Journal of Global Resources Volume 5 (02) July 2019 Page 91-97
ISSN: 2395-3160 (Print), 2455-2445 (Online)

13

SCOPE OF MACHINE LEARNING IN WATER CONSUMPTION AND POPULATION FORECASTING

Noopur Rathore[1], Rajeshwari Acharya[2] and **M. P. Punia[2]**
[1]B.Tech Student, Information Technology, Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur (Rajasthan), India
[2]JRF, Department of Remote Sensing, Birla Institute of Scientific Research, Jaipur, India
[2]Head, Department of Remote Sensing,
Birla Institute of Technology, Jaipur (Rajasthan), India
Email: nprr1305@gmail.com, rajeacharya@yahoo.com, punia.rsd@gmail.com

Abstract: *Water is of paramount importance for the existence of life on Earth. The causes of water depletion are both natural and anthropogenic. On Earth, amount of freshwater has remained persistent over span but the population has mushroomed. Therefore strive for freshwater intensifies day by day. Proper management and forecasting are required for better and effective water usage plans. Water demand and population forecasting are the major parameters for an Urban Water Management. Machine learning is among the best-known techniques for such forecasting. Machine learning is a data analytics technique that provides machine the potential to learn without being comprehensively programmed. Unlike the traditional methods of demand forecasting that were not suitable for historical unstructured and semi structured data, machine learning takes into account or has the capabilities for analysing such data. This paper focuses on supervised machine learning techniques i.e. Artificial Neural Networks, Linear Regression and ARIMA modelling for assessing water consumption and population forecasting for urban water management.*

Key words: Water Consumption, Population Forecasting, Demand Forecasting, Machine Learning, Artificial Neural Network, Linear Regression

Introduction

On Earth freshwater constitutes a very small fraction of the total available water and therefore need for sustainability of this natural resource is required. The water crisis is commonly seen and many steps are already been taken to avoid this problem but some are still uncertain. In today's data-driven world, massive quantities of data can be analyzed using machine learning (ML) with faster and more accurate results. ML is a great replacement for earlier statistical techniques. It can be done using three approaches that are Supervised Learning, Unsupervised Learning, semi-supervised and Reinforcement Learning. Existing machine learning works in the water and wastewater industry includes investigations into potential outbreaks through the water system, environmental impacts of wastewater treatment and disposal and estimates the effluent quality of wastewater treatment plants. Dexterity to handle data, ML proves to be an ideal tool for water asset management. Water consumption is a portion of water that is being withdrawn but never returned to the source while population forecasting is a field of demography that deals with projecting future population. Therefore interrelationship between them is used for predictive analysis. Described the advances in urban water demand modelling over the past three decades by considering parameters like water demand, climate, population, reservoirs, pumping stations, precipitation, and temperature and water price with the help of interactions across spatial and temporal scales.

Water consumption depends on several factors like topography, industrial presence and infrastructure, climatic conditions, quality of water, socio-economic elements, population growth and density, pressure in the water distribution system, water prices, policies of water meters, and efficiency of waterworks administrations. This study also proposes a time series analysis of water consumption. Described a time series model to forecast daily water consumption considering four parameters namely trend, seasonality, climatic correlation, and autocorrelation. This paper provides a frame of reference for machine learning algorithms for the use of practitioners who want to make decisions and for researchers who want to contribute in this field and moreover for the end users it provides a variety of new knowledge.

Objective

This paper focuses on the scope of machine learning algorithms (artificial neural networks, linear regression, and ARIMA model) for population growth and water consumption forecasting considering different parameters and hence assessing the interrelationship between the two. It also evaluates the applicability and accuracy of forecast models.

Assumptions

i. Data published by the government bodies are reliable.
ii. Open source software has been used for validation.

Limitations

i. Limited access to India's historical water consumption data.
ii. Focuses on Toronto city's water consumption data.

