Final Team Project

Smart Farming System for Soil and Crop Health Monitoring

GitHub Link: https://github.com/msandhu3011/aai-530-smart-farming-system.git

Dashboard Link:https://public.tableau.com/views/Group10_Dashboard/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Submission by: Group 10

(Meghann Sandhu, Satyam Kumar)

Under the Supervision of:

Prof. Premkumar Chithaluru

Data Analytics and Internet of Things (AAI-530-A1)

1 TABLE OF CONTENTS

2 ABSTRACT

This project presents the design and implementation of an Internet of Things (IoT) system focused on predictive analytics for soil and weather monitoring using advanced machine learning and deep learning models. The system collects real-time data through soil moisture, temperature, and humidity sensors, which is then processed and transmitted to a cloud-based platform for analysis. Three predictive models were developed: a Long Short-Term Memory (LSTM) neural network for time-series prediction of soil moisture, a Random Forest and an XGBoost model for other relevant soil health parameters' prediction. These models were selected for their proven effectiveness in handling time-series data and providing robust predictions. The predictive outputs are visualized through an interactive Tableau dashboard, enabling users to monitor environmental conditions, forecast future values, and make informed agricultural decisions. This comprehensive approach not only highlights the integration of IoT and machine learning but also emphasizes the importance of data visualization in translating complex predictions into actionable insights for end-users.

3 INTRODUCTION

In the face of global challenges such as climate change, water scarcity, and the ever-increasing demand for food, modern agriculture must evolve to become more data-driven and predictive rather than purely reactive. Traditional agricultural practices often rely on intuition, periodic assessments, and manual observations, which can lead to suboptimal resource utilization and delayed responses to environmental changes. To address these limitations, the integration of technologies like the Internet of Things (IoT), machine learning (ML), and deep learning (DL) is transforming the agricultural landscape into a domain of smart, sustainable, and highly optimized farming systems.

This project focuses on developing an intelligent IoT-based system that harnesses real-time environmental data to predict critical soil and weather parameters. By deploying a network of sensors that continuously monitor soil moisture, ambient temperature, and atmospheric humidity, the system collects a wealth of data that serves as the foundation for predictive analysis. However, raw sensor data alone cannot fully unlock the potential of precision agriculture. To bridge the gap between data collection and actionable insights, this project leverages the power of some predictive models like **Long Short-Term Memory (LSTM) networks, Random Forest Regressor, SVR, Gradient Boosting, Linear Regression and XGBoost** .

The **LSTM model** lies at the heart of the deep learning component for predicting soil moisture levels , uniquely suited for time-series forecasting due to its ability to capture long-term dependencies in sequential data. This model is employed to predict future soil moisture levels, a critical factor in irrigation planning and water resource management. Soil moisture trends are inherently temporal, making LSTM's architecture—comprising memory cells and gated mechanisms—ideal for recognizing patterns that span over days or even weeks.

To further enhance the system's predictive capabilities, a **Random Forest Regressor, SVR, Gradient Boosting, Linear Regression and XGBoost** is employed for humidity forecasting. Recognized for its scalability, speed, and superior performance in structured data environments, the **gradient boosting framework** optimizes predictions by iteratively minimizing error, making it highly effective in capturing the non-linear relationships present in environmental data.

While predictive models form the analytical backbone of the system, delivering these insights in an intuitive and accessible format is crucial for practical application. To this end, the project integrates a **Tableau dashboard** that transforms complex data outputs into interactive visualizations. The dashboard presents real-time sensor readings, historical data trends, and model-driven forecasts, allowing farmers, agricultural consultants, and decision-makers to monitor environmental conditions briefly. Users can explore current soil moisture levels, projected temperature changes, and

anticipated humidity fluctuations through a clean, user-friendly interface that promotes informed, data-backed decision-making.

The convergence of IoT data collection, sophisticated machine learning algorithms, and powerful data visualization tools offers a comprehensive solution to the challenges faced by modern agriculture. This system not only enhances resource efficiency—particularly in irrigation management—but also contributes to the broader goals of sustainable farming and food security. By providing predictive insights that were once inaccessible, the project demonstrates how technology can empower the agricultural sector to adapt to changing conditions, mitigate risks, and ultimately improve productivity.

