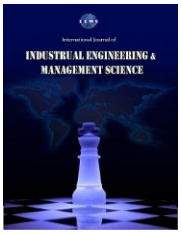
International Journal of Industrial Engineering & Management Science, Vol. 7, Issue 1, (2020) 105-114

University of Hormozgan

International Journal of Industrial Engineering & Management Science

Journal homepage: www.ijiems.com




# Water Consumption Analysis Using Data Mining Techniques A Case study: Fasa

Abbas Dehghan Bahraghani[a], Mohammad Reza Feylizadeh[b,*]

[a] *Persian Gulf International Branch, Islamic Azad University*
[b] *Department of Industrial Engineering, Shiraz Branch, Islamic Azad University*

***Keywords*:**

1 Behavior analysis
2 Water and wastewater
3 Consumers
4 Data Mining

**A B S T R A C T**

Optimizing the pattern of water consumption and institutionalizing the right structure in the consumer community will lead to improving the quality life parameters, eliminating the unnecessary financial costs and paving the path to global economic justice in the country. The first step to reach this goal is to analyze the behavior and detect the interests of the consumer community. Such an informational background will make it easier to find appropriate solutions to improve the consumption pattern. Data Mining (DM) is one of the most efficient methods to detect and dig out useful structures existing behind raw information with minimum human error. The retrieved conclusions can help analysts solve decision-making problems in various organizations. The goal of this research is to determine the behavior patterns of water consumers in a given period. Data mining method will be used to analyze the consumers' behavior in between 2015 and 2017. The extracted results about the total consumption of the sample community, the average amount of each user type and their changes will all be shown and reported.

## 1. Introduction

The competition over dominating the limited resources of water will be one of the critical challenges of the third millennium. Rapid population growth, industrial and social development, and climate changes have all put heavy pressure on the limited resources of accessible freshwater in the world. Even though freshwater has always been an invaluable resource, the false impression of abundance has led people in some parts of the world to neglect the resource management of water and consequently pushed them to the edge of the water crisis. The crisis has cast its shadow over Iran before many other countries. The water scarcity in recent years, despite being considered temporary, can be the beginning of an great crisis which in the absence of coherent attention can undoubtedly cause severe damage to the country's vulnerable economic structure [1].

Looking for a new management approach can be an effective way to face the challenge. In other words, we should extract the consumption models in given samples, analyze them and then find the relations between that processed information to move towards finding out the consumption pattern for the entire consumer community under study [2]. One of the methods to be used for this purpose is "Data Mining". Data Mining is a process which benefits from various analytic means for extracting pattern and relations that can be used to make valid predictions [3].

Subsequently, the need for collecting and managing significant amounts of temporal and spatial high granularity data becomes of decisive importance. Significant progress in information and communication technologies (ICTs), DM and social network, enable new paths of urban planning, empowering the resilience of urban infrastructure and the adequacy of resources and systems [4]. The rapid development of big data and cloud computing technology combined with advanced ICTs make traditional energy system been digitalized [5,6]

Data Mining (introduced in the early 90s) is an entirely scientific method to extract hidden knowledge and find the existing relations in a database. A method is a new approach to data handling and processing [7]. The birth of this branch of science can be traced back to 1989 and 1991 when Yanteski and colleagues ran the knowledge extraction workshops [8]. Various studies using Data Mining method have been conducted and carried out by water administration organizations on the subject of water consumption patterns and related problems. As an example, in a survey done by Christine and colleagues [9], the clustering algorithm was applied to bills issued by various social service providers in order to categorize customers and understand their characteristics. The survey was aimed at solving decision-making problems regarding logistics and customer feedback methods.

Considering the importance of freshwater administration problems, In urban areas with relatively dense industries and population, the supply-demand conflict of water resources is hugely significant, making water resources a critical factor that restricts the development of the social economy in the 21st century [10], Dutta [11] also used Data Mining techniques and ANN and fuzzy algorithms as well as the database of Geographic information system (GIS) to prepare a model for categorizing water pollutants.

