USMAN INSTITUTE OF TECHNOLOGY

Department of Computer Science

CS321 Artificial Intelligence

Research Paper

TITLE

Data Visualization Techniques in Smart Agriculture

Group Members

HAFSA SHOAIB	21B-054-SE
MAHREEN KHAN	21B-010-SE
ALIZA KHAN	21B-205-SE

Abstract

This paper shows how we can use data visualization to help farmers improve their yields and plan the timing of irrigation. In the cotton farming region of Matiari, we used sensors to collect data on soil moisture, temperature, humidity and more. By turning this information into easy-to-understand charts and graphs, we provide farmers with valuable information to make better decisions. Our results suggest that these visual tools can contribute to early detection of crop health issues and water use efficiency and help them decide when and how much to irrigate their crops and keep their plants healthy, thereby supporting sustainable agricultural practices.

Using IoT sensors, we can collect real-time information about environmental conditions in different locations. We used Python programming to clean the data and create visualizations using libraries.

This study shows that by applying machine learning algorithms how farmers can manage crop health and plan water use. This insight can lead to better management, less water waste and more crop yields.

Introduction

In modern agriculture, water management and crop health management are essential to achieve sustainable agricultural practices. The arrival of the Internet of Things (IoT) has transformed agriculture through the environment encountering vital information that will enable real-time collection of vital environmental data by sensors installed on the monitored farm boats, which can be used to optimize agricultural operations.

Traditionally, farmers have relied on manual monitoring and planned systems to irrigate their crops. But these processes may not always be responsible for changes in environmental conditions, leading to overflows or underflows, which can harm and damage crops.

Data visualization helps us visualize complex information in simple ways, such as through charts and graphs, so we can quickly identify patterns and problems. These observations can identify key factors affecting crop health and help predict optimal irrigation times.

The main objective of this study is to investigate the feasibility of data visualization for crop health monitoring and water irrigation planning. We believe that by making environmental issues more conscious, farmers can identify issues early and optimize their irrigation systems, resulting in healthier crops and optimal watering times.

Methodology

A. Data Collection:

In the cotton farming region of Matiari, we used sensors to collect data. The analytical parameters included humidity, temperature, heat index, soil moisture, soil temperature, salinity, TDS (Total Dissolved Solids), battery capacity, water requirement and more were regularly recorded as per time stamp. The collection spanned from February 2023 to August 2023.

B. Data Preprocessing:

The collected data were processed and analyzed using the following steps:

- <u>Data cleaning:</u> Handling missing values, removing duplicates, detecting outliers and ensuring data consistency.
- <u>Descriptive Statistics:</u> A summary of data using statistical basics such as mean, median, and standard deviation.

C. Machine Learning Models:

Three models were selected to predict water scheduling and crop health based on various agricultural parameters:

- <u>K-Nearest Neighbors (KNN):</u> Chosen for its simplicity and effectiveness in classification tasks. KNN is intuitive and works well with smaller datasets, making it suitable for initial model evaluations in our agricultural data context.
- <u>Naive Bayes:</u> Known for its efficiency and performance with large datasets. Naive Bayes assumes independence among features, which simplifies computations and is particularly useful in scenarios where this assumption holds true.
- <u>Support Vector Machine (SVM):</u> SVM works well with smaller datasets and is useful for finding the optimal hyperplane that separates different classes in the dataset, which is critical for accurate predictions in complex agricultural datasets.

D. Visualization Techniques:

The following visualization techniques were used:

- Histogram: To show trends over time for parameters such as temperature, humidity and soil moisture and more.
- Bar charts: To compare classes or groups in the data.
- Scatter plots: To examine the relationship between variables, such as soil moisture and soil temperature.
- Heatmaps: To identify patterns and correlations between features.
- Line charts: To tune adjustments over periods of time.
- Pie charts: To illustrate proportions within a dataset.
- 3D visualizations: To explore relationships between three variables and add intensity to the evaluation.
- Area charts: To show cumulative totals over the years

E. **Tools and Technologies**
    - Sensors: For data collection in the field.
    - Excel: For initial data storage and preprocessing.
    - Jupyter Notebook: For data analysis, predictions and visualization.
    - Python Libraries: Pandas for data manipulation, sklearn for models, Matplotlib and Seaborn for visualization and more.

Results

A. **Histogram:**

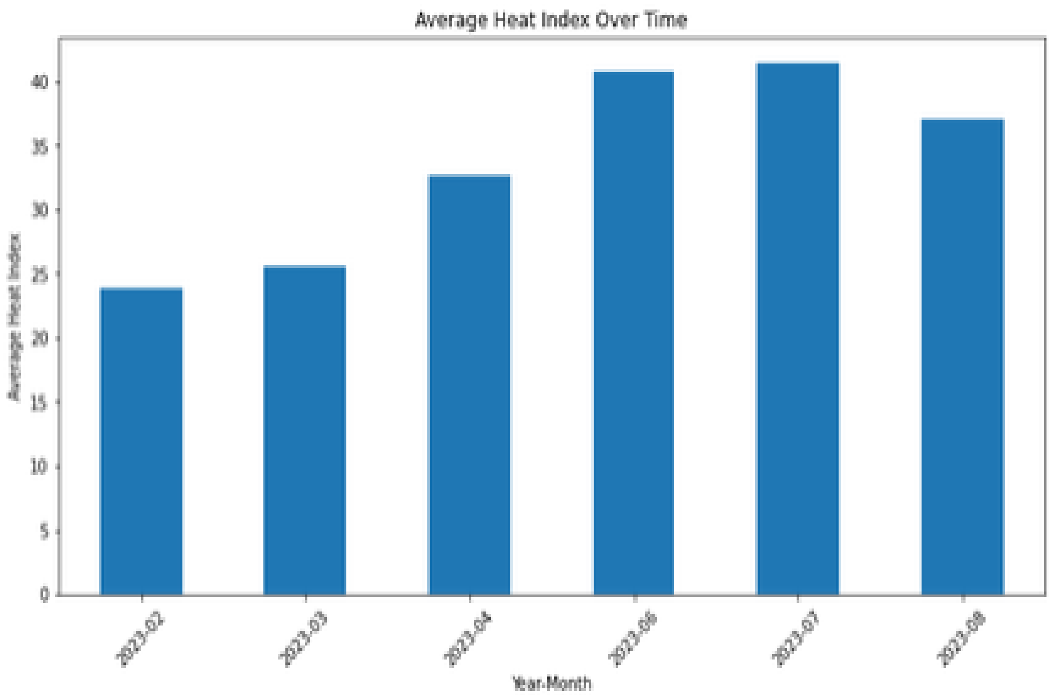

Figure 1: Histogram of Heat Index

The bar graph displays the average heat index from February to August 2023, showing a rising trend with peaks in July and August. The heat index notably increases from March onwards, reaching its highest point in July.