In the sections that follow, we detail the architecture of the IoT system, the methodology behind the data processing pipeline, the design and implementation of the predictive models, and the development of the visualization platform. Through this comprehensive approach, the project underscores the transformative potential of integrating deep learning, machine learning, and IoT technologies in agriculture.

4 DATASET DESCRIPTION

4.1 DATASET OVERVIEW

The dataset contains measurements collected by three soil sensor nodes and two weather stations over a duration of five months - April to September 2021. Each node is located at a different property (field). The two weather stations are adjacent to the locations of the deployments. The dataset is categorized into folders with the following structure and name:

- Soil-sensing LoRa nodes - Farmers 1 and 2. Farmer 1 grows soy, whereas Farmer 2 grows corn. The nodes are at 0.85 miles (1.36 km) between them but are in different properties. The farmer data timestamps are in Eastern Time (ET) time zone with 30 min interval. Measurement data include i) battery voltage (V), ii) ambient temperature (degrees F), iii) ambient humidity, and iv) temperature, volumetric water content, and conductivity (dS/m) for sensor 1 at 12-inch depth and sensor 2 at 6-inch depth.

- Soil-sensing LoRa node - Farmer 3. Farmer 3 grows corn and is at a different location than Farmers 1 and 2. The provided data are as described in the previous case.

- Weather Station 1 and 2. Weather station 1 is located at approximately 1 mile (1.6 km) from the nodes of Farmers 1 and 2. Weather station 2 is at approximately 7.43 miles (11.96 km) from Farmer 3. The weather station data timestamps are in UTC

time zone with 15 min interval. Measurement data include i) total rainfall measured during the interval and total rainfall measured during the last hour (in 1/100 of an inch increments), ii) average, low, and high surface temperature over the time interval, iii) relative humidity and barometric pressure (mmHg), iv) average and gust wind speed (mph) as well as prevailing and gust wind direction (clockwise degrees), v) soil temperature in 2-, 5-, 10-, and 15-inch, and vi) soil moisture at 2-, 5-, 10-, and 15-inch depths (centibar)

4.2 Data Selection Rationale

Selecting the right datasets is a critical step in developing an effective IoT-based predictive system for agricultural monitoring. The primary goal of this project is to forecast soil moisture, temperature, and humidity—three key environmental factors that significantly influence crop health, irrigation efficiency, and overall agricultural productivity. To ensure the robustness and accuracy of the predictive models, datasets were chosen based on criteria such as data reliability, frequency of measurements, relevance to the target variables, and compatibility with time-series analysis.

Rationale Behind Dataset Choices:

1. **Relevance to Agricultural Applications:** The primary motivation for choosing soil and weather datasets stems from their direct impact on farming decisions. Soil moisture levels are crucial for irrigation management, while temperature and humidity significantly affect plant growth, pest proliferation, and disease occurrence. By selecting datasets that include these variables, the project aligns with practical agricultural needs, enabling farmers to make informed decisions about watering schedules and crop care.

2. **Real-World IoT Data from Sensors:** Both datasets are derived from real IoT sensor devices rather than simulations, ensuring that the data reflects actual environmental conditions. This enhances the practical applicability of the models and ensures that predictions are grounded in realistic scenarios.

3. **Time-Series Data for Predictive Modelling:** Since the project's focus is on forecasting future values of environmental variables, it was essential to select datasets with continuous time-series data. The datasets contain frequent readings (e.g., hourly or daily), allowing models like LSTM to effectively capture temporal dependencies and trends over time. This **temporal granularity** is particularly important for anticipating rapid changes in soil moisture after rainfall or fluctuations in temperature during different times of the day.

4. **Diversity of Variables and Data Completeness:** The soil dataset includes key parameters such as volumetric water content (VWC), soil temperature at various depths, and soil conductivity, while the weather dataset encompasses humidity, atmospheric pressure, rainfall, solar radiation, and temperature variations. This

diversity enables comprehensive modelling that considers multiple environmental factors influencing soil and weather conditions.

