ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

Predictive Analytics for Optimal Water Management in Smart Irrigation Systems Using Node-MCU Data

Dr. Sreekanth[1], A. Abhinaya[1], B. Varshini[2], C. Niharika[2], Ch. Sathwika[2]

[1]Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering
[1,2]Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad, Telangana, India

Abstract

Water management is a crucial aspect of modern agriculture, and the adoption of smart irrigation systems has gained significant attention due to its potential to optimize water usage while maximizing crop yields. Predictive analytics plays a vital role in smart irrigation systems, enabling farmers to make data-driven decisions based on real-time and historical data. In this study, we propose a predictive analytics approach for optimal water management in smart irrigation systems using machine learning algorithms with temperature and humidity data acquired from Node-MCU. The trained machine learning models are used to forecast future water requirements based on real-time data, allowing the system to predict the optimal irrigation schedule for each crop.

Keywords: Irrigation system, smart agriculture, data collection, water management, predictive analytics.

1. INTRODUCTION

Predictive analytics is revolutionizing water management within smart irrigation systems, offering a comprehensive and data-driven approach to optimize water usage in agriculture and landscaping. This innovative technology relies on the seamless integration of various data sources, including weather forecasts, soil moisture measurements, crop-specific data, and historical irrigation patterns. By harnessing the power of predictive analytics, smart irrigation systems can make informed decisions and recommendations for efficient water management. One of the core aspects of this approach involves leveraging real-time and forecasted weather data to anticipate weather conditions, such as rainfall, temperature, humidity, and wind patterns. This information enables the system to proactively adjust irrigation schedules, ensuring that water is used judiciously and preventing overwatering when rain is expected. Furthermore, predictive analytics continuously monitors soil moisture levels through embedded sensors in the soil. This data is then analyzed to determine the optimal timing and quantity of irrigation needed to maintain ideal soil conditions, thus avoiding water wastage and the risk of waterlogging. Moreover, the system takes into account the specific water requirements of different crop types, tailoring irrigation schedules to each crop's needs. This precision not only conserves water but also maximizes crop yields, contributing to sustainable agriculture practices. Historical irrigation data is another vital component. Predictive analytics mines this data to identify long-term trends, such as seasonal variations or crop-specific preferences, enabling the system to fine-tune irrigation strategies for improved efficiency.

Resource allocation is also optimized through predictive analytics, considering factors like energy and labor. This ensures that resources are used efficiently, leading to cost savings and reduced environmental impact. Smart irrigation systems with predictive analytics capabilities are capable of sending real-time alerts and recommendations to users. These alerts can include suggestions for adjusting irrigation schedules based on predicted weather conditions, helping users make informed decisions to prevent water waste. Moreover, these systems often offer remote control capabilities, allowing users to make real-time adjustments to irrigation settings based on predictive insights, even when they are not physically present on-site. This feature enhances convenience and responsiveness in managing water resources effectively. So, predictive analytics is a game-changer in the realm of

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

smart irrigation, providing precise, data-driven solutions for optimizing water management. By integrating real-time data, predictive modeling, and historical trends, these systems promote water conservation, reduce operational costs, improve crop yields, and contribute to sustainable and responsible water management practices. In a world where water resources are increasingly scarce, this technology holds immense promise for ensuring the efficient and sustainable use of this vital resource in agriculture and landscaping.

