International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

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

Golla Chakrapani[2], Ippi Sumalatha[2], Dr. Vittapu Manisarma[1]

[1]Professor, [2]Assistant Professor, [1,2]Department of Computer Science Engineering
[1,2]Malla Reddy Engineering College and Management Sciences, Medchal, Hyderabad

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.

Keywords: Smart irrigation, Node MCU data, Optimal 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 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].

The waste and overuse of water resources for crop irrigation is a relevant topic that has been addressed by precision agriculture from different perspectives [11]. In this sense, automatic irrigation systems aim to optimize water utilization while helping farmers to improve crop yields by providing the right amount of water, at the right time, in the right place in the field [12]. To control the amount of water used during irrigation, typically these systems conduct measurements of soil moisture levels (volumetric water content), environmental parameters (solar radiation, wind speed, air temperature, air humidity), and crop conditions (canopy temperature, chlorophyll content, trunk diameter) [6].

3. PROPOSED SYSTEM

Overview

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. Figure 4.1 shows the proposed system model. The detailed operation illustrated as follows:

step 1: 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.

step 2: Data Preprocessing: Raw data collected from Node MCU devices may contain inconsistencies, outliers, or missing values. Data preprocessing involves cleaning and transforming the dataset to ensure it is suitable for analysis. This may include removing duplicates, handling missing data, and addressing outliers.

step 3: 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.

step 4: 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.

step 5: 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.

step 6: 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.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

step 7: 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.

step 8: 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.

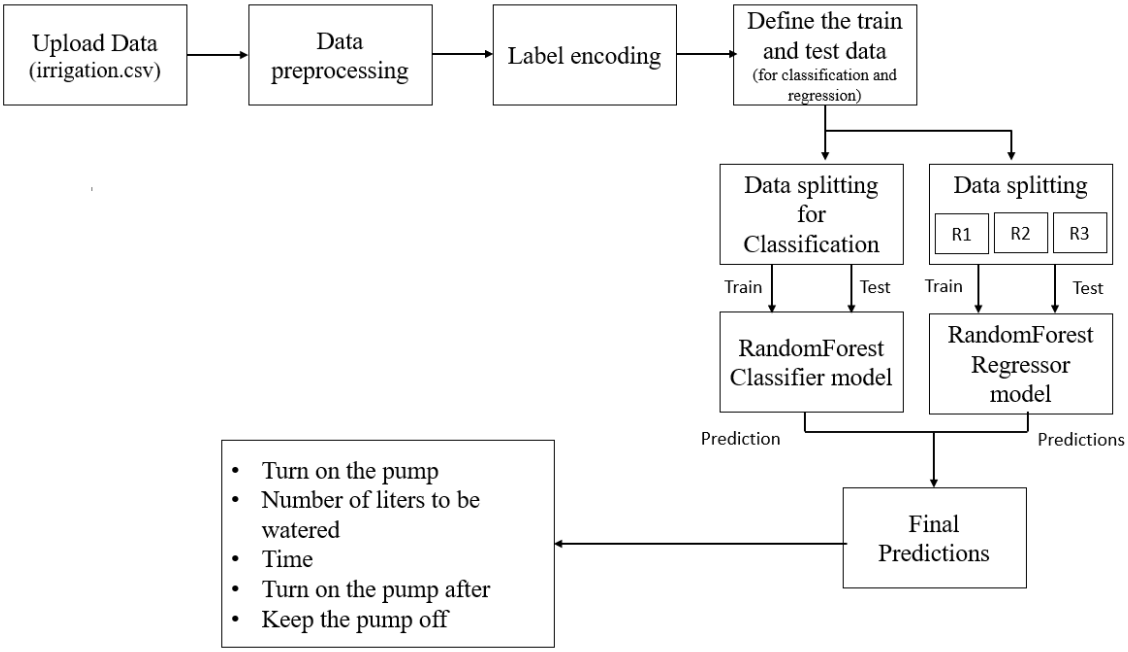

Figure. 1: Proposed methodology.

step 9: 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.

Data preprocessing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

- Getting the dataset
- Importing libraries

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

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.

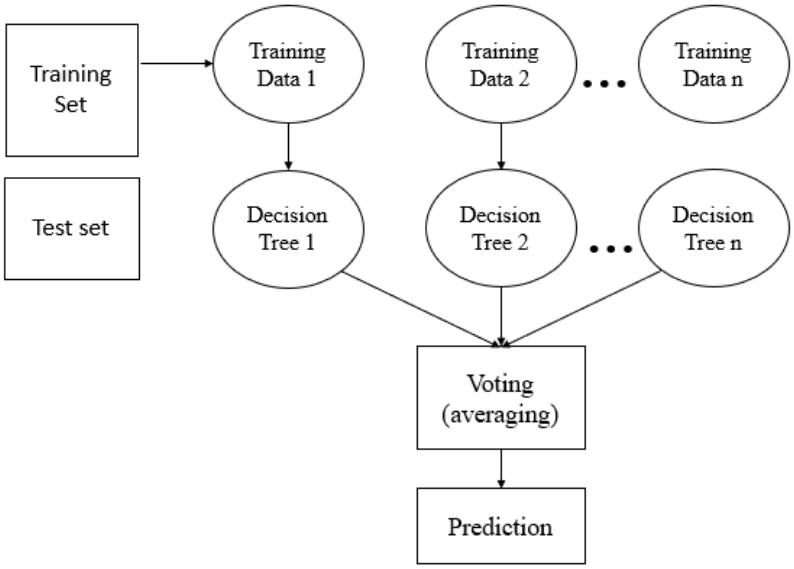

Fig.2: Random Forest algorithm.

