MACHINE LEARNING BASED PREDICTION OF WATER LEVEL AND VOLUME IN FARM WELLS FOR SUSTAINABILITY

Kavin Santhosh G.
Final year B.Tech (EEE) Student
Department of EEE
B.S. Abdur Rahman Crescent Institute of Science and Technology
kavinsanthoshg@gmail.com

Roopesh S.
Final year B.Tech (EEE) Student
Department of EEE
B.S. Abdur Rahman Crescent Institute of Science and Technology
roopesh.crescent@gmail.com

Udhayashankar U.
Final year B.Tech (EEE) Student
Department of EEE
B.S. Abdur Rahman Crescent Institute of Science and Technology
udhayashankarumapathy@gmail.com

Shaik Mohammad Siddiq
Third year B.Tech (AI & DS) Student
Department of Computer Science
B.S. Abdur Rahman Crescent Institute of Science and Technology
mohammadsiddiqshaik@gmail.com

Dr. R. Hannah Lalitha
Assistant Professor, Department of EEE
B.S. Abdur Rahman Crescent Institute of Science and Technology
hannah.eee@crescent.education

***Abstract*—Water resources are crucial for agriculture, and wells located near agricultural fields are often used as water sources. However, there is significant temporal variability in water use patterns, which can make it difficult to predict future water availability. This study focuses on the analysis of water level and volume in wells located near agricultural fields, using water level and volume data collected over a one-year period. Daily water level and volume patterns were well described by autoregressive integrated moving average (ARIMA) models. Model development was supported by unsupervised clustering analysis that revealed similarities in water use patterns and confirmed the time-series water use model attributes. The inclusion of ambient temperature as a model attribute improved ARIMA model performance for daily water level and volume in a well. The study aims to provide insights into the behavior of the water system and enable better decision-making in agricultural operations that depend on this water source. The ARIMA model can help predict future water availability, which can be critical for planning and managing agricultural activities such as irrigation and crop selection. The study can contribute to more efficient and sustainable use of water resources in agriculture.**

I. INTRODUCTION

Water is an essential resource for agricultural operations, and crop production depends on its availability. Because it enables farmers to schedule irrigation and choose crops wisely, accurate water availability forecasting is crucial for sustainable farming methods. At the same time, the proposed methodology should be easy to use and cost-effective. In this study, we use high-resolution water use data gathered over a year to evaluate the water use patterns of wells close to agricultural fields. Our aim is to develop an accurate time-series model that can predict the daily water level and volume in the well, taking into account the effect of ambient temperature. We employ the Autoregressive Integrated Moving Average (ARIMA) model to capture the temporal variability in water usage patterns. We also use unsupervised clustering analysis via simple moving average (SMA), exponential moving average (EMA), and cumulative moving average (CMA) to identify similarities in water usage patterns and confirm the model attributes. The inclusion of ambient temperature as an exogenous variable in the ARIMA model improves the model's performance in predicting the daily water level and volume in the well. The results of this study can contribute to more efficient and sustainable use of water resources in agriculture, enabling better decision-making in agricultural operations that depend on this critical water source.

II. RELATED WORK

Mohammad Emami discusses the importance of measuring the water demand for irrigation. The authors proposed the Gaussian process regression model to predict the water demand for irrigation [1]. Iqbal Singh discusses the importance of monitoring the water level on a farm. The authors proposed a topology under the control of WSN for precision agriculture [2]. Fabrizio Balducci discusses the importance of forecasting the harvesting period of crops. The authors proposed a time-series model to forecast the harvesting period of crops [3].

III. EXISTING SYSTEM

In the existing irrigation water management system, the water usage, and the recharge rate of the farm well are not analyzed accurately, but instead relied on blind human estimations. This meant that the usage of water was either underestimated or overestimated. Both of which can be problematic because an overestimation of the water present in the farm well could lead to a plantation of crops which consumes more water than the recharge rate of the farm well which can deplete the well and cause drought in the farm field. An underestimation of the volume of water can lead to the planting of low-water-intensity crops rather than the crop of the farmer’s choice.

IV. PROPOSED SYSTEM

In the proposed system the water level, volume, and usage from the well are measured using ultrasonic and flow rate sensors respectively. Then the compiled data, along with the weather data will be fed to the Machine Learning model. The Model forecasts the water level, and volume for the next 3 months, which allows farmers to decide which type of crop they can plant. Then the actual water level, usage, and volume data are stored and used for subsequent water level and volume predictions. This process repeats itself providing the farmers with the ability to plan things in advance for the type of crop to be planted.