2. LITERATURE SURVEY

According to the Food and Agriculture Organization (FAO) of the United Nations, it is estimated that around 70% of all water withdrawal worldwide is due to agricultural applications [1], contrasting the industrial sector at 20% with municipalities' local infrastructure for services and domestic water use taking the remaining 10%. This seems a logical percentage distribution given that around 2000 to 3000 L of water are required to grow food per person daily [2]. Nonetheless, what is more concerning regarding this volume of water is that 93% never returns to its original source, signifying an apparent complete loss of the resource. Irrigation efficiency refers to the ratio of water the crop uses to the total amount of water extracted from the source [3]. Different factors affect irrigation efficiency, like water run-off, evaporation, and deep percolation. Water efficiency mostly depends on the hydraulic infrastructure and irrigation method, while surface irrigation has a water efficiency from 50% to 65%, sprinklers range from 60% to 85%, and drip irrigation from 80% to 90% [4]. Surface irrigation implies surface evaporation, which contributes to water loss. Sprinkler technology reduces water loss but, still, the applied water evaporates off the leaves of the crop canopy. In contrast, drip irrigation delivers water directly to the plant's root zone, reducing losses due to run-off and evaporation [5]. In any case, water efficiency can be considerably improved when a sensor-based smart irrigation system is installed over the hydraulic infrastructure [6]. Notwithstanding, food production is stated to rise in the following ten years and for many decades to come. In [7], the author states that the demand for food and agricultural products is projected to further increase by up to 70% by 2050 in order to satisfy the requirements for an estimated 10-billion-person population by then. That, in addition to the growing effect of climate change on water shortage worldwide, can have terrible consequences in the near future regarding resource allocation and availability for agricultural purposes. Vulnerable communities in arid regions would potentially suffer the consequences of water scarcity and global warming more [8]. Moreover, severe social conflicts have already occurred in rural communities due to the unfair assignation of water resources for agricultural activities [9]. Therefore, technology and data-driven solutions for water management are required to improve resource efficiency, reduce water waste, and contribute to sustainable agriculture practices [10].

3. PROPOSED SYSTEM

The comprehensive methodology outlines the step-by-step process for conducting research on predictive analytics for smart irrigation systems. It involves data collection, preprocessing, model training, evaluation, and ultimately, the application of predictive insights to optimize water management practices. This research can significantly enhance the efficiency and sustainability of water usage in agriculture and landscaping. The detailed operation illustrated as follows:

- Node MCU Dataset: The research begins by collecting data from Node MCU devices installed within the smart irrigation system. These devices likely capture data such as soil moisture levels, weather conditions, and potentially other relevant parameters. This dataset serves as the foundation for the subsequent analysis and modeling.
- Data Preprocessing: Raw data collected from Node MCU devices may contain inconsistencies, outliers, or missing values. Data preprocessing involves cleaning and

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

transforming the dataset to ensure it is suitable for analysis. This may include removing duplicates, handling missing data, and addressing outliers.

- Label Encoding: In predictive analytics, it's crucial to convert categorical variables, into numerical values that machine learning models can understand. Label encoding assigns a unique numerical label to each category.
- Defining of Training and Testing Data for Classification and Regression: The dataset is split into two subsets: one for classification tasks and another for regression tasks. Classification tasks may include predicting whether the irrigation pump should be turned on or off based on current conditions, while regression tasks may involve predicting the number of liters to be watered or the time required to supply water.
- Data Splitting for Classification: For classification tasks, the dataset is further divided into training and testing sets. The training set is used to train the machine learning model, while the testing set is reserved to evaluate its performance. A common split might be, for example, 70% of the data for training and 30% for testing.
- Data Splitting for Regression: Similarly, for regression tasks, the dataset is divided into training and testing sets. Regression models are trained on the training set and evaluated on the testing set to predict continuous values like the number of liters to be watered or the time needed for water supply.
- Random Forest Classifier Model: For classification tasks (e.g., determining when to turn the pump on or off), a machine learning model like Random Forest Classifier is employed. Random Forest is an ensemble learning method that combines multiple decision trees to make accurate predictions. It's suitable for tasks where data may have complex relationships.
- Random Forest Regression Model: For regression tasks (e.g., predicting liters of water needed or time for water supply), a Random Forest Regression model is used. Similar to the classifier, it's an ensemble method that can handle both linear and non-linear relationships in the data, making it effective for predicting continuous values.
- Final Predictions: Once the models are trained and evaluated, they are used to make predictions. For classification, the model predicts whether to turn the pump on or off based on current conditions. For regression, it predicts the number of liters to be watered and the time required for water supply. Additionally, it can be used to determine when to turn the pump on and off based on the predicted supply time.