Methodology

Data and method

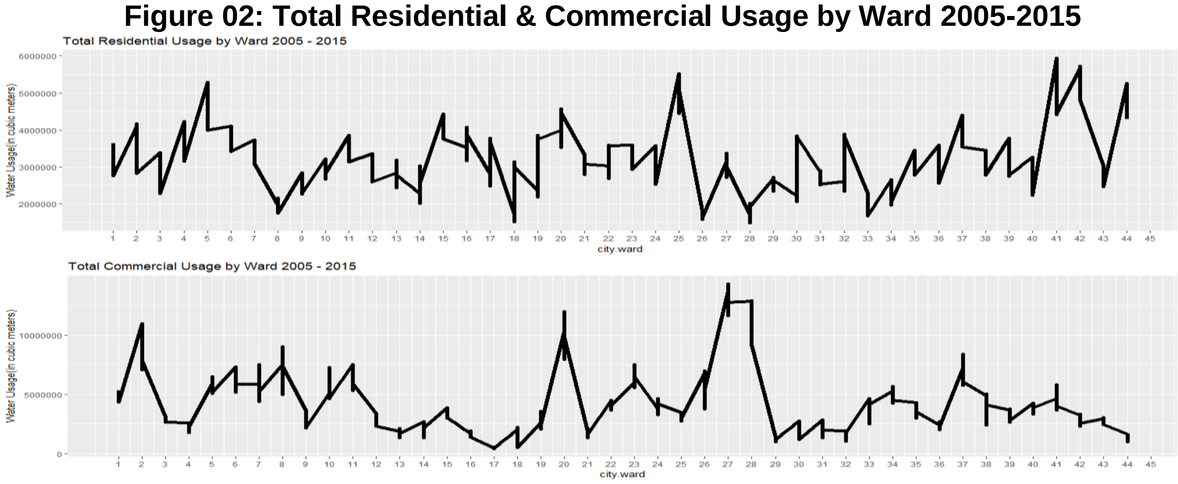
A model is built on Toronto city's water consumption and population data. Here, water consumption data from the year 2000-2015 and population census data from 2001-2016 is used (data source: [8]). RStudio software (version 1.2.1335) has been used. Figure 1 given below depicts the workflow process of data analysis.

Figure 01: Work Flow Process

Algorithm and Analysis

In data collection information from different sources was gathered and evaluated, followed by data modification which included remoulding of data in such a scheme that it can be feasible to work with by formatting and cleaning. After that data visualization was applied, in which data was modified into some form of plots and inspected further. By visualizing the data from 2005-2015, it was found that total residential usage was maximum in ward 41 i.e. Scarborough-Rouge River and total commercial usage was maximum in ward 27 i.e. Toronto Centre Rosedale. Average residential usage was maximum in ward 20 i.e. Scarborough Southwest and average commercial usage was maximum in ward 11 i.e. York South West. Ward 27 has consistently consumed the most water commercially every year from 2005 to 2015. Figure 2 shows the total residential usage by ward and total commercial usage by ward from 2005-2015; from this plot minimum and maximum water consuming wards can be identified.

Figure 02: Total Residential & Commercial Usage by Ward 2005-2015



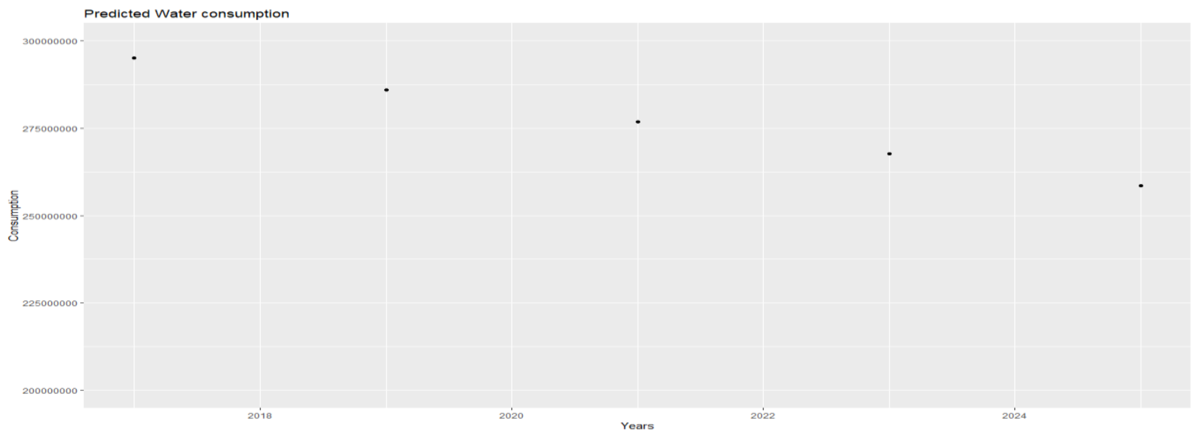
Linear Regression

Linear regression is a supervised machine learning technique, used for predictive analysis of dependent variables based on independent variables. Water demand forecasting year-wise was done using linear regression. The dependent variable for this algorithm was total consumption and the independent variable was year. Therefore equation so formed was "*Total Consumption ~ Year*" with slope and intercept as -4581081.66 and 9535168992.29 respectively.

