MACHINE LEARNING BASED PREDICTIVE ANALYTICS FOR OPTIMAL WATER MANAGEMENT IN SMART IRRIGATION SYSTEM

K Swathi, Dr GNV Vibha Reddy, P Chandrakantha
Assistance Professor, Associate Professor, Assistance Professor
Dept of CSE
Sree Dattha Institute of Engineering and Science

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. Traditional irrigation methods often rely on fixed schedules or manual observations, which may not accurately represent the actual water requirements of crops. Additionally, some existing smart irrigation systems use rule-based approaches that consider only basic environmental factors, potentially leading to suboptimal water allocation. These methods may not adapt well to changing environmental conditions and may not fully exploit the potential of predictive analytics. In this study, we propose a predictive analytics approach for optimal water management in smart irrigation systems using machine learning algorithms with temperature, 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.

INTRODUCTION

1.1 Overview

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 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.

1.2 Research Motivation

The research motivation for implementing predictive analytics for optimal water management in smart irrigation systems is driven by a confluence of critical challenges and opportunities.

First and foremost, water scarcity has become a pressing global concern. Climate change, population growth, and urbanization are placing unprecedented demands on freshwater resources, and agriculture accounts for a substantial portion of global water consumption. As a result, there is an urgent need to maximize the efficiency of water usage in agriculture to ensure food security while minimizing environmental impact. Predictive analytics offers a promising solution by enabling precision irrigation, which reduces water wastage and promotes responsible water management.

Furthermore, the unpredictability of weather patterns due to climate change exacerbates the challenge of water management in agriculture. Predictive analytics can help mitigate this uncertainty by integrating real-time weather data into irrigation decisions. This allows farmers and landowners to adapt their irrigation practices proactively, conserving water during periods of expected rainfall and optimizing irrigation during dry spells.

Additionally, the economic implications of water management cannot be overlooked. Agriculture is a significant sector in many economies, and inefficient water usage translates into higher operational costs. Predictive analytics can optimize resource allocation, leading to cost savings for farmers and promoting economic sustainability.

Moreover, sustainability is a growing concern in agriculture, driven by consumer demand for environmentally responsible practices. Smart irrigation systems with predictive analytics align with these sustainability goals by reducing water consumption, minimizing runoff and soil erosion, and promoting crop health. This, in turn, enhances the overall sustainability of agricultural practices.

Furthermore, advancements in technology have made predictive analytics more accessible and affordable for farmers and landowners, making it a viable solution for a broader range of agricultural settings. As these systems become more widespread, there is a growing need for research to refine and optimize their capabilities, tailoring them to specific crops, climates, and regions.

So, the research motivation for predictive analytics in smart irrigation stems from the urgent need to address water scarcity, adapt to changing climate conditions, reduce economic costs, and embrace sustainability in agriculture. This technology holds immense promise in revolutionizing water management practices, ensuring the efficient and responsible use of water resources, and ultimately contributing to the long-term viability of agricultural systems worldwide.

1.3 Problem Statement

The problem statement for implementing predictive analytics for optimal water management in smart irrigation systems can be articulated as follows:

The global agriculture sector faces a multifaceted challenge characterized by the increasing demand for food production, escalating water scarcity due to climate change, and the imperative to reduce environmental impact. This confluence of issues underscores the critical need for efficient water management practices in agriculture. Despite technological advancements in smart irrigation systems, there remains a substantial gap in achieving precision and sustainability in water usage. The current problem lies in the lack of comprehensive and data-driven approaches to address this challenge. Existing irrigation systems often rely on rudimentary scheduling methods that do not adapt to real-time weather variations, leading to over-irrigation, resource inefficiency, and environmental degradation. Moreover, the absence of crop-specific insights and historical trend analysis further exacerbates the problem, hindering optimal water allocation.

The unpredictable nature of climate change exacerbates this issue, necessitating the need for a proactive and adaptive solution. Farmers and landowners require tools that integrate real-time weather data, soil moisture measurements, and crop-specific information to make informed irrigation decisions. Additionally, the economic implications of inefficient water usage, including escalating operational costs, underscore the urgency of this problem. The lack of sustainable water management practices not only threatens agricultural viability but also poses environmental and social risks.

Hence, the problem statement at hand is the development and implementation of predictive analytics within smart irrigation systems to address these pressing issues comprehensively. This entails integrating diverse data sources, such as weather forecasts, soil moisture data, and historical irrigation patterns, into predictive models that can optimize water allocation and scheduling. The challenge lies in designing and fine-tuning predictive algorithms that are adaptable, accurate, and cost-effective, thus bridging the existing gap between traditional irrigation practices and sustainable, data-driven water management solutions.

