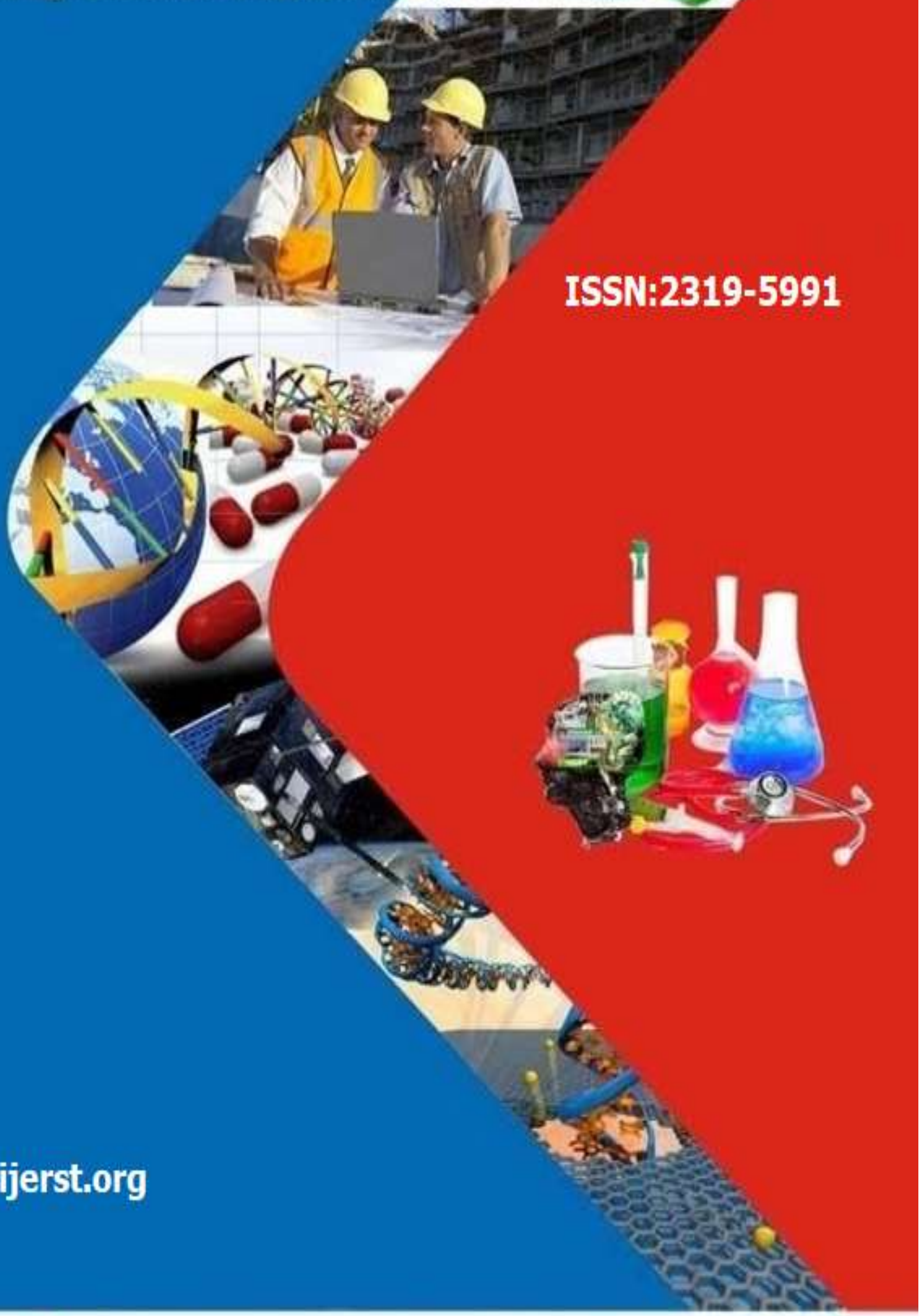
IJERST

International Journal of Engineering Research and Science & Technology



ISSN:2319-5991

www.ijerst.org

E-mail: editor@ijerst.org or ijerst.editor@gmail.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

