PROJECT REPORT

FOR

**SCOPE OF MACHINE LEARNING IN WATER CONSUMPTION AND POPULATION FORECASTING**



Submitted to: Submitted by:

Noopur Rathore (16ESKIT064)

Department of Information Technology

Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur

Rajasthan Technical University, Kota

TABLE OF CONTENTS

1. Chapter 1: Introduction………………...……………………………………………….....9-17
   1. Water……………………………………………………………………………………..9
   2. Machine Learning……………………………………………………………………..9-17
2. Chapter 2: Literature Review.……………………………………………………………18-20
3. Chapter 3: Objective………………………………………………………………………....21
4. Chapter 4: Workflow……………………………………………………………………..22-24
   1. Data……………………………………………………………………………………..22
   2. Software used…………………………………………………………………………..23
   3. Flow chart depicting analysis…………………………………………………………...24
5. Chapter 5: Analysis………………………………………………………………………25-39
   1. Data collection…………………………………………………………………………..25
   2. Data modification…………………………………………………………………....25-26
   3. Data visualization……………………………………………………………………26-31
   4. Data modeling……………………………………………………………………….31-39
      1. Water demand forecasting ward-wise………………………………………..31-34
      2. Water demand forecasting year-wise………………………………………...34-36
      3. Population forecasting………………………………………………………..36-37
      4. Water consumption l/c/d forecasting…………………………………………38-39
   5. Relationship graph………………………………………………………………………39
6. Chapter 6: Results and conclusion…………………………………………………………...40

References………………………………………………………………………………..41-43

**FIGURES**

1.1. Machine learning process……………………………………………………………………9

1.2.Types of machine learning…………………………………………………………………..10

1.3. Linear Regression……………………………………………………………………….......11

1.4.Artificial neural networks…………………………………………………………………...13

1.5.Activation function………………………………………………………………………….14

1.6.Feedforward ANN………………………………………………………………………......14

1.7.Feedback ANN………………………………………………………………………………15

1.8. Report flow…..……………………………………………………………………………...16

4.1. RStudio interface……………………………………………………………………………22

4.2. Flowchart depicting analysis………………………………………………………………...23

5.1. Code for data collection……………………………………………………………………..24

5.2. Code for data modification………………………………………………………………….25

5.3. Total Residential Usage by Ward and TotalCommercial

Usage by Ward (2011-2015)…...............................................................................................26

5.4. Average Residential Usage by Ward and Average Commercial

Usage by Ward (2011-2015)……………….………………………………………….........26

5.5. Total Water Usage…………………………………………………………………………..27

5.6. Average Residential Usage………………………………………………………………….27

5.7. Averge Commercial Usage………………………………………………………………….28

5.8. Toronto Total Residential Water Usage by Ward in 2015………………………………….28

5.9. Toronto Total Commercial Water Usage by Ward in 2015…………………………………29

5.10. Toronto Average Residential Water Usage by Ward in 2015……………………………..29

5.11. Toronto Average Commercial Water Usage by Ward in 2015…………………………….30

5.12. Total Consumption in Ward 1 with Total Count…………………………………………..31

5.13. Artficial Neural Network model for Ward 1 Prediction…………………………………32

5.14. Actual Water Consumption in Ward42…………………………………………………….32

5.15. Forecasted Water Consumption In Ward 42…………...…………………………………..33

5.16. Actual Water Consumption from 2005 to 2015……………………………………………34

5.17. Forecasted Water Consumption…………………………...……………………………….35

5.18. Population Trend of Toronto………………………………………………………………36

5.19. Forecasted Population of Toronto………………………………………………………..36

5.20.Water Consumption l/c/d trend of Toronto………….……………………………………37

5.21.Forecasted Water Consumption l/c/d trend of Toronto……………………………………38

5.22. Relationship between Forecasted Water Consumption and Forecasted Population….….38

**TABLES**

5.1. Ward 42 Water Consumption Forecasting…….…………………………………………….33

5.2. Water Consumption Forecasting Year-wise………………………………………………...34

5.3. Population Forecasting………………………………………………………………………36

5.4. Water Consumption l/c/d Forecasting………………………………………………………37

**ABBREVIATION**

AI: Artificial Intelligence

ANN: Artificial Neural Networks

ARIMA: Auto-Regressive Integrated Moving Average

BBLAWN: Battle of Background Leakage Assessment for Water Networks

CIS: Consumer Information System

CMMS: Computerized Maintenance Management System

EPANET: Environmental Protection Agency Network

GAMS: General Algebraic Modeling System

GIS: Geographic Information System

ICT: Information and Communications Technology

KNN: k-Nearest Neighbors

KPI: Key Performance Indicator

l/c/d: Litres per capita per day

LIMS: Laboratory Information Management System

M2M: Machine-to-Machine

MCP: Model Conditional Processor

ML: Machine Learning

MM: Metropolitana Milanese

MOODLE: Modular Object-Oriented Dynamic Learning Environment

NQT: Normal Quantile Transform

PRV: Pressure Reduction Valve

RMSE: Root Mean Square Error

SAGIS: Source Apportionment GIS system

SCADA: Supervisory Control and Data Acquisition

SSP: Successive Shortest Path

SVM: Support Vector Machine

SQL: Structured Query Language

TS: Time Series

UWOT: Urban Water Optioneering Tool

WDN: Water Distribution Networks

**ACKNOWLEDGEMENT**

On the very outset of this report, I would like to extend my sincere and heartfelt oligation towards all the personages who have helped me in this endeavor. Without their active guidance, help, cooperation and encouragement, I could not have made headway in the project.

First and foremost, I would like to express my deep sense of gratitude and indebtedness to my mentors Dr. M. P. Punia, Director, Department of Remote Sensing, Birla Institute of Scientific Research, Jaipur and Rajeshwari Acharya, JRF, Department of Remote Sensing, for providing their inspiring guidance, constructivecriticism and valuable suggestion throughout the course of the project.

I would like to express my special gratitude and thanks to my college faculties, SKIT, Jaipurwho have patiently helped me in accomplishing this undertaking. I would also like to express my gratitude towards my parents and members of BISR, Jaipur for their kind co-operation and encouragement which helped me in completion of this project.Last but not least, my sincere thanks to all my friends whohelped me a lot in finalizing this project within the limited time frame.

Thanking you

Noopur Rathore

**ABSTRACT**

Water is one of the crucial elements for the existence of life on earth. So it is our responsibility to conserve this life nectar.On Earth freshwater makes up a very small fraction of all water and thus water scarcity is faced.Nowadays water scarcity is commonly seen, it's due to the growing use of freshwater and thereby exhaustion of usable freshwater. Freshwater is broadly classified as groundwater and surface water. On Earth, the amount of freshwater has remained fairly constant over the period but the population has escalated. Therefore strive for freshwater intensifies day by day. For a nation like Canada water is of great significance, since water usage in Canada is high compared to Europe as water tariffs are low there. Though Canada is considered as a freshwater rich country and has less than 1% of the world's population still its rational use is very essential. This project is on Toronto, the provincial capital of Ontario, and the fourth most populous city in North America. Talking about the physical structure of Toronto, it covers an area of 630 km2, from the south it is bounded by Lake Ontario which is the 13th largest lake in the world and primary source of water for the city. Apart from Lake Ontario, Etobicoke Creek, Rouge River, Humber River, and Don River are some other adjacent water bodies to the city. After analyzing the situation this project focuses on the water consumption trend of Toronto using Machine learning algorithms i.e. artificial neural networks, linear regression, ARIMA modeling.Machine learning is the technology that provides machine the potential to learn without being comprehensively programmedand water consumption is basically water that is used and not returned to its source.Here, different forecast methods are implemented and therefore interrelationship between population and water consumption is assessed. The goal is to predict future water consumption and population and thereby serving the government in taking certain measures to avoid any future critical situation in advance as demand forecasting helps in resource management for the government. For this project, data includes some parameters like residential consumption, commercial consumption, water consumption litre per capita per day, etc. which influence the water management system. This project is built on language R and software RStudio Version 1.2.1335. Since for data analytics, it is one of the best-known languages.

**CHAPTER-1**

**INTRODUCTION**

**1.1 Water**

Water is the most required element that is surely a need for life on earth.It is one of the basic resources needed to sustain plant and animal life and ultimately human existence.So judicious use of this resource should be the concern of all.About 71% area of the earth is covered by water which is also referred as Hydrosphere of earth, in which 96.5% water is found in seas or oceans which are salted water and not useful for anyone. Also, 1.7% is in groundwater, 1.7% in glaciers and 0.001% in the air in the form of vapors or clouds. 2.5% of the total water on earth is fresh water in which 98.8% is ice and groundwater; thus, only around 1% is found as potable water. By seeing this calculation one can imagine how much water is available which is useful for us.

