Innovative Smart Water Management System Using Artificial Intelligence

Aamir Khambati[a], Keagen Pinto[b], Drumil Joshi[c], Dr. Sunil H Karamchandani[d]

[a,b,c,d]Department of Electronics and Telecommunication, Mumbai University, India

_____

Abstract: The research paper proposes an effective solution to the critical problem of management of water resources. With the growing awareness of the need for water conservation, the world is gearing to accommodate and implement latest technology for the optimum utilisation of the drying up reservoirs. This paper aims at developing a cost effective, portable and ready to plug mechanism labelled as "Innovative Smart Water Management System Using Artificial Intelligence" which shall monitor the proportion of water usage per household and keep a tracking on metric usage on the water usage on a weekly, monthly or a yearly basis. The data generated will be established and collected in the Firebase server. Based on the collected data, it shall also make predictions on the usage and hence allocate resources in a controlled fashion as per requirement. Later the data will be modelled in Time Series fashion to generate real time prediction of water consumption for the household respectively. The water consumer can track the usage by a custom-made android application on MIT App Inventor. Extensive, detailed instructions have been provided on the initial setup procedure and installation. The water monitoring module is used on a daily basis and a tab is kept on the amount of water spilled. This will keep a check on the casual approach of people towards water. It will also encourage a more systematic method to handle water resources and hence would result in better conservation efforts

Keywords: cost effective; metric usage; Time Series Modelling; MIT App Inventor; Firebase; water monitoring module

_____

1.   Introduction

Water is one of the most valuable resources available for evolution and the survival of the human race and will continue to do so forever. This is supported by historical evidence that most ancient civilizations thrived near a vast area bodies of freshwater such as rivers and lakes. It consists of growing population as well and as a result of the growing demand for water resources, water conservation is becoming increasingly important due to oppressive anxiety. This difficult situation is compounded by unsupervised and neglected exploitation of water resources already limited. That is why it is so important that we develop the reliable technology that ensures the efficient use of water and saves it for generations to come. The current system that is ubiquitous in buildings and houses is a complex pipe network, designed only to deliver water. The supply price is limited to the price of water available at local supplier dams. In other words, a lucky family found in an area with a lot of water cannot be stopped using water as long as the water is running away from their taps. This often leads to excessive use of water requirement. In addition, if such a family has a faulty faucet, or a person who has he said the house forgets that it leaves the tap on for a long time - the amount of damage that has been done here it will be much higher than a house that is already receiving a limited supply. In conclusion, the best way to tackle the situation of mismanagement of water is to limit the usage itself. The first and most crucial step in attaining this is to maintain a thorough check on the regular usage statistics. The approximate amount of the daily requirement can be estimated by keeping a log of the outgoing water from taps. Keeping this in mind, the Smart Water Flow Monitor module has been developed using data science methodologies. This paper introduces a review of relevant issues relating to the development and management of water resources in India. This will lead to public awareness of water conservation measures in order to create a sustainable environment for water conservation.

2.   Literature Survey

The OECD observed in 2012 [1][2] that smart water, despite being a part of water management in various forms for the past decade, has not been practically implemented with great success. It is not designed with an objective to replace existing services in operation, rather it improves them and therefore enables innovation rather than being the innovation itself. Four main users of smart water related information have emerged in recent years:

1.   Utilities seeking to optimise the efficiency of their operations (for example, lower distribution losses, less energy used, improved billing and maximising the operating lives of their assets.)

2.   Domestic customers seeking to manage their water and energy bills

_____

3.   Irrigators, seeking to lower the amount of water used and where appropriate, to optimise crop yields

4.   Those involved in monitoring and managing water in the natural environment, such assessing inland water quality and assessing flood vulnerability.