Moreover, Wen and colleagues [12] applied the cluster techniques, K-Means algorithms and Time Series Method (as another important Data Mining technique) to their database and managed to create some computer software, which could predict future consumption amounts. Similarly, in another survey performed in the United States, metaheuristic algorithm was applied to the domestic water use by breaking the pattern down to different user types such as washing machine, dish machine, and shower. Then, the amount of wastewater was calculated with Data Mining methods [13].

As another example, in a research performed in China, DBSCAN[1] and K-Nearest Neighbors (KNN) algorithms, as well as subspace clustering method, were used to analyze and evaluate water pipe bursts and other accidents in water networks, after which Density-Based clustering was applied to determine areas of the network which had the highest risk of accidents and pipe bursts [14].

Agha-Babaie and colleagues [15] used Data Mining technique to deal with indebted customers and debtors in the city of Mashhad (zone 3). In their research, the Neural Network Method was used to identify debtors and discover the pattern of their behavior. Decision trees were also used to detect the useful parameters that cause customers to fall behind with payments. The research opened up new horizons towards reducing financial costs in water administration departments. One of the specific and essential results of the research was the fact that if the customers received their bills with 15 days or more delay, the chance that they would not pay it at all would meaningfully increase.

In the scope of the mentioned background, the purpose of this research is to discover and introduce management and control strategies for detecting consumption patterns and structure as well as finding methods to improve them. For this purpose, the first step is to discover the existing water consumption structures and customer behavior, followed by researching the possibility of appropriate measures to optimize those patterns. We will use data mining methods to find out the

[1] Density Based Spatial Clustering of Applications with Noise

consumption patterns and find strategies to decrease the overall amount of water use. In the first part of the article (which follows the foreword), we will discuss the procedure of collecting information and data cleansing. In the second part, we will show how Data Mining is to be applied to the database to analyze the information and extract knowledge from them. Finally, in the third part, we will summarize the results and offer suggestions for further work and future research.

## 2. Data Collection and Cleansing

The goal of this article is to analyze the information about water consumption in order to discover the structures and patterns of water use for the community of customers. Data collecting means and methods for this purpose include library research (related articles and books), web database, related documents obtained from water administration organizations, mass media reports and interviews with experts, managers, and professionals.

Since various cultural and environmental parameters can affect the water consumption patterns, we needed to choose a sample community that acquired all requirements for such a research. That would include the possibility of generalizing the methods and the results for other communities of a given province or even the entire country. After reviewing previous researches and consulting with professionals in Water and Wastewater Company of Fars province, the types of data needed for the research were identified. Accordingly, since the city of Fasa counted as one of the four largest cities in Fars province and covered a wide variety of consumer types, it had the potentials to be used as the reference sample community for this research. Another reason to choose Fasa was the appropriate informational foundations and easy to access database in that city.

To discover the water customers' behavior in Fasa, all of the confirmed data had to be analyzed. We used the information of the years 2015 to 2017 from the documents and database of the customers' department of Water and Wastewater Company. According to the company's article of association, all such data have to be verified by legal auditors at the end of each financial year. Since the data we chose to use belonged to previous years, they had already been confirmed by auditors and had a very high level of validity and reliability.

On the other hand, since these data were collected from various sources and databases, they were of a low level of quality. In order to achieve data warehouse that could be effectively used for the discovery of structures and patterns, unrelated attributes were omitted, and various functions in the SQL environment were applied to convert data. Once the data cleansing procedure was completed, the results were tested for validity with S and SC criteria.

## 3. Applying data mining techniques

As mentioned before, one of the challenges of the entire country is the scarcity of freshwater, and it seems that changing the approach to consumption management can be considered a fundamental solution to the problem. After identifying the patterns of consumptions and analyzing them and recognizing the relations between those results and studying the consumption behavior in the sample community, it may be possible to find and present strategies to improve the situation. In this part, we use data mining techniques to study the behavior of customers in a certain period. The analysis has two stages, as follows:

### 3.1. Data preparation

In order to identify the consumption patterns, the relevant fields of information were determined by consulting with professionals, and the relationship diagram was prepared in the "Access" (software) environment.

