iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47797

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Water Demand Prediction Using Machine Learning

Mr. Rohit Patil[1], Ms. Priyadarshani Alandikar[2], Mr. Vaibhav Chaudhari[3], Ms. Pradnya Patil[4], Prof. Swarupa Deshpande[5]
[1, 2, 3, 4]*Student,* [5]*Project Mentor and Guide, Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, India*

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, the amount of freshwater has remained persistent over span but the population has mushroomed. Therefore, striving 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 machines 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 analyzing such data.***
Index Terms: Autoregressive Integrated Moving Average (ARIMA), Machine Learning, Time Series Analysis, Water Demand Prediction.

I. INTRODUCTION

Water is the basic source of life and an important natural economic resource. Water covers almost 70% of earth's surface, and it has been taken for granted that it will always be there for us, however, water shortage already affects multiple areas across different continents, according to the UNESCO recent study it's expected that by 2025, 1.8 billion people living in multiple areas will face severe water shortage, and about 33% of the world population may be under water stress conditions. The sustainability of economy & society development is to a large extent depending on rationalizing the utilization of water resources. For the last couple of decades desalination has become a vital alternative for water supply. It opens the door to tackle unconventional water resources that have great potential to provide sustainable water supply.

Desalination offers just about 1% of the world's drinking water, but this amount is rising year-on-year. State of Kuwait has a total area of 17,818 km2, Kuwait has a population of 4,62 million (2018), roughly 98 percent of Kuwait Metropolitan Area, 810 km2 or 4.5 percent of the total area, Kuwait is one of few countries in the world without rivers or natural lakes, Kuwait was entirely dependent on distillation plants for its freshwater supplies. For about 30 years multi-stage flash distillation plants have been used successfully in Kuwait.

The highest global water consumption per capita was recorded in Kuwait at 500 liters per person per day. As it requires significant energy consumption, desalinating seawater is more expensive than other natural resources as in groundwater or rivers, on the other hand water recycling and water conservation costs $1.09 to $2.49 per thousand gallons, the water demand Forecasting reducing capture, treatment, and distribution costs. Water demand forecasts allowed the Water Distribution Network to minimize energy consumption by 3.1% meanwhile reducing energy costs by 5.2% Different machine learning methods have been widely used lately in the implementation of effective short-term water demand forecasting such as neural networks, support vector machines, k-nearest neighbors and random forests which helps the operators of water distribution systems make decisions about pumping schedules, storage, treatment, and water distribution.

II. MOTIVATION

Water resource management tools include simulation, optimization and multi-objective analysis. The question is how to design simulation tools to support water decision making satisfying multiplicity of goals including multi-objective decisions. Computerized models were used for many years to support water related decision making and water resource management. A different approach is proposed in this paper, in which models are replaced with real dynamically changing environments with all of their complex intricacies and societal dependencies. This idea has been successfully applied to development of autonomous robots that interact with dynamically changing environments and learn proper interactions without building environment models, but rather using the environment.

©IJRASET: All Rights are Reserved

The main objective of this is to suggest an integrated modeling framework that may assist with the time consuming and difficult tasks of decision making by water management practitioners and to harmonize economic uses of water resources.
Integrated and effective machine learning platform may help to build effective partnerships between modelers and practitioners in the development and application of water management models and observe them in handling simulated crisis situations. Motivated machine learning that provides seamless support for intelligent decision making processes in dynamically changing environments could be applied to consider alternative water management policies. This framework uses a goal creation approach in intelligence (EI)that motivates machines to develop into a useful research tool through active interaction with the real environment. It integrates modeling with planning, decision making, policy implementation and evaluation, using dynamic feedback from the field to modify models and decision making processes. The method adapts to changes in the environment conditions, and resistance to policy implementation, and human factors, showing robustness under uncertain parameters, imperfect data, and imperfect models.