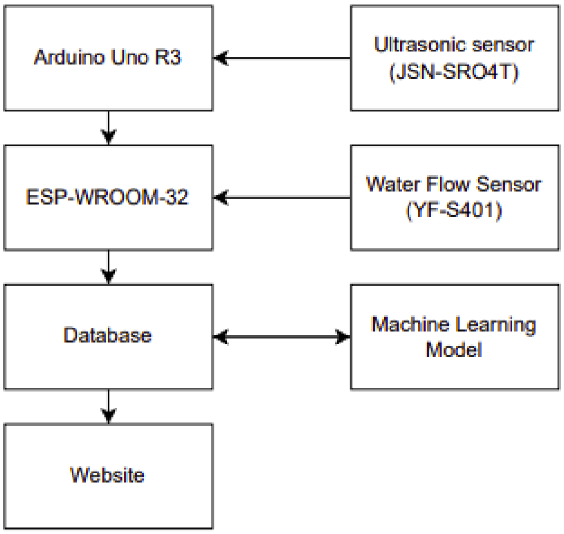

Figure 1. Workflow of the proposed system

Instrumentation Used for Data Collection

In this study, for data collection ultrasonic sensor, is used to sense the water level, and a water flow sensor is used to collect the water usage. Two microcontrollers were used, Arduino uno r3 and ESP32. An ultrasonic sensor is connected to the Arduino uno r3, and the water flow sensor is connected to the ESP32. Arduino uno r3 and ESP32 will be connected through serial communication for data transmission from Arduino uno r3 to ESP32. ESP32 has an in-built wifi module through which the compiled data will be transmitted to the database.

A. Ultrasonic Sensor

An Ultrasonic sensor is an electronic device that measures the distance of water from the surface of a well by emitting ultrasonic waves and converting the reflected sound into electrical signals. Ultrasound travels faster than audible sound (that is, sound that humans can hear).

The basic principle of an ultrasonic sensor is similar to a sonar or radar, which uses sound or radio waves to detect objects. An ultrasonic sensor consists of two main components: a transmitter and a receiver. The transmitter generates high-frequency sound waves using a piezoelectric crystal and sends them toward the water's surface. The receiver detects the sound waves that bounce back from the water surface and converts them into electrical signals.

B. Water Flow Sensor

A water flow sensor is an electronic device that measures the rate of flow of water in a pipe or a hose. It works by using a hall-effect sensor and a rotor that rotates when water passes through it. The hall-effect sensor detects the changes in the magnetic field caused by the rotation of the rotor and outputs a pulse signal proportional to the flow rate. The flow rate can be calculated by counting the number of pulses per unit time.

C. Database

Firebase is an app development platform that helps to build applications rapidly. Firebase provides a fully managed backend infrastructure that allows you to spin up servers without managing them. One can use Firebase for creating and managing databases, machine learning, etc. The water usage, level, and volume can be stored in a database through Firebase with the help of ESP32, which has an inbuilt wifi module.

Machine Learning Model

A. Workflow of the model

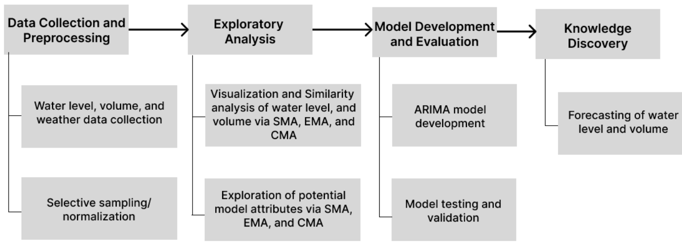

Figure 2. Workflow of water level, and volume data compilation, preprocessing, exploration, and model development.

Water level and volume in a well in the agricultural region of Guduvancheri, Tamil Nadu were explored along with time-series models developed as per the workflow described in Figure. Water level and volume data were obtained via a customized setup from a well, over a one-year period. The time-series water use data were initially explored by SMA, EMA, and CMA. Subsequently, autoregressive integrated moving average (ARIMA) predictive models were developed for water use, level, and volume patterns at different temporal scales. The water use patterns were analyzed to (i) assess the similarity of water use, level, and volume patterns in a well; (ii) evaluate the relevant attributes for describing water use, level, and volume patterns, including climate metrics, i.e., daily and low/high temperature (◦C); and (iii) establish predictive time-series models for forecasting water level and volume patterns.