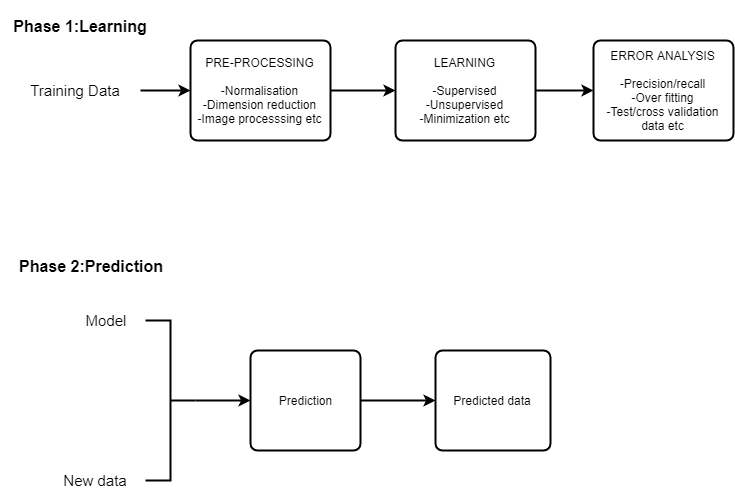
Water is extremely important for developed nations, such as Canada. Water is a basic resource which supports economic growth and maintains daily life. In light of environmental concerns and infrastructure costs, water management continues to be a priority for many Canadian municipalities. While Canada has one of the largest renewable supplies of water in the world, Canadians are also among the highest per capita water users.

Toronto is the provincial capital of Ontario and the most populous city in Canada. Toronto is a city built on a lake that contains about 1640 cubic kilometers of water. Lake Ontario is the city’s primary source of water and residents use nearly a billion litres each day, according to data from the city of Toronto.

So in this project, we estimate water usage in the city of Toronto by using different machine learning algorithms.

**1.2 Machine Learning**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs**that can access data and use it to learn for themselves.  As it is evident from the name, it gives the computer which makes it more similar to humans: The ability to learn. The process of machine learning is depicted in the figure below:

Figure 1.1. Machine learning process

**Broadly, there are 3 types of Machine Learning Algorithms:**

* **Supervised Learning:**

This algorithm consists of a target variable (or dependent variable) which is to be predicted from a given set of predictors (independent variables). Using these set of variables, we generate a function that map inputs to desired outputs. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of Supervised Learning: Regression, Decision Tree, Random Forest, KNN, Logistic Regression, etc.

* **Unsupervised Learning:**

In this algorithm, we do not have any target or outcome variable to predict / estimate.  It is used for clustering population in different groups, which is widely used for segmenting customers in different groups for specific intervention. Examples of Unsupervised Learning: Apriori algorithm, K-means.

* **Reinforcement Learning:**

Using this algorithm, the machine is trained to make specific decisions. It works this way: the machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions. Example of Reinforcement Learning: Markov Decision Process

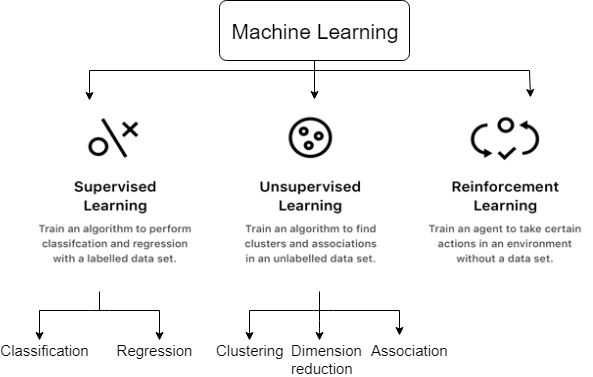


Figure 1.2. Types of machine learning

Most common machine learning algorithms include:

* **Naïve Bayes Classifier Algorithm (Supervised Learning – Classification)**
* **K Means Clustering Algorithm (Unsupervised Learning - Clustering)**
* **Support Vector Machine Algorithm (Supervised Learning - Classification)**
* **Linear Regression (Supervised Learning/Regression)**
* **Logistic Regression (Supervised learning – Classification)**
* **Artificial Neural Networks (Reinforcement Learning)**
* **Decision Trees (Supervised Learning – Classification/Regression)**
* **Random Forests (Supervised Learning – Classification/Regression)**
* **Nearest Neighbours  (Supervised Learning)**

**Among all of these linear regression, artificial neural networks and ARIMA (**AutoRegressive Integrated Moving Average**) model are majorly used in this project.**

* **Linear Regression:**

**Linear Regression** is a machine learning algorithm based on **supervised learning.** It performs a **regression task.** Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on the kind of relationship between the dependent and independent variables, they are considering and the number of independent variables being used.

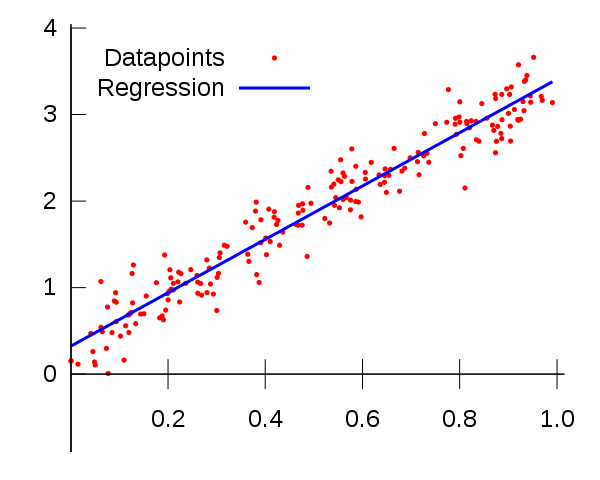
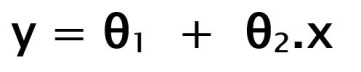


Figure 1.3. Linear Regression

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x(input) and y(output).Hence, the name is Linear Regression. **Hypothesis function for Linear Regression:**  


While training the model we are given:

**x:** input training data

**y:** labels to data

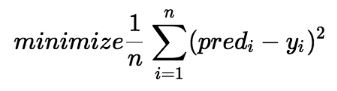
When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best θ1 and θ2 values.

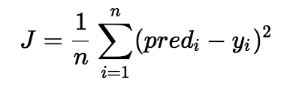
**θ1:** intercept  
**θ2:** coefficient of x

Once we find the best θ1 and θ2 values, we get the best fit line. So when we are finally using our model for prediction, it will predict the value of y for the input value of x.

**How to update θ1 and θ2 values to get the best fit line?**

**Cost Function (J):**  
By achieving the best-fit regression line, the model aims to predict y value such that the error difference between the predicted value and true value is minimum. So, it is very important to update the θ1 and θ2 values, to reach the best value that minimizes the error between predicted y value (pred) and true y value (y).





Cost function(J) of Linear Regression is the ***Root Mean Squared Error (RMSE)*** between predicted y value (pred) and true y value (y).

* **Artificial neural networks :**

Artificial Neural Network (ANN) is a computational non-linearmodel which is widely used in Machine Learning and is considered to be a prominent component of futuristic Artificial Intelligence.Artificial Neural Networks (ANN) are multi-layer fully-connected neural nets. They consist of an input layer, multiple hidden layers, and an output layer. Every node in one layer is connected to every other node in the next layer. The network is made deeper by increasing the number of hidden layers.

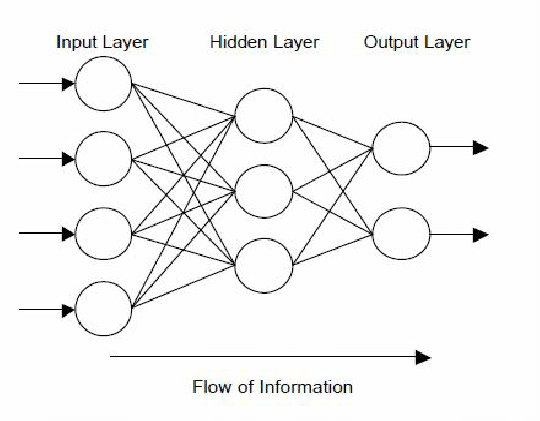


Figure 1.4. Artificial neural networks

**Elements of a neural network:**

**Input Layer:-**This layer accepts input features. It provides information from the outside world to the network, no computation is performed at this layer, and nodes here just pass on the information to the hidden layer.