$$\mathbf{Y_{estimated} = \theta_1 x + \theta_2}$$

Where $\mathbf{Y_{estimated}}$ is labels to data, **x** is training data, $\boldsymbol{\theta_1}$ is slope or gradient: -4581081.66 and $\boldsymbol{\theta_2}$ is intercept or value of y when x is 0.9535168992.29. Mean squared error (MSE) is the mean of the squared errors and the root mean square error (RMSE) is the square root of the mean of the squared errors and is given by:"$\sqrt{(1/n)\sum(y_i - y_i')^2}$". For this study RMSE value hence obtained was 0.7. Figure 3 illustrates the forecasted water consumption from 2017-2025.

Figure 03: Forecasted Water Consumption



Artificial Neural Networks

Water demand prediction ward-wise was done using ANN, which is a non-linear model. Feed forward ANN and feedback ANN are its two types. The configuration used here is,

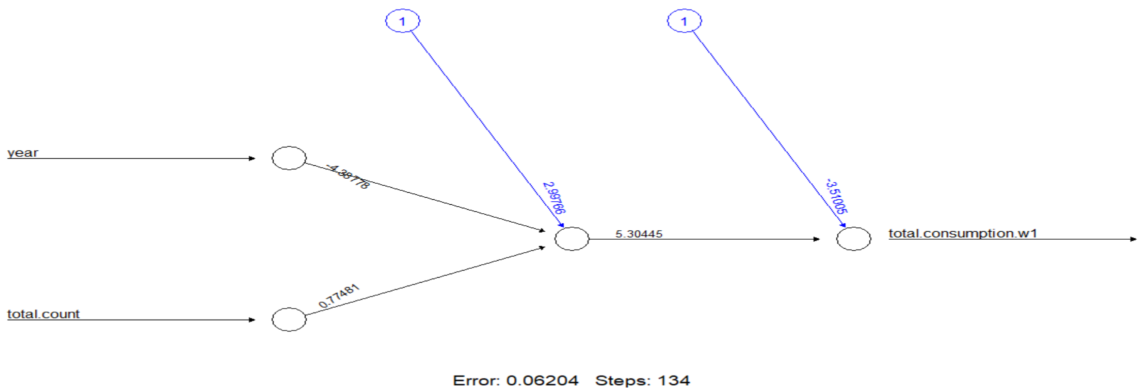
$$\mathbf{f(x)=K(\textstyle\sum_i w_i g_i(x))}$$

Where, **K** is activation function, $\mathbf{g_i(x)}$ is collection of functions, $\mathbf{w_i}$ is weight assigned to a connection. Here, feed forward, ANN is used with a single hidden node. Initially, data normalization was achieved using min-max normalization and then the model was built using the equation *"Total Consumption of ward1 ~ Year+ Total Count"* (Ward1 is used as a dummy). Total consumption was a dependent variable on independent variables year and total count. The accuracy of the model is quite evident by the plot, with an error of just 0.06204 and total steps 134. Output on the last node i.e. node 4 was also shown using the sigmoid function which was equal to the predicted value of total consumption.

$$\textbf{Sigmoid function} = \mathbf{1/(1 + e^{-x})}$$

Where, **x** is input to that node. Figure 4 represents the artificial neural network model of ward 1 prediction.