B. Study Area and Data Collection

Water usage, volume, and level data are collected from a well used for agricultural purposes in Guduvancheri, Tamil Nadu over a one-year period (January 2022 -December 2022). The describe function in Pandas generates descriptive statistics of a DataFrame, including measures of central tendency, dispersion, and shape of the distribution.
The collection of water use, level, and volume data was achieved via customized hardware (with ultrasonic sensor, and water flow sensor) installed at the well and periodic water usage, level, and volume data was transmitted to a centralized data storage server at regular 10 seconds intervals. Daily and hourly Temperatures for the study region were obtained from Kaggle.

C. Preprocessing

Calculating SMA, and EMA

In time series analysis, the SMA (Simple Moving Average) is a statistical measure that calculates the mean of a fixed window of observations in a time series to identify trends and patterns in the data.
The EMA is similar to the SMA, but it places more weight on the most recent data points, giving a greater emphasis on real-time activity. The EMA is calculated by taking a weighted average of the past volume data, with the weights decreasing exponentially as we move back in time.
The resulting EMA values from “ewm” function are then stored in a new column called “df_ema_10_days_volume” in the df DataFrame. This new column represents the EMA

of the volume data over a 10-day period and can be used to track the trend of volume over time, similar to the SMA.

Performing Adfuller Test

The ADF (Augmented Dickey-Fuller) test is a statistical test used to determine whether a given time series is stationary or non-stationary, by testing the null hypothesis that the time series has a unit root (non-stationary) against the alternative hypothesis that the time series is stationary.
If the p-value is less than or equal to 0.05, the function prints "Reject the Null Hypothesis", which means that there is sufficient evidence to conclude that the time series is stationary. If the p-value is greater than 0.05, the function prints "Accept the Null Hypothesis", which means that there is not enough evidence to conclude that the time series is stationary.

D. *Arima Model*

The ARIMA models were based on three polynomials (i.e., the first polynomial as auto-regressive, the second as integrated, and the third as moving averages). The auto-regressive (AR) polynomial constitutes the autoregressive model at a predefined order p describing the dependence of the variable (e.g., water usage, level, and volume in the well over a specified time period) on its values in a previous time. The integrated polynomial is used to model the non-stationary nature of the time series data. It involves differencing the time series data to make it stationary. The moving average (MA) polynomial describes the linear dependence of the forecast errors resulting from the autoregressive model on the third predefined order q. Model development for daily water volume and level was based on training and test data for January 2022 to December 2022 respectively. Depending on the model resolution, input attributes included the day, week, month of the year, and low and high daily ambient temperatures.

The combined mathematical expression for the ARIMA model is, $y_t = c + \varphi_1 y_{t-1} + \theta_1 e_{t-1} + \varepsilon_t$

Where, y_t is the volume of water in the well at time t, c is the constant term in the model, φ_1 is the autoregressive parameter of order 1, which measures the degree of dependence of y_t on its past values, y_{t-1} is the volume of water in the well at time t-1, θ_1 is the moving average parameter of order 1, which measures the degree of dependence of y_t on the error term at t-1, e_{t-1} is the error term at time t-1, ε_t is the error term at time t, which represents the deviation of the actual value of y_t from its predicted value, respectively.

E. *Fitting the ARIMA Model*

To forecast the water volume and level in the well for the next three months, we fit an ARIMA model to the daily water level and volume data collected from January 1 to December 31, 2022. The ARIMA model was based on three polynomials, namely the autoregressive (AR), integrated (I), and moving average (MA) polynomials. The AR polynomial was of order 1 and represented the dependence of the water volume and level on their past values. The I polynomial was of order 1 and was used to make the time series data stationary. The MA polynomial was of order 1 and represented the linear dependence of the forecast errors resulting from the AR model on the previous error term.