MACHINE LEARNING BASED IRRIGATION SCHEDULING FOR SMART FARMING SYSTEMS

Dr M Madhusudhan Reddy[1], T.Chandana[2] , T.Maheswari[2], N.Srujan[2], K.Akshith[2]

[1]Professor, [2]UG Student, [1,2] Department of Computer Science and Engineering(DS)

Sree Dattha Group of Institutions, Sheriguda, Hyderabad, Telangana

ABSTRACT

This 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.

Keywords: Irrigation Scheduling, Smart Farming, Machine Learning, Soil Moisture Sensors, Real-time Data, Bernoulli Naive Bayes, Ridge Classifier, Water Optimization, Automated Irrigation, Climate-aware Agriculture.

1.INTRODUCTION

1.1 History

The history of irrigation dates back thousands of years, with early civilizations developing ingenious methods to manage water for agriculture. [1] 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. [2] These early techniques laid the foundation for modern irrigation practices, demonstrating humanity's innate desire to harness water for agricultural purposes.

[3] Throughout history, irrigation has played a vital role in supporting agricultural development and sustaining civilizations. [4] The advent of irrigation allowed farmers to cultivate crops in arid regions and increase food production, leading to population growth and societal advancement. [5] In ancient Rome, sophisticated aqueducts were constructed to transport water over long distances, enabling large-scale farming and urbanization.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

During the Middle Ages, Islamic scholars made significant contributions to irrigation technology, developing innovative techniques such as qanats and water wheels. [6] 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. [7] 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. [8] 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.

In recent decades, the focus has shifted towards sustainable irrigation practices and the integration of technology into agricultural water management. [9] 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. [10] 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. [10] 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. [11] 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 [12]. Hassan-Esfahani et al. [13] 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

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

growth stage to water stress.In Shen et al. [14], 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. [15] 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. [16] 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 [17]. 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. [18] 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 [19] 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. [20], an enhanced data driven and monitoring modelling of the dynamics of variables affected the irrigation of mustard leaf plants is proposed.

3.PROPOSED SYSTEM

3.1 Overview

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

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

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.

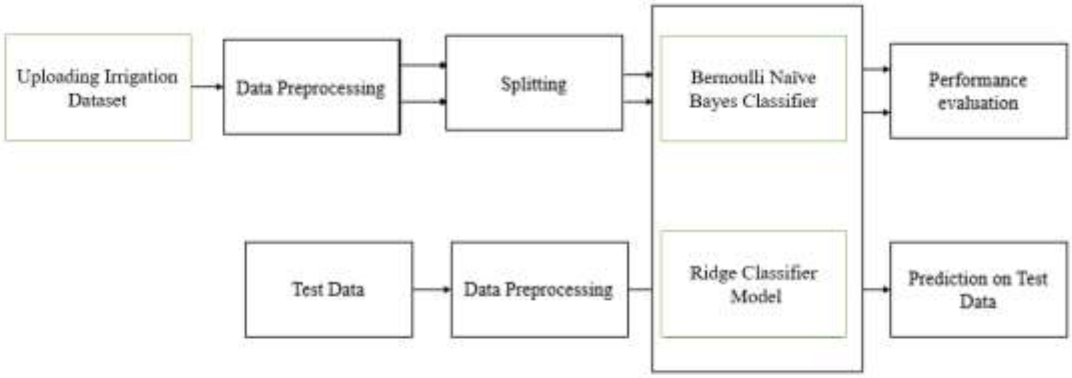


Fig. 1 Block Diagram of the Proposed System.

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

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- Splitting dataset into training and test set

Importing Libraries: To perform data preprocessing using Python, we need to import some predefined Python libraries. These libraries are used to perform some specific jobs. There are three specific libraries that we will use for data preprocessing, which are:

Numpy: Numpy Python library is used for including any type of mathematical operation in the code. It is the fundamental package for scientific calculation in Python. It also supports to add large, multidimensional arrays and matrices. So, in Python, we can import it as:

import numpy as nm

Here we have used nm, which is a short name for Numpy, and it will be used in the whole program.

