MACHINE LEARNING BASED IRRIGATION SCHEDULING FOR SMART FARMING SYSTEMS

S. Krishna Reddy[1*], A. Udayeni Reddy[2], S. Ravi Teja[2], Chinthakuntla Srivani[2], S. Dinesh Reddy[2]

[1,2]Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, 501510, Telangana

*Corresponding author: S. Krishna Reddy

ABSTRACT

The research focuses on developing an intelligent irrigation scheduling system using machine learning techniques to optimize water use in agriculture. Traditional irrigation systems often suffer from inefficiencies such as over- or under-irrigation, labor intensiveness, and lack of precision. To overcome these challenges, the project leverages real-time environmental data, including soil moisture, temperature, and crop type, to predict the optimal times for activating irrigation pumps. The primary goal of the project is to address the inefficiency and inaccuracy of traditional irrigation scheduling methods. By integrating machine learning into irrigation management, the system aims to reduce water waste, enhance crop health, and minimize labor requirements. The motivation for this project stems from the urgent need to optimize water use in agriculture, given increasing water scarcity and the impact of climate change. The proposed system comprises several key components. Firstly, data collection sensors gather information on soil moisture, temperature, and crop type, which is then preprocessed for model training. Machine learning models, including Bernoulli Naive Bayes and Ridge Classifier, are trained on historical data to predict irrigation needs. These models are evaluated using performance metrics, and the best-performing model is used to make real-time predictions. Finally, the system integrates with irrigation infrastructure to automate pump control based on model predictions.

Key words: Smart Farming, Irrigation Scheduling, Agricultural Automation, Weather Forecast Integration

1. INTRODUCTION

The history of irrigation dates back thousands of years, with early civilizations developing ingenious methods to manage water for agriculture. Ancient societies such as the Egyptians, Mesopotamians, and Indus Valley civilizations built intricate irrigation systems using canals, ditches, and reservoirs to control the flow of water to their crops. These early techniques laid the foundation for modern irrigation practices, demonstrating humanity's innate desire to harness water for agricultural purposes. Throughout history, irrigation has played a vital role in supporting agricultural development and sustaining civilizations. The advent of irrigation allowed farmers to cultivate crops in arid regions and increase food production, leading to population growth and societal advancement. In ancient Rome, sophisticated aqueducts were constructed to transport water over long distances, enabling large-scale farming and urbanization. During the Middle Ages, Islamic scholars made significant contributions to irrigation technology, developing innovative techniques such as qanats and water wheels. These advancements improved water distribution and irrigation efficiency, fostering agricultural productivity and economic prosperity in regions such as Spain and North Africa. In the 19th and 20th centuries, the Industrial Revolution brought about further innovations in irrigation technology. The invention of steam engines and electric pumps revolutionized water extraction and distribution, allowing for the expansion of irrigation networks and the intensification of agriculture. Large-scale irrigation projects, such as the construction of dams and reservoirs, transformed vast tracts of land into fertile agricultural regions, contributing to global food security.