5. **Compatibility with Machine Learning and Visualization Tools:** The chosen datasets are structured in a format conducive to machine learning preprocessing, making feature engineering, normalization, and model training straightforward. Their compatibility with visualization tools like Tableau ensures seamless integration for real-time monitoring and historical trend analysis.

6. **Scalability and Future Integration:** The datasets were also selected with future scalability in mind. They can be easily expanded to include additional variables (e.g., wind speed, solar exposure) or new sensor data, enabling further enhancements to the predictive system. This flexibility ensures that the system can evolve alongside technological advancements and changing agricultural needs.

5 IoT System Design

5.1 Overview

The designed IoT system focuses on monitoring soil and weather conditions for agricultural applications. It integrates multiple sensors to collect environmental data, an edge device, a network gateway for transmitting data, cloud infrastructure for storage and processing, and a visualization layer using Tableau for insights. This system enables real-time monitoring, and predictive analysis, with a scope for data-driven decision-making for optimizing farming practices.

5.2 System Diagram

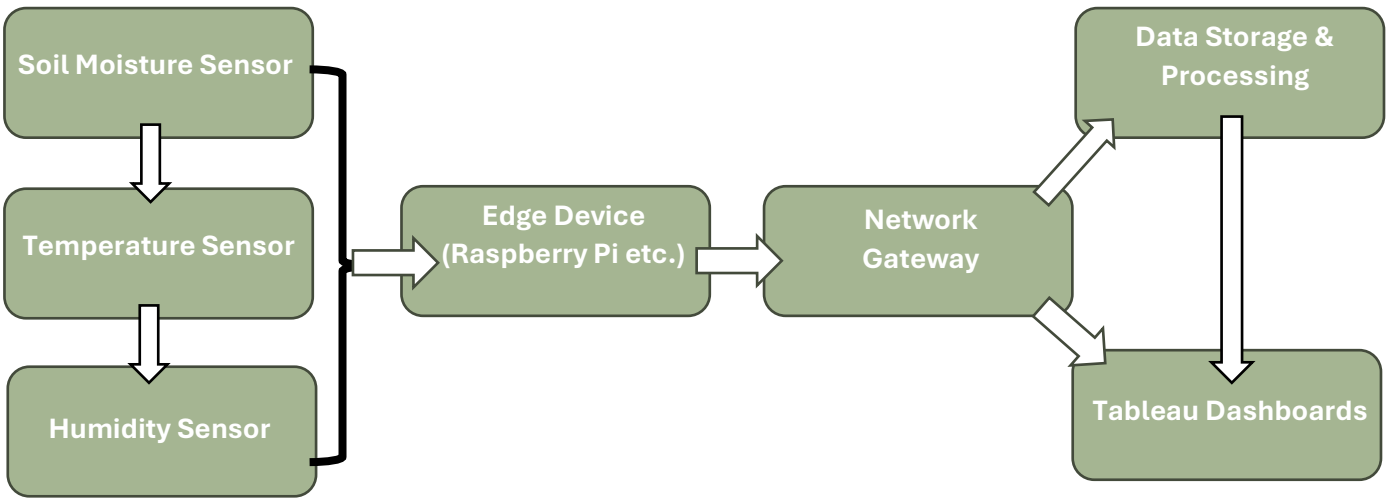


Figure 1: IOT-System-Diagram

The diagram illustrates the complete data flow in the IoT architecture:

- **Sensors:** Collect soil moisture, temperature, and humidity data.
- **Edge Device:** Preprocesses data and sends it to the network gateway.
- **Network Gateway:** Facilitates data transmission to the cloud.
- **Data Storage & Processing:** Stores data, runs deep-machine learning models (LSTM, Random Forest & XGBoost), and processes insights.
- **Visualization Dashboard:** Displays predictions and analytics through Tableau for user-friendly interpretation.