**Hidden Layer:-**Nodes of this layer are not exposed to the outer world, they are the part of the abstraction provided by any neural network. Hidden layer performs all sort of computation on the features entered through the input layer and transfer the result to the output layer.  
**Output Layer:-**This layer brings up the information learned by the network to the outer world.

If one of the hidden or output nodes is zoomed in, what is encounter is the figure below. A given node takes the weighted sum of its inputs and passes it through a non-linear activation function. This is the output of the node, which then becomes the input of another node in the next layer. The signal flows from left to right, and the final output is calculated by performing this procedure for all the nodes. Training this deep neural network means learning the weights associated with all the edges.

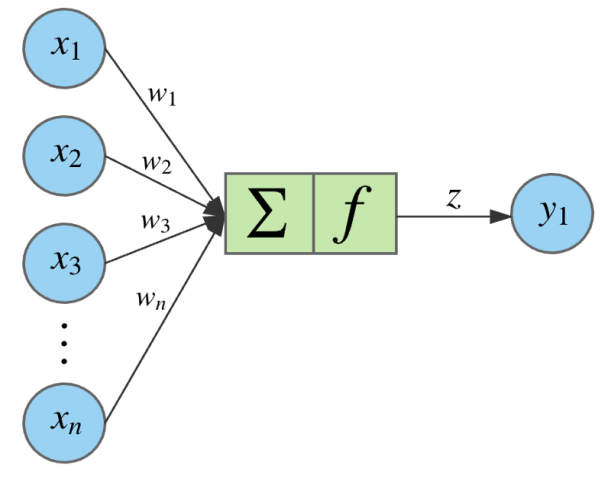
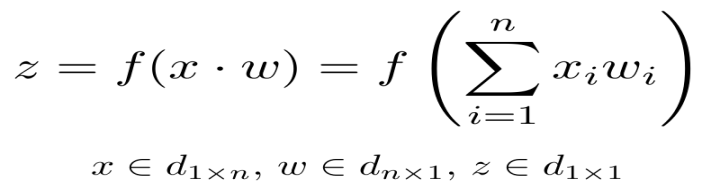


Figure 1.5. Activation Function

The equation for a given node looks as follows. The weighted sum of its inputs passed through a non-linear activation function. It can be represented as a vector dot product, where n is the number of inputs for the node.



**Types of Artificial Neural Networks:**

Generally, there are two types of ANN. Such as Feed Forward and Feedback.

**a.) Feed Forward ANN**

In this network flow of information is unidirectional. A unit used to send information to another unit that does not receive any information. Also, no feedback loops are present in this. Although, used in recognition of a pattern. As they contain fixed inputs and outputs.

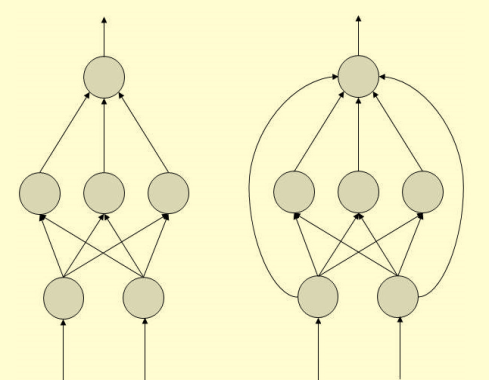


Figure 1.6. Feed Forward ANN

**b) FeedBack ANN**

In this particular Artificial Neural Network, it allows feedback loops. Also, used in content addressable memories.

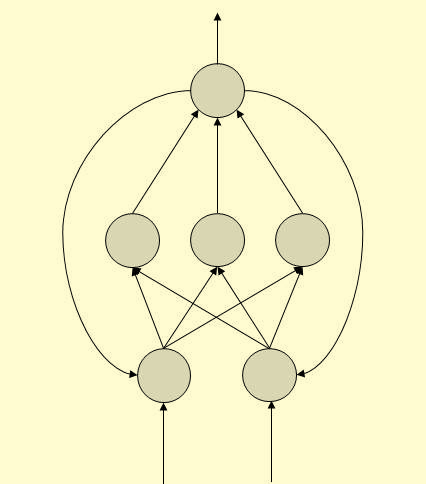


Figure 1.7.FeedBack ANN

* **ARIMA Model (Auto-Regressive Integrated Moving Average):**

An ARIMA Model is a class of statistical models for analyzing and forecasting time series data.ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

* **AR**: *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations.
* **I**: *Integrated*. The use of differencing of raw observations (e.g. subtracting an observation from observation at the previous time step) in order to make the time series stationary.
* **MA**: *Moving Average*. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components is explicitly specified in the model as a parameter. Standard notation is used of ARIMA(p,d,q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

* **p**: The number of lag observations included in the model, also called the lag order.
* **d**: The number of times that the raw observations are differenced also called the degree of differencing.
* **q**: The size of the moving average window, also called the order of moving average.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