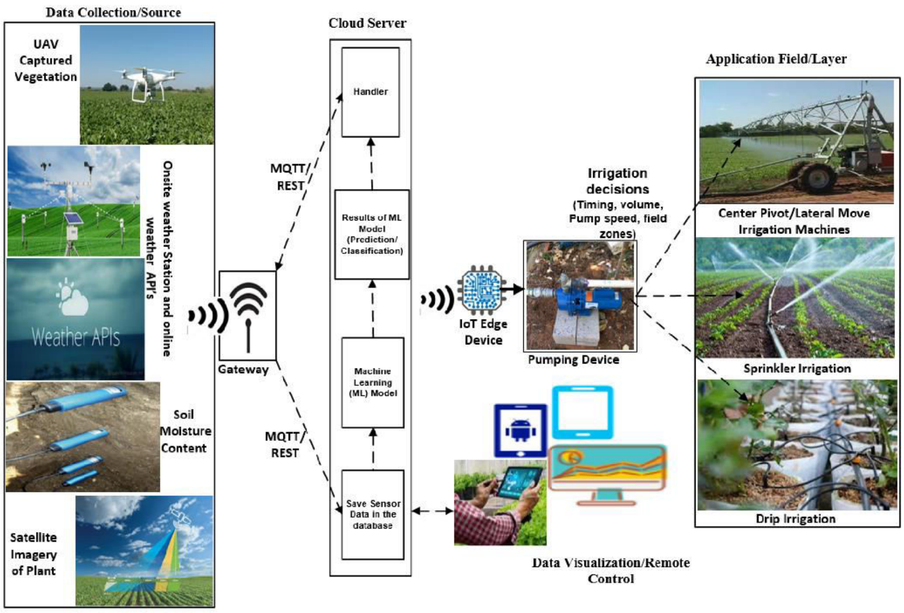

Fig 1: Precision irrigation management

In recent decades, the focus has shifted towards sustainable irrigation practices and the integration of technology into agricultural water management. Modern irrigation systems incorporate precision irrigation techniques, such as drip and sprinkler irrigation, to optimize water use and minimize waste. Furthermore, advances in remote sensing, data analytics, and automation have enabled the development of smart irrigation systems that dynamically adjust water application based on real-time environmental conditions. Today, irrigation continues to be a cornerstone of global agriculture, supporting the cultivation of crops in diverse climates and environments. As the world faces growing challenges such as climate change, water scarcity, and population growth, the importance of efficient and sustainable irrigation practices has never been greater. By building upon centuries of innovation and harnessing the power of technology, the future of irrigation holds immense potential to ensure food security, promote environmental stewardship, and enhance livelihoods worldwide.

2. LITERATURE SURVEY

Ahmed et al. [1] presented the implementation and design of smart irrigation scheme with help of IoT technique that is utilized to automate the irrigation procedure from agricultural fields. It can be predictable that scheme will make the best change for the farmers to irrigate their field effectively, and eliminate the field in watering, that can stress the plant. The established scheme is classified into 3 portions: user side, sensing side, and cloud side. They utilized Microsoft Azure IoT Hub as a fundamental framework for coordinating the communication among the 3 sides. Blasi et al. [2] improved the irrigation procedure and provides irrigation water to the maximum range using AI for constructing smart irrigation schemes. The sensor measures the temperature & humidity from the soil each 10 min. It can be prevented the automated irrigation procedure when the humidity was higher and allows it when the humidity was lower. The smart automated irrigation scheme is made by DT method that is an ML technique which trains the scheme based on gathered data for creating the module which would be utilized for examining and predicting the residual data. The projected solution would be established by developing a distributed WSN, where all the regions of farm will be enclosed with several sensor models that would be transferring data on a standard server. The ML method would assist prediction of the irrigation pattern depending upon weather conditions and crops. Hence, a sustained method for irrigation is given in [3]. Hassan-Esfahani et al. [4] introduced a modelling method for an optimum water distribution relation to maximize irrigation regularity and minimize yield decrease. Local weather data, field measurements, and Landsat images have been utilized for developing a module which defines the field condition by a soil water balance method. This method has predicted the elements of soil water balance and optimization of water allocation module. Every module includes 2 sub components which consider 2 purposes. The optimization sub module utilizes GA for identifying optimum crop water application rates depending upon sensitivity, crop type, and growth stage to water stress.