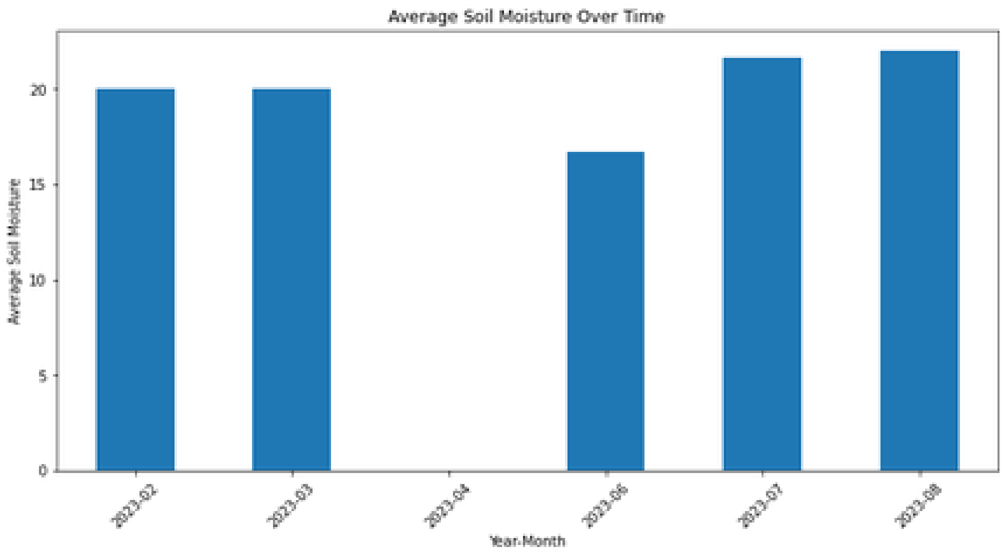

Figure 2: Histogram of soil moisture

The bar graph illustrates the average soil moisture from February to August 2023, with values fluctuating over the period. Soil moisture peaks in March, July, and August, while it dips in April and June.

B. Bar Graph:

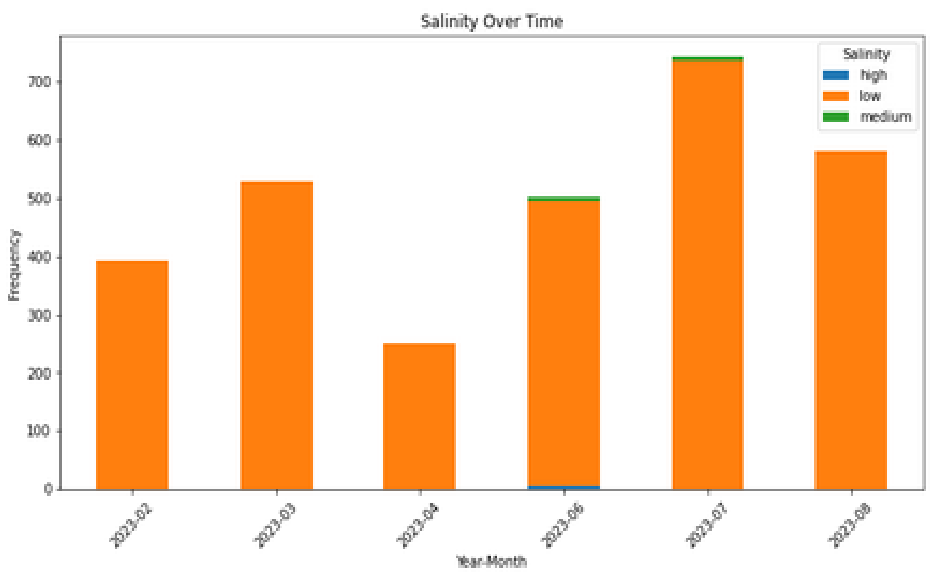

Figure 3: Bar Graph Of Salinity Over Time

The top bar graph shows the frequency of different salinity levels over time, with "medium" salinity being most common and "high" salinity spiking in July 2023.

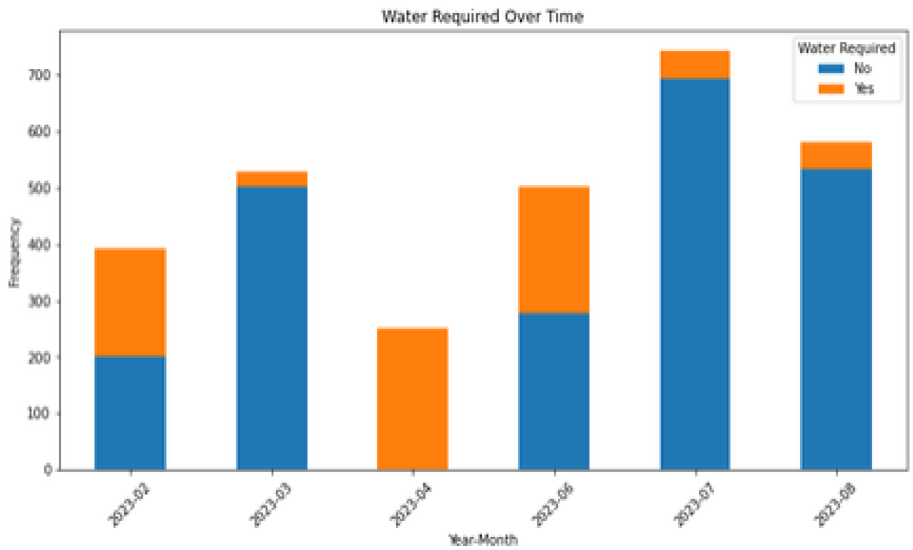

Figure 4: Bar Graph of Water Required

The bottom bar graph displays the frequency of water required over time, with a noticeable increase in water requirements from July 2023 onwards.