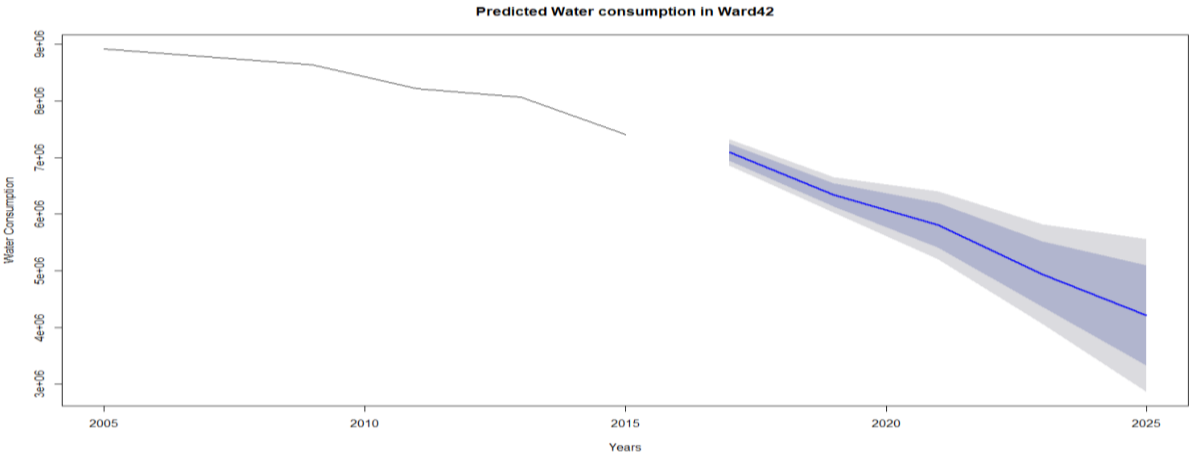
Figure 04: Artificial Neural Network model for Ward 1 prediction



ARIMA Model

Ward-wise water demand forecasting was done using the ARIMA (Auto-Regressive Integrated Moving Average) model, it is a type of statistical model used for forecasting using time series dataset. ARIMA (p,d,q) was used with the appropriate value of parameters to form a specific model. When two out of the three terms are zeros, the model is based on non-parameter releasing 'AR', 'I' or 'MA'. Water demand forecasting on ward 42 was done using the ARIMA model as it had the highest residential usage from the year 2011-2015. At the outset, data was converted into time series and then forecasted. Figure 5 given below shows the forecasted water consumption in ward 42.

Figure 05: Forecasted water consumption in ward 42



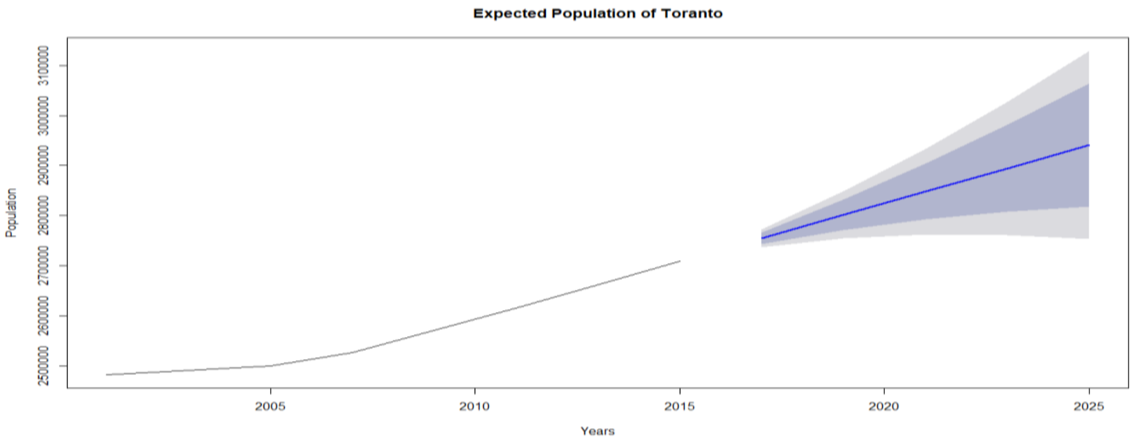
Population Forecasting

Similar to water demand ward-wise forecasting, population projection was also done using the ARIMA model. Data initially was available with a gap of 5 years. Using the population estimation formula, the population per year was calculated and then converted into time series and forecasted with parameters (1,2,0).

$$\mathbf{P_{estimated} = P_1 + (n/N)\,(P_2 - P_1)}$$

Where, $\mathbf{P_{estimated}}$ is the population estimate for a given year, **n** is the number of months from P_1 census to the date of estimate, **N** is the number of months between census periods, $\mathbf{P_2}$ is the last census taking, and $\mathbf{P_1}$ is the second to last census taking. Figure 6 portrays the expected population of Toronto from 2017-2025.