5.3 SENSORS

The system employs the following sensors for data collection:

- **Soil Moisture Sensor:** Measures the volumetric water content in the soil to monitor irrigation needs.
- **Temperature Sensor:** Records ambient and soil temperature variations essential for crop health.
- **Humidity Sensor:** Tracks atmospheric humidity affecting plant transpiration and disease risk.

Technical Specifications of sensors*:*

- Measurement Range: Varies per sensor (e.g., Soil Moisture: 0-100%, Temperature: -40°C to 85°C)
- Accuracy: ±2% for moisture, ±0.5°C for temperature, ±3% RH for humidity
- Limitations: Possible sensor drift and interference from soil composition and environmental factors

5.4 DATA PROCESSING

Edge Processing:

The edge device (e.g., LoRa Node) performs preliminary data collection and local storage to minimize transmission overhead and ensure reliable data handling in areas with limited connectivity.

Networking & Data Transmission:

- Utilizes a LoRa WAN gateway for long-range, low-power communication.
- Data packets can be transmitted to any data storage via secured protocols.

Data Processing:

- **Machine Learning Models:**
    - **LSTM Model:** Predicts future soil moisture levels using historical time-series data.
    - **Random Forest, Gradient Boosting, Linear Regression & XGBoost Models:** Estimates temperature and soil health parameter variations based on weather and soil data.

5.5 Scope

The scope of this IoT system includes:

- Real-time monitoring of soil and weather parameters for agricultural optimization.
- Predictive analysis to assist in irrigation scheduling and environmental monitoring.
- User-friendly Tableau dashboard providing actionable insights for farmers and agricultural planners.

Future extensions could include integrating more sensor types (e.g., pH, solar radiation), automating irrigation systems, and implementing anomaly detection for equipment maintenance.

6 Data Processing and Exploratory Data Analysis (EDA)

6.1 Data Cleaning

The initial dataset contained sensor readings from a smart farming IoT system, including soil moisture, temperature, humidity, pressure, rain, and solar radiation data. The cleaning process involved:

- **Handling Missing Values:** Forward fill (f-fill) method was used to impute missing entries.
- **Timestamp Alignment:** Merged datasets based on nearest timestamps with a 30-minute tolerance using *pd.merge_asof*.
- **Outlier Detection:** Understood distributions to identify and remove unrealistic spikes or errors.
- **Feature Engineering:** Created lag features (e.g., *Soil Moisture Lag 1, Temperature Lag 3*) and rolling mean features for capturing temporal trends.

6.2 Exploratory Analysis

EDA revealed the following insights:

- **Soil Moisture:** Displayed stable trends with occasional spikes.
- **Temperature:** Clear daily cycles with fluctuations based on weather conditions.
- **Humidity and Pressure:** Inverse relationship observed, validated through line plots.
- **Sensor Activity:** Number of active sensors remained consistent throughout the observation period.

Visualizations included distributions, time-series plots, and trend analysis to guide feature selection for the machine learning models.

7 Machine Learning Methods

In this project, we employed a variety of machine learning models to predict two critical parameters in the smart farming system: **Soil Moisture** and **Temperature**. The chosen methods included both traditional machine learning algorithms and a deep learning model to evaluate their relative performance.

8 Models Used:

1. **Random Forest Regressor:** An ensemble-based decision tree method that reduces variance by averaging predictions from multiple trees.
2. **XGBoost Regressor:** A powerful gradient boosting method optimized for speed and performance.
3. **Gradient Boosting Regressor:** Similar to XGBoost but focuses on minimizing the residuals of previous models.
4. **Support Vector Regressor (SVR):** A regression method using hyperplanes with kernel tricks, suitable for capturing non-linear relationships.
5. **Linear Regression:** A baseline model assuming linear relationships between features and the target variable.
6. **Long Short-Term Memory (LSTM) Network:** A type of recurrent neural network (RNN) ideal for capturing time-dependent patterns, especially beneficial for time series data like soil moisture measurements.