*Note: Project report will be preceded in the following order:*

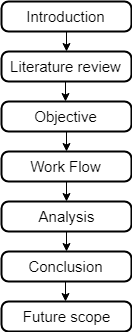


Figure1.8. Report flow

**CHAPTER-2**

**LITERATURE REVIEW**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Paper** | **Objective** | **Parameter** | **Methodology** | **Software used** | **Conclusion** |
| Making water systems smarter using M2M Technology | Improving water systems with the help of sensor technologies, firmware application, big data tools and Machine-to-machine(M2M) technology | water quality, flow, pressure | Improved sensor technologies and firmware applications, big data analytics tools and embedded Machine-to-Machine communications are creating new opportunities to collect data that have been manually collected and rarely evaluated. | M2M technology, SCADA system | M2M allows significantly more data to be gathered and analyzed, enabling better information to be provided to the human operators serving different functions at the utility. |
| An e-learning Approach for Improving Household Water Efficiency | Presents the development of a Moodle-based e-learning platform that aims to support end users to understand, manage and hopefully change their water and energy consumption profile. | Water calculator, water planner | The methodology involves structure and component of the platform, including, inter alia, FAQs, quizzes, and tips, more focus is given on the two most advanced web application for the exploration of domestic water demand profile. | Moodle (Modular object-oriented dynamic learning environment) -based web service termed as an e-learning platform, UWOT(urban water optioneering tool) developed using C whereas GUI was developed using Python 2.7 and PyQt4.8 | Paper presented a series of online tools and applications that aim to bring new awareness and increase endusers’ cognitions on household water practices. |
| Forecasting daily urban water demand: a case study of Melbourne | To forecast daily water consumption for Melbourne, Australia a time series model is formulated as a set of equations representing the effects of four factors on water use namely, trend, seasonality, climatic correlation, and autocorrelation. | trend,  seasonality, climatic correlation, autocorrelation | The methodology adopted was based on time series analysis in which daily water consumption is considered to be the sum of base consumption and seasonal consumption. | statistical software | Describes a computer-based mathematical model which relies on past demand data and weather forecasts to estimate consumption 24h ahead |
| Assessment of the predictive uncertainty within the framework of water demand forecasting by using the model conditional processor | An application of the Model Conditional Processor (MCP) is presented to assess the predictive uncertainty in water demand forecasting related to water distribution systems. | Water distribution system | The method entails converting historical observation and the corresponding forecasted values into a normal space using the Normal Quantile Transform (NQT). | Model conditional processor | The Model Conditional Processor enables to combine the (short-term) demand forecasts of two or more models and provides a probability distribution of the real future demand which is conditional on the values predicted by the individual forecasting models. |
| Water analytics and intelligent sensing for demand optimised management: The WISDOM Vision and Approach | The project aims to achieve a step change in water (and energy) savings via the integration of innovative Information and Communication Technologies (ICT) frameworks to optimize water distribution networks. | rainwater harvesting, grey-water recycling, wastewater treatment plants, internaland external water pipes | The technical design and implementation will adopt a system-of-systems approach that operates at three levels of abstraction : (a)Acquisition / Aggregation/ Actuation (b)Persistence, semantics and analytics (c)Services, and decision support | AQUASIM, ELODIE software | WISDOM will deliver an ICT platform that binds the base water production, distribution, and end-user consumption technologies together, augmented with smart ICT components for real-time management. |
| Multivariate statistical analysis for water demand modelling | This study has two objectives: (1) to propose a procedure based on a multivariate statistical analysis of the main features of the water consumption process at domestic level (2) to define a more realistic demand pattern with a given return period. | Pressure, flows, water quality, pipes friction coefficients, nodal demand, water consumption data | The method based on coupla recently introduced in hydrology is applied here | statistical software | The present paper proposed a statistical methodology for the definition of water consumption patterns based on the return period and multivariate probabilistic approach |
| Smart water in urban distribution networks: limited financial capacity and big data analytics | This paper presents the approach of combining time series clustering, for the identification of typical daily urban water demand patterns, and Support Vector Regression for performing a short term forecast. | The data considered in this study has been retrieved from the SCADA system of the WDN in Milan which is managed by Metropolitana Milanese (MM). | The approach consists of :(1)clustering together daily demand patterns, represented by the volume of water delivered at each hour,in order to identify the most typical patterns in consumption (2)aims at identifying prediction model, based on the Support Vector Regression | SCADA system | The approach presented in this paper, developed within the EU-FP7-ICT project ICeWater, proposes a predictive analytics solution for short term water demand forecast. |
| Leveraging big data to improve water system operations | The goal is to avoid inundating operators with voluminous raw data, instead of presenting solutions to automatically link and interpret data streams,instead of providing insight and actionable information to respond to a problem. | Big data in water and wastewater industries | Methodology opted here is data analytics. Data analytics, when applied to big data, examines large amounts of data of a variety of types to uncover hidden patterns, unknown correlations, and other useful information. | Computerized maintenance management system (CMMS), real-time data in a supervisory control and data acquisition (SCADA) system, regulatory data in a laboratory information management system (LIMS), customer complaint data from a consumer information system (CIS), geographical information system(GIS) map. | The use of big data techniques will be required to reduce the burden on staff, transform the data into information, and achieve the “Value of Now” to provide the maximum benefit to utilities. |
| Optimal Water System Operation Using Graph Theory Algorithms | The water system minimum-cost flow problem is solved using the successive shortest path (SSP), graph theory algorithm, by representing the network as a directed graph. The graph nodes represent water sources, junctions, tanks and consumers. The edges represent pipes, pumping stations water tanks. | Flow, water tank volume, source node, pump edge, junction node, water tank node, consumer demand node | The methodology includes: (a)examined water system (b)linear programming (c)directed graph (d) Successive shortest path algorithm | General algebraicmodeling system (GAMS) | A pump scheduling optimal operation problem was solved using the successive shortest path graph theory algorithm. The results were compared to the solution of the problem using linear programming and algebraic constraints. |
| A Comparison of Population-based Optimization Techniques for Water Distribution System Expansion and Operation | This paper presents a water distribution system expansion and operation methodology employing population-based optimization algorithms applied to the Battle of Background Leakage Assessment for Water Networks competition. | Population, cost, pipe diameter | The optimization software developed closely couples a number of population-based optimization techniques implemented in C++ with the EPANET2 hydraulic solver to model the effect on the performance of the hydraulic network when considering pipe replacement and duplication and the modification of pump and pressure reduction valve (PRV) operations. | C++ with the EPANET2 hydraulic solver | An optimization methodology for the Battle of Background Leakage Assessment for Water Networks(BBLAWN) problem has been formulated and solved. The BBLAWN leakage model has been directly incorporated into the EPANET hydraulic solver to maximize the efficiency of the leakage evaluation. |
| Leveraging Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector | This paper examines how the adoption of new Big Data tools and computational technologies can offer a great advantage to the water utility sector in addressing the environmental challenge. | Water, pollutants, chemicals, population, natural resources | This paper examines the challenge associated with the applicability of existing computational analytical approaches, applied to data sources, which may currently be limiting the value of the information that can be derived. | SAGIS (Source Apportionment GIS system), SCADA system, Apache software | The aim was to gain a better understanding of the sources, behavior and control measures used in rational decision making within the water industry. |
| Enhancing water system models by integrating big data | In this paper, a generic framework is presented to complete the data cycle for a water system. The data cycle presents an approach for integrating high-frequency data into existing water-related models and analyses while highlighting some of the more helpful data management tools. | Flow, pressure, land use, climate, population, pumps, tanks | A broad framework is proposed to enhance the current water computer models with a new API that enables near real-time dynamicmodelling and completes the data cycle. | Apache Spark, cloud computing, SQL, Hadoop, hydraulic model(EPANET) | The aim of this manuscript is to encourage the developmentand enhancement of water computer models by integrating bigdata. High-frequency data is collected from heterogeneous sources across environmental systems. |
| Analytics-driven water management system for Bangalore city | Introduction to a new water management software that is centered around powerful dash boarding, background analytics, management through exception and codifying standard operating procedures. | Water flow, real-time reporting on geospatial visual, population information | Making water management software that includes presentation layer, business layer, analytics layer, data transformation and storage and data acquisition with the help of a geospatial map, mobility interface, custom KPIs, database, SCADA, etc. | GSM technologies, SCADA server, web interface, mobile applications | Automation and water management for aging and complex water systems according to geography through the use of smart software. |
| Urban Water Management: Innovations and Paradigm Shifts to Address 21st Century Needs | The paper examines the status and evolution of innovations within urban water management (with particular reference to stormwater management) including impacts of regulation, barriers to innovation, current trends and paradigms. | Stormwater, sewer system, drainage channels, population climate, water quality, water quantity | Its methodology includes parameters like water quality, water quantity, sustainable water management, and soft and hard solution, along with comparative study in order to address current needs. | \_\_\_ | This study concluded that there are more effective and more sustainable approaches to the effectual draining of our increasingly urbanizing catchments than the conventional approach that has been the norm in the period leading up to the 21st century. |
| Water demand forecasting: review of soft computing methods | This study focuses on soft computing methods of water consumption forecasting published 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. | \_\_\_ | This paper has reviewed articles of water demand forecasting models from 2005 to 2015 with a focus on soft computing methodologies. |

*Note: After assessing research papers, language R and algorithm ANN found to be quite suitable and therefore are used in this model.*

**CHAPTER-3**

**OBJECTIVE**

This project focuses on the following objectives:

1. To assess the interrelationship between population and water consumption at ward level using language R.
2. To develop forecast methods using different machine learning algorithms (artificial neural networks, linear regression, ARIMA model) for population growth and water consumption.
3. To evaluate the applicability and accuracy of forecast models for the prediction of water consumption.

**Assumptions and Limitations**

1. Reports, data published by the government body are reliable.
2. Open source software may have used for validation.
3. It has been taken as a public service industry not as society reformation.
4. Limited access to data, so focuses on Toronto city.

**CHAPTER-4**

**WORKFLOW**

**4.1. Data**

There are basically two types of data used in this project of Toronto city, the provincial capital of Ontario, and the most populous city of Canada;

1. Data in tabular form i.e. excel sheets of water consumption from the year 2000 to 2015 and population census data from 2001 to 2016. Water consumption data was obtained from the site of *Statistics Canada*(https://www.statcan.gc.ca/eng/start) that is the official site by the Government of Canada,  government agency commissioned with producing statistics to help better understand Canada, its population, resources, economy, society, and culture. With this data, water demand forecasting is done. Whereas population census data was obtained by *Demographics of Toronto* (https://en.wikipedia.org/wiki/Demographics\_of\_Toronto). Here the data is according to 5-year census so it was converted to get per year population with the help of population estimation formula and then forecasted.
2. Data in raster form (grid data) i.e. shapefile of Toronto from *Statistics Canada* (https://www.statcan.gc.ca/eng/start) on which ward-wise water consumption was plotted and from that plot ward with highest and lowest water consumption can be easily drawn, therefore future steps can be taken accordingly.

Toronto city data has opted for this project as it is available on Canada site, though the same procedure can be applied on data of India access to that data is quite a tedious task i.e. by the survey since this was a time-bounded task so data of Toronto served the purpose.