OBSERVATION:

Effect of Soil Moisture on Water Requirement over Time:

1. February 2023
    a. Water Required (Yes):
        i. Soil Moisture: 14-15 & 27-34
    b. Water Not Required (No):
        i. Soil Moisture: 21-26

    In February 2023, water was required when soil moisture ranged from 14-15 & 27-34. When soil moisture was between 21-26, water was not needed.

2. March 2023
    a. Water Required (Yes):
        i. Soil Moisture: 20
    b. Water Not Required (No):
        i. Soil Moisture: 18-19 & 21
        ii. In March 2023, water was required when soil moisture level was 20. When soil moisture ranged from 18-19 & 21, water was not required.

3. April 2023
    a. Water Required (Yes):
        i. Soil Moisture: 0

    b. Water Not Required (No):
        i. Soil Moisture: -

In April 2023, soil moisture levels of 0 indicated the need for water. And there is no range when it did not require additional water.

4. June 2023
    a. Water Required (Yes):
        i. Soil Moisture: 60-11
    b. Water Not Required (No):
        i. Soil Moisture: 17-38

    In June 2023, water was needed when soil moisture was between 60-11. When soil moisture was ranged from 17-38, water was not required.

5. July 2023
    a. Water Required (Yes):
        i. Soil Moisture: 14-16
    b. Water Not Required (No):
        i. Soil Moisture: 17-32

    In July 2023, soil moisture levels of 14-16 required water, while levels between 17-32 did not need additional water.

6. August 2023
    a. Water Required (Yes):
        i. Mean Soil Moisture: 15-16
    b. Water Not Required (No):
        i. Mean Soil Moisture: 17-32

    In August 2023, water was required when soil moisture was between 15-16. When soil moisture was between 17-32, water was not needed.

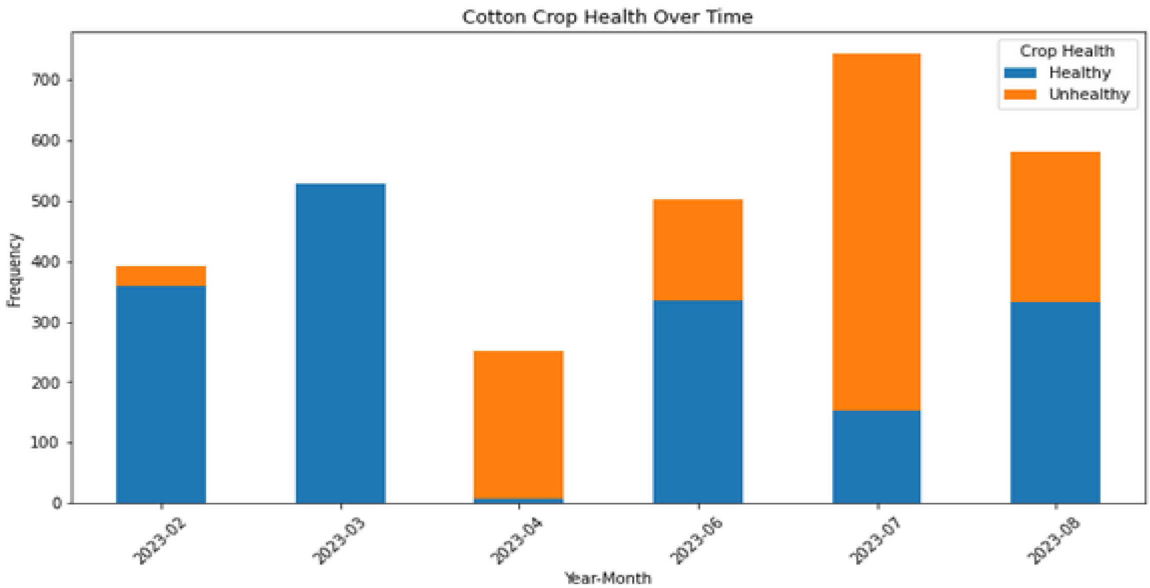

Figure 5: Bar graph of cotton crop health

The chart shows a dramatic increase in unhealthy cotton crops in July 2023, peaking at around 600, while healthy crops peaked in March 2023 at about 500. This suggests a significant adverse event affecting crop health mid-year

C. Scatter plot:

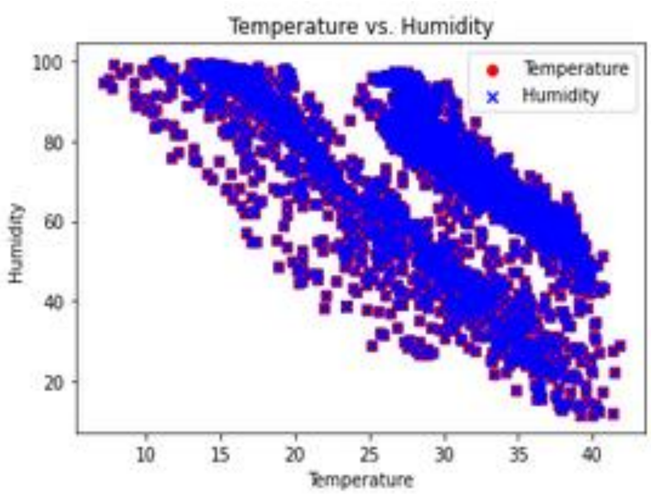

Figure 6: Scatter plot of temperature vs humidity

The scatter plots indicate a negative correlation between temperature and humidity, showing that as temperature increases, humidity decreases.