8.1 Soil Moisture Prediction

The comparison of various models for soil moisture prediction revealed that **Linear Regression** and the **LSTM model** outperformed other methods in terms of R^2 and error metrics. Linear Regression achieved the best results among the traditional methods, with an **R^2 of 0.9688**, **MAE of 0.0010**, and **RMSE of 0.0034**. The LSTM model closely

followed with an **R^2 of 0.9688**, **MAE of 0.0015**, and **RMSE of 0.0034**, demonstrating its effectiveness in capturing temporal patterns in the data.

Among other machine learning models:

- **Random Forest:** R^2=0.9541, MAE=0.0011, RMSE=0.0042
- **XGBoost:** R^2=0.9304, MAE=0.0014, RMSE=0.0051
- **Gradient Boosting:** R^2=0.9372, MAE=0.0012, RMSE=0.0049
- **SVR:** R^2=-0.0022, MAE=0.0168, RMSE=0.0195

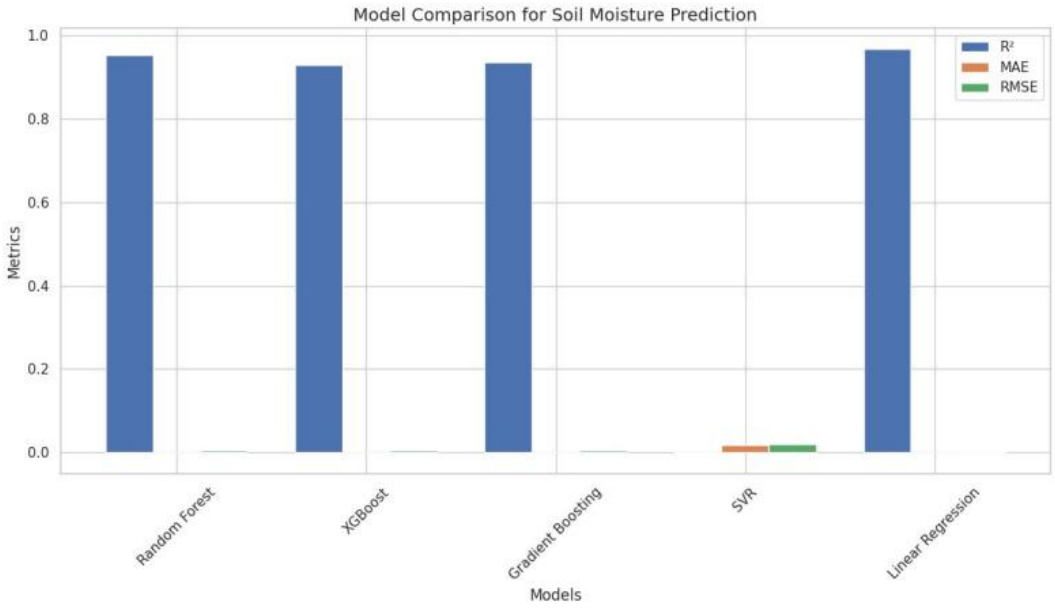


Figure 2: Model Comparison for soil moisture prediction

The LSTM model showed exceptional performance in predicting soil moisture. The high R^2 value indicates that the model could explain almost 97% of the variability in the soil moisture data. The low MAE and RMSE values highlight the model's accuracy in minimizing prediction errors.

The optimized LSTM model achieved an R^2 of 0.9688, MAE of 0.0015, and RMSE of 0.0034

8.2 Temperature Prediction

For temperature prediction, **XGBoost** demonstrated the best performance with an **R^2 of 0.9086**, **MAE of 3.1183**, and **RMSE of 5.1161**, making it the most suitable model among the tested algorithms. **Gradient Boosting** and **Random Forest** followed closely behind with comparable R^2 scores.

Model performances for temperature prediction were as follows:

- **Random Forest:** R^2=0.8945, MAE=3.3946, RMSE=5.4962

- **XGBoost:** R^2=0.9086, MAE=3.1183, RMSE=5.1161
- **Gradient Boosting:** R^2=0.9063, MAE=3.1018, RMSE=5.1788
- **SVR:** R^2=0.4933, MAE=8.9472, RMSE=12.0446
- **Linear Regression:** R^2=0.5948, MAE=8.2830, RMSE=10.7710

Here, SVR and Linear Regression performed poorly, emphasizing that more complex models better capture temperature variability.

