International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning Model on Soil-Plant-Water relationships for predicting Plant-Water requirement for Smart Irrigation

N Poojitha[1], M S L N Uday[2], K Hemasundara Rao[3], M Ravi[4]

GMR Institute of technology

ABSTRACT:

Smart irrigation systems have revolutionized how we manage water resources in agriculture, improving crop yields, reducing water wastage, and enhancing overall sustainability. However, predicting plant-water requirements remains complex, requiring a deep understanding of the complex relationships between soil, plants, and water. This study proposes a machine learning model that uses advanced algorithms and data analytics to predict plant-water requirements based on soil-plant-water relationships. The model is built on a range of data sources, including soil parameters, temperature, and other weather data, to provide a real-time understanding of plant-water requirements. The proposed model is evaluated using a comprehensive dataset and compared to traditional irrigation methods, showing significant improvements in water efficiency, and overall sustainability. The results of this study highlight the potential of machine learning models in optimizing irrigation practices and promoting sustainable agriculture and provide a framework for future research and development in this area.

Keywords: Machine Learning, Soil-Plant-Water Relationships, Smart Irrigation, Plant-Water Requirement, Water Efficiency, Crop Yields, Sustainability.

Introduction:

Smart watering systems are revolutionizing agriculture by optimizing water usage and enhancing crop production. These innovative systems intelligently assess the water needs of plants, allowing farmers to use water more responsibly and efficiently. By leveraging advanced computer programs, farmers can predict how much water is needed based on various factors such as soil conditions and the specific requirements of different crops. There are several types of smart irrigation methods that cater to different farming needs. Drip irrigation, for instance, delivers water directly to the roots of plants, minimizing water loss and ensuring that each plant receives the hydration it needs without excess waste. Sprinkler irrigation, which uses overhead systems to distribute water evenly across larger fields, is popular in extensive farming landscapes. Micro-spray irrigation focuses on watering specific areas with precision, while subsurface drip irrigation works below the soil's surface to nourish roots directly. Precision irrigation systems elevate the concept of smart irrigation by incorporating sensors and data analytics to monitor soil conditions continuously. This real-time feedback allows farmers to adjust their water usage dynamically, ensuring maximum efficiency in water application. The idea of managing water for farming dates back thousands of years, with ancient civilizations such as Mesopotamia, Egypt, and the Indus Valley employing rudimentary irrigation systems. These early techniques laid the foundation for the more sophisticated methods we have today, such as those developed in ancient China, where farmers utilized buried clay pots to gradually feed their crops. Throughout the 20th century, advancements like plastic emitters made irrigation systems more effective, paving the way for innovations in the 21st century, including micro-spray technology and precision irrigation. The early 2000s saw the introduction of machine learning techniques to predict plant water needs, marking a significant shift in agricultural practices. As sensor technology and data analysis capabilities advanced, machine learning models began to optimize irrigation by analyzing real-time data, which dramatically reduced water waste and boosted crop yields. These models can determine the best watering schedules based on historic and current data, ensuring crops receive the ideal amount of water when they need it the most. This not only leads to healthier and more productive plants but also promotes sustainability by minimizing energy use and conserving water. Furthermore, by reducing the reliance on manual labor, machine learning models cut down on operational costs for farmers, making smart irrigation systems a financially viable option. In summary, smart irrigation presents an exciting area for research and innovation due to its numerous advantages over traditional techniques. By incorporating real-time weather and soil moisture data, these systems conserve water and enhance crop yields, which is vital as we face challenges like food security and climate change. The flexibility of smart irrigation allows systems to be tailored to varied farming operations, making them relevant for farmers of all sizes. Moreover, the integration of technologies like the Internet of Things (IoT) and drones can create a comprehensive farming ecosystem that maximizes efficiency. As smart irrigation continues to evolve, it holds the potential to change the landscape of agriculture in a sustainable and profitable way, providing farmers with the tools they need to thrive in a challenging environment.

Methodology:

1. DEFINING THE PROBLEM:

Develop a machine learning model that accurately predicts plant-water requirements for smart irrigation systems, considering the complex interactions between soil, plants, and water. The model should be able to analyze data from various sources, including soil moisture, temperature, humidity, rainfall, and plant growth metrics, to provide real-time insights into optimal watering schedules.

2. DATA COLLECTION:

Relevant data on soil N, P, K (Nitrogen, Phosphorus, Potassium) values, temperature, soil moisture, soil type, soil pH, and humidity is collected from trusted sources.

3. DATA PREPROCESSING:

3.1. Data is carefully studied and understood to identify its structure, quality, and potential issues.

3.2. Statistical tools like boxplot, heatmap, pair plot, and hist plot are used to visualize the data to find the co-relation between the features and to find the outliers and missing values i.e. discontinuity in the data.

3.3. Duplicate records are deleted to ensure data integrity and prevent biases in the analysis.

3.4. Outliers are detected and handled by IQR (Interquartile Range)-based outlier treatment method.

3.5. Categorical data is encoded into a format that machine learning algorithms can process by converting them into numerical representations.