III. RELATED WORKS

Water demand prediction is useful for encouraging water savings. In [1] authors discuss the great progress in predicting water consumption that has been noted in recent years as being caused by the intensive development and implementation of algorithms. methods that have been used so far are universal enough to be used for water consumption forecasts under all water operating conditions. [2] Water Demand Prediction Using Machine Learning Methods: (A Case Study). A prediction model combines economic and social factors with geographic information to forecast water demand. There are two main steps used in prediction modeling: explanatory variables selection and model building. [3] Water Demand Prediction Using Support Vector Machine Regression. Water demand methodology: Data analysis helps to understand the nature of the data which can be used for finding out more information from the historical data.
These data can be compared with real data for predicting. [4] Water demand forecasting by trend and harmonic analysis. Harmonic analysis: It is based on the assumption that a time series consists of sine and cosine waves with different frequencies.[5] Performance comparison of techniques for water demand forecasting. TSA: A time series analysis considers the data to have an internal structure around which the model will be developed. ARMA model and its variations are used in numerous forest tasks like water consumption, under water reserves, drought and analyze seasonal trends. [6] Forecasting surface water level fluctuations of lake Serwy by artificial neural networks and multiple linear regression. Several methods and statistical tools are used to forecast water level fluctuations.
One of the most commonly used is Auto Regressive Moving Average (ARMA). [7] An Intelligent Approach to Demand Forecasting. Time-Series Algorithms: The actual data is fed into the model consisting of the time-series algorithms. The algorithms which are used in the time-series algorithms are Autoregressive integrated moving average (ARIMA). [8] Forecasting of Industrial Water Demand Using Case-Based Reasoning. The industrial water demand accurately is crucial for sustainable water resource management.
Keywords: Industrial water demand; forecast; case-based reasoning; water resources management, Zhangye city. [9] Domestic water demand forecasting under different socioeconomic scenarios for a Central Himalayan watershed, India. Key Words: Water demand; Water demand forecasting; Water Policy; Socioeconomic scenarios; Consumption pattern. Factors that affect water demand are identified as population, population growth rate, per capita water demand and per capita water consumption growth rate.[10] Prediction of Future Agriculture Water Demand in Thailand Using Multi Bias Corrected Climate Models. An agricultural water demand estimation can be calculated from relevant input variables including effective rainfall, reference evapotranspiration, percolation, and cultivated areas.

IV. ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

The SDLC model applied in our system is the Waterfall Model. Waterfall model is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping in the phases. The Waterfall model is the earliest SDLC approach that was used for software development. The waterfall Model illustrates the software development process in a linear sequential flow. This means that any phase in the development process begins only if the previous phase is complete. In this waterfall model, the phases do not overlap. Waterfall approach was the first SDLC Model to be used widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate phases. In this Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

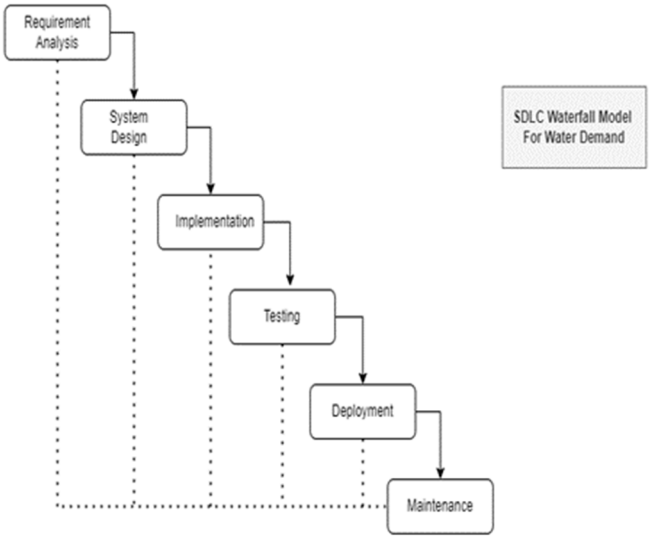


Figure: Waterfall Model

Sequential Phases:

1) *Requirement Analysis:* All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.
2) *System Design:* The requirement specifications from the first phase are studied in this phase and the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.
3) *Implementation:* With inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.
4) *Integration and Testing:* All the units developed in the implementation phase are integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.
5) *Deployment of System:* Once the functional and non-functional testing is done; the product is deployed in the customer environment or released into the market.
6) *Maintenance:* There are some issues which come up in the client environment. To fix those issues, patches are released. Also, to enhance the product some better versions are released. Maintenance is done to deliver these changes in the customer environment.

V. ALGORITHM ANALYSIS AND DIAGRAM

A. Arima Model (Autoregressive Integrated Moving Average)

ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

An ARIMA model is characterized by 3 terms: p, d, q where, p is the order of the AR term q is the order of the MA term d is the number of differencing required to make the time series stationary.

The most common approach to make a series stationary is to subtract the previous value from the current value. Sometimes more than one differentiation may be required, depending on the complexity of the series. Therefore, the value of d is the minimum amount of differentiation needed to render the sequence stationary, and if the time series is stationary already, then d = 0. The 'p' is the order of 'Auto-Regressive' (AR) terms; it refers to the number of Y lags which should be used as predictors. The 'q' is the order of the 'Moving Average' (MA) term; it refers to the number of lagged errors in the forecast that should go into the ARIMA model. The principal objective of the fitting ARIMA model is to correctly recognize the stochastic mechanism of the time series and forecast future values. Such approaches have also proven useful in other types of scenarios in which models for discrete-time series and dynamic systems are created.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

This method is, however, not appropriate for lead times or seasonal series with a broad random variable [3;4]. This univariate time series model has been used in the present study to forecast the water . The steps of the ARIMA model building methodology are presented in a flowchart below in Fig.

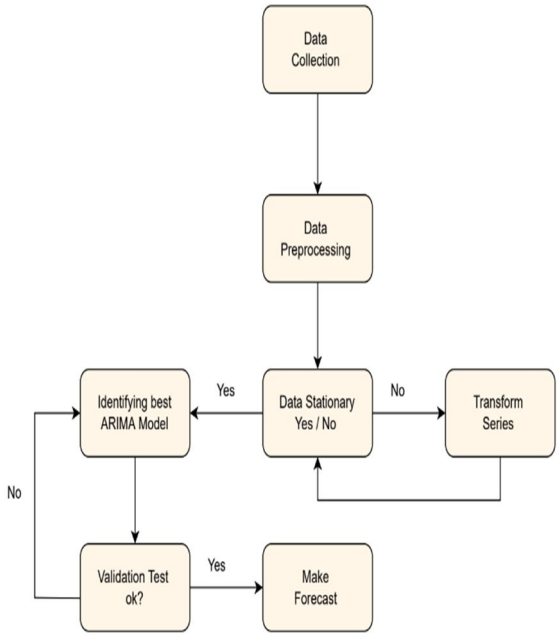


Figure: Algorithm Analysis Diagram

B. Multivariable Linear Regression

Multiple linear regression refers to a statistical technique that is used to predict the outcome of a variable based on the value of two or more variables. It is sometimes known simply as multiple regression, and it is an extension of linear regression. The variable that we want to predict is known as the dependent variable, while the variables we use to predict the value of the dependent variable are known as independent or explanatory variables.

Where,

$y = \beta o + \beta_1 X_1 + \ldots + \beta n X n + \varepsilon$

1) y is the dependent or predicted variable

2) β0 is the y-intercept, i.e., the value of y when both xi and x2 are 0.

3) β1 and β2 are the regression coefficients that represent the change in y relative to a one-unit change in xi1 and xi2, respectively.

4) βp is the slope coefficient for each independent variable

5) ϵ is the model's random error (residual) term.