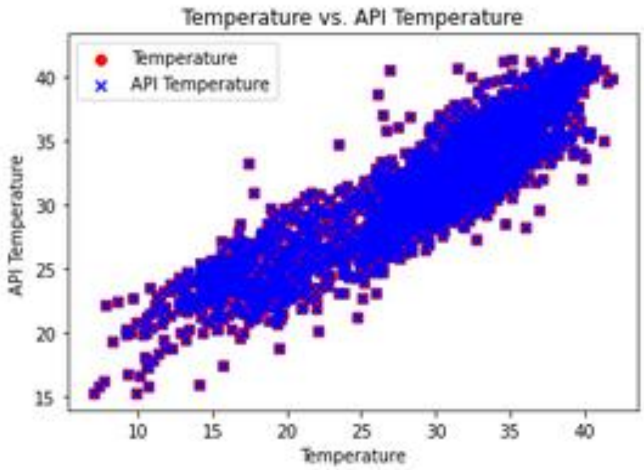

Figure 7: Scatter plot of Temperature vs API Temperature

Conversely, there is a positive correlation between temperature and API temperature, demonstrating that both variables increase together.

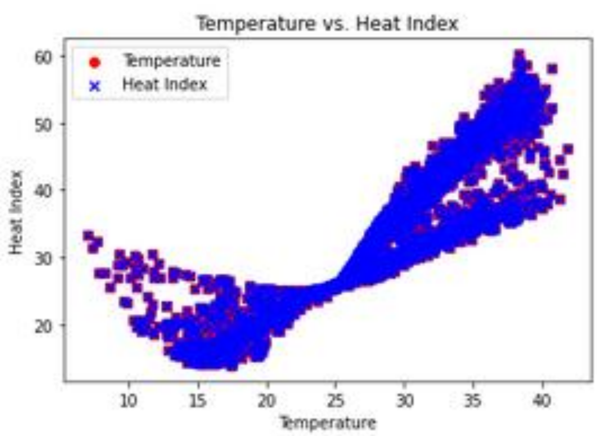

Figure 8: Scatter Plot of Temperature and Heat index

The scatter plots show a strong positive correlation between temperature and heat index, indicating that heat index increases with temperature.

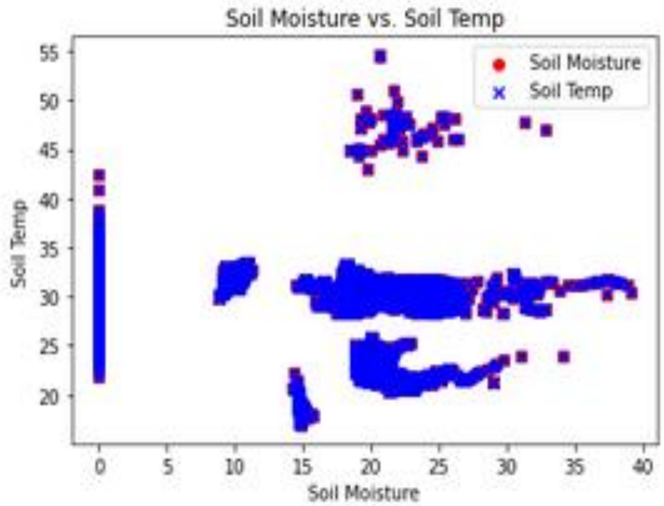

Figure 9: Scatter plot of soil temperature and soil moisture

The soil moisture vs. soil temperature plot displays a clustered pattern, suggesting complex interactions without a clear linear relationship.

D. Correlation Matrix:

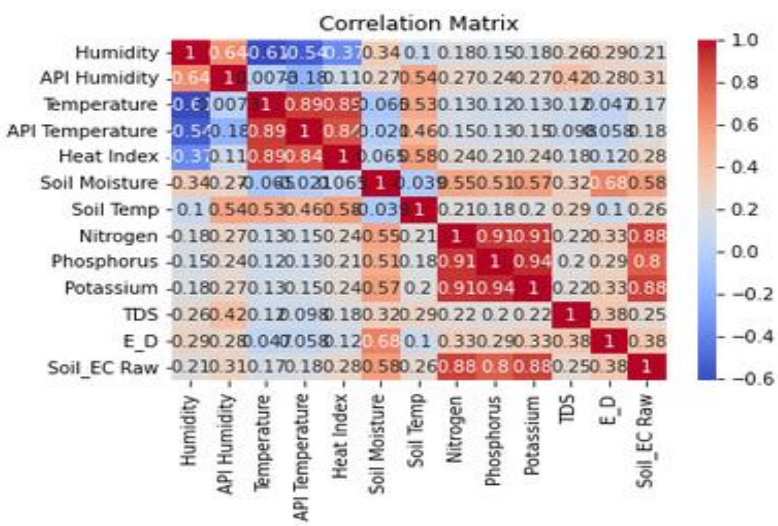

Figure 10: Correlation Matrix

This correlation matrix shows the relationships between different variables, with values closer to 1 indicating a strong positive correlation and values closer to -1 indicating a strong negative correlation. The color gradient from red to blue visually represents the strength and direction of these correlations.