Random Forest Classifier

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

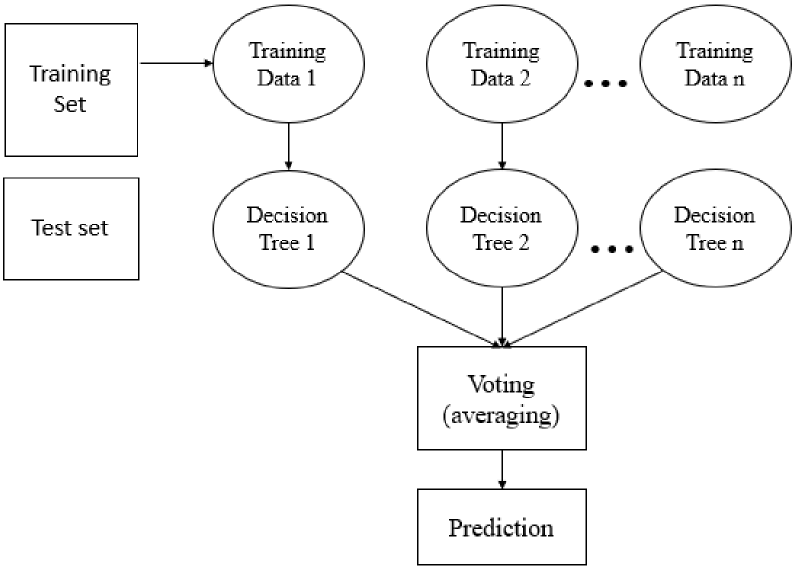

Fig. 1: Random Forest algorithm.

Random Forest algorithm

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Important Features of Random Forest

- Diversity- Not all attributes/variables/features are considered while making an individual tree, each tree is different.
- Immune to the curse of dimensionality- Since each tree does not consider all the features, the feature space is reduced.
- Parallelization-Each tree is created independently out of different data and attributes. This means that we can make full use of the CPU to build random forests.
- Train-Test split- In a random forest we don't have to segregate the data for train and test as there will always be 30% of the data which is not seen by the decision tree.
- Stability- Stability arises because the result is based on majority voting/ averaging.

Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random Forest classifier:

- There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
- The predictions from each tree must have very low correlations.
- Below are some points that explain why we should use the Random Forest algorithm
- It takes less training time as compared to other algorithms.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

- It predicts output with high accuracy, even for the large dataset it runs efficiently.
- It can also maintain accuracy when a large proportion of data is missing.

Types of Ensembles

Before understanding the working of the random forest, we must look into the ensemble technique. Ensemble simply means combining multiple models. Thus, a collection of models is used to make predictions rather than an individual model.

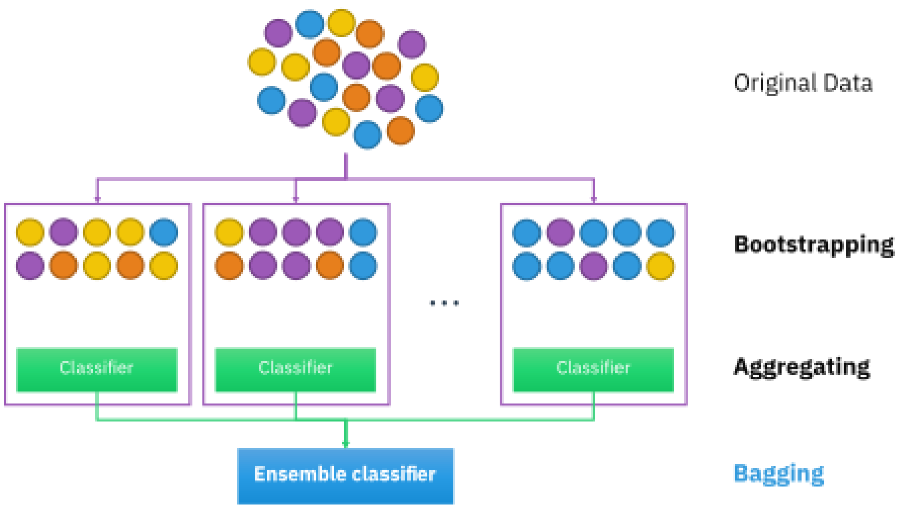

Fig. 2: RF Classifier analysis.