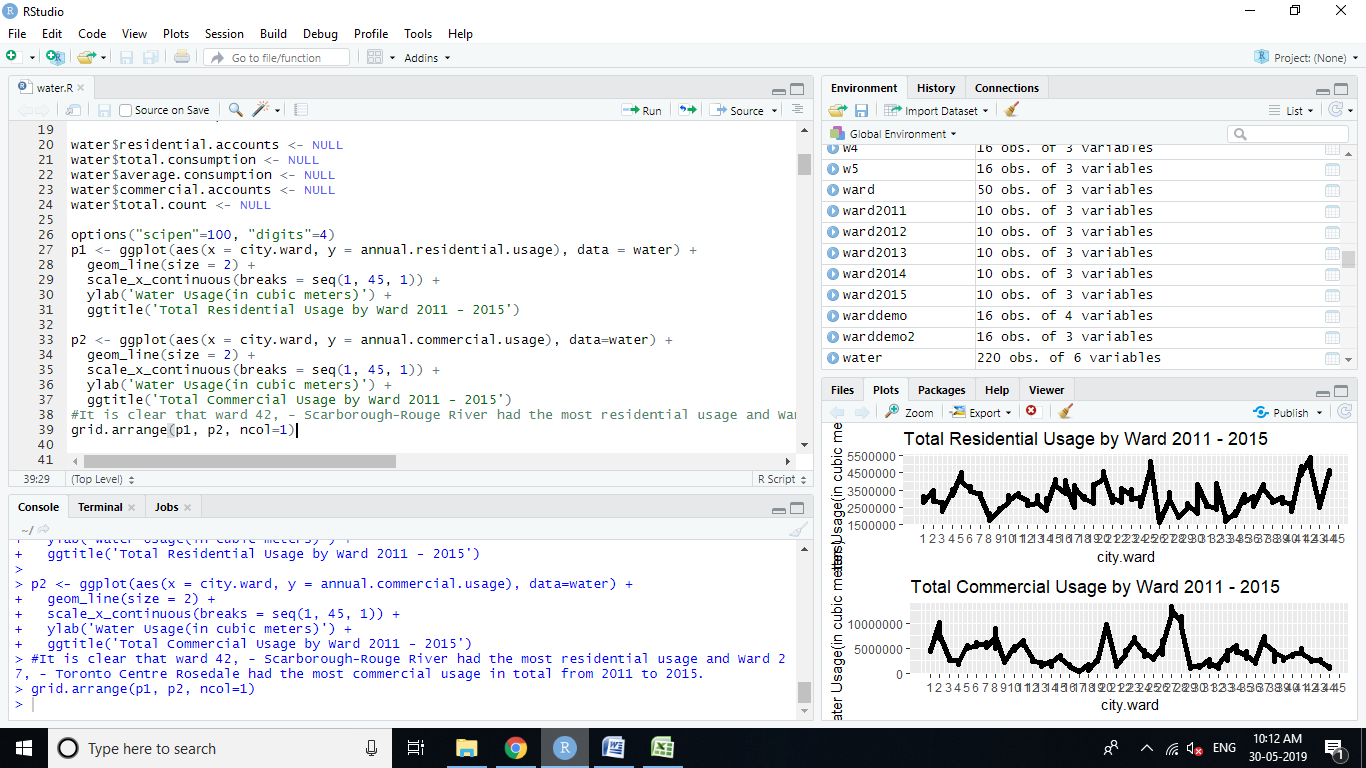
**4.2.Software used**

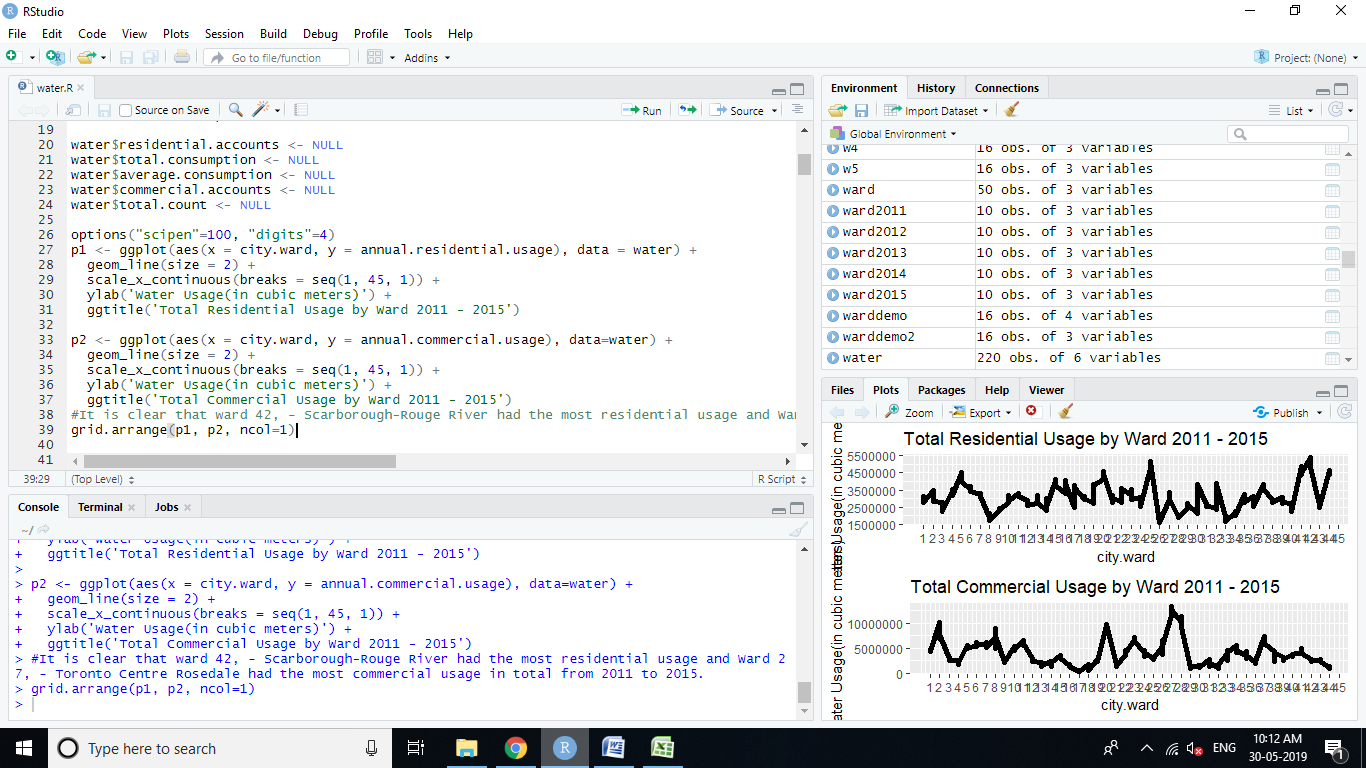
This project is majorly built using *Language R* and *Software RStudio*.

R is an open source language and software environment for statistical computing and graphics, supported by the R Foundation for Statistical Computing. Since R is mainly used for:Statistical inference, Data analysis, and Machine learning algorithm so R has opted for this project. Moreover, R possesses many more benefits:

1. The style of coding is quite easy.
2. It’s open source. No need to pay any subscription charges.
3. Availability of instant access to over 7800 packages customized for various computation tasks.
4. High-performance computing experience ( require packages)
5. One of highly sought skill by analytics and data science companies.

RStudio is a free, open source IDE (integrated development environment) for R. Its interface is organized so that the user can clearly view graphs, data tables, R code, and output all at the same time.

******

**R script**

**R environment**

**R console**

**Graphical Output**

**Figure 4.1.RStudio interface**

The interface of R Studio:

1. **R Console:** This area shows the output of the code you run. Also, you can directly write codes in console. The code entered directly in R console cannot be traced later. This is where R script comes to use.
2. **R Script:** As the name suggests, here you get space to write codes. To run those codes, simply select the line(s) of code and press Ctrl + Enter. Alternatively, you can click on the little ‘Run’ button located at the top right corner of R Script.
3. **R environment:** This space displays the set of external elements added. This includes data set, variables, vectors, functions, etc. To check if data has been loaded properly in R, always look at this area.
4. **Graphical Output:** This space displays the graphs created during exploratory data analysis. Not just graphs, you could select packages, seek help with embedded R’s official documentation.
   1. **Flow chart depicting analysis**

This project study includes the following steps:

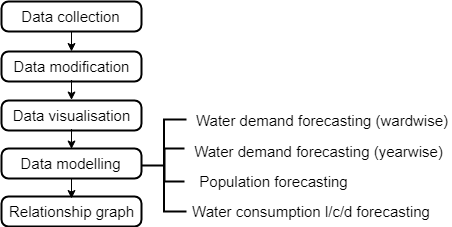


Figure 4.2. Flow chart

**CHAPTER-5**

**ANALYSIS**

**5.1. Data collection**

*Data collection* is the process of gathering and measuring information from different sources.