E. Bar Chart (Accuracy Comparison of Models):

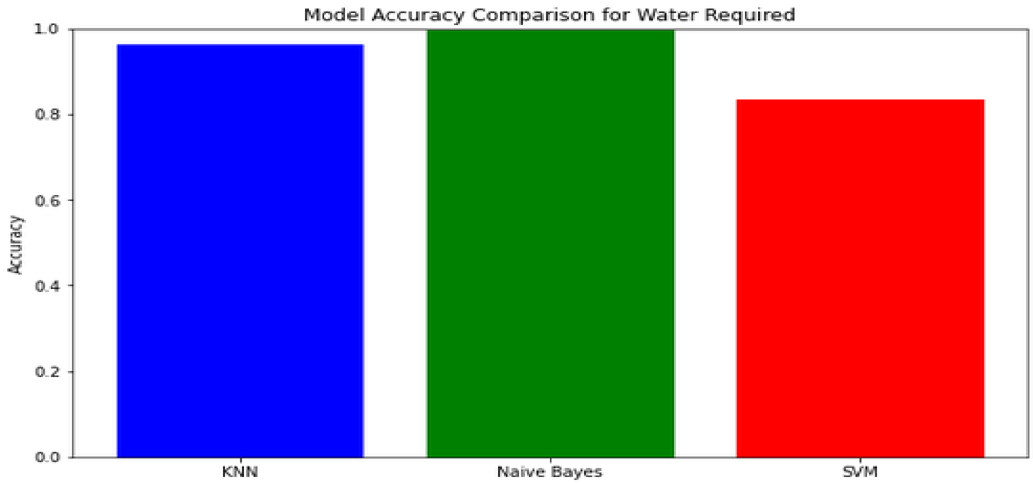

Figure 11: Bar chart of model accuracy comparison for water required

This bar chart compares the accuracy of three models (KNN, Naive Bayes, and SVM) for predicting water requirements, showing Naive Bayer's with the highest accuracy, followed by KNN, and SVM with the lowest accuracy.

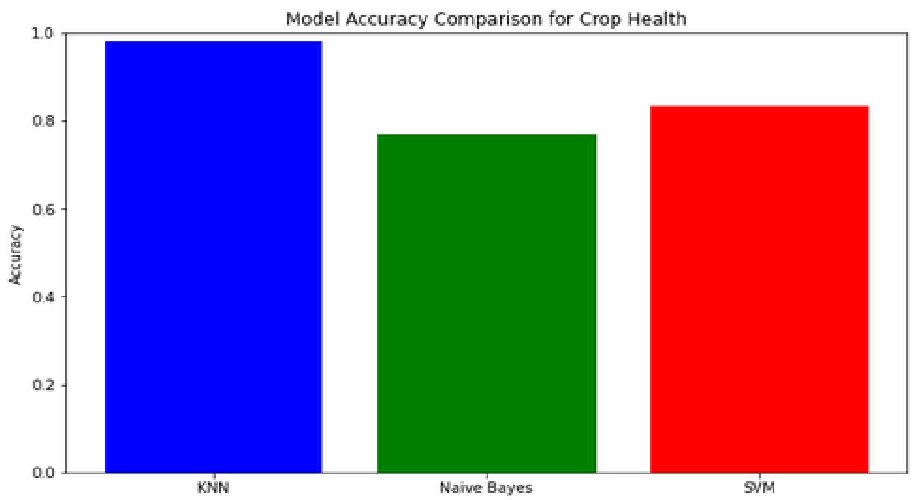

Figure 12: Bar chart of model accuracy comparison for crop health

This bar chart compares the accuracy of three models (KNN, Naive Bayes, and SVM) for predicting crop health, showing KNN with the highest accuracy, followed by SVM, and Naive Bayes with the lowest accuracy.

F. Performance Metrics Comparison Between Models:

Water Required:

Models	Accuracy	Precision		Recall		F1 Score		Support	
		0	1	0	1	0	1	0	1
KNN	0.96	0.96	0.97	0.99	0.87	0.97	0.92	451	149
Naive Bayes	0.99	0.99	1.00	1.00	0.98	1.00	0.99	451	149
SVM	0.83	0.82	1.00	1.00	0.33	0.90	0.49	451	149

Crop Health:

Models	Accuracy	Precision		Recall		F1 Score		Support	
		Healthy	Unhealthy	Healthy	Unhealthy	Healthy	Unhealthy	Healthy	Unhealthy
KNN	0.98	0.98	0.98	0.98	0.98	0.98	0.98	333	267
Naive Bayes	0.76	0.71	1.00	1.00	0.48	0.83	0.65	333	267
SVM	0.83	0.82	1.00	1.00	0.33	0.90	0.49	333	267

G. Line chart:

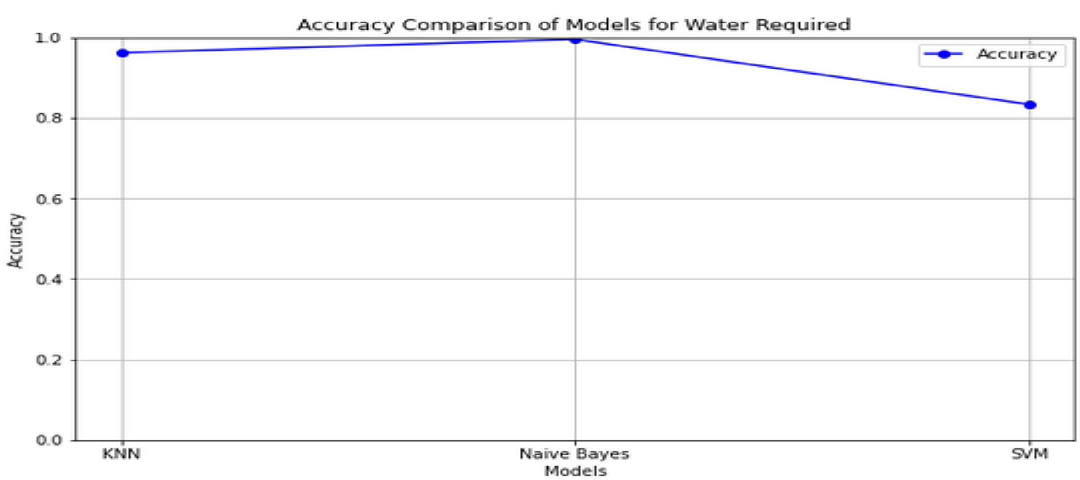

Figure 13: Line chart of accuracy comparison of model for water required

This line chart compares the accuracy of three models (KNN, Naive Bayes, and SVM) for predicting water required, showing Naive Bayes with the highest accuracy, followed by KNN, and SVM with the lowest accuracy.