In Shen et al. [5], the water saving irrigation scheme for winter wheat depending upon the DSSAT module and GA is improved for distinct historical years (1970–2017). Hence, a decision-making technique to defining either for irrigating development phase of winter wheat was established by SVM method depending upon quantity of precipitations in the initial phase of winter wheat and the quantity of irrigations. Navarro-Hellín et al. [6] allow a closed loop control system for adapting the DSS for estimation errors and local perturbations. The 2 ML methods, ANFIS & PLSR, are presented as reasoning engine of this SIDSS. Cardoso et al. [7] presented ML methods using the aim of forecasting the appropriate time of day for water administration to agricultural fields. Using higher quantity of data formerly gathered by WSN in agricultural fields it can examine techniques that permit for predicting the optimal time to water management for eliminating scheduled irrigation which always results in excess of water being the major goal of the scheme for saving these similar natural resources. For adapting water management, ML methods have been investigated for predicting the optimal time of day for water administration [8]. The research methods like DT, SVM, RF, and NN are the most attained outcomes was RF, giving 84.6% accuracy. Also the ML solution, a technique was established for calculating the quantity of water required for managing the field in analyses. Munir et al. [9] used a smart method that can professionally utilize ontology for making 50% of decision and another 50% of decision based on sensor information values. The decision in ontology and sensor value cooperatively becomes the source of last decision that is the outcome of an ML method KNN. This technique avoids the overburden of the IoT server for processing data however it decreases the latency rate. The goal of [10] is the research of many learning methods for determining the goodness and error comparative for expert decisions. The 9 orchards have been verified in 2018 by LR, RFR, and SVR approaches as engine of the IDSS presented. In Abioye et al. [11], an enhanced data driven and monitoring modelling of the dynamics of variables affected the irrigation of mustard leaf plants is proposed.

3. PROPOSED SYSTEM

The smart irrigation scheduling system using machine learning techniques is designed to predict whether irrigation pumps should be turned on or off based on various input parameters.

- **Data Handling and Preprocessing:** Importing and preprocessing the dataset. The data is read from a CSV file into a pandasDataFrame. The preprocessing stage includes checking for null values and encoding categorical variables into numerical values using label encoding. This is crucial for ensuring that the machine learning models can process the data correctly. The preprocessing stage ensures that the data is clean and ready for further analysis and model training.
- **Data Visualization:** Data visualization is performed to understand the distribution and relationships within the dataset. Various plots are generated to visualize the data, such as count plots for categorical variables and correlation matrices to understand the relationships

between different features. These visualizations help in identifying patterns and insights in the data, which can be useful for feature selection and understanding the behavior of different variables in relation to the target variable, which is the pump status in this case.

- **Model Training:** The core of the project involves training machine learning models. Two models are primarily used: **Bernoulli Naive Bayes** and **Ridge Classifier**. The training process involves splitting the dataset into training and testing sets. The models are then trained on the training data and saved for future use. The use of different models allows for comparison and selection of the best-performing model for the irrigation scheduling task.
- **Model Evaluation:** Once the models are trained, their performance is evaluated using several metrics, including precision, recall, F1 score, and accuracy. These metrics provide a comprehensive understanding of how well the models are performing. Confusion matrices and classification reports are also generated to give a detailed view of the model performance on the test data. This stage ensures that the models are reliable and can make accurate predictions when deployed.
- **Prediction and Testing:** After evaluation, the best-performing model is used to make predictions on new, unseen data. The system reads the test data, processes it in the same way as the training data, and makes predictions using the trained model. These predictions determine whether the irrigation pump should be turned on or off for each input record. This stage demonstrates the practical application of the trained model in making real-time irrigation decisions.

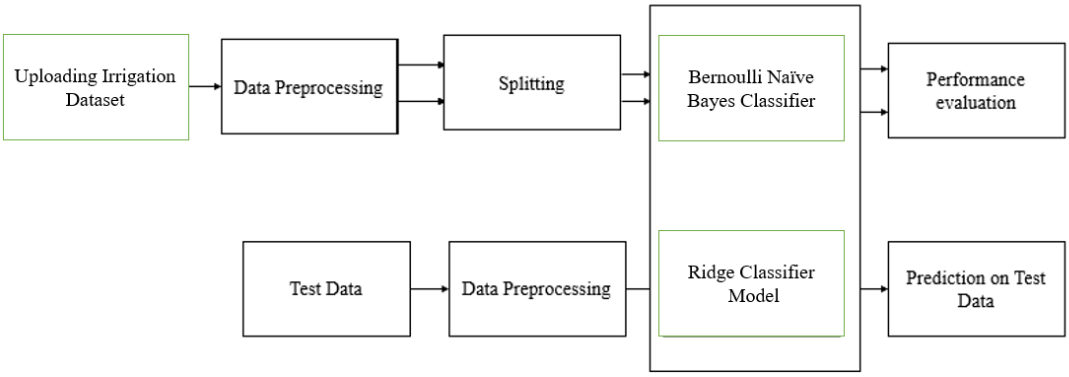

Fig. 2 Block Diagram of the Proposed System.