A lot of approaches have been reviewed in the past for the controlled consumption of water resources. The researchers in [3] has explained the design parameters of smart metering systems for a variety of domestic meter networks. Their findings can be suitably supplemented by the work of the researchers at [4] who described deployment of such meters in urban localities. Optimal network sensor placement, another important requirement in such complex systems, has been delineated by the developers at [5]. The use of machine learning and neural networks in the handling of smart water related data also has a few precedents [6], although it is a highly underexplored domain. Neural networks provide an immense amount of potential in database management, coupled with the rapid development of cutting-edge software and emerging technologies that are designed to optimise the existing firmware. This can potentially find application in several fields in modern times. Another important aspect of modern-day smart water management is smart leakage detection. A large amount of wastage can be minimised by providing alerts in case of undesirable drips and leakages. Leakage sensors have been developed and as explained by [7], the operational 4 life of these sensors can be maximised to great extents. Combining these projects in a smart metering system can be instrumental in conservation of water resources. A few innovative strategies based on the use of cutting-edge technology have been developed for the efficient management of water. Researchers at [8] have combined the use of Cloud Computing technology, Big Data Analysis and Internet of Things to improve the efficiency of management of water. The availability and preservation of the resources is ensured. They also implemented the use of a Smart Water Grid for the smart distribution of water. It is wise to maintain a Water Information System to keep track of all hydrological data so that it can be worked on using various technologies, as explained in [9]. Such data can be outsourced to third party software tool developers for handling data in a particular manner. Once the data is collected, filtered and standardized, it has to be stored or transmitted. Data loggers can be used to collect and store the data which in turn will transmit the data to the central system via direct or remote access. The processes of generating or capturing data however need to be managed and monitored and reference data and how they are shared within the information system has to be defined. But, above all, the target system should be fully aligned with the water business processes. The approach has been already promoted for example within the Smart Water Management Initiative taken by K-water. [10]

3.   Methodology

Every Problem in the world can be solved by an appropriate machine learning workflow. The aim of the project is to predict the water consumption of a building in a respective borough by machine learning as well as to predict thewater consumption for the next 10 days using the seasonality trend by means of time series analysis. The proposed algorithm for this is as follows:
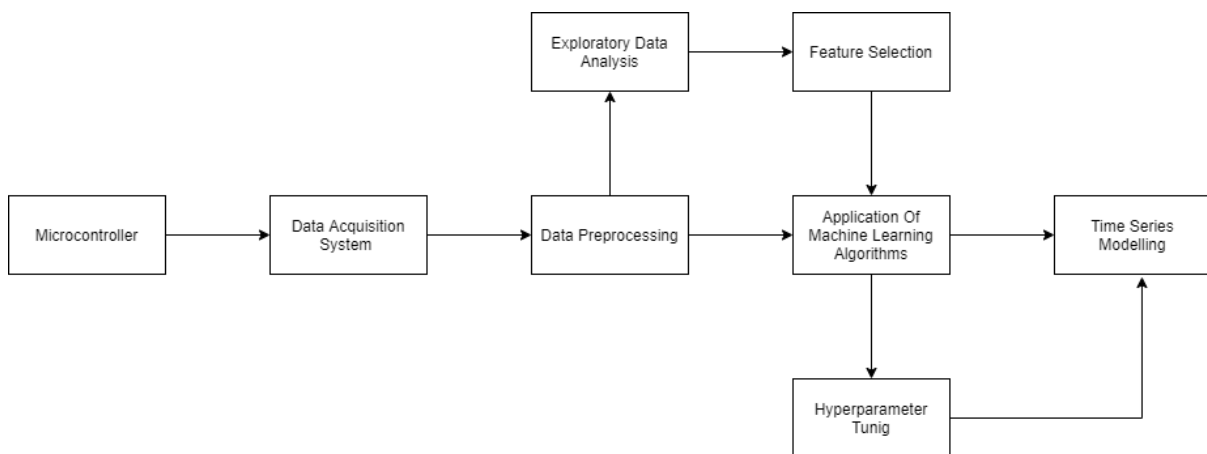