3.6. NumPy, Pandas, and Scikit-learn libraries in Python are used for efficient data preprocessing tasks such as data cleaning, transformation, and encoding.

3.7. Pre-processed data is split into training and testing sets using the technique of train-test split to evaluate the model's performance.

3.8. A balance is maintained between training and testing data to ensure the model's generalization and robustness.

4. FEATURE SELECTION:

4.1. Key features that influence plant-water requirements are identified.

4.2. Relevant features that contribute to accurate predictions in the machine learning model are selected and others are dropped.

5. MODEL DEVELOPMENT

5.1. Suitable machine learning algorithms namely SVM, Logistic Regression, KNN, Random Forest, Decision tree, Bagging, and Boosting algorithms are selected based on the characteristics of the dataset and the complexity of the problem.

5.2. These models are trained with the already split training dataset on soil-plant-water relationships.

6. MODEL EVALUATION

6.1. Cross-validation is performed to assess the model's performance on different subsets of the data and ensure its robustness and generalization capability.

6.2. The performance of the model is evaluated using evaluation metrics F1, score, recall, precision, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and accuracy to assess its predictive capabilities.

7. RESULT ANALYSIS

7.1. The model's predictions are compared with actual plant-water requirements to measure its effectiveness in smart irrigation applications.

7.2. machine learning models are interpreted to understand how they make predictions and the underlying patterns in the data.

7.3. The model's decision boundaries, feature importance, and prediction outcomes are visualized to gain insights into the plant-water relationship predictions.

8. MODEL FINALIZATION:

Based on the Model evaluation results, their boundaries, and their prediction outcomes, the most suitable model is selected for deployment.

9. DISCUSSION

9.1. Implications of the machine learning model for optimizing plant-water requirements for smart irrigation systems are discussed.

9.2. The strengths, limitations, and potential applications of the model in sustainable water management practices are addressed.

9.3. Key findings of the report are summarized and the significance of using machine learning in predicting plant-water requirements for smart irrigation is emphasized.

Results & Discussion:

Exploratory Data Analysis (EDA)