Matplotlib: The second library is matplotlib, which is a Python 2D plotting library, and with this library, we need to import a sub-library pyplot. This library is used to plot any type of charts in Python for the code. It will be imported as below:

import matplotlib.pyplot as mpt

Here we have used mpt as a short name for this library.

Pandas: The last library is the Pandas library, which is one of the most famous Python libraries and used for importing and managing the datasets. It is an open-source data manipulation and analysis library. Here, we have used pd as a short name for this library. Consider the below image:

```
# importing libraries
import numpy as nm
import matplotlib.pyplot as mtp
import pandas as pd
```

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.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

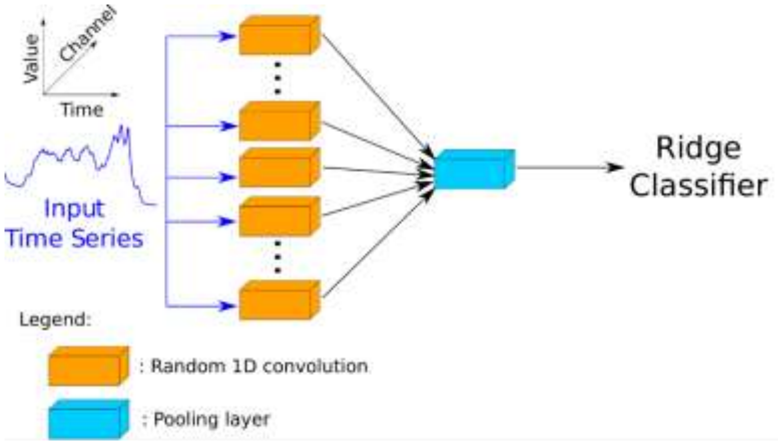


Fig. 2: Architectural diagram of Ridge Classifier model.

Principle:

The principle behind the Ridge Classifier lies in its regularization technique, which adds a penalty term to the objective function. This penalty term, based on the L2-norm of the coefficient vector, constrains the magnitude of the coefficients. By doing so, the Ridge Classifier reduces the model's complexity, mitigating overfitting and enhancing generalization performance.

Working:

- **Objective Function**:
    - The objective function of the Ridge Classifier combines the logistic loss function with the regularization term. Mathematically, it can be expressed as:

$$J(\theta) = LogisticLoss(\theta) + \alpha \sum_{i=1}^{n} \theta_i^2$$

    - Where:
        - $J(\theta)$ is the total cost function.
        - $LogisticLoss(\theta)$ represents the logistic loss function, which measures the difference between the actual and predicted probabilities.
        - $\alpha\ \alpha$ is the regularization parameter (lambda), controlling the degree of regularization.
        - $\sum i = 1 n\ \theta\ i\ 2$ is the L2-norm of the coefficient vector $\theta\ \theta$, representing the sum of squared coefficients.
- **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.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

- **Regularization Term**:
    - The regularization term $\alpha \sum i = 1 n \, \theta \, i \, 2$ is added to the objective function to penalize large coefficients. This term is based on the L2-norm of the coefficient vector θ θ. 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 α α 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 α α.
- **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

4.1 Implementation Description

The project focuses on developing a machine learning-based irrigation scheduling system for smart farming, utilizing various machine learning algorithms to predict whether an irrigation pump should be turned on or off based on different environmental and crop-related factors. The primary goal is to optimize water usage, thereby enhancing crop yield and promoting sustainable farming practices.

Initially, the dataset is imported and loaded into a pandas Data Frame. The dataset contains multiple features, including environmental conditions and crop types, along with a target variable indicating the pump status (ON or OFF). Basic exploratory data analysis is conducted to understand the dataset's structure, summary statistics, and correlation between features. Label encoding is applied to the categorical variable 'crop' to convert it into numerical format, making it suitable for machine learning algorithms.