Figure 06: Forecasted Population of Toronto

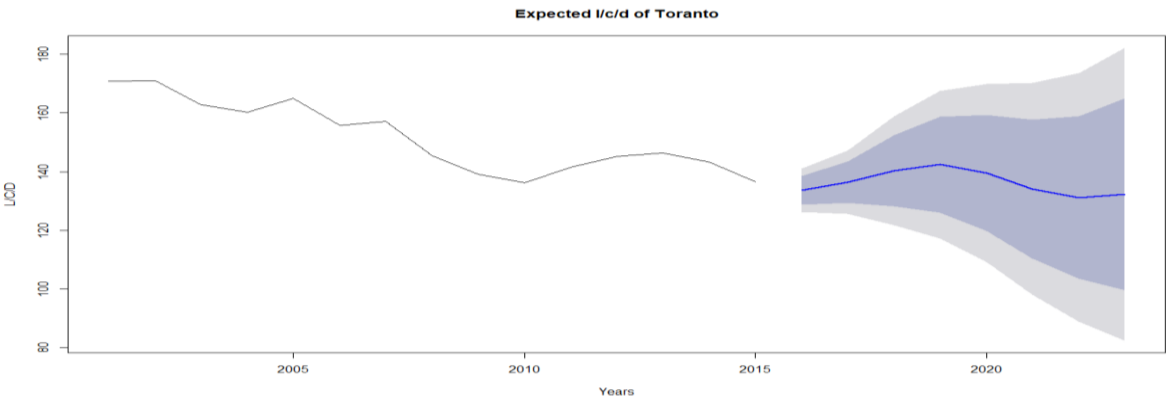


Lastly, projected water use in litres per capita per day (l/c/d) was calculated using residential water usage and population by the formula given below:

$$\mathbf{l/c/d = (WC * 1000) / (D * P)}$$

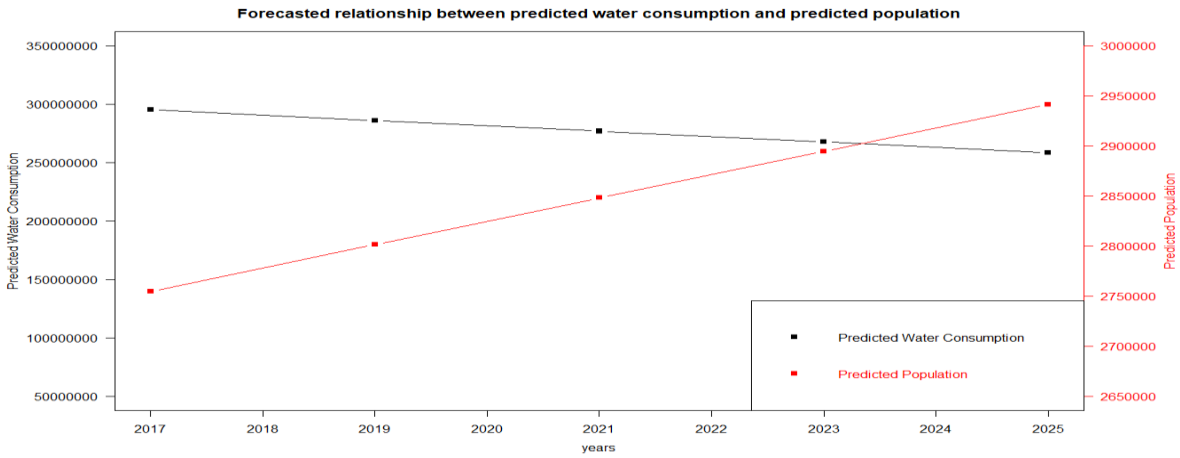
Where, **WC** is Total water consumption during the assessment period, **D** is Number of days, **P** is Total number of the served population. Similar to the previous implementation of ARIMA model l/c/d trend of Toronto was also forecasted and represented as shown in figure 7 given below:

Figure 07: Forecasted water consumption l/c/d trend of Toronto



When forecasted water consumption and forecasted population were drawn on a single plot then the graph so obtained depicted the relationship between the two, figure 8 drafts the same. From the graph thus obtained, certain conclusions were drawn which are mentioned in the result section.