Box Plot

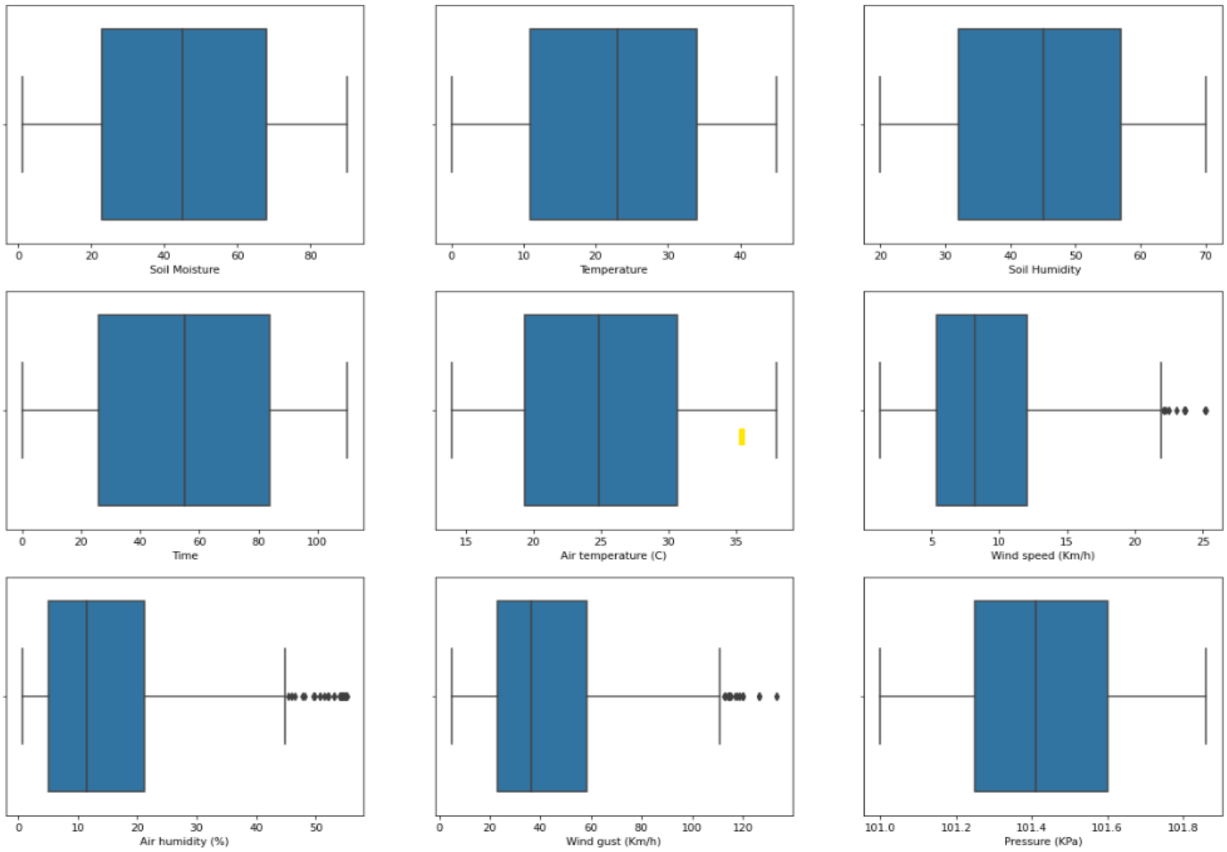

Fig.1.1: Boxplot for outlier detection

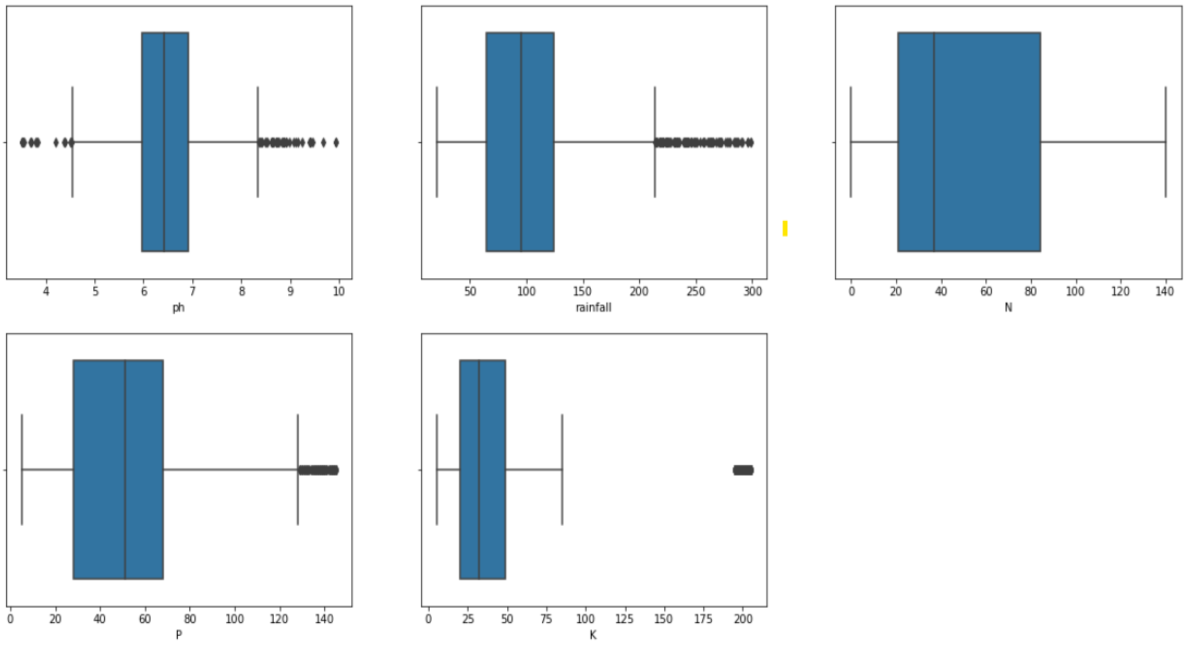

Fig1.2: Boxplot for outlier detection

Interpretations: A box plot, also known as a box-and-whisker plot, is a graphical representation of the distribution of a dataset based on five key summary statistics: minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum. The box in the plot represents the interquartile range (IQR), which spans from Q1 to Q3, with a line indicating the median. Whiskers extend from the box to the minimum and maximum values within a certain range, typically 1.5 times the IQR. Outliers, if present, are plotted individually. Box plots provide insights into the central tendency, spread, and skewness of the data distribution.