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.

4. RESULTS AND DISCUSSION

Dataset description

The dataset contains the following columns

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

- Crop: This column indicates the type of crop being cultivated, which is exclusively "cotton" in this dataset.
- Moisture: Represents the moisture content in the soil. The unit of measurement for moisture content is not specified in the provided data. It could be percentage or any other unit specific to soil moisture.
- Temperature (Temp): Indicates the temperature of the environment or soil. The unit of measurement for temperature is not specified. It could be in Celsius or Fahrenheit.
- Pump: A binary indicator (0 or 1) representing whether a pump is being used or not for irrigation. Here, value 0 indicates No pump on, and value 1 indicatrices pump is off.
- Water Liters: Represents the amount of water to be watered for irrigation.
- Unit: The unit of measurement for water volume is specified as Liters.
- Time: Represents the duration of time watering the plants in hours
- Days: Indicates the How much time will take to water the plants when the soil meets the moisture in days units

Results description

Figure 3represents the initial dataset employed in the context of smart irrigation. This dataset serves as the fundamental source of information for all subsequent analyses and modeling. It likely contains various data points, each capturing different aspects of the smart irrigation system, such as soil moisture levels, weather conditions, crop types, and potentially other pertinent variables. This raw dataset reflects the real-world measurements collected from sensors, such as those embedded in Node MCU devices deployed throughout the irrigation setup. These data points represent unprocessed observations and may contain missing values, outliers, or inconsistencies. Before meaningful analysis can occur, this raw data must undergo a crucial data preprocessing step, which involves cleaning the data to rectify errors, handling missing information, and ensuring that the dataset is structured for machine learning or statistical analysis.

	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. 3: Sample dataset used for Smart irrigation

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Figure 4 presents the same dataset as Figure 1 but after undergoing data preprocessing. This processed dataset has been cleaned and transformed to ensure that it is suitable for analysis. Preprocessing steps include handling missing values, removing duplicates, encoding categorical variables, and addressing any inconsistencies.

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

177 rows × 7 columns

Figure. 4: Sample dataset used for Smart irrigation after preprocessing.

Figure 5 presents a heatmap representation of the confusion matrix for the Random Forest classifier. The confusion matrix visually displays the true positive, true negative, false positive, and false negative values, allowing for a visual assessment of the model's classification accuracy and error distribution.

Figure 6 likely serves as a summary or visualization of the overall predictions and outcomes generated by the smart irrigation system. It integrates both the classification (pump on/off) and regression (water volume, time, and irrigation schedule) results, providing a holistic view of the system's recommendations for efficient water management in the context of smart irrigation.

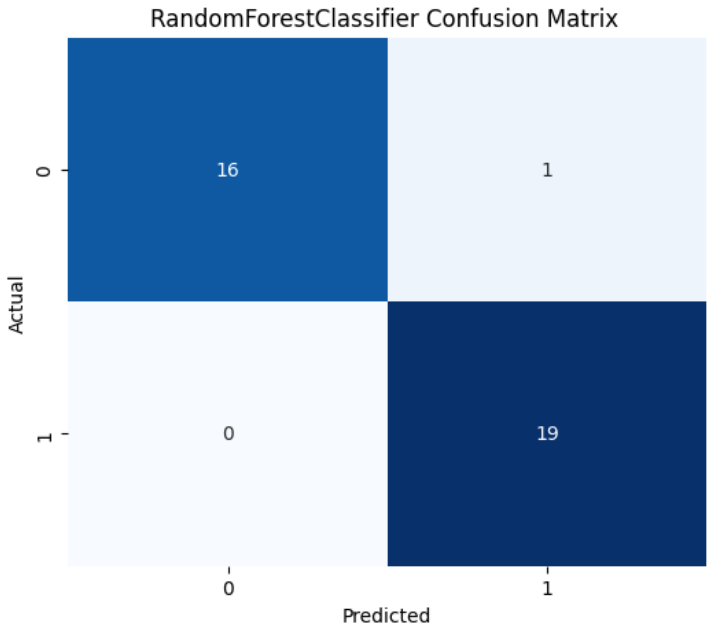

Figure 5: Confusion matrix heatmap for Random Forest classifier

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

```
# Use classification output to determine regression output
combined_output = []
for clf_pred, reg_pred, reg_pred1, reg_pred2 in zip(clf_predictions, reg_predictions1, reg_predictions2, reg_predictions3):
    if clf_pred == 1:
        combined_output.append((reg_pred, reg_pred1,reg_pred2))
        print('\n')
        print("*****************************************************************************************************")
        print("Turn on the pump")
        print("Number of Liters to be Watered:", reg_pred)
        print("time :", reg_pred1)
        print("Turn on the pump after:", reg_pred2)
        print("*****************************************************************************************************")
        print('\n')
    else:
        print('\n')
        print("Keep the pump off")
        print('\n')
```

```
*****************************************************************************************************
Turn on the pump
Number of Liters to be Watered: 7250.095174157625
time : 4.375265913617386
Turn on the pump after: 4.890218210477095
*****************************************************************************************************



Keep the pump off
```

Figure 6: Prediction results for Smart irrigation.

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.

[11]. Gutierrez, J.; Villa-Medina, J.F.; Nieto-Garibay, A.; Porta-Gandara, M.A. Automated Irrigation System Using a Wireless Sensor Network and GPRS Module. IEEE Trans. Instrum. Meas. 2014, 63, 166–176.

[12]. Lozoya, C.; Mendoza, C.; Aguilar, A.; Román, A.; Castelló, R. Sensor-Based Model Driven Control Strategy for Precision Irrigation. J. Sens. 2016, 2016, 9784071.