To find the best-fit line for each independent variable, multiple linear regression calculates three things.

[1] The regression coefficients that lead to the smallest overall model error. [2]The t-statistic of the overall model. [3]The associated p-value (how likely it is that the t-statistic would have occurred by chance if the null hypothesis of no relationship between the independent and dependent variables was true). It then calculates the t-statistic and p-value for each regression coefficient in the model. *Assumptions of multiple linear regression:* Multiple linear regression makes all of the same assumptions as simple linear regression. *Homogeneity of variance (homoscedasticity)* : The size of the error in our prediction doesn't change significantly across the values of the independent variable. *Independence of observations :* The observations in the dataset were collected using statistically valid methods, and there are no hidden relationships among variables.

In multiple linear regression, it is possible that some of the independent variables are actually correlated with one another, so it is important to check these before developing the regression model. If two independent variables are too highly correlated (r2 > ~0.6), then only one of them should be used in the regression model. ***Normality*** The data follows a normal distribution. ***Linearity*** The line of best fit through the data points is a straight line, rather than a curve or some sort of grouping factor.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

VI. DATA ANALYSIS

Data analysis helps to understand the nature of the data which can be used for finding out more information from the historical data. These data can be compared with real data for predicting. The multidisciplinary field of data science combines the study of data using a statistical method, machine learning, and database technology. Leveraging these factors to build the system which can effectively predict and estimate the water demand is the key application that describes the methodologies adopted.
By obtaining the proper data from the model, the processed data can be analyzed for further estimations. The data model steps for prediction are as follows:

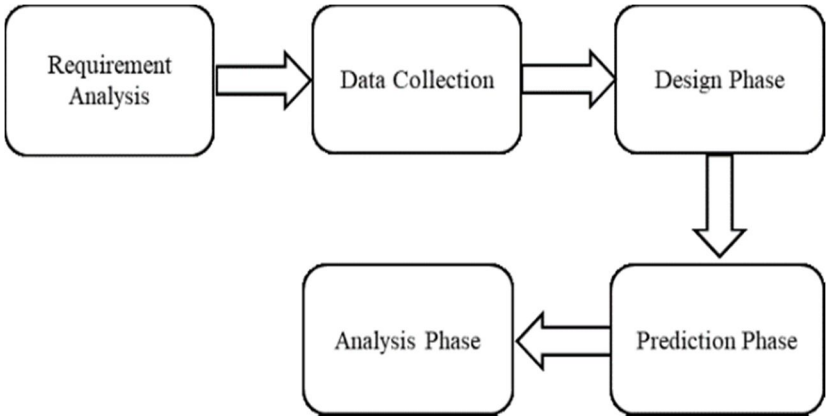


Figure: Data analysis

VII. SYSTEM ARCHITECTURE

Fig. 3.1 shows the system architecture diagram of water forecasting and prediction .In this architecture we use two methods [1] ARIMA model. [2] multivariable linear regression model. In this diagram we considered historic databases which contain year wise data like suppose it contain data from 1990 to 2021.using this database we can predict the future values like next year how much water is required for this prediction we used ARIMA model. Next we used a multivariable linear regression algorithm. Multivariable linear regression algorithms contain different types of prediction like industrial water prediction, domestic water prediction, garden water prediction , agriculture prediction are done. Multivariable linear regression is an algorithm which will predict multiple variables. After that all these prediction values are pre-processed and given to multivariable linear regression and it is predicted.