Ensemble uses two types of methods:

Bagging– It creates a different training subset from sample training data with replacement & the final output is based on majority voting. For example, Random Forest. Bagging, also known as Bootstrap Aggregation, is the ensemble technique used by random forest. Bagging chooses a random sample from the data set. Hence each model is generated from the samples (Bootstrap Samples) provided by the Original Data with replacement known as row sampling. This step of row sampling with replacement is called bootstrap. Now each model is trained independently which generates results. The final output is based on majority voting after combining the results of all models. This step which involves combining all the results and generating output based on majority voting is known as aggregation.

Boosting– It combines weak learners into strong learners by creating sequential models such that the final model has the highest accuracy. For example, ADA BOOST, XG BOOST.

4. RESULTS AND DISCUSSION

Table 1: Overall performance comparison of proposed ML models.

Model name	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Naive bayes Classifier	72	83	72	85
RFC classifier	97	97	97	97

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 12, Iss 4, 2023

Table 2: Class-wise performance comparison of proposed ML models.

Model name	Naive bayes Classifier		RFC classifier	
	Pump OFF	Pum ON	Pump OFF	Pum ON
Precision	0.73	0.88	100	95
Recall	0.86	0.97	94	100
F1-score	0.85	0.88	97	97

```
MAE: 41.20710363623162
MSE: 7463.521712095189
RMSE: 7463.521712095189
R-squared: 0.9993808656844166
MAPE: 0.00655841639669914
```

Fig. 4: Performances metrics for predicting number of Liters to be watered.

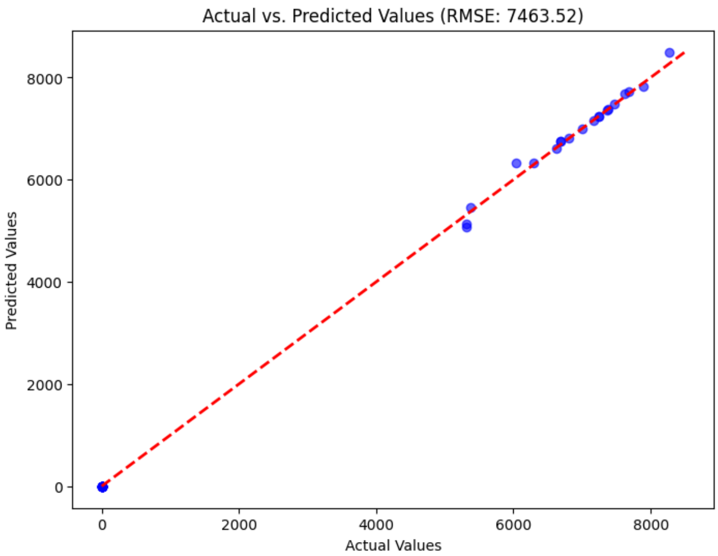

Fig. 5: Plot for representation of performance metrics.

```
RandomForestClassifier classification_report:

              precision    recall  f1-score   support

           0       1.00      0.94      0.97        17
           1       0.95      1.00      0.97        19

    accuracy                           0.97        36
   macro avg       0.97      0.97      0.97        36
weighted avg       0.97      0.97      0.97        36
```