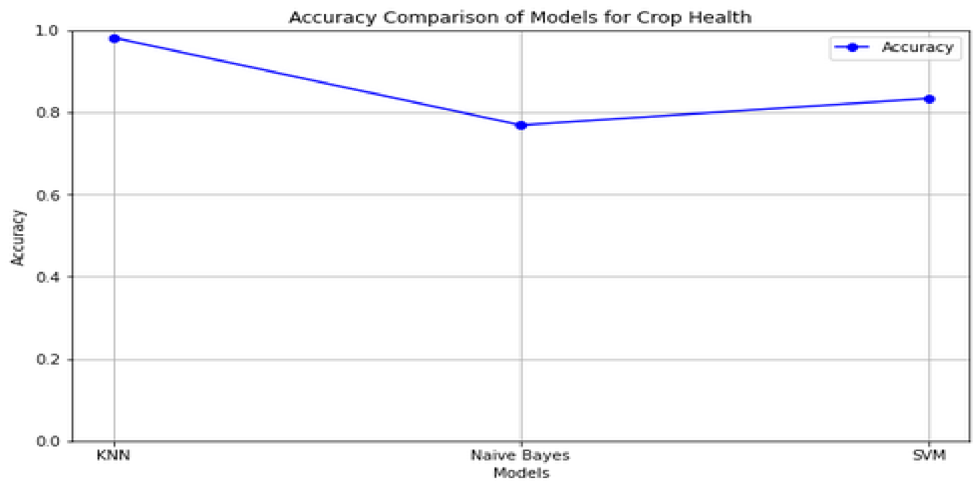

Figure 14: Line chart of accuracy comparison of models for crop health

This line chart compares the accuracy of three models (KNN, Naive Bayes, and SVM) for predicting crop health, showing Naive Bayes with the highest accuracy, followed by KNN, and SVM with the lowest accuracy.

H. Pie chart:

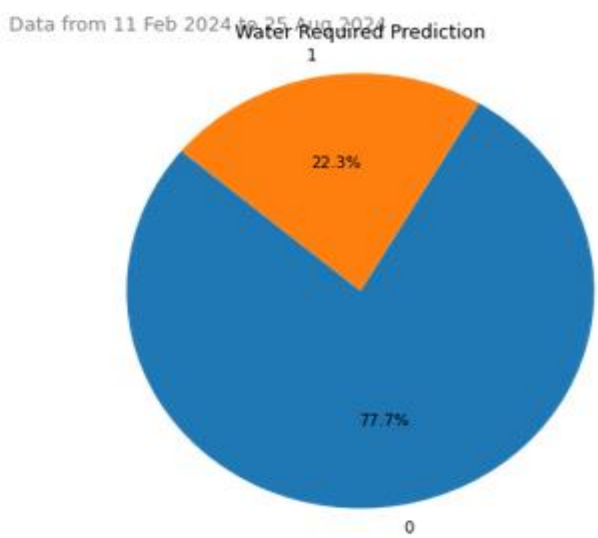

Figure 15:Pie chart of water required

This pie chart represents water requirement predictions from February 11, 2024, to August 25, 2024, showing that 22.3% predict a need for water (1), while 77.7% do not predict a need for water (0).

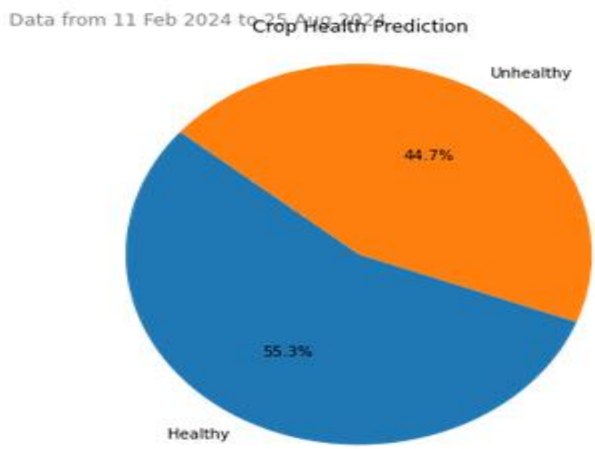

Figure 16: Pie chart of crop health

This pie chart illustrates crop health predictions from February 11, 2024, to August 25, 2024, indicating that 55.3% of the crops are predicted to be healthy, while 44.7% are predicted to be unhealthy.

I. 3dVisualization:

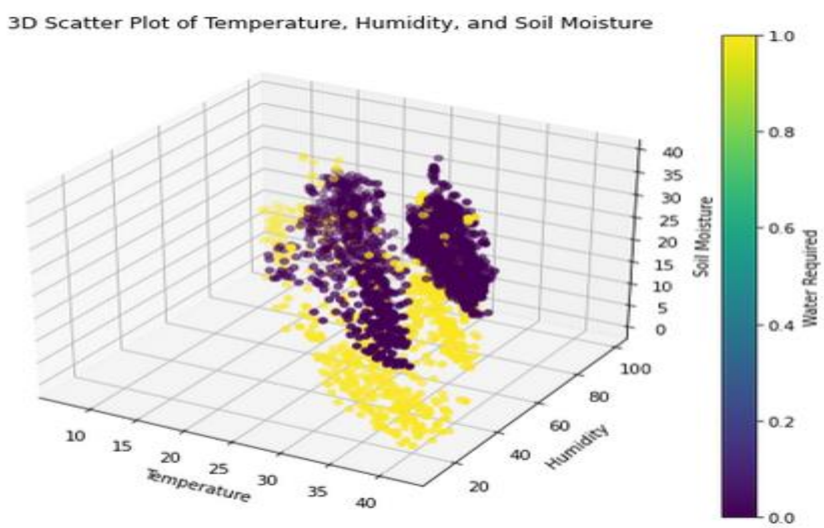

Figure 17: 3D scatter plot of Temperature, Humidity, and Soil Moisture

This 3D scatter plot visualizes the relationship between temperature, humidity, and soil moisture, with color indicating whether water is required (yellow for yes, purple for no). The plot shows data points clustered by different levels of these variables, providing insight into the conditions under which water is needed.