1.4 Applications

The application of predictive analytics for optimal water management in smart irrigation systems holds tremendous potential across various domains, including agriculture, landscaping, and environmental conservation. Here are four detailed application scenarios:

- Precision Agriculture: Predictive analytics in smart irrigation systems are a game-changer for precision agriculture. By integrating real-time weather data, soil moisture measurements, and crop-specific insights, these systems can precisely determine when and how much water each crop needs. This not only maximizes crop yields but also conserves water resources by avoiding over-irrigation. In addition, the ability to remotely control irrigation based on predictive insights allows farmers to fine-tune their irrigation strategies, even from afar, ensuring optimal crop health and resource efficiency.

- Landscaping and Turf Management: Beyond agriculture, smart irrigation systems find applications in landscaping and turf management. Predictive analytics can tailor irrigation schedules to the specific needs of lawns, golf courses, parks, and gardens. This not only enhances the visual appeal of these spaces but also reduces water consumption and maintenance costs. By accounting for weather forecasts and historical trends, these systems can optimize irrigation to create lush and healthy landscapes while minimizing waste.
- Water Resource Conservation: Predictive analytics in smart irrigation systems play a crucial role in conserving water resources. By avoiding unnecessary irrigation during rainy periods and adjusting schedules based on anticipated weather conditions, these systems contribute to responsible water management. This is particularly important in regions facing water scarcity, as it helps prevent over-extraction of groundwater and reduces the strain on local water sources, promoting environmental sustainability.
- Environmental Impact Mitigation: Smart irrigation systems with predictive analytics capabilities can have a positive environmental impact. By preventing excessive runoff and soil erosion resulting from over-irrigation, they help maintain soil quality and prevent water pollution. Additionally, the reduction in energy consumption associated with optimized irrigation practices contributes to lower greenhouse gas emissions, aligning with broader environmental conservation goals.

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

3.1 Overview

This project aims to create a web application for predicting irrigation requirements based on input features like crop type, moisture level, and temperature. Here's an overview of the provided code:

- **Importing Libraries**: The initial section imports necessary libraries including Flask for web development, pandas for data manipulation, and scikit-learn modules for machine learning tasks.
- **Initialization**: An instance of the Flask application is created.
- **Loading Models and Data**: Random Forest Classifier and Regressors are loaded as models. Additionally, a LabelEncoder is initialized to transform categorical features. The dataset is loaded from a CSV file and preprocessed by encoding the categorical feature 'crop' into numerical values.
- **Routes**:
    - **'/' Route**: Renders the home page, typically an HTML template named 'index.html'.
    - **'/predict' Route**: Handles POST requests containing form data for prediction. It extracts input features like crop type, moisture level, and temperature from the form. The categorical feature 'crop' is encoded using the LabelEncoder. Then, the classification model predicts whether to turn on or off the pump. If the pump is predicted to be turned on, regression models predict the amount of water required (liters) and the number of days until the next watering. The results are formatted into a dictionary and passed to the 'result.html' template for display.
- **Rendering Templates**:
    - **'index.html'**: Represents the homepage containing a form to input crop type, moisture level, and temperature.
    - **'result.html'**: Displays the prediction results including the action to take (turn on or off the pump), predicted water requirements in liters, and days until the next watering.
- **Running the Application**: Finally, the application runs in debug mode allowing for easy debugging during development.