Figure 08: Relationship between Forecasted Water Consumption and Forecasted Population



Results and Conclusion

This paper presented the scope of machine learning in urban water management and aims to bring advance knowledge and understanding about the same to develop methodologies and tools for further study. This paper culminated from observations that, water use in liters per capita per day went down gradually which is below the standard water consumption i.e. 150 l/c/d, it signifies that water supply w.r.t. to population is not up to the required quantity. From the relationship plot between forecasted water consumption and population, it is observed that time around 2024 will be a very critical one since the population will increase and water supply will decrease; as depicted in the plot, two curves cut each other. This is a matter of concern and should be kept in mind by the government and steps should be taken accordingly to avoid such conditions. Moreover, it is proven by the study that ML algorithms like ANN, linear regression and ARIMA modeling outperforms on such datasets as the minimal error was obtained.

Further demand forecasting helps in resource management for the government, so many more machine learning techniques can be used to predict and forecast water consumption and population to attain sustainable life. This paper permits a reliable forecast of water demand which in turn can also help in optimizing energy costs, treatment, storage and distribution of water. Moreover, it can be beneficial for analysts in the field of hydrology and water resource management when combined with artificial intelligence and other cognitive technologies to give remarkable results.

Literature Review

S.No.	Objective	Parameter	Methodology	Reference
1	This paper presents the machine learning water demand forecasting models with more accurate predictions as compared to traditional strategies.	weather, seasonality, water demand	Forecasting methodology involves neural networks, random forests, support vector machines, and k-nearest neighbours.	1
2	This paper presents combining time series clustering, for the identification of daily urban water demand patterns, and Support Vector Regression for performing a short term forecast.	Data retrieved from the SCADA system	The approach consists of clustering together daily demand patterns, represented by the volume of water delivered at each hour, and aiming at identifying a prediction model, based on the Support Vector Regression.	2
3	This paper reviews the literature on urban water demand forecasting from 2000 to 2010 to identify methods and models necessary for specific water utility decision-making problems.	Leakage, water demand	The methodology included forecasting with qualitative methods, univariate time series analysis, time series regression models, decision support system, artificial neural networks, and composite models.	6
4	This paper focuses on real-time pump scheduling in a water distribution system using an optimizer.	Sensor data, flow data	The methodology involves the use of adaptive ANN training as an alternative to the back propagation method and making use of a multi-method optimization algorithm as an alternative to the Genetic algorithm.	7
5	This paper reviews the soft computing methods of water consumption forecasting between 2005 and 2015.	Population, water consumption, climate	The methodology includes artificial neural networks (ANNs), fuzzy and neuro-fuzzy models, support vector machines, metaheuristics, and system dynamics.	9
6	Water demand forecasting using a combination of techniques such as multilayer perceptrons, support vector machines, extreme learning machines, random forests, and adaptive neural fuzzy inference systems.	Climate, water demand	Presented the methodology of committee machines for creating a predictive model for urban water demand.	10
7	The goal was to avoid overwhelming operators with voluminous raw data and present the solutions to link and interpret data streams.	Big data from water and wastewater industries	The methodology used here involves data analytics. Data analytics is used here to determine probable future outcomes to identify relationships and detect multivariate data excursions from the normal ones.	13
8	Enhancing water system using sensor technologies, firmware application, big data tools, and M2M technology	Water quality, flow, pressure	It uses sensor technologies and firmware applications, big data analytics tools and embedded Machine-to-Machine communications for collecting data enabling better information to be provided to the human operators that were previously been manually collected and rarely evaluated.	14
9	A Model Conditional Processor (MCP) application is presented to assess the predictive uncertainty in water demand forecasting.	Water demand, water distribution system	The method involves converting historical observation and the corresponding forecasted values into normal space for the conditional distribution of the real values to the forecasted ones.	17
10	To forecast water consumption for Melbourne, Australia a time series model is formulated which relies on past demand data.	Trend, seasonality, climatic correlation, autocorrelation	The methodology adopted was based on time series analysis in which daily water consumption was taken as the sum of base consumption and seasonal consumption.	18