We used the "auto.arima" function in the "pmdarima" package in Python to automatically select the optimal values of p, d, and q for the ARIMA model based on the Bayesian Information Criterion (BIC). The input attributes used in the model development included the day, week, month of the year, and low and high daily ambient temperatures.

```
model = ARIMA(df['volume'],order=(0,1,0))
model_fit = model.fit()
print(model_fit.summary())

                               SARIMAX Results
==============================================================================
Dep. Variable:                 volume   No. Observations:                  365
Model:                 ARIMA(0, 1, 0)   Log Likelihood               -4023.709
Date:                Mon, 01 May 2023   AIC                           8049.419
Time:                        12:51:24   BIC                           8053.316
Sample:                             0   HQIC                          8050.968
                                - 365
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2      2.333e+08   2.74e+06     85.030      0.000    2.28e+08    2.39e+08
===================================================================================
Ljung-Box (L1) (Q):                   1.60   Jarque-Bera (JB):             91039.61
Prob(Q):                              0.21   Prob(JB):                         0.00
Heteroskedasticity (H):               0.70   Skew:                             5.36
Prob(H) (two-sided):                  0.05   Kurtosis:                        79.73
===================================================================================
```

Figure 3. Code snippet of fitting the ARIMA model

V. RESULT

The customized hardware setup with microcontrollers and sensors can be used to collect the required data and stored it in the database for future predictions. The volume of water forecast indicates an increase in volume in the future (for the next 3 months). The forecast for water level depth indicates a future increase in depth.

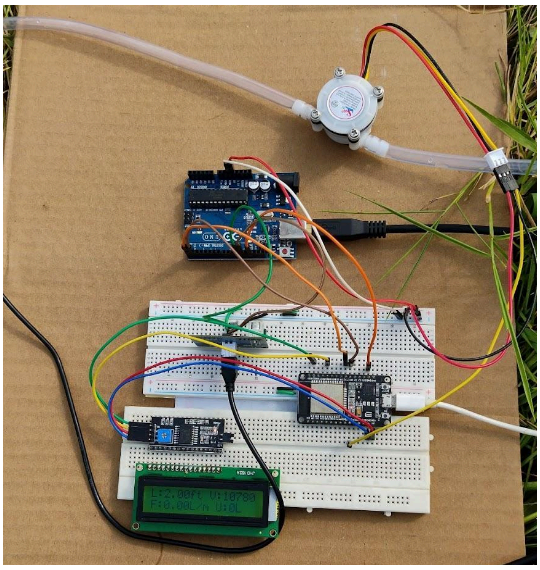

Figure 4. Working setup for data collection

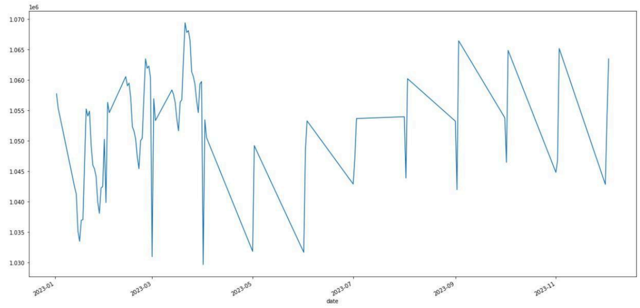

Figure 5. Water volume forecast

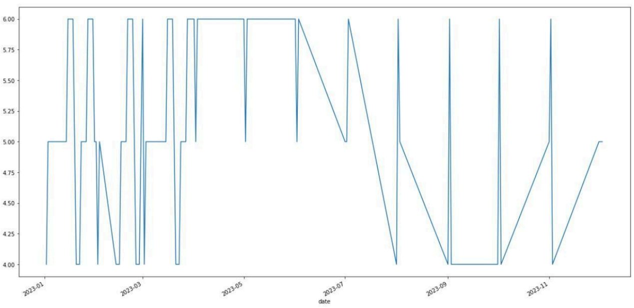

Figure 6. Water level forecast

REFERENCES

[1] Emami, M., Ahmadi, A., Daccache, A., Nazif, S., Mousavi, S. F., & Karami, H. (2022). County-level irrigation water demand estimation using machine learning: Case study of California. Water, 14(12), 1937

[2] Singh, I., & Bansal, M. (2011). Monitoring water level in agriculture using sensor networks. International Journal of Soft Computing and Engineering, 1(5), 202-204.

[3] Balducci, F., Impedovo, D., & Pirlo, G. (2018). Machine learning applications on agricultural datasets for smart farm enhancement. Machines, 6(3), 38. summarize the research paper.