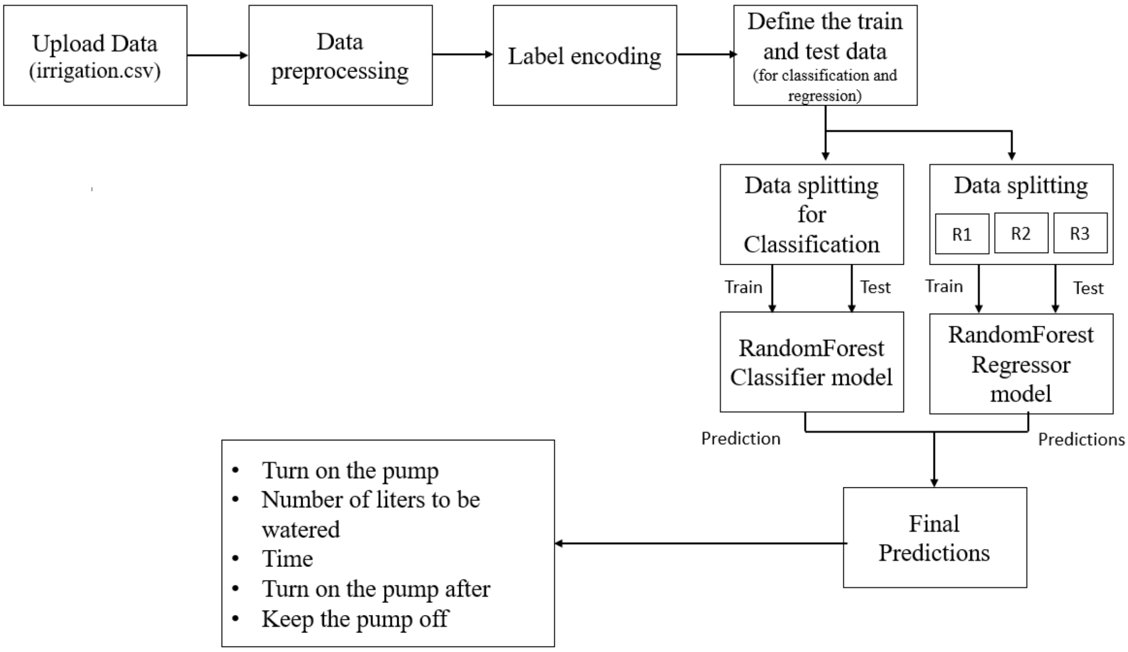

Figure 3.1 Proposed methodology.

3.2 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.

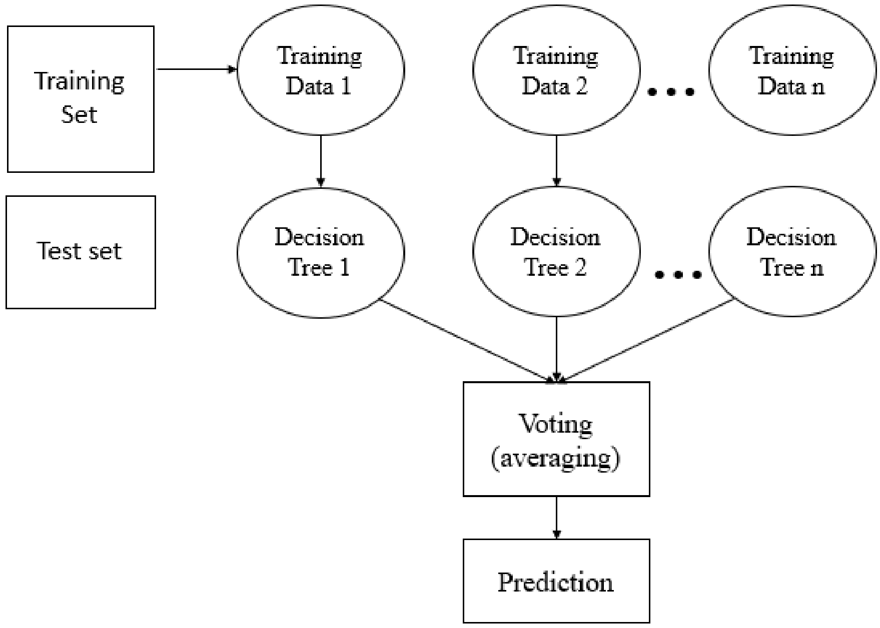

Fig. 3.2: Random Forest algorithm.

3.3 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.

3.3.2 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.

3.3.3 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.
- 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.

3.3.4 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. 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.

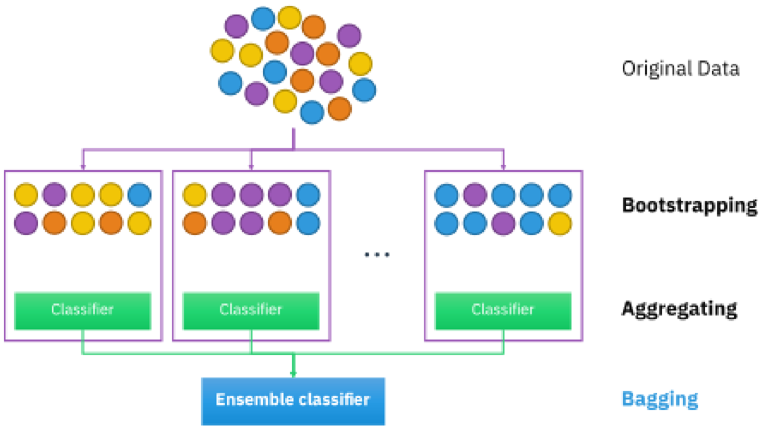

Fig. 3.3: RF Classifier analysis.

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.