Fig. 6: Classification report for Random Forest classifier

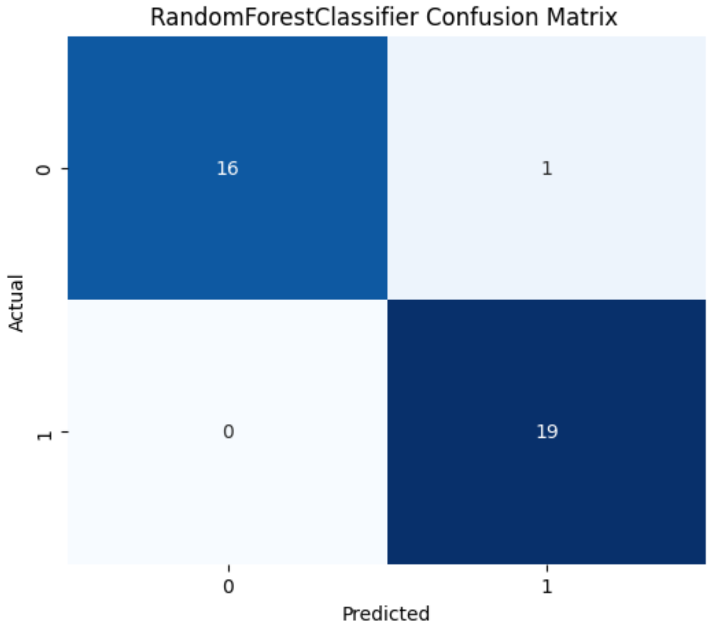

Fig. 7: Confusion matrix heatmap for Random Forest classifier

5. CONCLUSION

In conclusion, the implementation of predictive analytics for optimal water management in smart irrigation systems represents a transformative approach to address the pressing challenges of water scarcity, resource efficiency, and sustainability in agriculture and landscaping. This methodology, built upon the collection of data from Node MCU devices, data preprocessing, machine learning models, and precise decision-making, offers a host of advantages. It enables data-driven, precision irrigation, leading to reduced water wastage, resource efficiency, and cost savings. The adaptability to changing weather conditions, remote monitoring, and environmental benefits contribute to responsible water management and reduced environmental impact. Moreover, the potential for increased crop yields and scalability makes this approach invaluable to both small-scale farmers and large commercial operations. As global concerns about water resources and sustainable agriculture intensify, the application of predictive analytics in smart irrigation systems stands as a promising solution to address these challenges effectively.

REFERENCES

[1] Koncagül, E.; Tran, M.; Connor, R. The United Nations World Water Development Report 2021: Valuing Water; Facts and Figures. Technical Report, UNESCO. 2021. Available online: https://www.unesco.org/reports/wwdr/2021/en/download-report (accessed on 3 May 2023).

[2] Omran, H.A.; Mahmood, M.S.; Kadhem, A.A. A study on current water consumption and its distribution in Bahr An-Najaf in Iraq. Int. J. Innov. Sci. Eng. Technol. 2014, 1, 538–543.

[3] Grafton, R.Q.; Williams, J.; Perry, C.J.; Molle, F.; Ringler, C.; Steduto, P.; Udall, B.; Wheeler, S.A.; Wang, Y.; Garrick, D.; et al. The paradox of irrigation efficiency. Science 2018, 361, 748–750.

[4] Munir, M.S.; Bajwa, I.S.; Naeem, M.A.; Ramzan, B. Design and Implementation of an IoT System for Smart Energy Consumption and Smart Irrigation in Tunnel Farming. Energies 2018, 11, 3427.

[5] Hunter, M.C.; Smith, R.G.; Schipanski, M.E.; Atwood, L.W.; Mortensen, D.A. Agriculture in 2050: Recalibrating Targets for Sustainable Intensification. BioScience 2017, 67, 386–391.

[6] El-Fakharany, Z.M.; Salem, M.G. Mitigating climate change impacts on irrigation water shortage using brackish groundwater and solar energy. Energy Rep. 2021, 7, 608–621.

[7] Pluchinotta, I.; Pagano, A.; Giordano, R.; Tsoukiàs, A. A system dynamics model for supporting decision-makers in irrigation water management. J. Environ. Manag. 2018, 223, 815–824.

[8] Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. IEEE Access 2021, 9, 4843–4873.

[9] Zhai, Z.; Martínez, J.F.; Beltran, V.; Martínez, N.L. Decision support systems for agriculture 4.0: Survey and challenges. Comput. Electron. Agric. 2020, 170, 105256.

[10] Bu, F.; Wang, X. A smart agriculture IoT system based on deep reinforcement learning. Future Gener. Comput. Syst. 2019, 99, 500–507.