Heat Map

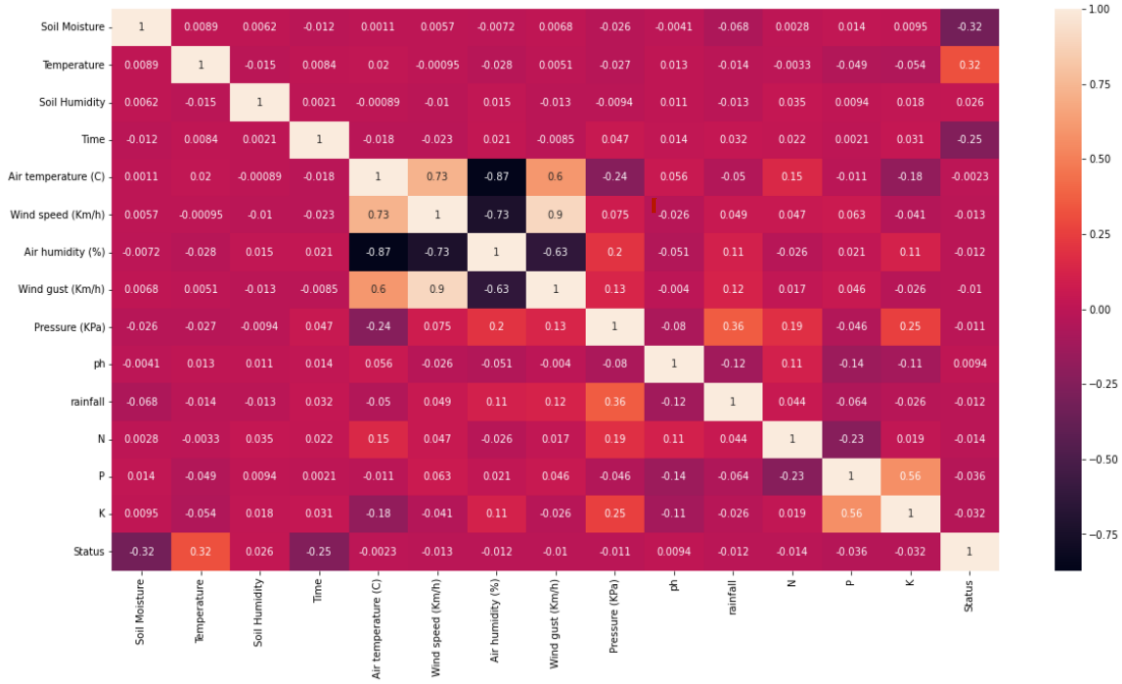

Fig.2: Heat map for correlation analysis

Interpretations: A heatmap is a graphical representation of data where values in a matrix are represented as colors. It's often used to visualize the magnitude of a variable in two dimensions. In a heatmap, each cell in the matrix is assigned a color based on its value. Typically, a color gradient is used, where low values are represented by lighter colors and high values by darker colors, making it easy to identify patterns and variations in the data. Heatmaps are commonly used in various fields like data analysis, biology, and finance to represent complex data sets and identify trends or correlations visually.