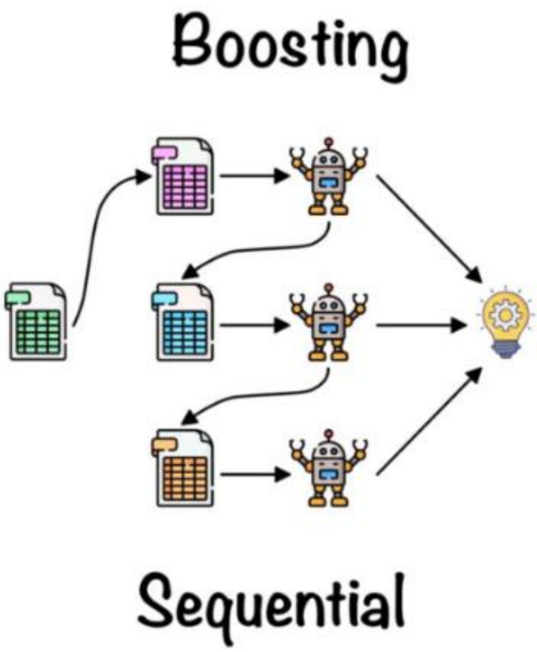

Fig. 3.4: Boosting RF Classifier.

3.4 Random Forest Regressor

A Random Forest regression is a powerful ensemble learning technique used in machine learning for both regression and classification tasks. It is based on the concept of decision trees and combines the predictions of multiple decision trees to improve the overall accuracy and robustness of the model.

Data Preparation: The process begins with a dataset that contains input features (independent variables) and corresponding target values (the variable you want to predict).

Bootstrapping (Random Sampling): Random Forest creates multiple subsets of the original dataset through a process called bootstrapping. This means that for each tree in the forest, a random sample of the data is taken with replacement. Some data points may appear multiple times in a subset, while others may be omitted.

Tree Building: For each subset of data, a decision tree is constructed. The construction of each tree involves selecting the best feature to split the data at each node. This selection is typically based on criteria like Gini impurity or mean squared error reduction.

Random Feature Selection: An essential aspect of Random Forest is that it introduces randomness during the tree-building process. Instead of considering all features at each node, it randomly selects a subset of features to choose from. This helps to decorrelate the trees and make them more diverse.

Tree Growing: The decision trees are allowed to grow until a stopping criterion is met. This may involve specifying a maximum depth for the tree or setting a minimum number of samples required to split a node.

Ensemble Aggregation: Once all the trees are constructed, they are used to make predictions on new data points. For regression tasks, the predictions of individual trees are averaged (or sometimes weighted) to obtain the final ensemble prediction. In the case of classification tasks, a majority vote is typically used.

Random Forest Prediction: The final prediction from the Random Forest ensemble is the average (or weighted average) of the predictions made by individual trees. This prediction tends to be more accurate and less prone to overfitting compared to a single decision tree.

Evaluation: The performance of the Random Forest regression model is evaluated using appropriate metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared (R2) on a separate validation or test dataset.

3.5 Advantages

The outlined methodology for predictive analytics in smart irrigation systems offers several notable advantages, making it a powerful approach for optimizing water management in agriculture and landscaping:

- Data-Driven Decision Making: By utilizing data collected from Node MCU devices and employing machine learning models, this methodology enables data-driven decision-making in water management. This means that irrigation decisions are based on real-time and historical data, leading to more informed and precise actions.
- Precision Irrigation: One of the primary advantages is the ability to achieve precision in irrigation. Machine learning models, such as Random Forest Classifier and Regression, can determine exactly when and how much water is needed for different crops and conditions. This precision minimizes over-irrigation, reducing water wastage and associated costs.
- Resource Efficiency: The methodology optimizes the allocation of resources, including water, energy, and labor. By accurately predicting water requirements and automating the irrigation process, it maximizes resource efficiency, leading to reduced operational costs and improved sustainability.
- Adaptation to Changing Conditions: Smart irrigation systems using predictive analytics can adapt to changing weather conditions. They can delay irrigation when rain is expected or increase it during dry spells, ensuring that crops receive the right amount of water regardless of unpredictable weather patterns.
- Remote Monitoring and Control: The ability to remotely monitor and control irrigation based on predictive insights offers convenience and flexibility to farmers and landowners. They can make real-time adjustments to irrigation settings, even when not physically present on-site, enhancing operational efficiency and responsiveness.
- Environmental Benefits: The methodology promotes environmentally responsible water management. By preventing over-irrigation and runoff, it reduces soil erosion and minimizes the pollution of local water sources. Additionally, by conserving water and energy, it contributes to a reduction in greenhouse gas emissions.
- Economic Savings: Smart irrigation systems powered by predictive analytics can lead to significant cost savings. Reduced water consumption, energy usage, and labor costs translate into lower operational expenses for farmers and landowners, improving their economic sustainability.
- Improved Crop Yields: Precision irrigation ensures that crops receive the optimal amount of water, promoting healthier growth and potentially increasing crop yields. This can have a positive impact on agricultural productivity and food security.