Fig. 1. Block Diagram Of Working Of Proposed System

Data Acquisition System

The system consists of the ESP32 microcontroller module at its core. Two water level sensors are used to detect the amount of water consumed and this data is communicated with the microcontroller. The solenoid valve is responsible for regulating the flow of water. It is controlled with the help of the relay module. Furthermore, a

low voltage dropout regulator (LM1117) is used to convert the 12V battery supply to a workable voltage of 5V necessitated by the water level sensors. This voltage is also applied to the ESP32 module. With the help of IoT, the ESP32 microcontroller sends commands to the cloud storage for given items. Cloud storage objects will receive commands and indicate the response. The program uses the developer of the MIT app to display sensor status. The proposed framework examines sensor information, such as temperature, gas, light, motion sensors, but also activates the process purposefully. For example, it alternates with the globule when it becomes dull. In addition, it stores the sensor information in the cloud. It will assist the client by knowing the different parameter regions at home anytime and anywhere.
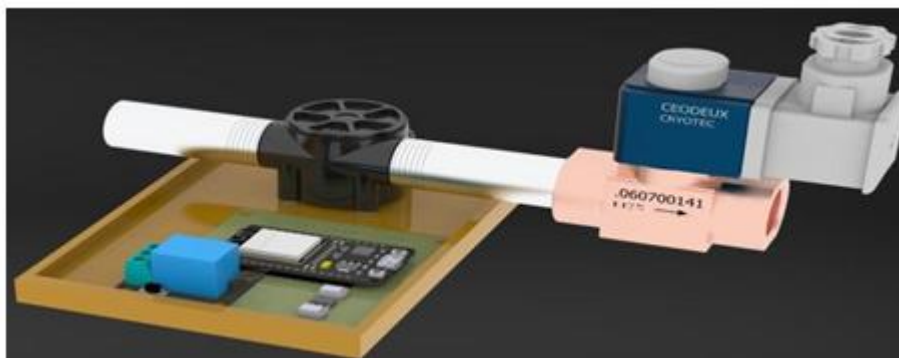

Fig. 2. 3D Model for Data Acquisition System

Data Preprocessing

The real-world data of any origin is very grimy in nature. The dataset used here was issued by the government of United States of America for research purposes. The environment for research was created by anaconda where all the essential libraries were installed. The data was available in the form of excel sheet where it was loaded in the environment using pandas library. When the data is loaded, there are 30 features which comprises of meter numbers, different types of charges and the consumption. The dataset comprises of different data types so it makes analysis difficult. Hence to overcome this problem we have to convert different string types into numerical format especially in float format. The first step of preprocessing consists of removing missing values. We remove columns having more than 50% missing values. The next step here is outlier detection and its removal. Any outlier can manipulate the results of the model developed, hence their removal is very necessary.

Exploratory Data Analysis

Exploratory Data Analysis is a truthful, open-ended process done to find the trends, anomalies and patterns from the given dataset. The role of matplotlib library plays an important role in analysis. Using matplotlib we plot histogram of water consumption which clearly shows the presence of outliers. After using the formula mentioned above we can see the binomial distribution of the consumption. The next step in EDA process is to plot density plot also known as smoothed histogram to get variations of consumption with respect to different boroughs as shown. The boroughs do not have significant impact on water consumption. The next step of Exploratory Data Analysis is to find correlations between the features of the dataset with the help of heatmap a feature of seaborn library.
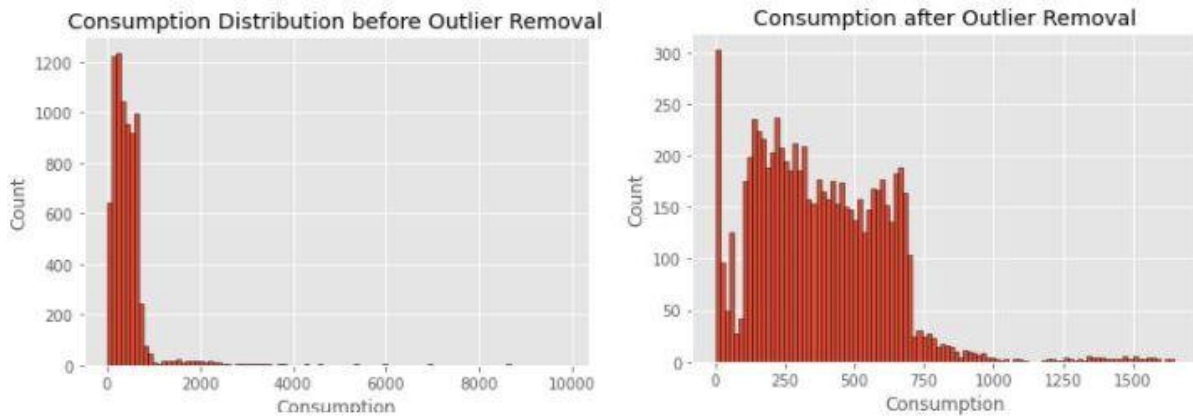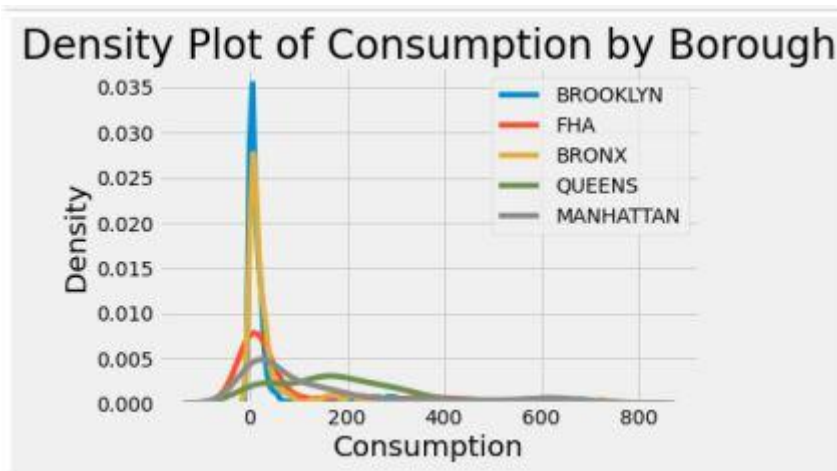
Fig. 3. The output disturbed by Outliers**.**