The table is called “Information retrieved from the company” in the data warehouse, and it is the most important table in the database. The table includes over one million entries, which show the water consumption of all customers in Fasa between 2015 and 2017. Those data are fixed and definite and have a very high analytical value. This table, as the central core of relations in the database, can provide the information required to create the relational database shown in the figure.1 below.

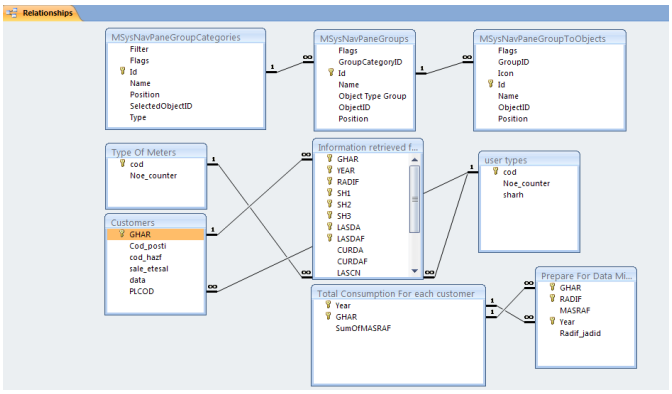



**Fig. 1.** Relational Database

Beside this table, another topic is introduced as user types (or use types) in which different categories and sub-categories of water use are classified. Different user types identified with their different counter types are as follows: domestic, commercial, office buildings, public services (such as hospitals and parks), universities, non-governmental educational organizations, shrines, public baths, bakery, non-traditional bakery, industrial buildings (such as plants and factories), governmental cultural centers (such as libraries and clubs), educational centers, non-governmental cultural centers, charity, landscapes, non-potable water (divided to non-potable for domestic, office buildings, commercial; these refer to the non-potable water mainly used as building material), non-governmental services (such as newspaper office and notary public office), rural, military, second home (such as holiday cottage and garden), holy places, etc.

This information, together with the existing tools in “Access” software and SQL programming language, were used to create Data Warehouse, which is the most vital part of Data Mining method. Data Warehouse includes several sections, all of which have played their specific roles in data preparation for Data Mining. The data warehouse is shown below in figure.2.

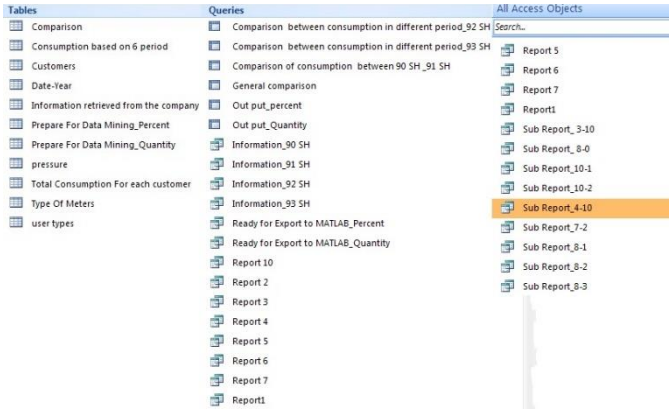



**Fig. 2.** Objects of Data warehouse

### 3.2. Data analysis

After creating the road map and knowing what kind of information is needed, the research was commenced. The results of data analysis are categorized in separate reports. These reports which will be discusses below separately include the report of annual consumption, the report of the number of customers with counters that had been defective for more than 4 months, the report of average consumption of each customer, the report of the consequences of having defective counters for over 4 months, the report of the changes of consumption amount of each user type in between 2015 and 2017, spectral chart for different user types and overuse.

After creating the road map and knowing what kind of information is needed, the research was commenced. The results of data analysis are categorized in separate reports. These reports which will be discusses below separately include the report of annual consumption, the report of the number of customers with counters that had been defective for more than 4 months, the report of average consumption of each customer, the report of the consequences of having defective counters for over 4 months, the report of the changes of consumption amount of each user type in between 2015 and 2017, spectral chart for different user types and overuse.