Pair Plot

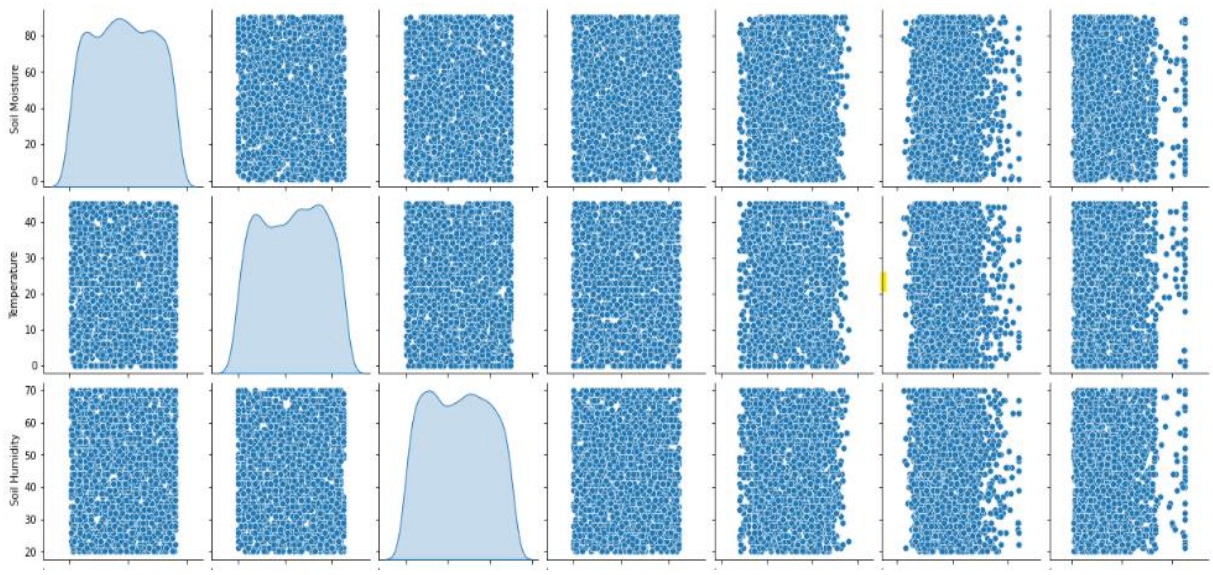

Fig.3.1: Pair plot for correlation analysis

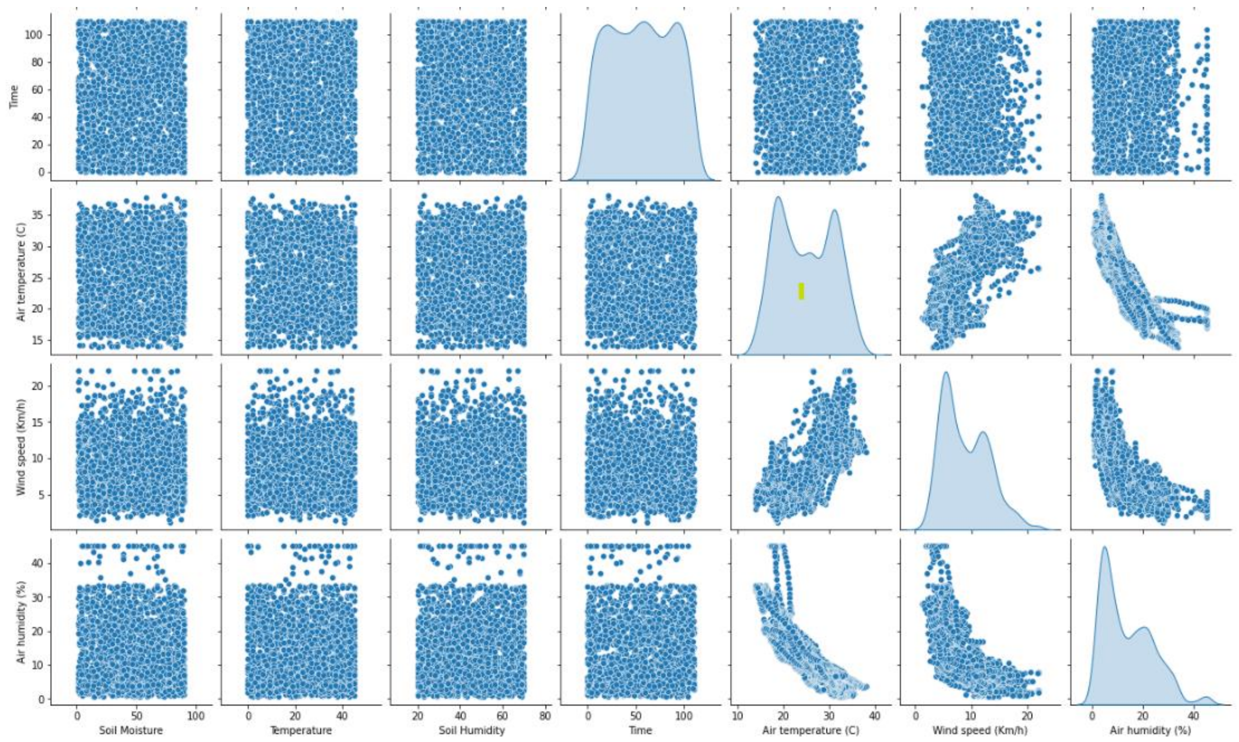

Fig.3.2: Pair plot for correlation analysis

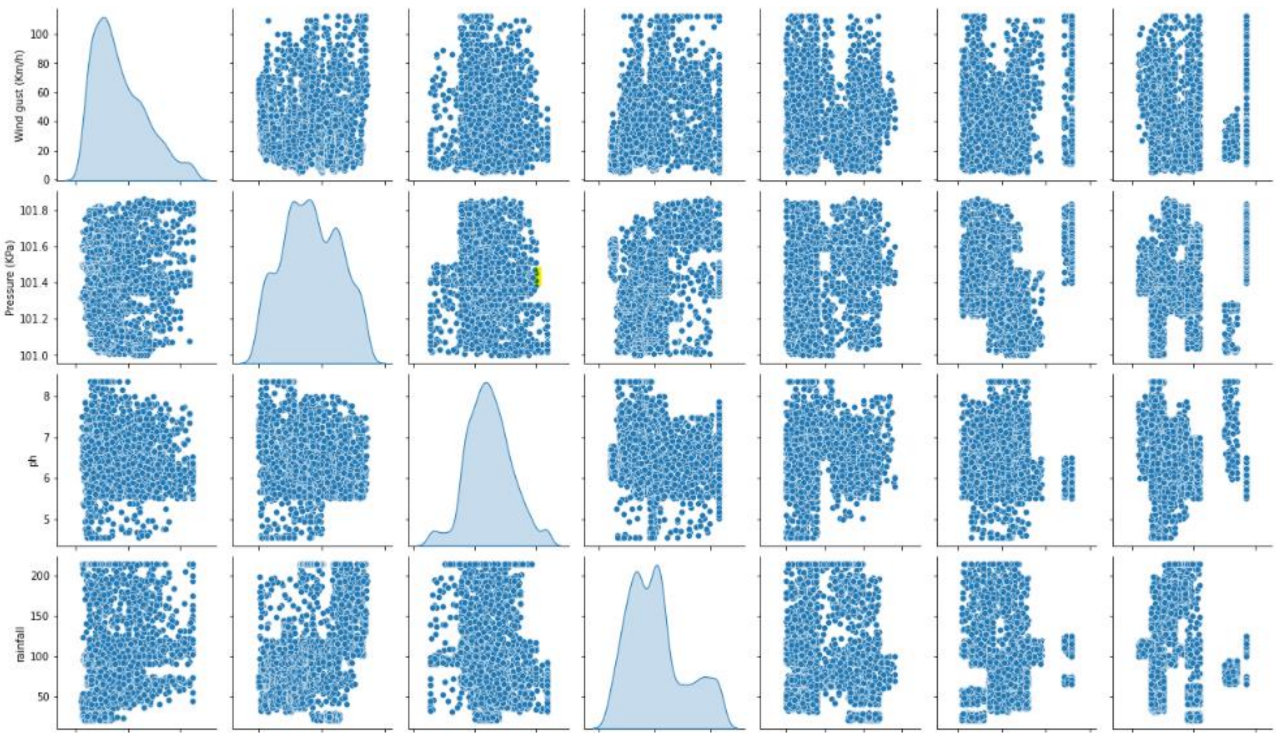

Fig.3.3: Pair plot for correlation analysis

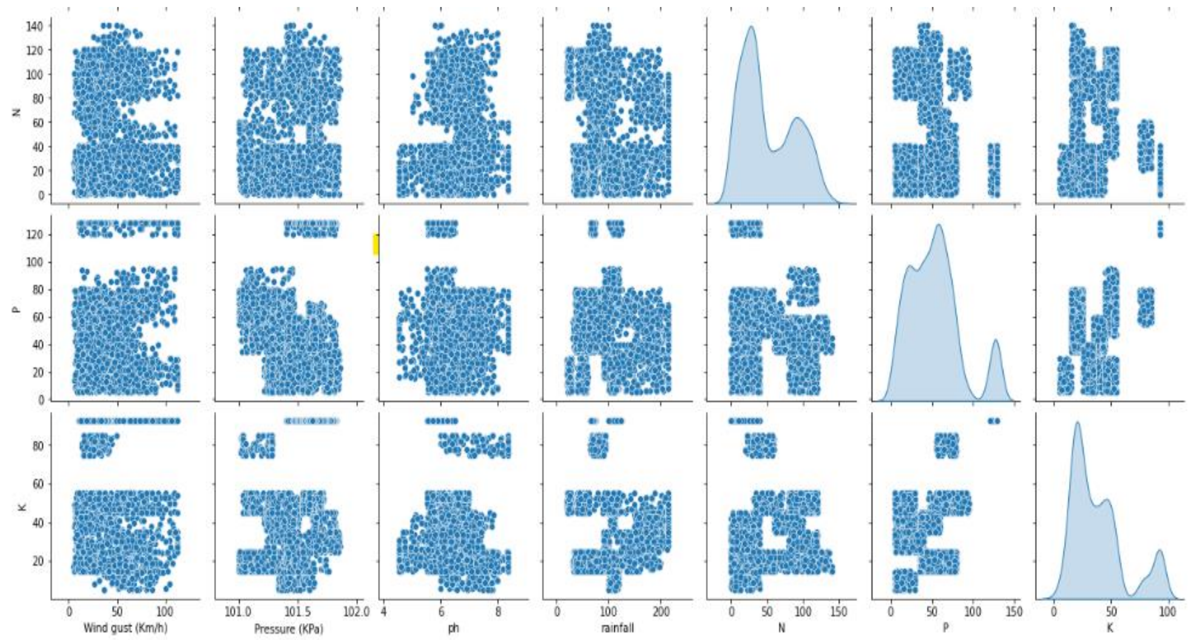

Fig.3.4: Pair plot for correlation analysis

Interpretations: A pair plot, also known as a scatterplot matrix, is a graphical representation of pairwise relationships between variables in a dataset. It consists of a grid of scatterplots where each variable is plotted against every other variable. Along the diagonal of the grid, histograms or kernel density plots may be shown to visualize the distribution of each individual variable. Pair plots are useful for identifying patterns, correlations, and potential outliers in multivariate data, providing a quick and comprehensive overview of the relationships between variables in a dataset. They are commonly used in exploratory data analysis (EDA) to gain insights into the structure of the data before further analysis.