Fig. 4. Density Plot Of Consumption by Borough

Application of Machine Learning Models

The main aim of this research paper is to develop a machine learning model that has high accuracy and high model interpretability. The role of sklearn library plays an important role in reference of machine learning. The dataset consisting of 30,000 entries will be divided in the ratio of 70/30 in the form of train and test set. The first step for application of machine learning algorithms is to make the data compatible so that the machine could learn it efficiently. The main problem occurring with respect to large dataset is the test data leakage so to overcome this problem we impute all the nan values in train and test data set. The next step is to standardize the data which means to establish a range for the given dataset so that algorithms such as Random Forest and Support Vector Machine can gain maximum accuracy.

Standardization is done on numerical as well as categorical columns. In our research paper, we will be doing one hot encoding on categorical values and natural log transformations on numerical columns. After this process every feature in the dataset has a range of 0 to 1 and the dataset is ready for machine learning algorithms. The machine learning algorithms implemented here are Linear Regression, Lasso Regression, Elastic Net Regression, Decision Tree, Random Forest, Support vector machine and Gradient Boosting. The model algorithms always depend upon bias-variance trade-off which usually makes model overfit or underfit. The accuracy metric used

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

here which is Root Mean Square Error which is denoted by the formula as follows:

The primary algorithm implemented in this research paper is linear regression. The algorithm perfectly fits in the current scenario as we are regressing consumption on total charges. The linear regression implemented shows high model interpretability and low accuracy.The next algorithm implemented was lasso regression ,which helps in deals with multicollinearity. Multicollinearity is a problem where all the dependent variables are inter-linked with each other. Lasso regression is a type of regularization regression which helps in getting more accuracy but decreases the model interpretability. It increases the high biasness between the predictor variables.
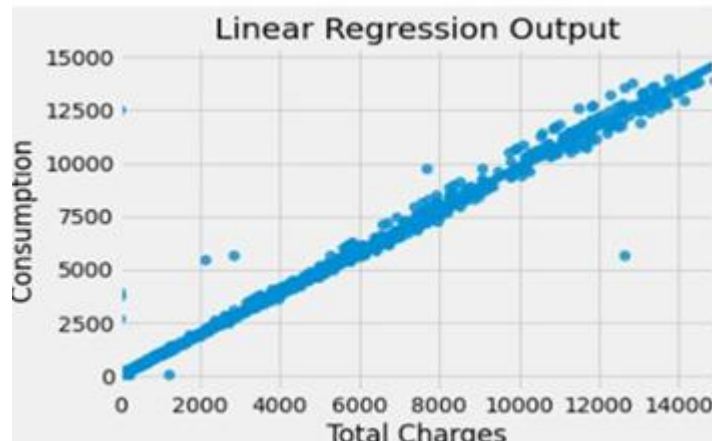