Visualization techniques, such as count plots, are used to illustrate the distribution of the pump status, providing insights into the dataset. This is essential for understanding the balance of classes in the target variable, which can impact the performance of the classifiers.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

The dataset is then split into training and testing sets using the **train_test_split** method from Scikit-learn, with 70% of the data allocated for training and 30% for testing. This split ensures that the models are trained on a substantial portion of the data while reserving enough data to evaluate their performance effectively.

Two machine learning models are employed: Bernoulli Naive Bayes (BNB) and Ridge Classifier. These models are chosen for their distinct characteristics. The Bernoulli Naive Bayes classifier is suitable for binary and categorical data, making it a good fit for this problem. On the other hand, the Ridge Classifier, which is a linear model, can handle multiclass classification and is effective when the dataset has a high-dimensional feature space.

For each model, performance metrics such as precision, recall, F1-score, and accuracy are calculated using the test set predictions. These metrics provide a comprehensive evaluation of the model's performance, considering both the precision of the predictions and the recall, or how well the model identifies the positive class. Additionally, confusion matrices are plotted to visualize the performance, highlighting the number of true positives, true negatives, false positives, and false negatives.

The models are saved to disk using the **numpy.save** function, allowing for their reuse without retraining. This is particularly useful for deploying the models in a production environment where quick and efficient predictions are necessary.

The Ridge Classifier outperforms the Bernoulli Naive Bayes classifier in terms of precision, recall, F1-score, and accuracy, indicating its suitability for this particular application. This performance is further validated by testing the model on an unseen test dataset, where the Ridge Classifier's predictions are analyzed, and the corresponding rows from the dataset are printed along with the predicted pump status.

4.2 Dataset Description

The dataset is designed to facilitate the development of a machine learning-based irrigation scheduling system for smart farming. It comprises four primary columns, each representing crucial aspects of the farming environment and the system's operational parameters. Below is a detailed description of each column:

Crop: The "crop" column contains categorical data indicating the type of crop being cultivated. Each crop type is typically encoded numerically to facilitate the application of machine learning algorithms. Different crops have varying water requirements, and this feature helps the model to consider crop-specific irrigation needs when predicting the pump status. For instance, crops like rice may require more frequent watering compared to crops like wheat.

Moisture: The "moisture" column contains numerical data representing the soil moisture level. This is a critical feature for irrigation scheduling, as soil moisture directly influences the need for watering. The moisture level is often measured using sensors placed in the soil, providing real-time data that helps in making precise irrigation decisions. Lower moisture levels indicate the need for the pump to be turned on to water the crops, while higher levels suggest sufficient moisture, indicating the pump can remain off.

Temperature: The "temperature" column holds numerical data reflecting the ambient temperature in the farming area. Temperature affects evapotranspiration rates, which in turn influences soil moisture levels. Higher temperatures can lead to increased water loss from both the soil and plants,

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

necessitating more frequent irrigation. Conversely, cooler temperatures reduce water loss, potentially decreasing the need for irrigation. This feature is essential for adjusting irrigation schedules based on weather conditions to optimize water use.

Pump: The "pump" column is the target variable in the dataset, containing binary data indicating the status of the irrigation pump. It has two possible values:

- 0 (OFF): Indicates that the pump is not currently operating.
- 1 (ON): Indicates that the pump is operating to irrigate the crops.

This column is used as the dependent variable that the machine learning models aim to predict based on the input features (crop type, soil moisture, and temperature). Accurate predictions of this column help in automating the irrigation process, ensuring that crops receive the right amount of water at the right time.

4.3 Results Description

The figure 1 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 3: Presents the Sample Dataset of this project.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

	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 2: Label Encoded Dataset.