- Scalability: The methodology is scalable and adaptable to various agricultural settings and crop types. Whether applied to small-scale farming or large commercial operations, the approach can be customized to suit the specific needs of the users.
- Sustainability: Overall, the methodology contributes to sustainable agriculture practices. It aligns with global sustainability goals by conserving water resources, reducing environmental impact, and promoting responsible water management in a world where water scarcity is an increasing concern.

4.RESULTS

4.1 Results description

The result consists of a prediction based on the input provided by the user through the web form. If the prediction indicates that the pump should be turned on (clf_prediction[0] == 1), the app uses the regression models to predict the amount of water (liters_prediction) to pour and the number of days (days_prediction) before pouring water again. If the prediction indicates that the pump should be kept off (clf_prediction[0] == 0), the app simply suggests keeping the pump off without providing further predictions.

The application combines classification and regression models to make decisions about whether to turn on the irrigation pump and how much water to pour. Random Forest is used as it's a versatile algorithm suitable for both classification and regression tasks, and it handles non-linear relationships well. Label Encoding is used to transform categorical data into numerical format, which is necessary for machine learning models.

The application is designed to be user-friendly, as it provides input fields for the user to interact with and receive predictions. Flask is chosen as the web framework due to its simplicity and flexibility for building small to medium-sized web applications. The code lacks error handling and input validation, which are important for ensuring the robustness of the application. The dataset path (your_dataset_path) should be replaced with the actual path to the dataset for the application to work properly.

	crop	moisture	temp	pump	water_liters	time	days
0	cotton	638	16	1	6380	5.0	2
1	cotton	522	18	1	5220	1.0	2
2	cotton	741	22	1	7410	1.8	3
3	cotton	798	32	1	7980	2.5	2
4	cotton	690	28	1	6900	3.2	3
...	...	...	...	...	...	...	...
172	cotton	853	29	0	0	0.0	0
173	cotton	922	23	0	0	0.0	0
174	cotton	998	28	0	0	0.0	0
175	cotton	966	16	0	0	0.0	0
176	cotton	950	13	0	0	0.0	0

177 rows × 7 columns

Figure 4.1: Sample dataset using NODE-MCU sensor

The above figure-1 represents the sample dataset using NODE-MCU sensor. The sensor is used to collect & measure the levels of weather conditions like moisture, temperature and type of the crop.