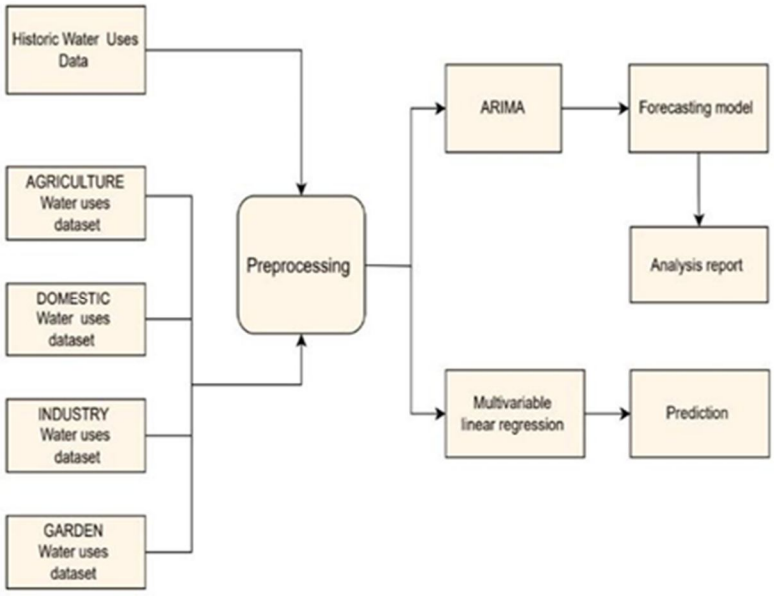


Figure: System Architecture

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

VIII. FUNCTIONAL MODEL AND DESCRIPTION

Functional Modeling gives the process perspective of the object-oriented analysis model and an overview of what the system is supposed to do. It defines the function of the internal processes in the system with the aid of Data Flow Diagrams.

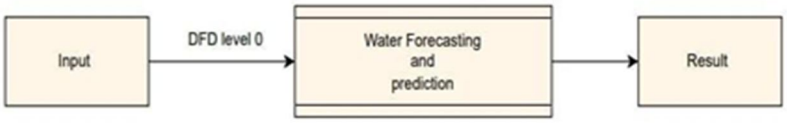


Figure: DFD Level 0

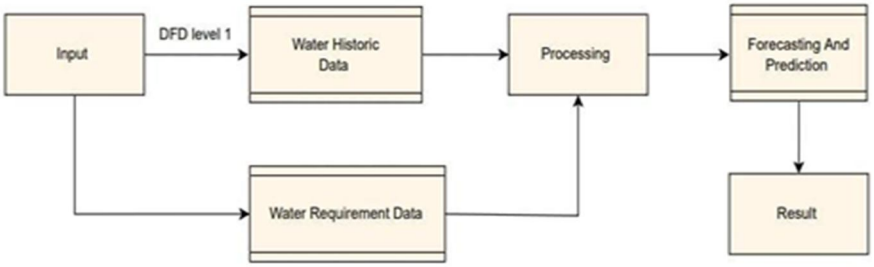


Figure: DFD Level 1

IX. RESULTS

On Earth, the amount of freshwater has remained persistent over span but the population has mushroomed. Therefore, striving 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 machines 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 analyzing such data.



Figure: Water Analysis Per Year

©IJRASET: All Rights are Reserved

X. ADVANTAGES & DISADVANTAGES

1) *Advantages:* [1] Although water managers use demand forecasting in a variety of ways, here, we focus on why demand forecasting is important for water rates. [2] To establish records of water distribution, wastage and consumption. [3] We can predict the water demand in a particular area.[4] Also alert when there is excess use of water and due to that there will be water shortage.[5] To identify and predict water availability through Machine learning.[6] Only requires the prior data of a time series to generalize the forecast. [7] Performs well on short term forecasts. [8] Models non-stationary time series.
2) *Disadvantages:* [1] New requirements can require new software or other project inputs. [2] Forecasts use only small no. of scenarios, there are some uncertainties considering hundreds of alternative representations of the future. [3] Difficult to predict turning points. [4] There is quite a bit of subjectivity involved in determining (p,d,q) order of the model. [5] Cannot be used for seasonal time series. [6] Power supply must be maintained in the monitor as high voltage may damage the sensors present inside the device. [7] Difficult to predict turning points. [8] There is quite a bit of subjectivity involved in determining (p,d,q) order of the model. [9] Cannot be used for seasonal time series.