Hist Plot

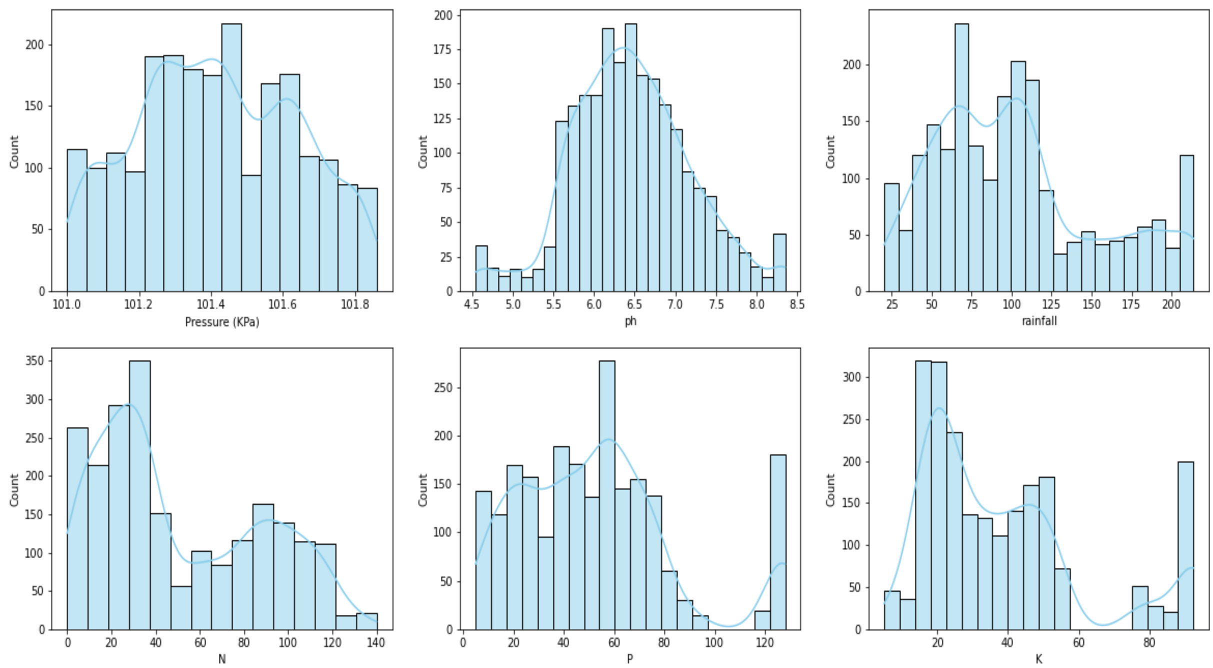

Fig.4.1: Hist plot for a range of values

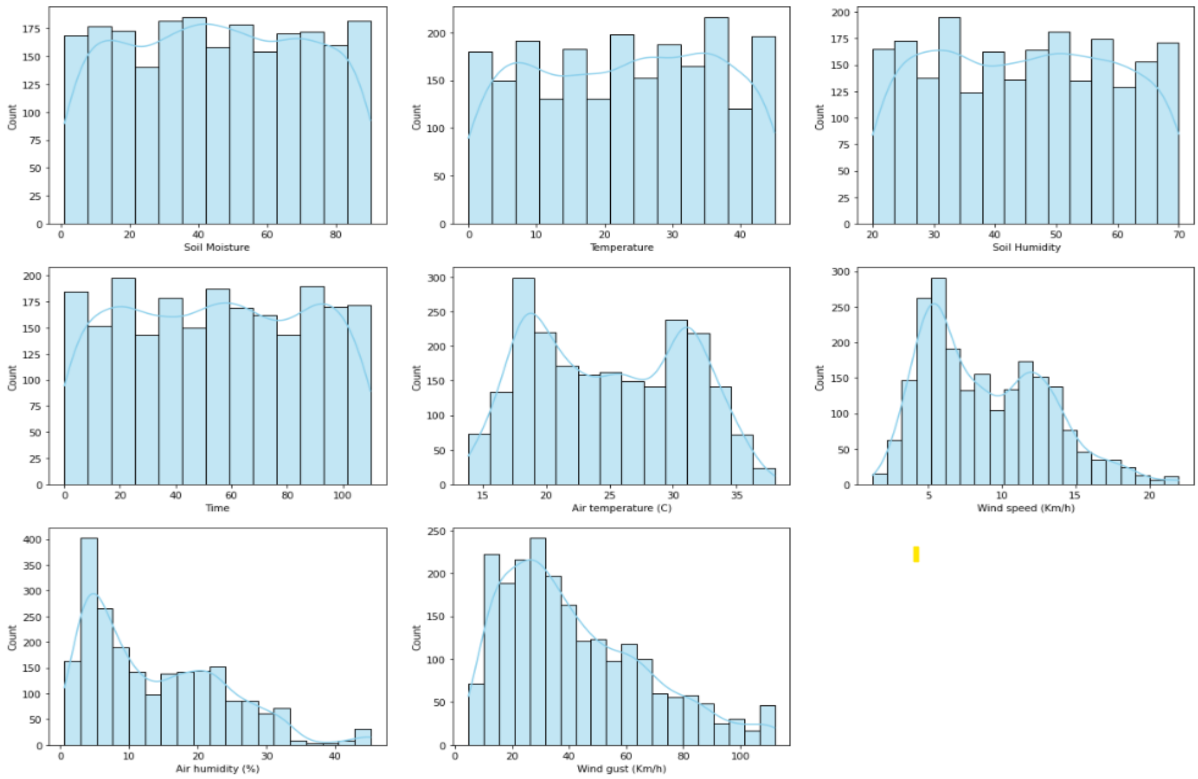

Fig.4.2: Hist plot for a range of values

Interpretations: A hist plot, short for histogram plot, is a graphical representation of the distribution of numerical data. It consists of a series of adjacent rectangles or bars, where the width of each bar represents a range of values (bin) and the height represents the frequency or count of observations falling within that bin. Hist plots provide a visual summary of the distribution of a dataset, allowing for easy identification of patterns such as central tendency, spread, skewness, and potential outliers. They are commonly used in exploratory data analysis (EDA) to understand the underlying structure of the data and to make initial assessments before further analysis or modeling.

Model Accuracy Comparisons

Machine learning models have exhibited significant improvements in water efficiency, and sustainability compared to traditional irrigation methods by predicting optimal watering schedules and detecting deviations in data, these models enable precise and sustainable water management practices that benefit both the environment and agricultural productivity. The accuracies of the respective models are as follows:

ALGORITHM	TRAINING ACCURACY	**TESTING ACCURACY**
SVM	71.48	72.27
Logistic Regression	71.53	72.05
Random Forest	90.52	89.32
KNN	84.83	76.82
Decision Tree	90.23	88.64
Bagging and Boosting	90.32	88.33
Bagging (Decision Tree Classifier)	91 %	88 %
Ensemble Boosting (Random Forest)	99 %	91 %
ADA Boosting	87 %	84 %
Gradient Boosting	93 %	89 %