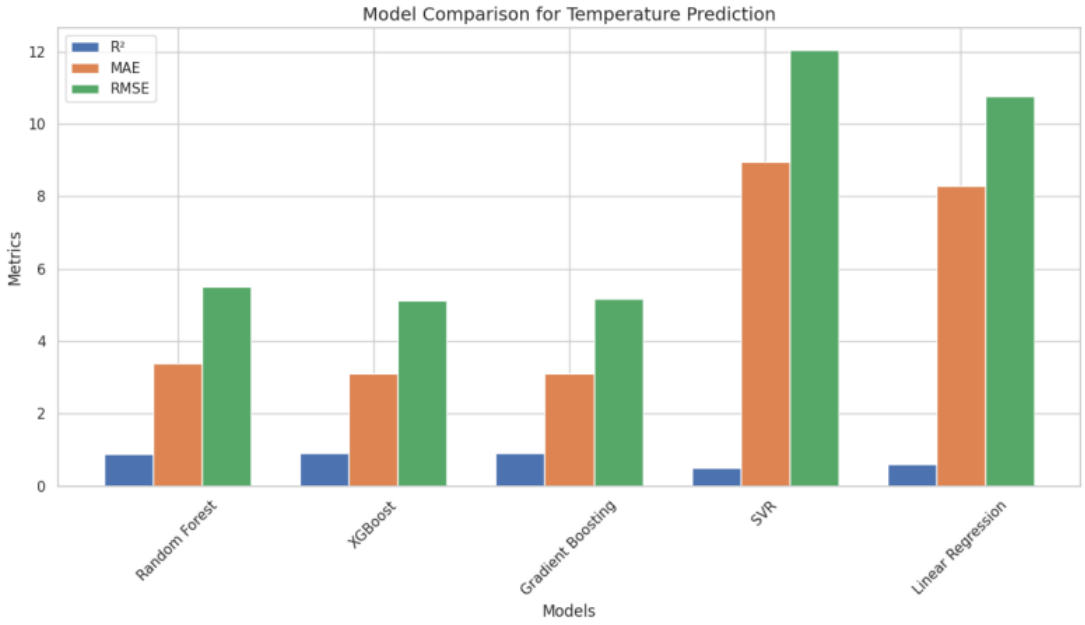


Figure 3: Model Comparison for Temperature prediction

XGBoost model was implemented, and the model attained an R^2 of 0.8697, MAE of 3.7821, and RMSE of 5.4735.