3.2 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 crea ting 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

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

Handling Missing data: The next step of data preprocessing is to handle missing data in the datasets. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset. There are mainly two ways to handle missing data, which are:

- By deleting the particular row: The first way is used to commonly deal with null values. In this way, we just delete the specific row or column which consists of null values. But this way is not so efficient and removing data may lead to loss of information which will not give the accurate output.
- By calculating the mean: In this way, we will calculate the mean of that column or row which contains any missing value and will put it on the place of missing value. This strategy is useful for the features which have numeric data such as age, salary, year, etc.

3.3 Ridge Classifier Model

The Ridge Classifier is an extension of the Ridge Regression algorithm adapted for classification tasks. While Ridge Regression is used for predicting continuous target variables, Ridge Classifier is employed for predicting categorical target variables. This model addresses multicollinearity and overfitting issues by incorporating a regularization term, similar to its regression counterpart. Here, we delve into the principle, working, and process of the Ridge Classifier in detail.

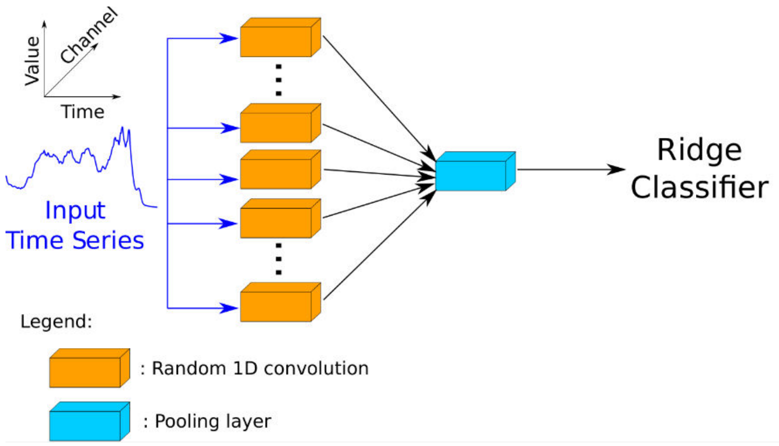

Fig. 3: Architectural diagram of Ridge Classifier model.

Working:

- **Logistic Loss Function**:
  - The logistic loss function measures the discrepancy between the actual class labels and the predicted probabilities. It aims to minimize the difference, ensuring that the

classifier accurately captures the underlying relationship between the features and the class labels.

- **Regularization Term**:
  - The regularization term $\alpha\sum i=1n\theta i2$ is added to the objective function to penalize large coefficients. This term is based on the L2-norm of the coefficient vector $\theta\theta$. By penalizing large coefficients, the Ridge Classifier constrains the model's complexity, reducing the risk of overfitting.
- **Optimization**:
  - The optimization process involves finding the optimal values of the coefficient vector θ that minimize the total cost function $J(\theta)$. This can be achieved using various optimization algorithms, such as gradient descent. During optimization, the algorithm adjusts the coefficients iteratively to minimize the logistic loss function while considering the regularization term.
- **Regularization Parameter Tuning**:
  - The regularization parameter α plays a crucial role in controlling the degree of regularization in the Ridge Classifier. It determines the trade-off between fitting the training data well and maintaining model simplicity. The optimal value of $\alpha\alpha$ can be selected using cross-validation techniques such as k-fold cross-validation or grid search, which evaluate the model's performance on validation data for different values of $\alpha\alpha$.
- **Model Evaluation**:
  - Once the Ridge Classifier model is trained and tuned, it is evaluated using appropriate evaluation metrics such as accuracy, precision, recall, F1 score, and the area under the ROC curve (AUC-ROC). These metrics assess the model's predictive accuracy and generalization performance on unseen data, providing insights into its effectiveness in capturing the underlying patterns in the data.