XI. LIMITATIONS

[1] Monitor failure: The system will not work properly as a result access is not given to the user properly. [2] To avoid this risk we will make sure to learn all the new unfamiliar things in depth. Get deep knowledge of that thing before implementing it. [3] Back up plan to Avoid: Probabilistic approach consists in running models repeatedly using uncertain input variables randomly chosen from defined probability distribution. Sample their values based on randomized techniques. Calculate forecast for large no. of sample.

XII. FUTURE WORK

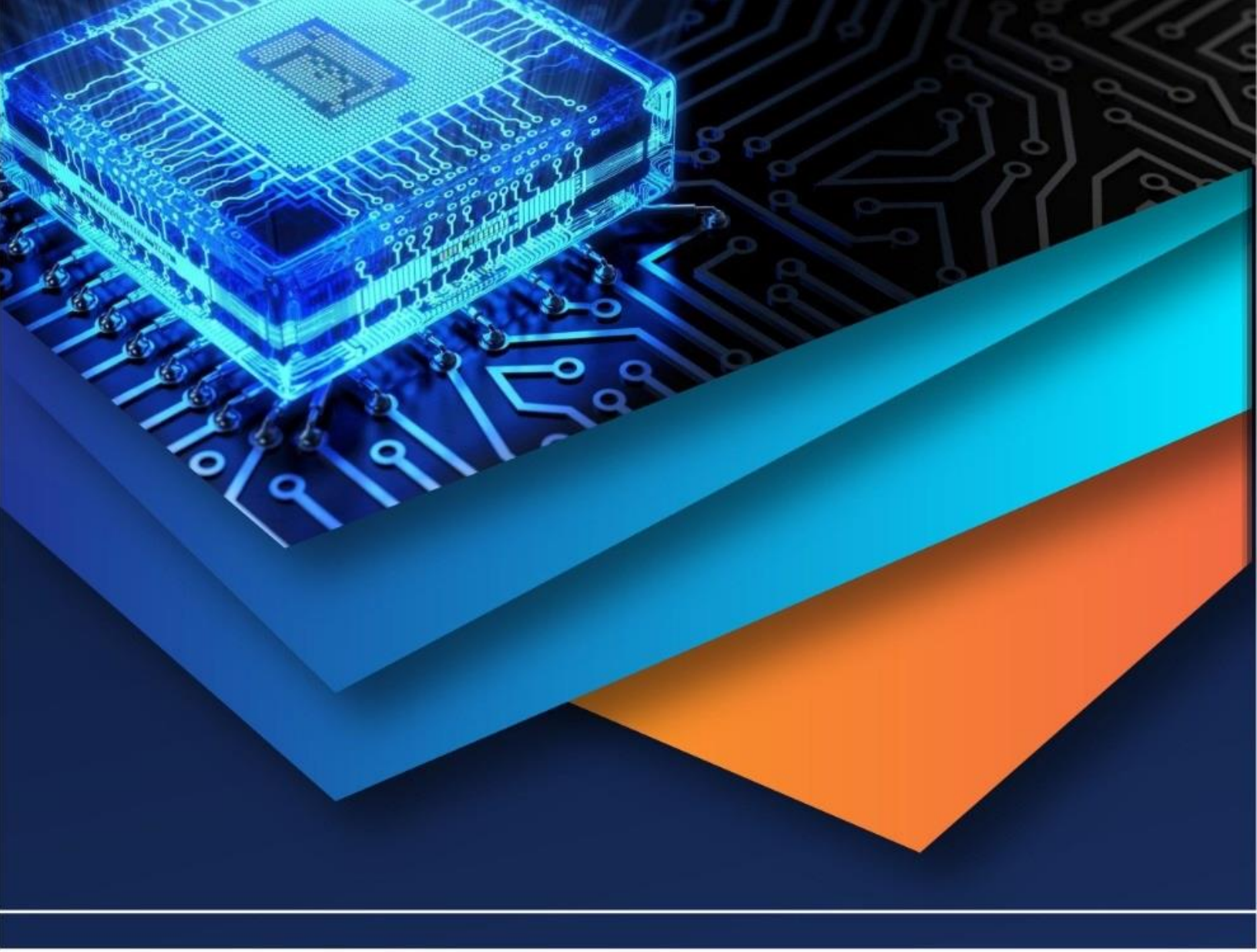
[1] In the future, we can predict the large amount of data and accurate water demand forecasting which is required for the city or area which has a large population. [2] In future, We require the various datasets that show how much water is provided in the Garden area, Irrigation Area and in Domestic Areas of particular Cities. [3] The System can be implemented in embedded processors such as raspberry PI. The System can be further extended to water forecasting like Alkaline water, Purified water, Distilled water