 In this project, data has been collected from different excel sheets (from the year 2011 to 2015) and all the parameters are combined into one variable named “water” that makes it feasible for processing.

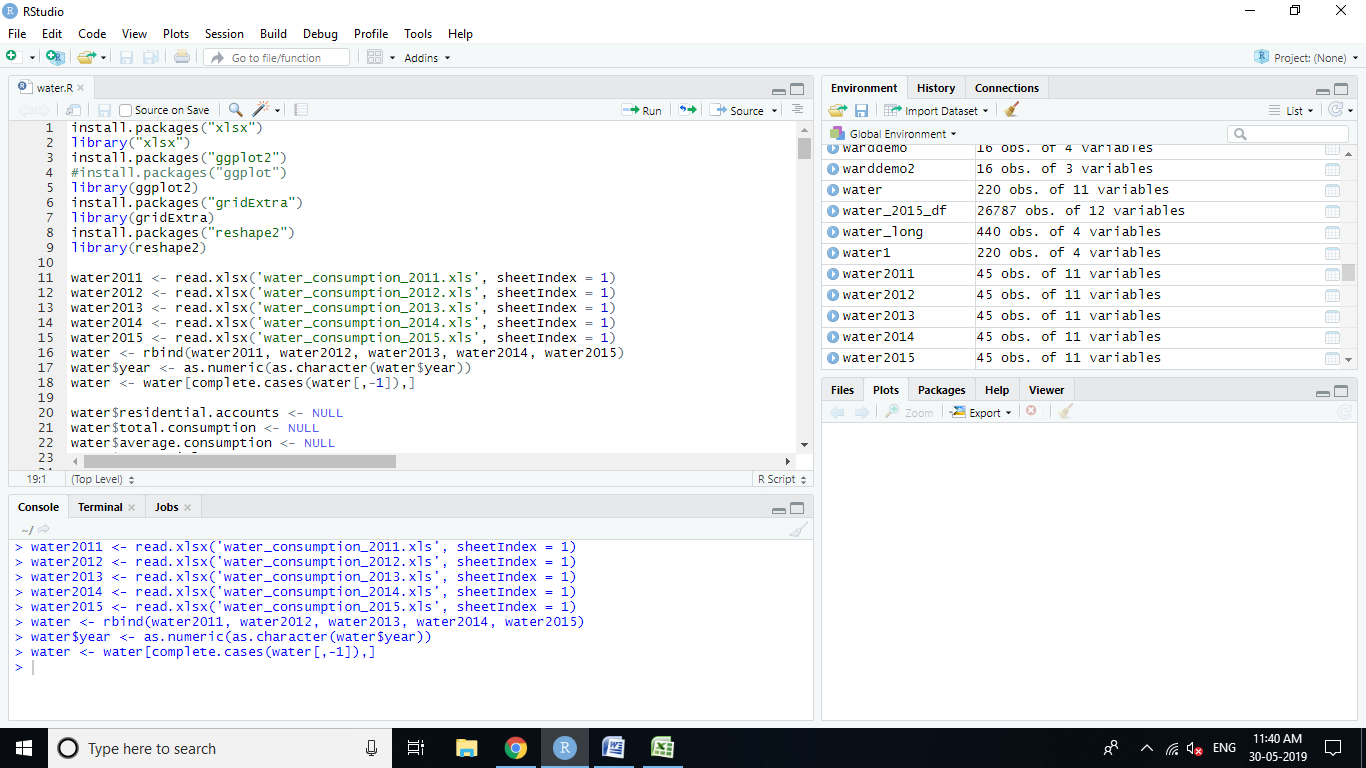


Figure 5.1.Code for Data collection

**5.2. Data modification**

*Data modification* includes converting the data in such a format that can be suitable to work with. It basically includes two steps:

**Formatting**: The data selected may not be in a format that can be suitable to work with. So to solve that purpose, here data has been formatted by combining all and changing it in numeric type.

**Cleaning**: Cleaning data is the removal or fixing of missing data. There may be data instances that are incomplete and do not carry the data so this problem needs to be addressed. These instances may need to be removed. For that “NULL” has been removed here.

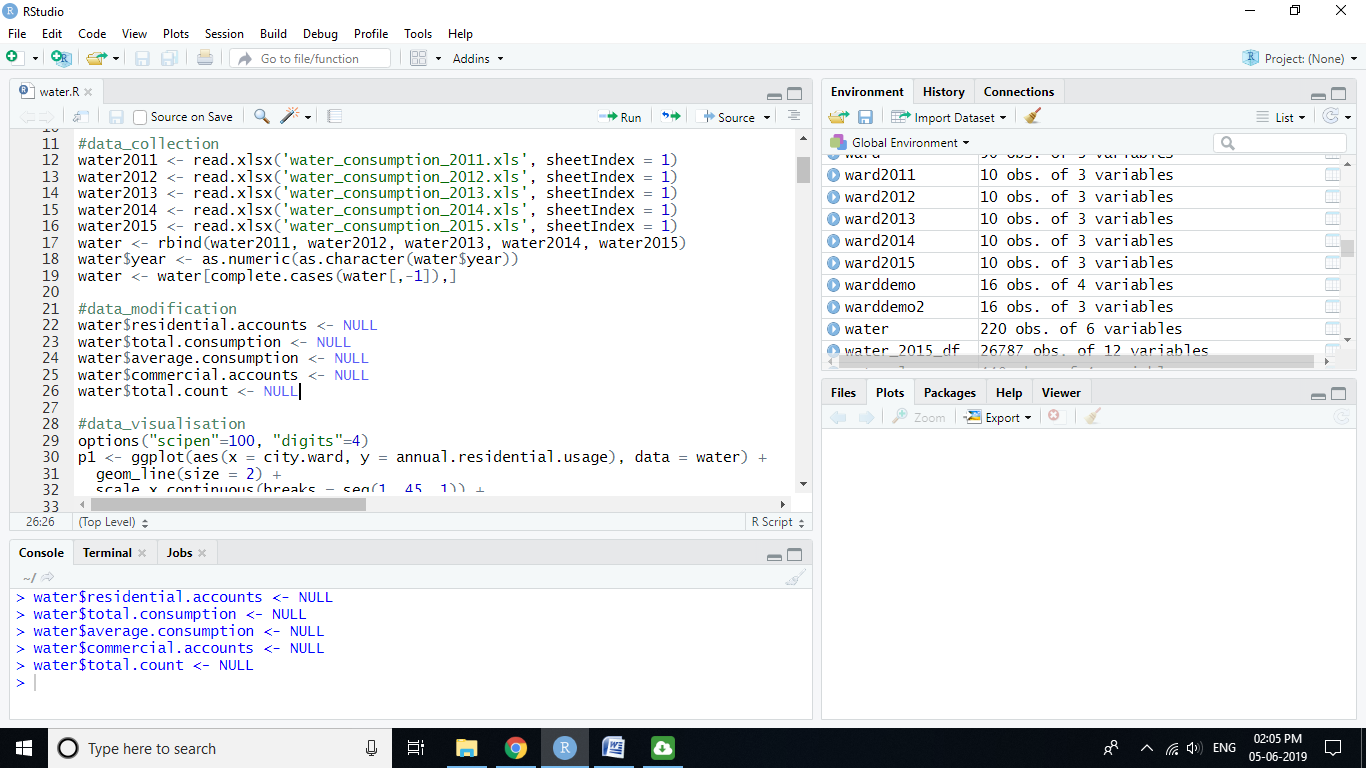


Figure 5.2. Code for Data Modification

**5.3. Data visualization**

*Data visualization,* wherein all the data is transformed into some form of plots and analyzed further from that. It is an important skill in applied statistics and machine learning as one can grasp a lot of information from diagrammatic representation than the counterparts.So this is helpful in exploring and knowing about datasets and moreover for identifying patterns, corrupt data, outliers, and much more.

Here total residential and total commercial usage by ward (2011-2015) is visualized and by that we concluded that ward 42, Scarborough-Rouge River had the most residential usage and Ward 27, Toronto Centre Rosedale had the most commercial usage in total from 2011 to 2015.

***C:\Users\Admin\Desktop\Final Graph\Graph1.tif***

Figure 5.3.Total Residential Usage by Ward and Total Commercial Usage by Ward (2011-2015)

Whereas when average residential and average commercial usage by ward (2011-2015) is visualized then we concluded that again,ward 27, Toronto Centre Rosedale had the most residential usage on average from 2011 to 2015, and ward 11, York South West consumed the most water commercially on average from 2011 to 2015.