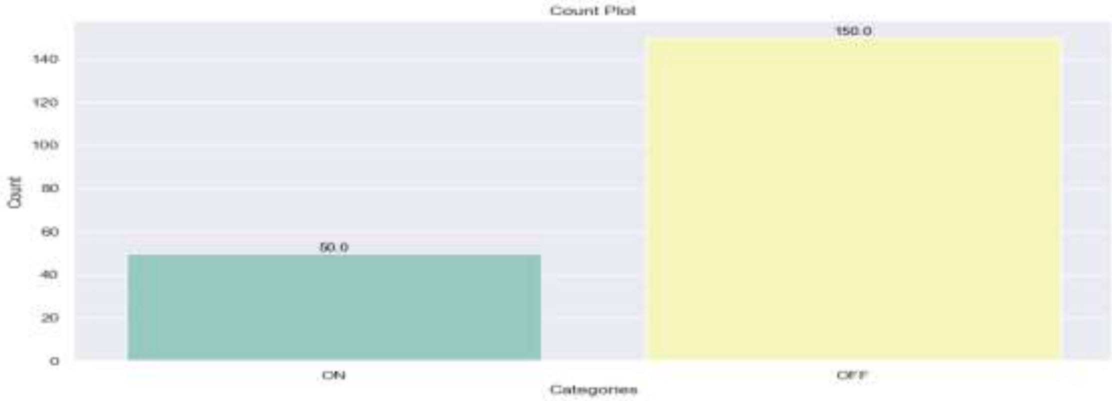


Fig 4: Count of each label in Dataset.

```
BernoulliNBClassifier Accuracy    : 75.0
BernoulliNBClassifier Precision   : 37.5
BernoulliNBClassifier Recall      : 50.0
BernoulliNBClassifier FSCORE      : 42.857142857142854

 BernoulliNBClassifier classification report
              precision    recall  f1-score   support

          ON       0.00      0.00      0.00         0
         OFF       1.00      0.75      0.86        60

    accuracy                           0.75        60
   macro avg       0.50      0.38      0.43        60
weighted avg       1.00      0.75      0.86        60
```

Fig. 5: Performance metrics of Bernoulli Naïve Bayes Classifier.

This figure 4 describes the performance metrics of Bernoulli Naïve Bayes Classifier.

Accuracy: 75% of the time, the model predicted the correct pump status (ON or OFF) for the test data.

Precision: Out of all the instances where the model predicted the pump to be ON, only 37.5% were truly ON based on the actual labels in the test data. There were many false positives (model predicted ON but actual label was OFF).

Recall: Out of all the actual ON instances in the test data, the model was only able to identify 50.0% of them correctly. There were many false negatives (model predicted OFF but actual label was ON).

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

F1-score: This metric is a harmonic mean of precision and recall, trying to balance between the two. A score of 42.86 indicates a moderate performance on the Bernoulli Naive Bayes model for this task.

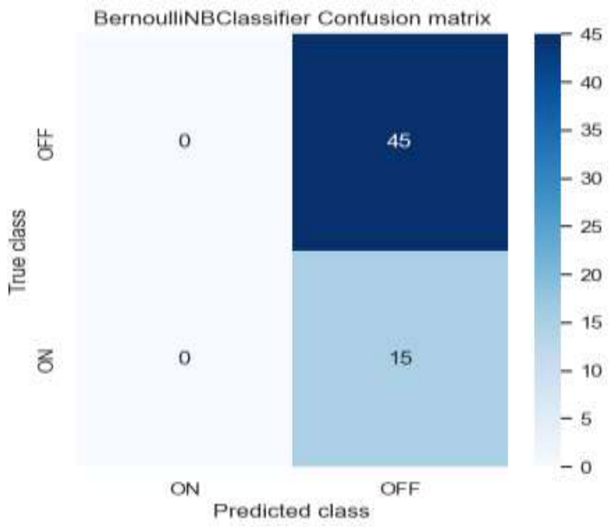


Fig 6: Confusion matrix of Bernoulli Naïve Bayes Classifier.