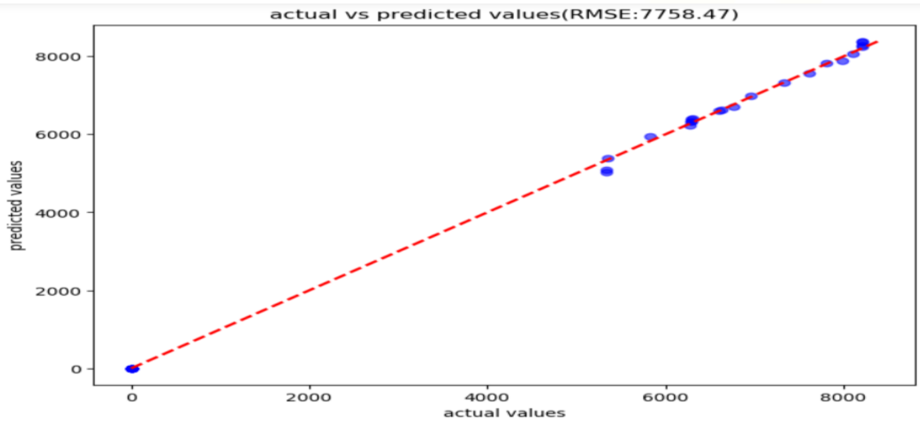

Figure 4.2: Performance Evaluation of Random Forest Regression

The above figure2 represents the Performance evaluation of the plot. We will calculate this evaluation using performance metrics such as mean absolute error(MAE), mean squared error(MSE), R2, mean absolute percentage error(MAPE). With the help of metrics, we will predict the performance of Actual values Vs predicted values.

```
randomforestclassifier classification_report:

              precision    recall  f1-score   support

           0       0.95      1.00      0.97        19
           1       1.00      0.94      0.97        17

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

Figure 4.3: Classification Report of Random Forest Classifier

Here, we are calculating the overall accuracy of the model by using Random forest classification algorithm. And also calculating the precision values, recall, f1-score, support etc.

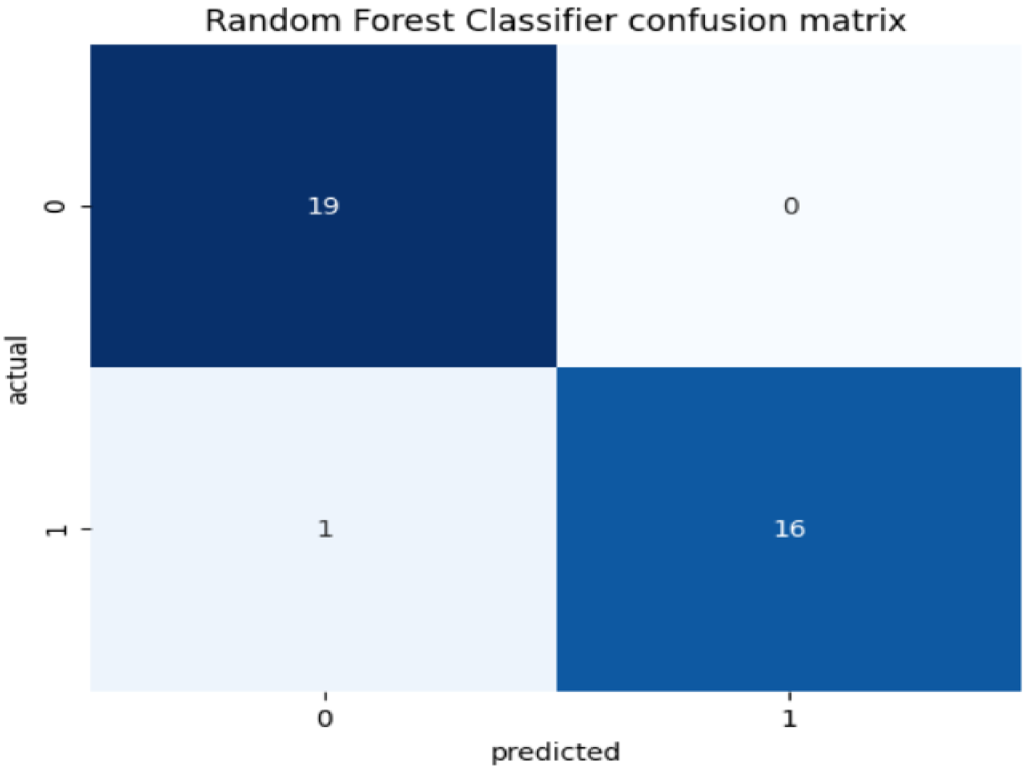

Figure 4.4: Confusion Matrix of Random Forest Classifier

We are predicting the values of confusion matrix such as TP, TN, FP, FN using Random Forest classification algorithm. It generates a heatmap visualization of the confusion matrix for a Random Forest Classifier's predictions compared to the actual labels.

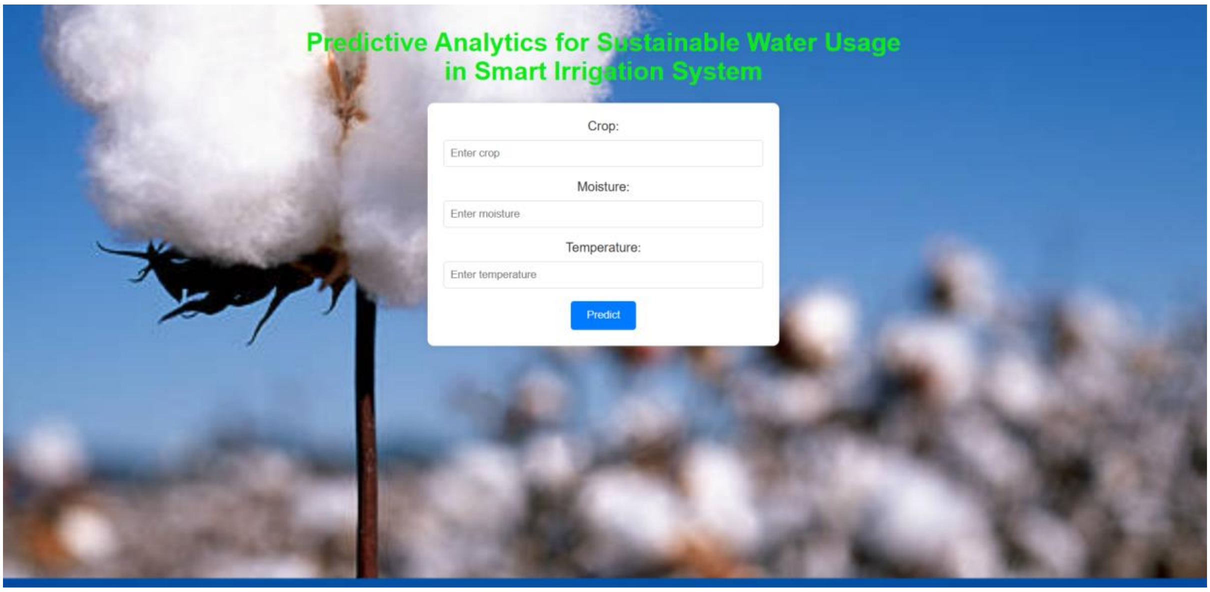

Figure 4.5: GUI Design of Smart Irrigation using flask server

The above figure-5 represents about the Design of graphical user interface (GUI) of our Smart irrigation project. It contains the input fields as Crop type, moisture, temperature. Based on the soil moisture levels, it predicts the output.

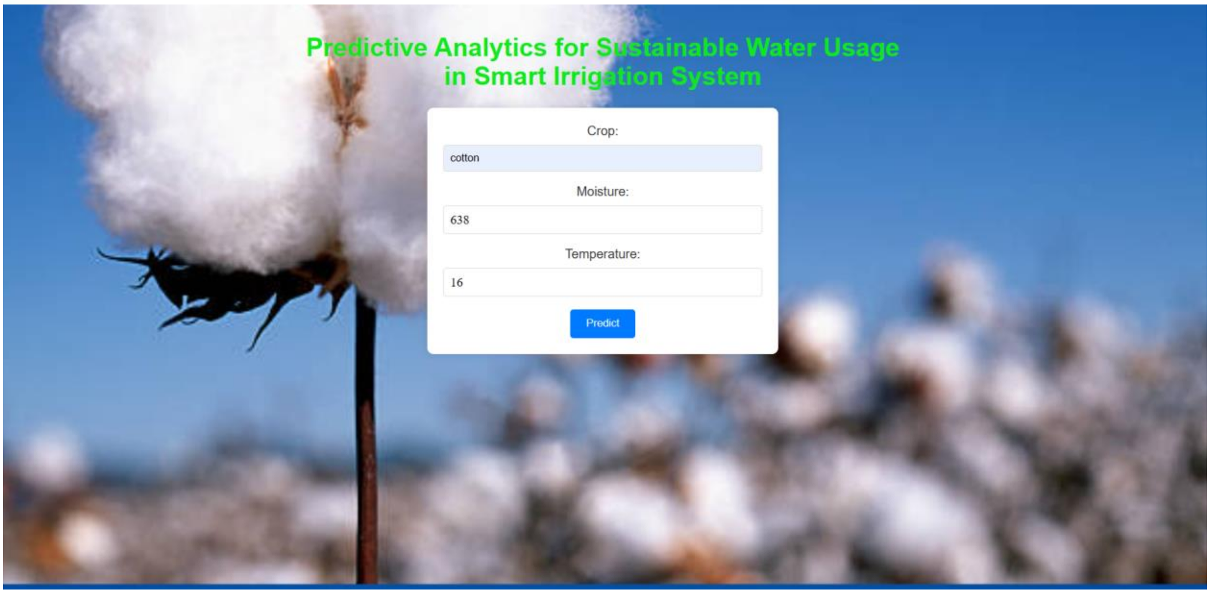

Figure 4.6: Home Page

The above figure-6 illustrates that, we will assign the input values in the required fields which are measured by the sensor, then it will predict the output based on that levels.

Figure 4.7: Prediction Result

It will predict the output as whether we should on/off the pump. If the pump is in ON mode, then it will also provide the amount of water that the crop required and the duration of period to pour again to the crops. It is based on the soil moisture conditions.

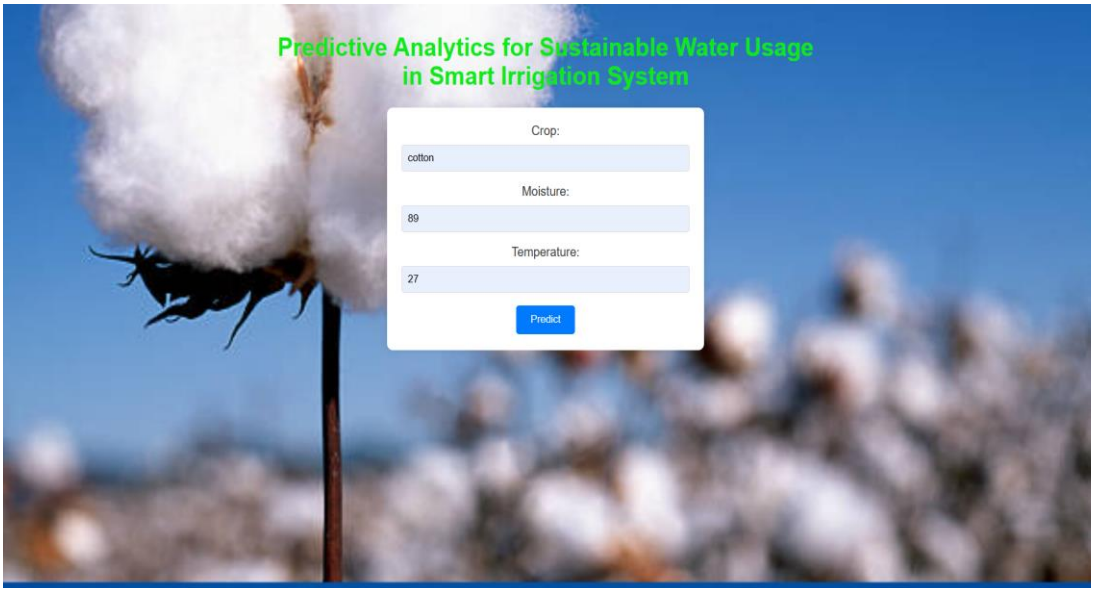

Figure 4.8: Home page

Here, we will give the values of the moisture, temperature, crop into the required fields. Then it will predict the output regarding those input values.