***C:\Users\Admin\Desktop\Final Graph\Graph2.tif***

Figure 5.4.Average Residential Usage by Ward and Average Commercial Usage by Ward (2011-2015)

From the plot of total water usage, it is evident that consistentlyward 27 has consumed the most water commercially every year from 2011 to 2015.

C:\Users\Admin\Desktop\Final Graph\Graph3.tif

Figure 5.5.Total Water Usage

It is also proven by the average residential usage plot the ward 27 has consumed the most water every year from 2011 to 2015 except for 2015. In 2015 maximum water consumption was in ward 20 due to some reason.

C:\Users\Admin\Desktop\Final Graph\Graph5.tif

Figure 5.6.Average Residential Usage

By average commercial usage plot, we concluded that consistently ward 11 has consumed the most water every year from 2011 to 2015.

C:\Users\Admin\Desktop\Final Graph\Graph4.tif

Figure 5.7.Average Commercial Usage

When data of 2015 is plotted on Toronto mapward-wise then certain observations are found:

* In 2015 whentotal residential water usagewas considered then maximum water consumption was seen in ward 25 i.e.50,87,109.21cubic meters and minimum in ward 26 i.e. 16,15,814.12 cubic meters.

Graph9.tif

Figure 5.8.Toronto Total Residential Water Usage by Ward in 2015

* In 2015, total commercial water usage was maximum in ward 27 i.e. 1,27,54,138.95cubic meters and minimum in ward 17 i.e.4,85,127.01 cubic meters.

Graph8.tif

Figure 5.9.Toronto Total Commercial Water Usage by Ward in 2015

* In 2015 average residential water usage was maximum in ward 20 i.e. 533.83 cubic meters and minimum in ward 36 i.e. 212.40 cubic meters

Graph11.tif

Figure 5.10.Toronto Average Residential Water Usage by Ward 2015

* In 2015 average commercial water usage was maximum in ward 11 i.e. 41,515.02 cubic meters and minimum in ward 18 i.e. 7,177.30 cubic meters.

Graph10.tif

Figure 5.11. Toronto Average Commercial Water Usage by Ward 2015

Similarly, for each year, data can be plotted on a map and thus corresponding water consumption can be seen.

**5.4 Data modeling**

**5.4.1. Water demand forecasting ward-wise:**

1. **Water demand prediction using Artificial Neural Networks:**

Here,theartificial neural network has been used to predict water consumption in ward 1.Initially, data was collected and plotted as given below;

C:\Users\NOOPUR\Documents\Final Graph\Graph12.tif

Figure 5.12. Total Consumption in Ward 1 with Total Count

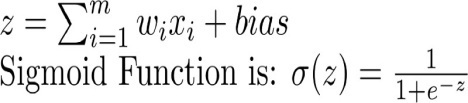
After collecting the data, data has been normalized using min-max normalization;

Normalized = (x-min(x))/max(x)-min(x))

Followed by data partitioning in training and testing datasets. Then the model was built using the following equation;

Total consumption w1 ~ Year + Total Count

It is evident by the plot that the model is quite accurate, as the error was just 0.06204 and total steps followed were 134 with a single hidden layer. Output on the last node i.e. node 4 was also shown using sigmoid function which was equal to the predicted value of total consumption.



C:\Users\NOOPUR\Documents\Final Graph\Graph13.tiff

Figure 5.13. Artificial Neural Network model for Ward 1 prediction

Likewise, water consumption can be predicted for any ward using artificial neural networks.

1. **Water demand forecasting using ARIMA model:**

Water demand forecasting is shown in ward 42 as it had the highest total residential usage from the year 2011-2015. Likewise for any ward water consumption can be forecasted.

After collecting the data for ward 42, data was converted into time series with an interval of 2 years starting from 2005 to 2015. Actual water consumption was depicted as:

C:\Users\NOOPUR\Documents\Final Graph\Graph16.tif

Figure 5.14. Actual Water Consumption in Ward42

Further when the ARIMA model was applied on time series data and forecasted for 5 years i.e. from 2017 to 2025 forecasted graph appeared withthe following values

**Table 5.1 Ward 42 Water Consumption Forecasting**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
| 2017 | 7089301 | 6937235 | 7241368 | 6856735 | 7321867 |
| 2019 | 6336995 | 6133399 | 6540591 | 6025622 | 6648368 |
| 2021 | 5806476 | 5414354 | 6198599 | 5206777 | 6406176 |
| 2023 | 4941011 | 4365915 | 5516107 | 4061477 | 5820544 |
| 2025 | 4214823 | 3331263 | 5098384 | 2863534 | 5566112 |

C:\Users\Admin\Desktop\Final Graph\Graph17.tif

Figure 5.15. Forecasted Water Consumption in Ward 42

5.4.2. Water demand forecasting year-wise using linear model:

Water demand consumption year-wise was forecasted using a linear model. Before that actual water consumption was plotted as

C:\Users\NOOPUR\Documents\Final Graph\Graph20.tif

Figure 5.16. Actual Water Consumption from 2005 to 2015

The equation used in the linear model was Total Consumption ~ Year. By this equation value of slope and intercept were calculated as -4581081.66 and 9535168992.29 respectively. Now using the function predict( ) forecasted values were

**Table5.2 Water Consumption Forecasting Year-wise**

|  |  |
| --- | --- |
| Year | Forecasted value |
| 2017 | 295127275 |
| 2019 | 285965112 |
| 2021 | 276802949 |
| 2023 | 267640785 |
| 2025 | 258478622 |

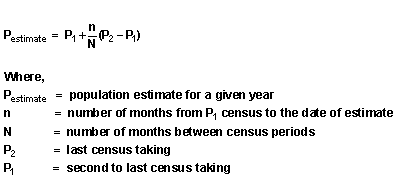
And the plot so obtained was

**C:\Users\NOOPUR\Documents\Final Graph\Graph21.tif**

Figure 5.17. Forecasted Water Consumption

**5.4.3. Population forecasting:**

Similar to water demand forecasting ward-wise, population forecasting was also done using the ARIMA model. Data initially was available with a gap of 5 years. By using the population estimation formula, population per year was calculated.



After reading the data from 2001-2015 it was then converted into time series using function ts() and then plotted

C:\Users\NOOPUR\Documents\Final Graph\Graph18.tif

Figure 5.18. Population Trend of Toronto

Further using the ARIMA model forecasted plot and values were

**Table 5.3 PopulationForecasting**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
| 2017 | 2754874 | 2743255 | 2766494 | 2737104 | 2772645 |
| 2019 | 2801480 | 2770941 | 2832019 | 2754774 | 2848186 |
| 2021 | 2848086 | 2792019 | 2904152 | 2762339 | 2933832 |
| 2023 | 2894691 | 2807539 | 2981844 | 2761403 | 3027979 |
| 2025 | 2941297 | 2818274 | 3064320 | 2753150 | 3129444 |

C:\Users\NOOPUR\Documents\Final Graph\Graph19.tif

Figure 5.19.Forecasted Population of Toronto

5.4.4. Water consumption l/c/d forecasting:

Firstly litres per capita per day (l/c/d) was calculated using residential water usage and population by the formula given below

Calculation Method:(Total water consumption during the assessment period \* 1000) / (number of days \* total number of served population)

Similar to the previous implementation of ARIMA model l/c/d was also forecasted i.e. first converting the model into time series then plotting it and then forecasting it

C:\Users\NOOPUR\Documents\Final Graph\Graph22.tif

Figure 5.20. Water Consumption l/c/d Trend of Toronto

Forecasted l/c/d plot and values were

**Table 5.4 Water Consumption l/c/d Forecasting**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
| 2016 | 133.6874 | 128.84086 | 138.5339 | 126.27527 | 141.0995 |
| 2017 | 136.4646 | 129.41711 | 143.5121 | 125.68639 | 147.2428 |
| 2018 | 140.2772 | 128.15588 | 152.3986 | 121.73923 | 158.8152 |
| 2019 | 142.3666 | 125.96465 | 158.7685 | 117.28200 | 167.4511 |
| 2020 | 139.4936 | 119.68566 | 159.3015 | 109.19999 | 169.7872 |
| 2021 | 134.0708 | 110.44326 | 157.6983 | 97.93561 | 170.2060 |
| 2022 | 131.2044 | 103.47404 | 158.9348 | 88.79447 | 173.6144 |
| 2023 | 132.3008 | 99.62729 | 164.9743 | 82.33097 | 182.2706 |

C:\Users\NOOPUR\Documents\Final Graph\Graph23.tif

Figure 5.21. Forecasted water consumption l/c/d trend of Toronto

**5.5 Relationship graph**

When the forecasted water consumption and the forecasted population were drawn on a single plot then a graph was obtained as shown below which depicted the relationship between the two. From this, certain conclusions were drawn which are mentioned in the next section.