Fig. 5. Linear Regression Output

The third model implemented was Decision Tree regressor, the regressor identifies the features of the dataset and produces meaningful continuous output. It uses flowchart/tree like structure which has all possible outcomes. The Decision tree regressor has high model interpretability and low accuracy. It gets influenced by bias-variance tradeoff. It has less accuracy as compared to linear regression.
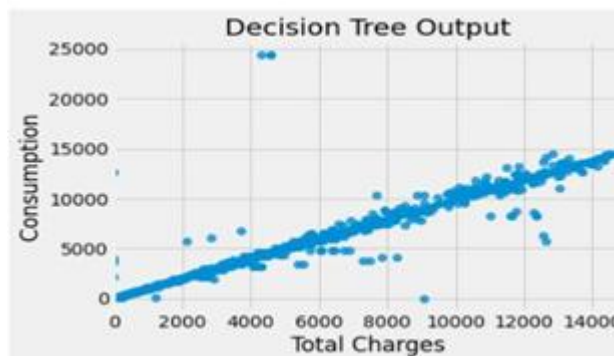

Fig. 6. Decision Tree Output

The fourth ML algorithm implemented was Random Forest. Random Forest algorithm is an ensemble learning algorithm. This algorithm is an additive model which ensembles all the decision tree models. The Random Forest also has high model interpretability, and is quite hampered by the bias-variance trade off. It is said that the more trees there are, the stronger the forest becomes. Random forests form decision-making trees from randomly selected data samples, find predictions for each tree and select the best solution for voting. It also provides a good indication of the importance of the feature. Unplanned forestsis considered the most accurate and powerful method due to the number of decision trees involved in this process. It does not suffer from overcrowding. The main reason is that it takes a measure of allthe predictions.
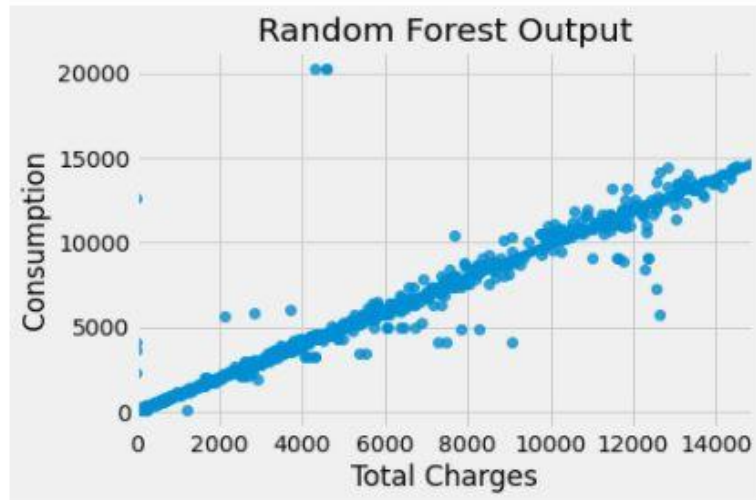
Fig. 7 Random Forest Output

The fifth algorithm implemented was Support Vector Machine Classifier. This algorithm doesn't give much accuracy as it undergoes formation of hyperplane. This algorithm finds two support vectors or corresponding points in the hyperplane and finds the euclidean distance between the next upcoming points. We find the appropriate hyperplane to separate the details.So by this, you should have understood that naturally, SVM can only do binary separation (i.e., choose between two classes). The constant $C > 0$ determines the relationship between f(flatness) and ε(deviations).
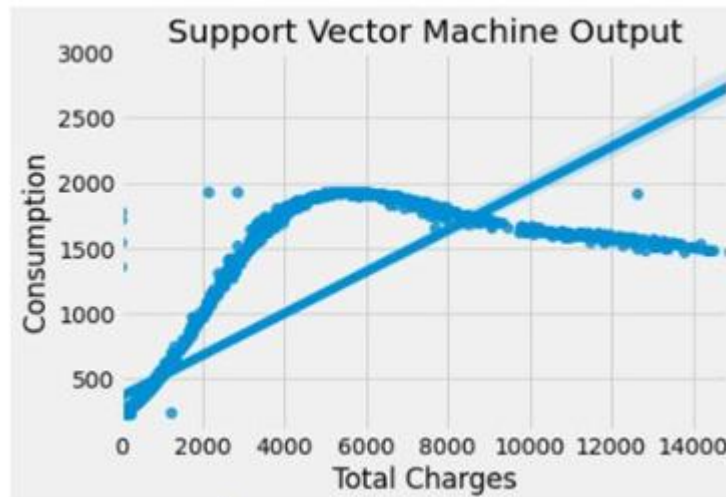