Figure 4.9: Prediction Result

It will give the output as keep the pump off. It means there is no requirement of water to the crops. The crops have the sufficient water, so based on the weather conditions it will predict that the pump is in OFF mode.

Table 2 compares the overall performance comparison of various ML models. Accuracy is a measure of how well a model correctly predicts both the positive and negative classes. In this table, the accuracy percentages indicate the overall correctness of the models' predictions. For example, the Naive Bayes Classifier achieves an accuracy of 72%, while the RFC (Random Forest Classifier) achieves an accuracy of 97%. This suggests that the RFC model performs significantly better in terms of overall accuracy.

Precision measures the proportion of true positive predictions among all the positive predictions made by the model. It indicates how well the model avoids false positives. For the Naive Bayes Classifier, the precision for the positive class is 83%, while for the RFC classifier, it is 97%. The RFC model demonstrates higher precision, meaning it has a lower rate of false positive predictions. Recall, also known as sensitivity, measures the proportion of true positive predictions among all actual positive instances. It indicates how well the model captures positive cases. In this table, the Naive Bayes Classifier has a recall of 72%, while the RFC classifier has a recall of 97%. The RFC model excels in capturing positive instances, resulting in a higher recall.

The F1-score is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance, considering both false positives and false negatives. For the Naive Bayes Classifier, the F1-score is 85%, whereas for the RFC classifier, it is also 97%. The RFC model demonstrates a better balance between precision and recall, leading to a higher F1-score.

Table 2: 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

Table 3 presents a detailed comparison of the class-wise performance metrics for two machine learning models: the Naive Bayes Classifier and the RFC. The models are evaluated based on their ability to classify instances into two classes: "Pump OFF" and "Pump ON."

- **Pump OFF**: This row represents performance metrics for the "Pump OFF" class.