XIII. CONCLUSION

This paper attempted to model to predict the water demand using ARIMA Model related with the water usage pattern. Developing and implementing forecasting algorithms can result in a cost reduction of 18% or more. When compared with the other algorithms, the ARIMA model provides greater accuracy. This project introduces models for the acquisition of learning tools that can produce accurate predictions compared to traditional strategies. It was found to be reliable in the use of water data for real data, if there were no abnormalities of the data used during the training. The model which will predict the future value will help for water usability and Multiple linear regression will predict different parameters like industrial water prediction, domestic water prediction, garden water prediction.

REFERENCES

[1] Edward Kozłowski, Beata et al. (2018) Kowalska, Dariusz Kowalski, Dariusz Mazurkiewicz "Water demand forecasting by trend and harmonic analysis" in ELSEVIER.
[2] Adam Piasecki, Jakub Jurasz & Rajmund Skowron ‘et al. (2017) Forecasting surface water level fluctuations of lake Serwy (Northeastern Poland) by artificial neural networks and multiple linear regression, Journal of Environmental Engineering and Landscape Management, 25:4, 379-388
[3] Praveen Vijay, Bhagavathi Sivakumar(2018) "Performance comparison of techniques for water demand forecasting" in ELSEVIER,(ICACC)
[4] Qing Shuang and Rui Ting Zhao (2021) , "WaterDemand Prediction Using Machine Learning Methods: A CaseStudy of the Beijing–Tianjin–Hebei Region in China" MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations, Water 2021, 13, 310.
[5] Justyna Stańczyk, Jonna., Kajewska- Szkudlarek, J., Lipiński, P. et al.(2022) Improving short-term water demand forecasting using evolutionary algorithms. Sci Rep 12, 13522 .
[6] Sneh Joshi (2019) "Domestic water demand forecasting under different socioeconomic scenarios for a Central Himalayan watershed, India" 'Research Gate'.
[7] Amrita Tamang and S. Shukla,(2019) "Water Demand Prediction Using Support Vector Machine Regression," 2019 International Conference on Data Science and Communication (IconDSC), 2019, pp. 1-5, doi:10.1109/IconDSC.2019.8816969.
[8] Adhikari, N.C.D. et al. (2019). An Intelligent Approach to Demand Forecasting. In: Smys, S., Bestak, R., Chen, JZ., Kotuliak, I. (eds) International Conference on Computer Networks and Communication Technologies. Lecture Notes on Data Engineering and Communications Technologies, vol 15. Springer, Singapore.
[9] Bohan Yang ID, Weiwei Zheng and Xinli Ke et al. (2017) "Forecasting of Industrial Water Demand Using Case-Based Reasoning—A Case Study in Zhangye City, China” 2017 ‘Researchgate’.
[10] Winai Chaowiwat, Kanoksri et al. (2019) Sarinnapakorn, Sutat Weesakul" Prediction of Future Agriculture Water Demand in Thailand Using Multi Bias Corrected Climate Models.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)