#### 3.2.1. Report 1: Annual consumption

In this report (figure 3) the amounts of water consumption from 2015 to 2017 are demonstrated. As shown in figure 3 below, the consumption amount has increased by 17% from 2015 to 2017 while the annual rainfall and the level of underground water have both diminished during those years. This consumption pattern rings an alarm for managers and strategists to modify their policies in the region. One of their priorities would undoubtedly be amending and improving water consumption structures.

| Year | Total cunsumption | Percent Of Growth |
|---|---|---|
| 2015 | 8,270,977.00 | 0 |
| 2016 | 9,170,504.00 | 11% |
| 2017 | 9,683,947.00 | 17% |

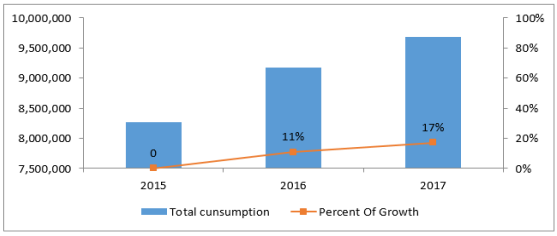



**Fig. 3.** The trend of rising consumption

#### 3.2.2. Report 2: Customers with defective counters (for over four months)

After considering more detailed information, we can observe that 2323 customers have had defective counters for over two counter reading periods (which is equal with four months). Interestingly, the average number of counter reading periods during which those customers' counter has remained defective is 7, which is a remarkable and drastic number, because defective counters affect the company's income very severely. On the other hand, the cluster graph (figure 4 below) shows that even though the highest density of clusters is between numbers 0 – 100 (on the vertical axis) and 0 – 10 (on horizontal axis), there are still observable cases in other areas of the graph, which warrant a quick consideration to control and fix the situation of those customers.

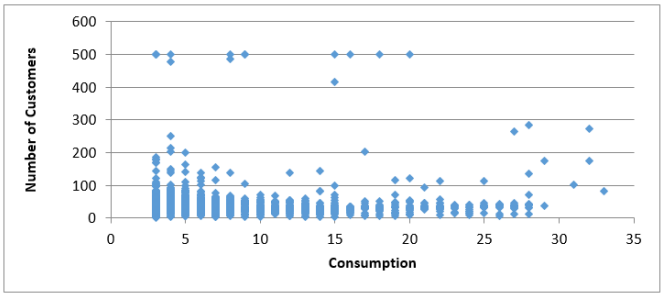



**Fig. 4.** The densityof clusters for defective counters

### 3.2.3. Report 3: Average consumption of each Customer

Report no. 2 demonstrated the number of customers with defective counters (for over two counter reading periods). Investigations showed the value of money the company was losing due to those customers. Operational investigations were used to determine the average values that would read on those defective counters if they were operative.

### 3.2.4. Report 4: consequences of defective counters (for over four months)

Next, after reports 2 and 3 reported 4, which compares the average amount of water consumption when the counter is defective and when it is operative. As stated before, in report 2, there were 2323 customers with defective counters. Calculating the amounts of consumption for those customers when they had operative counters and comparing the numbers with the time their counters were defective, it was shown that for the majority of those customers the approximated amount of consumption when they had defective counters was higher than the exact calculated amount when they had operative counters (which could have been financially good news for the company). However, for the entire 2323 customers, the total amount of approximated amount of consumption during the time the counters were defective was lower than the calculated amount when the counters were working, which indicates a significate financial loss for the company[2]. When those customers had defective counters, a total amount of 191376 $m^3$ was approximated for them, while their actual calculated use of water when they had operative counters was 22390 $m^3$.

### 3.2.5. Report 5: Changes in consumption amount of each user type

In this report, the amount of change and percentage change of consumption types in between the years 2015 and 2017 are examined and demonstrated. The report is created with Crystal Reports application. In this report, the percent changes have been calculated first, but then the amount changes have been shown as well to achieve a better analysis of the information. As it is evident in the figure. 5 below, the percentage change and the amount of change are not similar for domestic and non-potable domestic use types, which means the customer department of the company has never had a professional approach to the situation and have always used very general and raw reports instead.