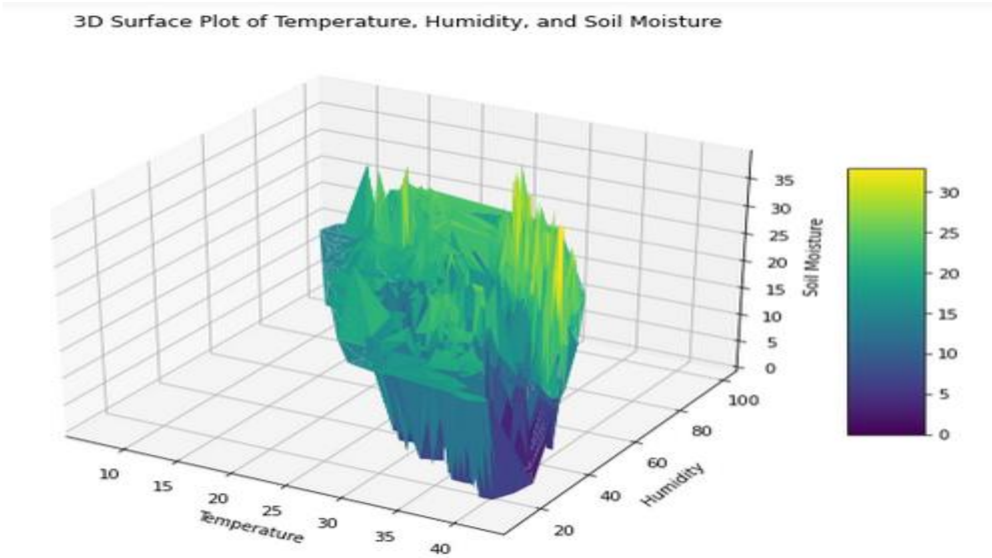

Figure 18: 3D surface plot of Temperature, Humidity, and soil Moisture

The 3D surface plot depicts the relationship between temperature, humidity, and soil moisture, where the z-axis represents soil moisture levels. The color gradient indicates varying soil moisture, with yellow representing higher moisture and purple representing lower moisture.

J. Area chart:

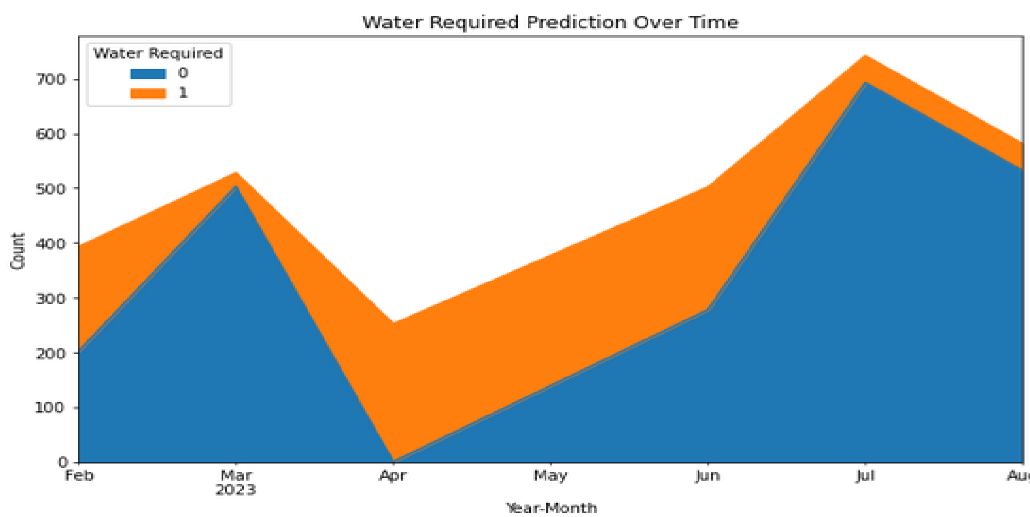

Figure 19: Area chart for water required prediction

The area chart illustrates the prediction of water required over time from February to August 2023, with counts on the y-axis. The blue area represents periods when no water is required (0), while the orange area indicates when water is needed (1).

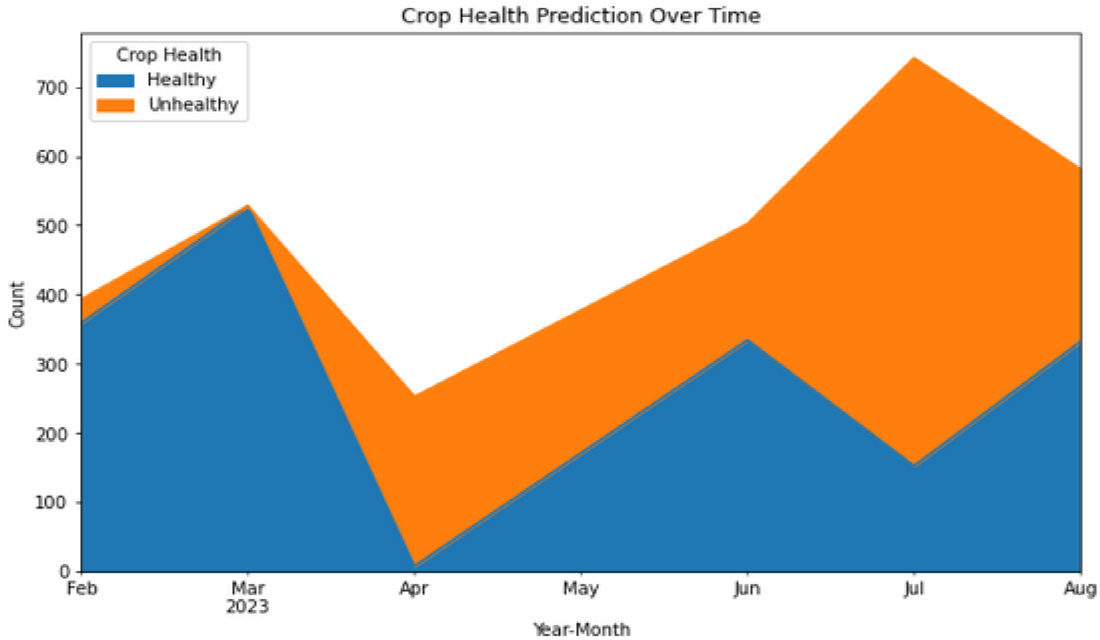

Figure 20: Area chart for crop health prediction

The area chart shows crop health predictions over time from February to August 2023, with counts on the y-axis. The blue area indicates the count of healthy crops, while the orange area represents the count of unhealthy crops.