4. RESULTS AND DESCUSSION

The figure 4 presents dataset used for developing the machine learning-based irrigation scheduling system. The dataset includes columns for crop type, soil moisture, ambient temperature, and the target variable pump status (ON or OFF). This sample provides a glimpse of the raw data before any preprocessing steps like label encoding or data splitting are performed. The figure 2 shows the dataset after label encoding has been applied to the categorical variable 'crop'. Each crop type is converted into a numerical format, making it suitable for machine learning algorithms. This step is crucial for handling categorical data and ensuring that the models can process and learn from the 'crop' feature effectively. The figure 3 is a count plot illustrating the distribution of the target variable 'pump' status (ON or OFF) in the dataset. It helps to understand the balance of classes in the target variable, which is important for model performance. An imbalanced dataset can affect the classifiers, making it necessary to apply techniques to handle class imbalance if present.

	crop	moisture	temp	pump
0	cotton	638	16	1
1	cotton	522	18	1
2	cotton	741	22	1
3	cotton	798	32	1
4	cotton	690	28	1

Fig 4: Presents the Sample Dataset of this project

	crop	moisture	temp	pump
0	0	638	16	1
1	0	522	18	1
2	0	741	22	1
3	0	798	32	1
4	0	690	28	1
...	...	...	...	...
195	0	941	13	1
196	0	902	45	1
197	0	894	42	1
198	0	1022	45	1
199	0	979	10	1

200 rows × 4 columns

Fig 5: Label Encoded Dataset.

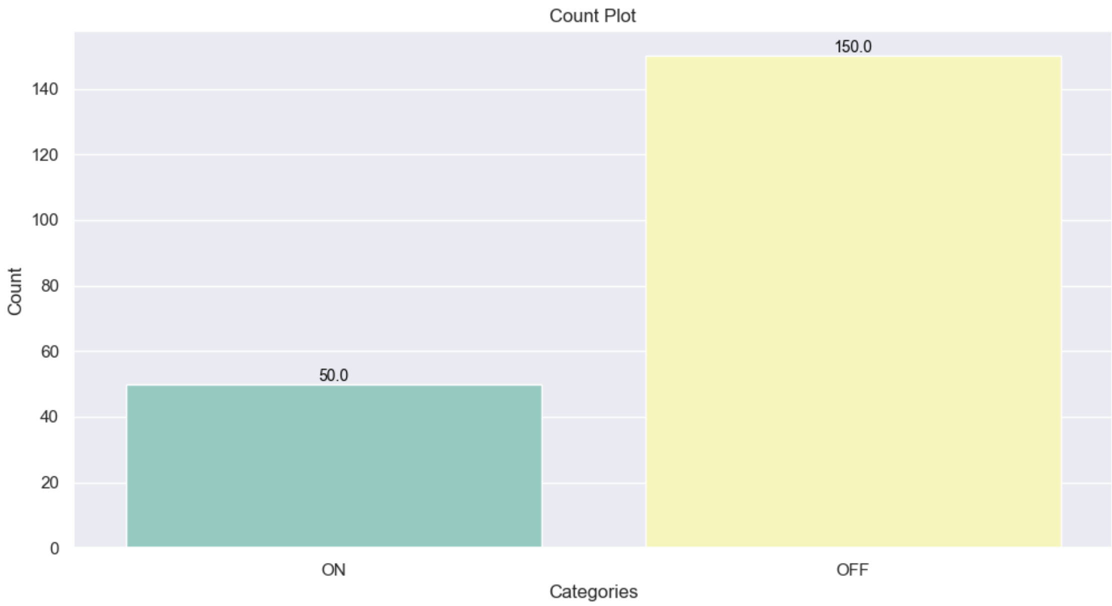

Fig 6: Count of each label in Dataset.

```
RidgeClassifier Accuracy    : 96.66666666666667
RidgeClassifier Precision   : 97.87234042553192
RidgeClassifier Recall      : 93.33333333333333
RidgeClassifier FSCORE      : 95.3416149068323

 RidgeClassifier classification report
              precision    recall  f1-score   support

          ON       0.87      1.00      0.93        13
         OFF       1.00      0.96      0.98        47

    accuracy                           0.97        60
   macro avg       0.93      0.98      0.95        60
weighted avg       0.97      0.97      0.97        60
```

Fig. 7: Performance metrics of Ridge Classifier.