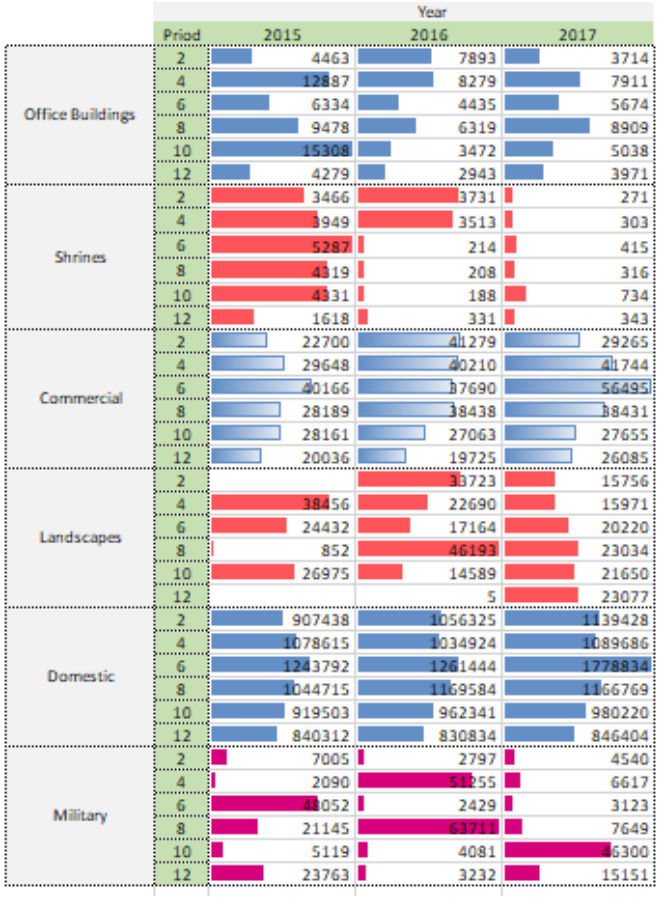
[2] Because based on the approximations, the company charged those customers far less than they actually would have had to pay (if they had operative counters).

| | Percent of Consumption | | | Consumption ($M^3$) | | | Trend Of Percen | Trend of Consumption |
|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2015 | 2016 | 2017 | | |
| Domestic | 72.96% | 68.87% | 72.30% | 6,034,375 | 6,315,452 | 7,001,341 | | |
| Rural | 5.12% | 8.44% | 8.68% | 423,648 | 773,780 | 840,776 | | |
| Non-Potable Domestic | 4.95% | 4.83% | 4.82% | 409,376 | 443,392 | 466,664 | | |
| Universities | 2.42% | 2.94% | 2.87% | 200,028 | 269,561 | 278,348 | | |
| Other | 3.98% | 3.03% | 0.07% | 329,248 | 277,786 | 6,362 | | |
| Commercial | 2.04% | 2.23% | 2.27% | 168,900 | 204,405 | 219,675 | | |
| Industrial Buildings | 2.14% | 1.98% | 2.25% | 177,397 | 181,216 | 217,987 | | |
| Landscapes | 1.10% | 1.47% | 1.24% | 90,715 | 134,364 | 119,708 | | |
| Military | 1.30% | 1.39% | 0.86% | 107,174 | 127,505 | 83,380 | | |
| Nongovernmental Services | 0.78% | 1.60% | 0.43% | 64,586 | 146,505 | 41,912 | | |
| Educational Centers | 0.23% | 1.19% | 1.21% | 18,795 | 109,258 | 116,830 | | |
| Grand Total | 100.00% | 100.00% | 100.00% | 8270977 | 9170504 | 9683947 | | |

**Fig. 5.** The trend of consumption for each user type

### 3.2.6. Report 6: Spectral Chart for different user types

The purpose of this report is to show the annual consumption of each user type in the form a spectral chart. The spectral color of the domestic type is more meaningful and logical than other user types, because for example, as it can be seen in the figure. 6 below, the maximum consumption for domestic users happens in the summer months of July and August (which is logical), while on the other hand, for landscape use type, the consumption amount remains the same in all periods (which does not make much sense).