Discussion

The data visualization techniques applied in this project provide insightful analysis into the parameters influencing smart agriculture. The project effectively demonstrates how different variables such as water requirements, soil moisture, and nutrient levels fluctuate over time.

A. Interpretation of Results:

1. Temporal Analysis: The histogram and bar plots indicate how various agricultural metrics change monthly. These visualizations can help in identifying seasonal patterns and trends that are crucial for optimizing crop yields and resource allocation.
2. Correlation Insights: The scatter plots and correlation heatmaps reveal relationships between variables, such as the correlation between soil moisture levels and crop yield. These insights can guide farmers in making informed decisions on irrigation and fertilization practices.
3. Anomaly Detection: Through data cleaning and exploration, anomalies like outliers and missing values were identified and addressed, ensuring the robustness of the analysis.

B. Limitations:

1. External Factors: The analysis primarily focuses on internal farm data. External factors such as weather conditions, market prices, and policy changes are not considered but can significantly impact agricultural outcomes.

2. Scalability: While the techniques used are effective for the given dataset, scaling these methods to larger datasets or different regions may require additional computational resources and advanced algorithms.

Conclusion

This project demonstrates the potential of data visualization in smart agriculture, highlighting key findings related to temporal trends and relationships between various agricultural parameters. These insights are valuable for improving decision-making processes.

A. Significance includes:

This research emphasizes the importance of integrating data analytics into agricultural practices. By providing actionable insights, data visualization helps farmers optimize resource use, reduce costs, and increase crop yields. This approach supports the broader goals of sustainable agriculture and food security.

B. Future Research:

Future research could explore:

1. Using More Data: Including weather data, economic factors, and satellite images to get a fuller picture.
2. Advanced Algorithms: Improving the machine learning algorithms used for predictions to enhance accuracy and efficiency..
3. Scalability: Adapting the techniques for larger datasets and different regions to ensure wider applicability.

You can find our dataset and code file at this link: Dataset and Code

References

[1] Nasim, S. F., Khurram, M., Kamran, A., Fatima, S. A., & Qaiser, A. (2023). Environmental Monitoring and Agricultural Insights: Analysis of Cotton Crop using PowerBI. *International Journal of Emerging Engineering and Technology (IJEET)*, 2(2), 13-20. https://grsh.org/journal1/index.php/ijeet/article/view/12/13

[2] Zafar SA, Noor MA, Waqas MA, Wang X, Shaheen T, Raza M, Mehboob-Ur-Rahman. Temperature Extremes in Cotton Production and Mitigation Strategies. In: Past, Present and Future Trends in Cotton Breeding. Open Access Peer-Reviewed Chapter. IntechOpen; 2018. DOI: 10.5772/intechopen.74648.

[3] Ahmad, F., Waris, M., Khuhro, S. N., & Sajjad, A. (2018). Spatial Field Survey of Cotton Whitefly and its Pupal Parasitism in Relation to Temperature and Humidity in Southern Pakistan. *Pakistan Journal of Zoology*, 50(4), 1231-1236. https://doi.org/10.17582/journal.pjz/2018.50.4.1231.1236

[4] Edmisten, K. (2022, February 26). Thoughts on cotton planting – soil temperatures. NC State Extension. NC State University. Retrieved from https://cotton.ces.ncsu.edu/2022/02/thoughts-on-cotton-planting-soil-temperatures/

[5] Ahmad, M. S. A., Ashraf, M., & Ali, Q. (2010). Effect of salt stress on growth and ion distribution in wheat (Triticum aestivum) accessions differing in salt tolerance. *Pak. J. Bot.*, 42(1), 273-279.

[6] Hodges, H.F., Whisler, F.D., Bridges, S.M., Reddy, K.R., McKinion, J.M. (1997). Simulation in crop management: GOSSYM/COMAX. In: Peart, R.M., Curry, R.B. (Eds.), Agricultural Systems Modeling and Simulation. Marcel Decker, New York, pp. 235–281.

[7] Jallas, E., Akins, D.C., Turner, S., Gourley, K.E., McKinion, J.M., Willers, J.L., Clouvel, P., Cretenet, M., Sequeira, R., Jenkins, J.N. (1999). Precision Farming, Myth or Reality: a Case Study in Mississippi Cotton Fields. Proceedings of Beltwide Cotton Production and Research Conferences, Orlando, FL, 3–7 January, 1999.

[8] McKinion, J.M., Jenkins, J.N., Akins, D., Turner, S.B., Willers, J.L., Jallas, E., Whisler, F.D. (2001). Analysis of a precision agriculture approach to cotton production. Computers and Electronics in Agriculture, 32(2), 213-228.

[9] Reddy, K.R., McKinion, J.M. (1997). GOSSYM/COMAX.

[10] Sequeira, R.A., Jallas, E. (1995). The simulation Model GOSSYM and its decision support system, COMAX: its applications in American agriculture. Agric. Dev. 8, 25–34.

[11] Steiner, J.L., Williams, J.R., Jones, O.R. (1987). Evaluation of the EPIC simulation model using a dryland wheat-sorghum-fallow crop rotation. Agron. J. 79, 732–738. [8]

[12] Wilkerson, G.G., Jones, J.W., Boote, K.J., Ingram, K.T., Mishoe, J.W. (1983). Modeling soybean growth for crop management. Trans. ASAE 26, 63–73