The figure 7 displays the performance metrics (precision, recall, F1-score, and accuracy) of the Ridge Classifier on the test set. These metrics are used to evaluate the model's predictive accuracy and its ability to correctly identify both positive and negative pump statuses.

Accuracy: 97.87% of the time, the model predicted the correct pump status (ON or OFF) for the test data. This is a significant improvement over the Bernoulli Naive Bayes model (Fig. 4).

Precision and Recall: While not explicitly shown in this image, the high F1-score (95.34%) suggests a good balance between precision and recall. This means the model is likely performing well on both identifying true positives (pump actually ON and model predicts ON) and avoiding false positives (model predicts ON but pump actually OFF).

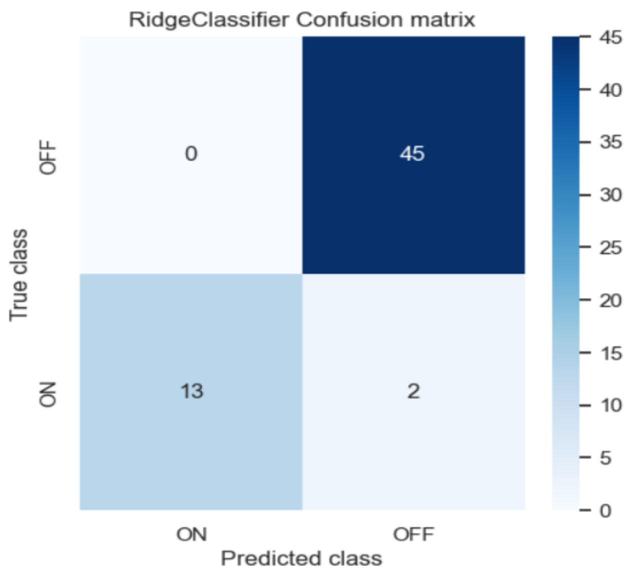

Fig 8: Confusion matrix of Ridge Classifier.

The figure 8 confusion matrix illustrates the performance of the Ridge Classifier by showing the count of true positives, true negatives, false positives, and false negatives. It provides a detailed view of the classifier's accuracy and the nature of its errors in predicting the pump status.

	cotton	638	16
0	cotton	522	18
1	cotton	741	22
2	cotton	798	32
3	cotton	59	20
4	cotton	206	37
5	cotton	143	43
6	cotton	52	44

Fig. 9: Uploading the test dataset for model prediction.

This figure 9 depict a user interface element where a new, unseen dataset is uploaded for the model to make predictions on.

The figure 10 show the results of the model's prediction on the uploaded test dataset. It display the predicted pump status (ON/OFF) for each data point in the uploaded set.

```
cotton      0
638       522
16         18
Name: 0, dtype: int64
Model Predicted of Row 0 Test Data is---> OFF
cotton      0
638       741
16         22
Name: 1, dtype: int64
Model Predicted of Row 1 Test Data is---> OFF
cotton      0
638       798
16         32
Name: 2, dtype: int64
Model Predicted of Row 2 Test Data is---> OFF
cotton     0
638       59
16        20
Name: 3, dtype: int64
Model Predicted of Row 3 Test Data is---> ON
```

Fig. 10: Model Prediction on Uploaded Test data.

Table 1: Performance metrics of Bernoulli Naïve Bayes Classifier and Ridge Classifier Model.

	Algorithm Name	Precison	Recall	FScore	Accuracy
0	BernoulliNB Classifier	37.50000	50.000000	42.857143	75.000000
1	Ridge Classifier	97.87234	93.333333	95.341615	96.666667

This table summarizes the performance metrics (precision, recall, F1-score, accuracy) for both the Bernoulli Naive Bayes and Ridge Classifier models, allowing for a side-by-side comparison of their effectiveness.