This table summarizes the performance of various machine learning algorithms in terms of training and testing accuracy. The training accuracy indicates how well each model learned from the training data, while testing accuracy shows its performance on new data. Ensemble methods like Ensemble Boosting with Random Forest achieved the highest training accuracy at 99%, and also excelled in testing accuracy with 91%, indicating strong model performance. Other high-performing algorithms include Gradient Boosting (93% training, 89% testing) and Random Forest (90.52% training, 89.32%

testing). Traditional models like SVM and Logistic Regression have lower but consistent accuracies between training and testing. Meanwhile, K-Nearest Neighbors (KNN) shows a noticeable drop from training to testing, suggesting overfitting, as does ADA Boosting with a slightly lower overall accuracy. The results suggest that ensemble methods, particularly Ensemble Boosting with Random Forest and Gradient Boosting, provide the best balance of high accuracy and generalizability to new data. Decision Tree-based methods like Bagging and Boosting, as well as Random Forest, also performed well, consistently achieving strong testing accuracy. In contrast, algorithms such as SVM, Logistic Regression, and KNN exhibited lower performance on both metrics, with KNN and ADA Boosting specifically showing signs of overfitting. Overall, ensemble approaches like boosting methods and Random Forest emerge as the most promising models for reliable predictions on this dataset, combining high accuracy with effective generalization.

Conclusions:

- Ensemble Boosting (Random Forest) and Gradient Boosting achieve the highest accuracy, with Ensemble Boosting reaching **99% on training and 91% on testing**. This makes ensemble methods the most powerful options for reliable predictions.
- Random Forest and Decision Tree models show strong performance with balanced training and testing accuracies (both close to **90%**), indicating they generalize well to new data without overfitting.
- SVM and Logistic Regression are consistent but have lower accuracies around **71-72%**. These models seem less effective for capturing the complexity in this dataset.
- K-Nearest Neighbors (KNN) and ADA Boosting show high training accuracies (84.83% and 87%, respectively) but drop on testing, indicating they may be overfitting and less reliable on new data.
- Boosting techniques, particularly Ensemble Boosting with Random Forest and Gradient Boosting, deliver top performance. These methods capture complex patterns in the data, making them ideal for high predictive accuracy.
- Ensemble and tree-based models, especially those using boosting, emerge as the best options for accurate and generalizable predictions. Traditional models, while stable, don't perform as well, and models like KNN and ADA Boosting may overfit.

REFERENCES:

[1]. K. S. Pratyush Reddy, Y. M. Roopa, K. Rajeev L.N. and N. S. Nandan, "IoT based Smart Agriculture using Machine Learning," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 130-134 https://doi: 10.1109/ICIRCA48905.2020.9183373.

[2]. Shetty, S., & Smitha, A. B. (2021). Smart agriculture using IOT and machine learning. Transactions on Computer Systems and Networks, 1–16. https://doi.org/10.1007/978-981-16-6124-2_1.

[3]. IORLİAM, A., BUM, S., AONDOAKAA, I. S., IORLIAM, I. B., & SHEHU, Y. (2022). Machine learning techniques for the classification of IOT-enabled smart irrigation data for agricultural purposes. Gazi University Journal of Science Part A: Engineering and Innovation, 9(4), 378–391. https://doi.org/10.54287/gujsa.1141575

[4]. Aditya Gor, Kandarp Joshi, Rushil Togadiya, & Dr. Warish Patel. (2023, March). (PDF) automation in irrigation using IOT and ML based crop ... researchgate.
https://www.researchgate.net/publication/369529211_Automation_in_Irrigation_using_IoT_and_ML_based_Crop_Recommendation_System

[5] Bouali, E.-T., Abid, M. R., Boufounas, E.-M., Hamed, T. A., & Benhaddou, D. (2022). Renewable Energy Integration Into Cloud & IOT-based Smart Agriculture. IEEE Access, 10, 1175–1191. https://doi.org/10.1109/access.2021.3138160
[6] Ramya, S., Swetha, A. M., & Doraipandian, M. (2020). IOT framework for smart irrigation using machine learning technique. Journal of Computer Science, 16(3), 355–363. https://doi.org/10.3844/jcssp.2020.355.363

[7] Smart irrigation system using machine learning and IOT - IRJET. (n.d.). https://www.irjet.net/archives/V9/i6/IRJET-V9I6376.pdf

[8] Hussain, R. (n.d.). Control of irrigation automatically by using Wireless Sensor Network.
https://www.ijsce.org/wp-content/uploads/papers/v3i1/A1390033113.pdf

[9] K.M, Dr. S., Chandra G, Dr. M., & Mrs. Vidya J. (2022, March 3). International Journal of Mechanical Engineering Sustainable Automated ... kalaharijournals. https://kalaharijournals.com/resources/39_MARCH%20ISSUE.pdf

[10c] Author links open overlay panelAnneketh Vij 1, A. V., 1, S. V., 2, A. J., 3, S. B., 4, A. B., & 5, A. S. (2020, April 16). IOT and machine learning approaches for automation of Farm Irrigation System. Procedia Computer Science.
https://www.sciencedirect.com/science/article/pii/S1877050920309078