C:\Users\NOOPUR\Documents\Final Graph\Graph24.tif

Figure 5.22.Relationship between Forecasted Water Consumption and Forecasted Population

**CHAPTER-6**

**RESULTS AND CONCLUSION**

From observations, we concluded that:

1. Maximum water consumption was found in ward 27 in past 5 years (both residential and commercial usage).
2. Litres per capita per day went down gradually from 170.750663 l/c/d in 2001 to 136.5072648 l/c/d in 2015 which is below the standard water consumption parameter i.e. 150 l/c/d, it signifies that water supply w.r.t to population is not up to the mark.
3. From the relationship plot between forecasted water consumption and forecasted population, it is observed that time around 2024 will be a very critical one as the threshold will arrive i.e. 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 condition.

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 project 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.

**REFERENCES**

1. A. Candelieri, F. Archetti, E. Messina*, Improving Leakage Management In Urban Water Distribution Networks Through Data Analytics And Hydraulic Simulation,* Water Resource Management, Volume 171, Pages 107-117, 2013.
2. A. Candelieri, F. Archetti, *Smart water in urban distribution networks: Limited financial capacity and big data analytics,* WIT Transaction on The Built Environment, Volume 139, Pages 11, 2014.
3. A. R. S.Diniz, J. G. DalfréFilho, *Physical Loss Evaluation In Water Distribution Systems In The Metropolitan Region Of São Paulo, Brazil,* Water Resource Management, Volume 171, Pages 11, 2013.
4. A.Knobloch, N.Guth, P.Klingel, *Automated Water balance Calculation for Water Distribution Systems*, Procedia Engineering, Volume 89, Pages 428-436, 2014.

## A.Merchant, M.S. Mohan Kumar, P.N.Ravindra, P.Vyas,U.Manohar, *Analytics-driven water management system for Bangalore city,* Procedia Engineering,Volume 70, Pages 1137-1146, 2014.

## A.Rodríguez, A.Cisneros*, Challenges and Benefits of an Open ICT Architecture for Urban Water Management,* Procedia Engineering,Volume 89, Pages 1073-1079, 2014.

1. A.Zarli, Y.Rezgui, D.Belziti, E.Duce, *Water analytics and intelligent sensing for demand optimized management: The WISDOM Vision and Approach,* Procedia Engineering, Volume 89, Pages 1050-1057, 2014.
2. B.J.Eck, E.Arandia, J.Naoum-Sawaya, F.Wirth, *A simulation-Optimization Approach for Reducing Background Leakage in Water Systems*, Procedia Engineering, Volume 89, Pages 59-68, 2014.
3. C.M.Fontanazza, V.Notaro, V.Puleo, G.Freni, *Multivariate statistical analysis for water demand modeling,* Procedia Engineering, Volume 89, Pages 901-908, 2014.
4. D.Páez, J.Saldarriaga, L.López, C.Salcedo, *Optimal Design of Water Distribution Systems with Pressure Driven Demands*, Procedia Engineering, Volume 89, Pages 839-847, 2014.
5. D.Savić, L.Vamvakeridou-Lyroudia, Z.Kapelan, *Smart meters, smart water, smart societies: The iWIDGET Project,* Procedia Engineering, Volume 89, Pages 1105-1112, 2014.
6. E.Price, A.Ostfeld, *Optimal Water System Operation Using Graph Theory Algorithms,* Procedia Engineering, Volume 89, Pages 502-508, 2014.

## Emmanuel A. Donkor, S.M.ASCE; Thomas A. Mazzuchi, RefikSoyer, and J. Alan Roberson, P.E., *Urban Water Demand Forecasting: Review of Methods and Models,* Journal of water resources planning and management, Volume 140, February 2014.

1. F. KeizoOdan, L. F. Ribeiro Reis, Z.Kapelan, *Use of Metamodels in Real-Time Operation of Water Distribution Systems*, Procedia Engineering, Volume 89, Pages 449-456, 2014.
2. F.Tscheikner-Gratl, R.Sitzenfrei, M.Hammerer, W.Rauch, M.Kleidorfer, *Prioritization of Rehabilitation Areas for Urban Water Infrastructure.A Case Study,* Procedia Engineering, Volume 89, Pages 811-816, 2014.
3. G.Anzaldi, E.Rubion, A.Corchero, R.Sanfeliu, X.Domingo, J.Pijuan, F.Tersa, *Towards an Enhanced Knowledge-based Decision Support System (DSS) for Integrated Water Resource Management (IWRM),* Procedia Engineering, Volume 89, Pages 1097-1104, 2014.
4. Iman Ghalehkhondabi, Ehsan Ardjmand, William A.YoungII, Gary R.Weckman, *Water demand forecasting: review of soft computing methods,* Springer International Publishing, 2017.
5. J.E. Van Zyl, *Theoretical modeling of pressure and leakage in water distribution systems,* Procedia Engineering, Volume 89, Pages 1420-1427, 2014.
6. Jason Levy, Ross Prizzia, *From Data Modeling to Algorithmic Modeling in the Big Data Era: Water Resources Security in the Asia-Pacific Region under Conditions of Climate Change,* Springer, Cham, 2018.
7. Juan Manuel Ponce Romero, Stephen H. Hallett, Simon Jude, *Leveraging Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector,* Sustainability, 23 November 2017.
8. K.Thompson, R.Kadiyala, *Leveraging big data to improve water system operations,* Procedia Engineering, Volume 89, Pages 467-472, 2014.
9. K.Thompson, R.Kadiyala, *Making water systems smarter using M2M Technology,* Procedia Engineering, Volume 89, Pages 437-443, 2014.z
10. KhoiA.Nguyen, Rodney A.Stewart, Hong Zhang, OzSahin, NilminiSiriwardene, *Re-engineering traditional urban water management practices with smart metering and informatics,*Environmental Modeling & Software, Volume 101, Pages 256-267, March 2018.
11. L.Tuhovčák, M.Tauš, T.Kučera, *The assessment of the technical condition of the water distribution systems,* Procedia Engineering, Volume 89, Pages 1420-1427, 2014.

## Lily A. House‐Peters, Heejun Chang, *Urban water demand modeling: Review of concepts, methods and organizing principles,* Water Resources Research,Volume 47, 28 May 2011.

1. M. EhsanShafiee, Zachary Barker, Amin Rasekh, *Enhancing water system models by integrating big data,* Sustainable Cities and Society, Volume 37, Pages 485-491, February 2018.
2. M.Deyi, J.vanZyl, M.Shepherd, *Applying the FAVAD Concept and Leakage Number to Real Networks: a Case Study in Kwadabeka, South Africa,* Procedia Engineering, Volume 89, Pages 1537-1544, 2014.
3. M.MazharRathore, Awais Ahmad, Anand Paul, Seungmin Rho, *Urban Planning and Building Smart Cities based on the Internet of Things using Big Data Analytics*, Computer Networks, Volume 101, Pages 63-80, 4 June 2016.
4. M.S.Morley,C.Tricarico, *A Comparison of Population-based Optimization Techniques for Water Distribution System Expansion and Operation*, Procedia Engineering, Volume 89, Pages 13-20, 2014.
5. P.Kossieris, A.Panayiotakis, K.Tzouka, P.Gerakopoulou, E.Rozos, C.Makropoulos, *An elearning Approach for Improving Household Water Efficiency,* Procedia Engineering, Volume 89, Pages 1113-1119, 2014.
6. Robert Y. G. Andoh, Michael G. Faram, KwabenaOsei*, Urban Water Management: Innovations and Paradigm Shifts to Address 21st Century Needs,* World Environmental & Water Resources Congress 2008, May 12-16, 2008.
7. S.Alvisi, M.Franchini, *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, 2014.
8. S.L.Zhou, T.A.McMahon, A.Walton, J.Lewis, *Forecasting daily urban water demand: a case study of Melbourne,* Journal of Hydrology, Volume 236, Pages 153-164, 30 September 2000.

|  |
| --- |
|  |