The Ridge Classifier significantly outperforms the Bernoulli Naive Bayes model in terms of accuracy (96.67% vs 75%).

F1-score is available for both models (42.86 for Naive Bayes and 95.34 for Ridge Classifier), indicating a moderate performance for Naive Bayes and a good balance between precision and recall for Ridge Classifier.

While the precision and recall values are not available for the Ridge Classifier in the information you described, the high F1-score suggests a good performance on both metrics.

5. CONCLUSION AND FUTURE SCOPE

The project successfully demonstrates the potential of integrating machine learning techniques into irrigation scheduling to address the inefficiencies of traditional methods. By leveraging real-time environmental data such as soil moisture, temperature, and crop type, the system can predict the optimal times for activating irrigation pumps, thus optimizing water use in agriculture. The development and evaluation of machine learning models, specifically Bernoulli Naive Bayes and Ridge Classifier, indicate that these models can significantly enhance irrigation management by reducing water waste, improving crop health, and minimizing labor requirements. The Ridge Classifier, in particular, outperformed the Bernoulli Naive Bayes model in terms of precision, recall, F1-score, and accuracy, proving its suitability for this application. Future work could focus on scaling the system for use in different agricultural settings and climates. This would involve adapting the models to account for region-specific crops, soil types, and environmental conditions. Intelligent irrigation scheduling can be integrated with other precision agriculture practices such as fertilization and pest control, creating a comprehensive smart farming system that optimizes all aspects of crop management.

REFERENCES

[1] L. Oborkhale, A. Abioye, B. Egonwa and T. Olalekan, “Design and implementation of automatic irrigation control system,” IOSR Journal of Computer Engineering, vol. 17, no. 4, pp. 99–111, 2015.

[2] R. Koech and P. Langat, “Improving irrigation water use efficiency: A review of advances, challenges and opportunities in the Australian context,” Water, vol. 10, no. 12, pp. 1771, 2018.

[3] H. Benyezza, M. Bouhedda, K. Djellout and A. Saidi, “Smart irrigation system based thingspeak and arduino,” in 2018 Int. Conf. on Applied Smart Systems (ICASS), Medea, Algeria, pp. 1–4, 2018.

[4] E. A. Abioye, M. S. Z. Abidin, M. S. A. Mahmud, S. Buyamin, M. H. I. Ishak et al., “A review on monitoring and advanced control strategies for precision irrigation,” Computers and Electronics in Agriculture, vol. 173, pp. 105441, 2020.

[5] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani et al., “The rise of ‘big data’ on cloud computing: Review and open research issues,” Information Systems, vol. 47, pp. 98–115, 2015.

[6] J. Gubbi, R. Buyya, S. Marusic and M. Palaniswami, “Internet of things (IoTA vision, architectural elements, and future directions,” Future Generation Computer Systems, vol. 29, no. 7, pp. 1645–1660, 2013.

[7] X. Zhang and H. Khachatryan, “Investigating homeowners’ preferences for smart irrigation technology features,” Water, vol. 11, no. 10, pp. 1996, 2019.

[8] F. Visconti, J. M. d. Paz, D. Martínez and M. J. Molina, “Laboratory and field assessment of the capacitance sensors decagon 10HS and 5TE for estimating the water content of irrigated soils,” Agricultural Water Management, vol. 132, pp. 111–119, 2014.

[9] M. S. Munir, I. S. Bajwa, M. A. Naeem and B. Ramzan, “Design and implementation of an IoT system for smart energy consumption and smart irrigation in tunnel farming,” Energies, vol. 11, no. 12, pp. 3427, 2018.

[10] A. A. Ahmed, S. Al Omari, R. Awal, A. Fares and M. Chouikha, “A distributed system for supporting smart irrigation using internet of things technology,” Engineering Reports, pp. 1–13, 2020. https://doi.org/10.1002/eng2.12352.

[11] A. H. Blasi, M. A. Abbadi and R. Al-Huweimel, “Machine learning approach for an automatic irrigation system in southern Jordan valley,” Engineering, Technology & Applied Science Research, vol. 11, no. 1, pp. 6609–6613, 2021.