Fig. 8. Support Vector Machine Output

The last algorithm implemented is Gradient Boosting. Boosting builds models by turning

making weak features more resilient features in a repetitive manner. Increasingly, individual models are not built randomly on data subsets and features but in sequence by setting too much weight in cases with wrong assumptions too high errors. As compared to other boosting algorithms, gradient boosting algorithms prevents more data loss and gives more accuracy. This completes the application of machine learning part in the project.
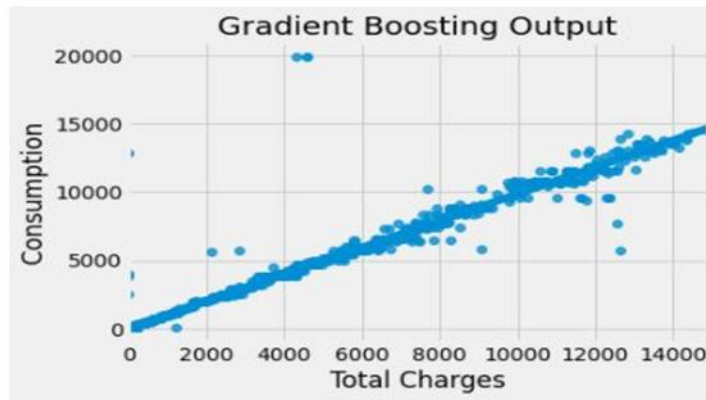
Fig. 8. Gradient Boosting Output

Hyperparameter Tuning for Model Optimization

Hyperparameters are the professors of the machine learning algorithms, which helps to train all the data and its features to gain more accuracy and model interpretability. As a coin has two sides even controlling the model hyperparameters can make the model perform overfitting or underfitting. When a model is underfitting it means it is undergoing high biasness and low variance that means the train data is controlling the features of the test data. In our research paper, we would be using Randomized search CV hyperparameter tuning concept where we will define a grid of all the features and then randomly sample different unknown combinations for better accuracy. As gradient boosting was the best performing machine learning model, we will be fine tuning the model hyperparameters.

Using Randomized Grid Search CV, we will be establishing 25 cross validations to get perfect accuracy and we will be substituting number of estimatiors as 800 to reduce overfitting. Random search generates random values in each hyperparameter tested, and then uses cross verification to check the accuracy of each compound. Random search is better than grid search because it can look at very different values for each hyperparameter. This is important because some hyperparamters are more important than others.
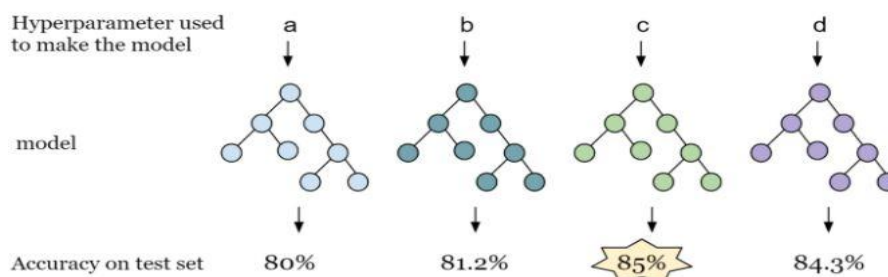