The figure 5 confusion matrix visualizes the performance of the Bernoulli Naive Bayes classifier by showing the number of true positives, true negatives, false positives, and false negatives. It helps in understanding the types of errors the model makes and the overall effectiveness in classifying the pump status

```
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 6 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).

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

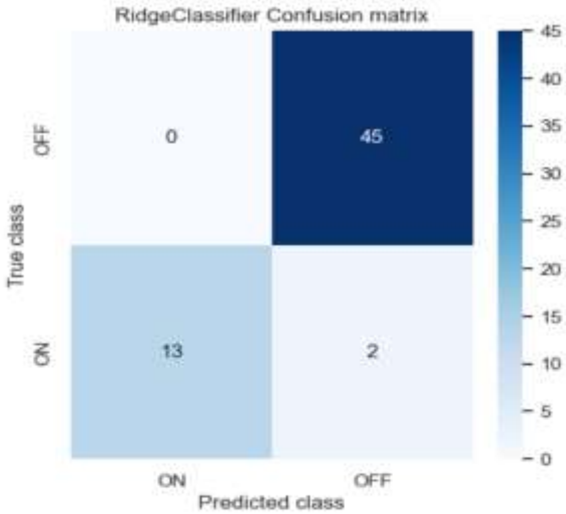


Fig :8 Confusion matrix of Ridge Classifier.

The figure 7 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 8 depict a user interface element where a new, unseen dataset is uploaded for the model to make predictions on.

This figure 9 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.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

```
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%).

5.CONCLUSION

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

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.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

[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.

[12] A. Vij, S. Vijendra, A. Jain, S. Bajaj, A. Bassi et al., “IoT and machine learning approaches for automation of farm irrigation system,” Procedia Computer Science, vol. 167, pp. 1250–1257, 2020.

[13] L. H. Esfahani, A. T. Rua and M. McKee, “Assessment of optimal irrigation water allocation for pressurized irrigation system using water balance approach, learning machines, and remotely sensed data,” Agricultural Water Management, vol. 153, pp. 42–50, 2015.

[14] H. Shen, K. Jiang, W. Sun, Y. Xu and X. Ma, “Irrigation decision method for winter wheat growth period in a supplementary irrigation area based on a support vector machine algorithm,” Computers and Electronics in Agriculture, vol. 182, pp. 106032, 2021.

ISSN 2319-5991 www.ijerst.com
Vol. 21, Issue 2, 2025

[15] H. N. Hellín, J. M. del-Rincon, R. D. Miguel, F. Soto-Valles and R. Torres-Sánchez, “A decision support system for managing irrigation in agriculture,” Computers and Electronics in Agriculture, vol. 124, pp. 121–131, 2016.

[16] J. Cardoso, A. Gloria and P. Sebastiao, “Improve irrigation timing decision for agriculture using real time data and machine learning,” in 2020 Int. Conf. on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, pp. 1–5, 2020.

[17] A. Glória, J. Cardoso and P. Sebastião, “Sustainable irrigation system for farming supported by machine learning and real-time sensor data,” Sensors, vol. 21, no. 9, pp. 3079, 2021.

[18] M. S. Munir, I. S. Bajwa, A. Ashraf, W. Anwar and R. Rashid, “Intelligent and smart irrigation system using edge computing and IoT,” Complexity, vol. 2021, pp. 1–16, Feb. 2021.

[19] R. Torres-Sanchez, H. Navarro-Hellin, A. Guillamon-Frutos, R. San-Segundo, M. C. Ruiz-Abellón et al., “A decision support system for irrigation management: Analysis and implementation of different learning techniques,” Water, vol. 12, no. 2, pp. 548, 2020.

[20] E. A. Abioye, M. S. Z. Abidin, M. S. A. Mahmud, S. Buyamin, M. K. I. AbdRahman et al., “IoT-Based monitoring and data-driven modelling of drip irrigation system for mustard leaf cultivation experiment,” Information Processing in Agriculture, pp. 1–32, 2020. https://doi.org/10.1016/j.inpa.2020.05.004.