| | | Year | | |
|---|---|---|---|---|
| | Priod | 2015 | 2016 | 2017 |
| Office Buildings | 2 | 4463 | 7893 | 3714 |
| | 4 | 12887 | 8279 | 7911 |
| | 6 | 6334 | 4435 | 5674 |
| | 8 | 9478 | 6319 | 8909 |
| | 10 | 15308 | 3472 | 5038 |
| | 12 | 4279 | 2943 | 3971 |
| Shrines | 2 | 3466 | 3731 | 271 |
| | 4 | 3949 | 3513 | 303 |
| | 6 | 5287 | 214 | 415 |
| | 8 | 4319 | 208 | 316 |
| | 10 | 4331 | 188 | 734 |
| | 12 | 1618 | 331 | 343 |
| Commercial | 2 | 22700 | 41279 | 29265 |
| | 4 | 29648 | 40210 | 41744 |
| | 6 | 40166 | 37690 | 56495 |
| | 8 | 28189 | 38438 | 38431 |
| | 10 | 28161 | 27063 | 27655 |
| | 12 | 20036 | 19725 | 26085 |
| Landscapes | 2 | | 33723 | 15756 |
| | 4 | 38456 | 22690 | 15971 |
| | 6 | 24432 | 17164 | 20220 |
| | 8 | 852 | 46193 | 23034 |
| | 10 | 26975 | 14589 | 21650 |
| | 12 | | 5 | 23077 |
| Domestic | 2 | 907438 | 1056325 | 1139428 |
| | 4 | 1078615 | 1034924 | 1089686 |
| | 6 | 1243792 | 1261444 | 1778834 |
| | 8 | 1044715 | 1169584 | 1166769 |
| | 10 | 919503 | 962341 | 980220 |
| | 12 | 840312 | 830834 | 846404 |
| Military | 2 | 7005 | 2797 | 4540 |
| | 4 | 2090 | 51255 | 6617 |
| | 6 | 48052 | 2429 | 3123 |
| | 8 | 21145 | 63711 | 7649 |
| | 10 | 5119 | 4081 | 46300 |
| | 12 | 23763 | 3232 | 15151 |

**Fig. 6.** Spectral chart for different user types

### 3.2.7. Report 7: Overuse

There is no acceptable and precise definition for "water overuse" in Water and Wastewater Company. The problem is that a certain fixed amount of consumption relative to the user family size is taken as a criterion and applied to all user types[3] and if a user consumes more than that amount, they will be found overusing. What we did in our analysis was to take all users between the years 2015 and 2017. Also, as shown in the figure. 7, we detected customers whose annual water consumption was increasing year by year as well as those whose annual consumption was decreasing. The result we deduced was that 19% of customers of those years had been overusing water and 22% of them had been reducing their water use every year. Due to the mentioned analysis, the following measurements can be taken into consideration:

- The consumption management should give priority to those customers whose annual consumption is increasing.

- The control and monitoring priority, on the other hand, should be placed on customers whose annual consumption is decreasing (to detect illegal splits, defective counters, management errors etc.)

It is worth noting that if we consider only the information about the years 2015 and 2017, and we try to discover the percentage of customers whose water consumption has increased by over 5%, we will find the figure 37%, which is very alarming for future water supply of the city of Fasa and demands for particular amending strategies regarding those customers.

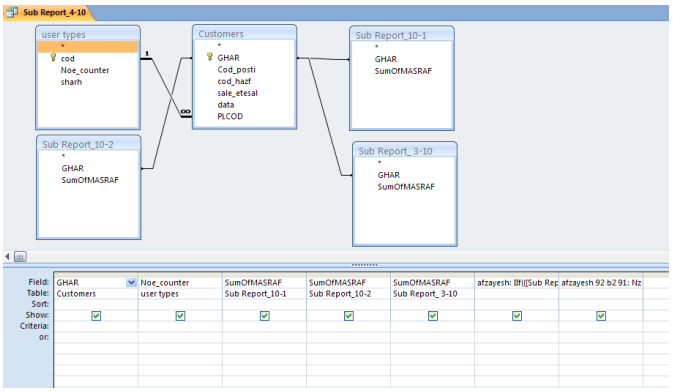



**Fig. 7.** Design view for select query

## 4. Summary and suggestions

Data mining as a development system is increasingly growing up. Owners and managers of many companies and organizations believe that they need to use these techniques for Knowledge Management (KM). Social services organizations such as Water and Wastewater Company whose ultimate goal is to present the best possible services with the lowest costs are no exception to this rule and need to move towards using these methods to keep pace with the fast development of knowledge production methods.