Fig. 10. Hyperparameter Tuning

Time Series Analysis

A series of time sequences of data in chronological order, each datapoint is assigned a specific point in time. The simplest example would be heat over time, with seasonal variations depending on the changing weather conditions. Flexible future predictions on these data sets, known as time series predictions, are an important goal of machine learning that aims to achieve.In our research paper we would be implementing univariate and multivariate analysis. We would be using time-stamps to produce the time series and predictions. We would decompose the data into three aspects which are trend, seasonal and irregualar. The main aim of the time series analysis is to make the entire distribution should not vary with respect to time. The given dataset continuously varies with time, so we have to do 1st order differencing on the main data.
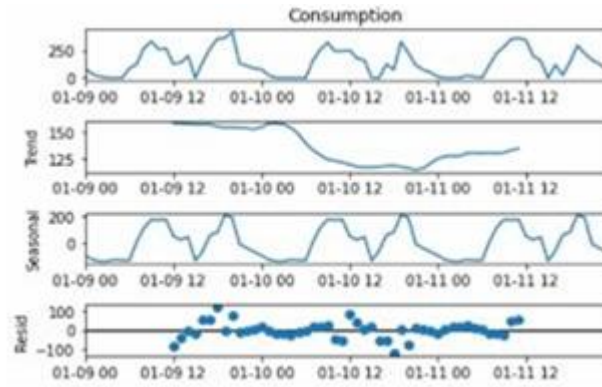
Fig. 11. Decomposition Of Analyzed Data

The concept of p-value comes in action when p-value is less than 0.005 , then that series is ready for time series which is done by Augmented Dickey Fuller Test. As the series is ready, we would be applying sarimax

$$\Delta P_t = c + \beta X + \phi_1 \Delta P_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$

model as the given distribution follows a certain seasonality. The formula for sarimax is as follows

We can use the Time Series to do a lot of research to predict the future such as circadian rhythms, annual trends, styles, changes, etc. The sarimax model is completely dependent on seasonality. The model developed has to be evaluated in new unforeseen conditions that is testing dataset. The accuracy metric used here is also root mean square error.

4.   Results and Discussions

The machine learning model deployed is very complex in nature and its interpretability and accuracy is explained by the rmse values. The bias-variance trade-off is the key aspect in the entire research paper. The residuals are followed in a binomial distribution. Every feature has its own relevance and contribution in the accuracy. The below horizontal bar graph depicts the machine learning algorithms and their accuracy scores.
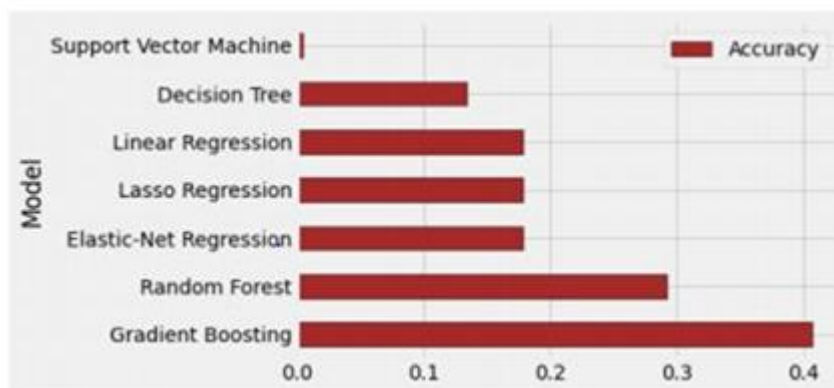