References

1. Antunes, A. Andrade-Campos, A. Sardinha-Lourenço and M. S. Oliveira (2018) "*Short-term water demand forecasting using machine learning techniques",* Journal of Hydro informatics, Volume 20, Issue 6.
2. Candelieri, F. Archetti (2014) "*Smart water in urban distribution networks: Limited financial capacity and big data analytics",* WIT Transaction on The Built Environment, Volume 139, Pages 1.
3. Adam Piasecki, Jakub Jurasz, Bartosz Kaźmierczak (2018) *"Forecasting Daily Water Consumption: a Case Study in Torun, Poland"*, Periodica Polytechnica Civil Engineering, Volume 62 No 3, pp. 818–824.
4. Ankit Kumar Nigam, Prof. D.C Rahi (2016) "*Analysis of Water Demand and Forecasting Water Demand for Year 2048 Jabalpur City"*, SSRG International Journal of Civil Engineering (SSRG – IJCE), Volume 3, Issue 7.
5. D. Savić, L. Vamvakeridou-Lyroudia, Z. Kapelan (2014) "*Smart meters, smart water, smart societies: The iWIDGET Project",* Procedia Engineering, Volume 89, Pages 1105-1112.
6. Emmanuel A. Donkor, Thomas A. Mazzuchi, Refik Soyer, J. Alan Roberson (2014) "*Urban Water Demand Forecasting: Review of Methods and Models",* Journal of Water Resources Planning and Management, Volume 140.
7. F. Keizo Odan, L. F. Ribeiro Reis, Z. Kapelan (2014) "*Use of Metamodels in Real-Time Operation of Water Distribution Systems"*, Procedia Engineering, Volume 89, Pages 449-456.
8. https://www.statcan.gc.ca/eng/start,https://en.wikipedia.org/wiki/Demographics_ofToronto
9. Iman Ghalehkhondabi, Ehsan Ardjmand, William A. YoungII, Gary R. Weckman (2017) "*Water demand forecasting: review of soft computing methods",* Springer International Publishing.
10. Julia K. Ambrosio, Bruno M. Brentan, Manuel Herrera, Edevar Luvizotto Jr., Lubienska Ribeiro, Joaquín Izquierdo (2019) "*Committee Machines for Hourly Water Demand Forecasting in Water Supply Systems",* Mathematical Problems in Engineering, Volume 2019, Article ID9765468, 11 pages, 8.
11. Junguo Liu, Hubert H.G. Savenije, Jianxin Xu (2013) *"Forecast of water demand in Weinan City in China using WDF-ANN model",* Physics and Chemistry of the Earth, Parts A/B/C, Volume 28, Issue4-5, Pages 219-224.
12. K.B. Khatri, K. Vairavamoorthy (2009) "*Water Demand Forecasting for the City of the Future against the Uncertainties and the Global Change Pressures: Case of Birmingham",* World Environmental and Water Resources Congress.
13. K.Thompson, R.Kadiyala (2014) "*Leveraging big data to improve water system operations",* Procedia Engineering, Volume 89, Pages 467-472.
14. K.Thompson, R.Kadiyala (2014) *Making water systems smarter using M2M Technology,* Procedia Engineering, Volume 89, Pages 437-443.
15. Khoi A. Nguyen, Rodney A. Stewart, Hong Zhang, Oz Sahin, Nilmini Siriwardene (2018) "*Re-engineering traditional urban water management practices with smart metering and informatics",* Environmental Modeling & Software, Volume 101, Pages 256-267.
16. Lily A. House-Peters, Heejun Chang (2011) "*Urban water demand modeling: Review of concepts, methods and organizing principles",* Water Resources Research, Volume 47.
17. S. Alvisi, M. Franchini (2014) "*Assessment of the predictive uncertainty within the framework of water demand forecasting by using the model conditional processor",* Procedia Engineering, Volume 89, Pages 893-900.
18. S.L. Zhou, T.A. Mc Mahon, A. Walton, J. Lewis (2000) "*Forecasting daily urban water demand: a case study of Melbourne",* Journal of Hydrology, Volume 236, Pages 153-164.