In this respect, the discovery of the customer behavior patterns will greatly help managers make strategic decisions.

[3] without taking their differences into account

The present research showed that water consumption has been increasing in Fasa during the past few years. The amount of consumption rose up to 17% from 2015 to 2017, while the annual rainfall and the level of underground water both diminished. This is alarming for the water and wastewater company of Fasa and necessitates new consumption management strategies. In other words, the knowledge discovered in this research leads us to the conclusion that we use more water every year while there is no increase in our water resources. Maybe the only way out of the dilemma is to use water correctly and amend the customers' consumption structure.

Furthermore, the analysis of various user types showed that the amount change and the percentage change of water use are not similar for some use types such as domestic and non-potable domestic. This problem may have two possible causes. Either counters are defective, or those people who read the counters are not doing their job correctly and accurately. This is another problem that requires the immediate attention of the customers' department.

The bottom line is that the customers' departments of Water and Wastewater Company need to give the highest priority to revising their approach and strategies and start professional workshops to define entirely new priorities.

As a suggestion for further and future researches, we recommend using other Data Mining methods such as "Clustering," "Classification," "Regression" and "Association Rule Learning" to extract hidden knowledge from the customers' water consumption database.

## **References**:

[1] Agha babaie, A. shograbi, j.,2012, Application of Data Mining in the face of debt consumer in the water and wastewater company of Mashhad, The Sixth of Data Mining Conference, (In Persian).

[2] Mobini, A., 2007, *Demand (waster) management guide in water and wastewater system*, Water and Wastewater Engineering Company, (In Persian).

[3] Berry, M., and Linoff, G., 2012, *Data Mining Techniques: For Marketing, Sales, and Customer Support*. 3rd edition. New York: John Wiley and Sons.

[4] Yang, L., Yang, S.-H., Magiera, E., Froelich, W., Jach, T. and Laspidou, C. 2017. *Domestic water consumption monitoring and behaviour intervention by employing the internet of things technologies*. Procedia Computer Science, 111, 367-375.

[5] Lund, P.D., Mikkola, J., Ypya, J., 2015. *Smart energy system design for large clean power schemes in urban areas*. J. Clean. Prod. 103, 437-445.

[6] Zhou, K., Fu, C., Yang, S., 2016. *Big data driven smart energy management: from big data to big insights*. Renew. Sustain. Energy Rev. 56, 215-225.

[7] An Introduction to Data Mining: http://www.thearling.com/, retrieved on December 2, 2014.

[8] Tabatabaie, A., 2009, *Review of Data Mining*, Islamic azad university of Ghazvin, p 10-15. (In Persian).

[9] Christine, B., Shadi, E., Wyatt, M., Hughes, Jeffrey, 2011, *Mining Water Billing Data to Inform Policy and Communication Strategies*, Journal –American Water Works Association, 45-58.

[10] Liu, K.; Fu, H.; Chen, H. *Research on the Influencing Mechanism of Traditional Cultural Values on Citizens' Behavior Regarding the Reuse of Recycled Water*. Sustainability 2018, 10, 165.

[11] Dutta, P., 2012, *A survey of Data Mining Applications in Water Quality Management*, CUBE International Information Technology Conference, USA, 470-475.

[12] Wen, Y., Huang, W., Wu, J., *2013, Water Consumption Analysis System Based On Data Mining*, Mechanics and Materials, 241-244, 1093-1097.

[13] R Cardell-Oliver, G Peach, 2013, a *data mining approach for discovering Water Use Pattern*, Smart metering Data, 2-6.

[14] Sun, J., Wanga, R., Wanga, X., Pinga, J., 2014, *Spatial cluster analysis of bursting pipes in water supply networks*, Procedia Engineering ,70,1610 – 1618.

[15] Badi'ie, Sh., 2011, *Drinking Water Crisis*, shahrab, 530, (In Persian)**.**