    - **Precision**: Precision for "Pump OFF" measures how accurately the model predicts instances when the pump should be turned off. For the Naive Bayes Classifier, the precision is 0.73, indicating that 73% of the predicted "Pump OFF" instances were correct. In contrast, the RFC classifier achieves a perfect precision of 100% for this class, meaning it correctly identifies all instances of "Pump OFF."
    - **Recall**: Recall (or sensitivity) for "Pump OFF" measures the model's ability to capture all actual instances when the pump should be turned off. The Naive Bayes Classifier has a recall of 0.86, signifying that it captures 86% of the actual "Pump OFF" instances. The RFC classifier has a recall of 94%, indicating that it captures 94% of the "Pump OFF" instances.
    - **F1-score**: The F1-score for "Pump OFF" is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. The Naive Bayes Classifier achieves an F1-score of 0.85 for "Pump OFF," while the RFC classifier attains an F1-score of 0.97. The RFC model demonstrates a more balanced performance in terms of precision and recall for this class.
- **Pump ON**: This row represents performance metrics for the "Pump ON" class.

- **Precision**: Precision for "Pump ON" measures how accurately the model predicts instances when the pump should be turned on. The Naive Bayes Classifier achieves a precision of 0.88, indicating that 88% of the predicted "Pump ON" instances are correct. The RFC classifier achieves a precision of 95% for this class.
- **Recall**: Recall for "Pump ON" assesses the model's ability to capture all actual instances when the pump should be turned on. The Naive Bayes Classifier has a recall of 0.97, signifying that it captures 97% of the actual "Pump ON" instances. The RFC classifier achieves a perfect recall of 100%, indicating that it captures all "Pump ON" instances.
- **F1-score**: The F1-score for "Pump ON" is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. The Naive Bayes Classifier achieves an F1-score of 0.88 for "Pump ON," while the RFC classifier attains an F1-score of 0.97. The RFC model demonstrates a more balanced performance for this class as well.

Table 3: 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

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 make 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.