Fig. 12. Horizontal Bar Graph Showing The Accuracy Of ML Algorithms

The best performing algorithm is gradient boosting and the least performing algorithm is support vector machine. We are able to forecast the prediction of water consumption for next 5 days using the concept of seasonality and by implementing the sarimax model. The output of the model is shown as follows:
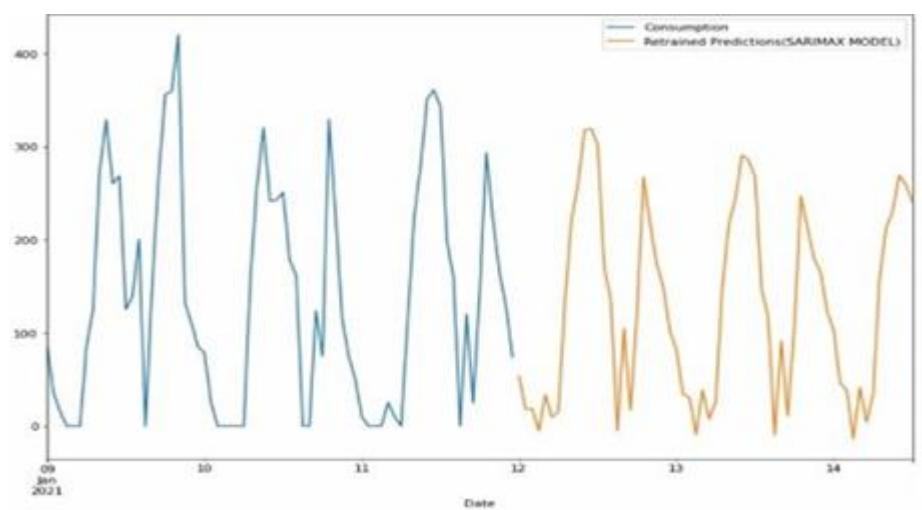
Fig. 15. Time Series Prediction Using SARIMAX Model

5. Conclusion and Future Scope

The proposed system warrants a well-maintained distribution system in line with the cuttingedge technology in the long run, since it is a low-cost project that works wirelessly. Data collection and further testing must be done consistently and the pattern of the water level, water flow and water distribution must be studied regularly to optimise the performance and capabilities for both the preliminary and field work. Furthermore, the implementation can be protracted to all tanks and distribution pipes and the data collected can be utilized for data analytics such as forecasting water consumption, flow rate, output and leakage detection. Since the water level sensors do not discriminate between substrate flowing through it as long as it is water-like (i.e., liquid or slurries), the system can prove to be useful to industries handling a variety of substances. One such potential application of the system lies in the food industry, where the system can ensure only measured quantities of water-like ingredients to be released into a food mixture. On similar terms, the system can be a great asset to chemical manufacture plants. Based on the required molarity or molality value of a solution, liquid chemicals or chemical slurries can be released as per the limit set..

References

OECD (2012) Policies to support smart water systems. Lessons learnt from countries experience. Background paper, ENV/EPOC/WPBWE(2012)6. OECD, Paris, France

OECD (2011), Infrastructure needs for the water sector. Background paper, ENV/EPOC/WPBWE(2011)14, OECD, Paris, France.

Savić D, Vamvakeridou-Lyroudia L, Kapelan Z (2014) Smart Meters, Smart Water, Smart Societies: The iWIDGET Project. Procedia Engineering 89 (2014) 1105 – 1112

Boyle T, Giurco D, et al (2013) Intelligent Metering for Urban Water: A Review. Water 2013, 5, 1052-1081; doi:10.3390/w5031052.

Christodoulou SE (2015) Smarting up water distribution networks with an entropy-based optimal sensor placement strategy. Journal of Smart Cities, 1(1): 47–58. doi.org/10.18063/JSC.2015.01.004.

Asnaashari A, McBean EA, Gharabaghi B, Tutt D (2013) Forecasting water main failure using artificial neural network modelling. Canadian Water Resources Journal, 2013, 38 (1) 24–33 doi.org/10.1080/07011784.2013.774153.

Sadeghioon AM, Metje N, Chapman DN, Anthony CJ (2014) SmartPipes: Smart Wireless Sensor Networks for Leak Detection in Water Pipelines J. Sens. Actuator Netw. 2014, 3, 64- 78; doi:10.3390/jsan3010064.

Abdeljebbar, Najat & Moussaid, Laila. (2019). Smart water management: Pillars and Technologies.

Gourbesville, Philippe & Ler, Lian. (2018). Framework Implementation for Smart Water Management. 3. 1139-1146. 10.29007/gb4g.

Choi, G. W., Chong, K. Y., Kim, S. J., & Ryu, T. S. (2016). SWMI: new paradigm of water resources management for SDGs. Smart Water (1:3).

Lloyd Owen, David. (2018). Towards Smart Water Management. 10.1002/9781119078678.ch10.

New York Water Detection Dataset, https://catalog.data.gov/dataset/waterconsumption-and-cost-2012-2016