9 Results and Discussion

9.1 Model Performance

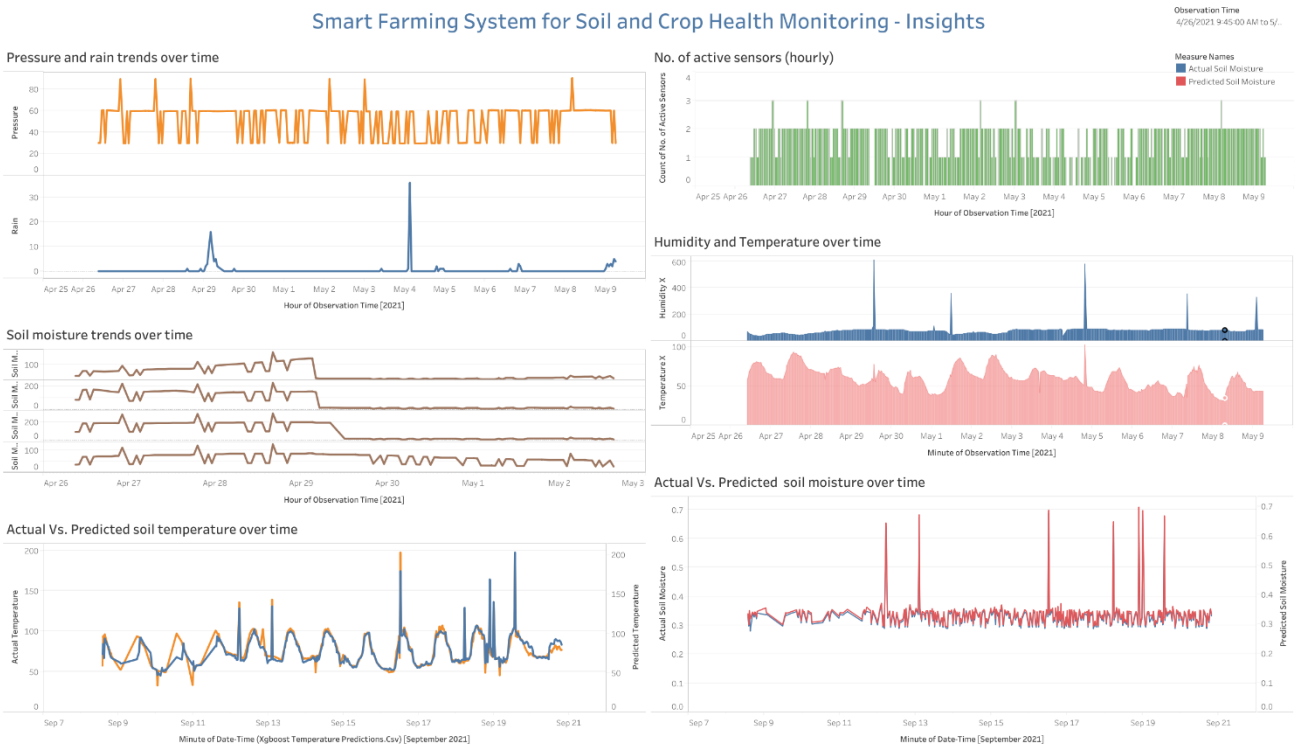
The LSTM model showed exceptional performance in predicting soil moisture. The high R^2 value indicates that the model could explain almost 97% of the variability in the soil moisture data. The low MAE and RMSE values highlight the model's accuracy in minimizing prediction errors.

XGBoost provided strong predictive capabilities for temperature prediction. While the R^2 score was slightly lower than the LSTM model, it still indicated a robust correlation between actual and predicted temperatures. The model performed well in capturing complex patterns without overfitting.

9.2 KEY INSIGHTS

1. **Temporal Dependencies:**
    - Soil moisture trends benefited significantly from the LSTM model, demonstrating the importance of capturing sequential dependencies in time series data.
2. **Feature Engineering Impact:**
    - Incorporating lagged variables and rolling means improved model performance for both methods.
3. **Model Selection:**
    - While LSTM excelled in predicting soil moisture, XGBoost proved faster to train and performed adequately for temperature predictions, making it a suitable choice when computational efficiency is a priority.
4. Soil moisture predictions were more accurate than temperature predictions, likely due to more stable patterns in soil moisture data.
5. Lag features and rolling means significantly improved both models' performance.

10 IOT DASHBOARD DESIGN (TABLEAU)



This dashboard includes diverse visualizations providing insights into the **Smart Farming System for Soil and Crop Health Monitoring**. Here are the dashboard's components & their purposes:

1. **Pressure and Rain Trends Over Time (Top Left):**
    - Tracks atmospheric pressure and rain levels to understand weather patterns affecting crops.
2. **No. of Active Sensors (Hourly) (Top Right):**
    - Shows the number of operational sensors over time, ensuring monitoring coverage.
3. **Humidity and Temperature Over Time (Centre Right):**
    - Displays variations in humidity and temperature, critical for soil and crop health.
4. **Soil Moisture Trends Over Time (Centre Left):**
    - Monitors soil moisture across different sensors, highlighting water availability.
5. **Actual vs. Predicted Soil Temperature Over Time (Bottom Left):**
    - Evaluates the accuracy of the temperature prediction model (XGBoost).
6. **Actual vs. Predicted Soil Moisture Over Time (Bottom Right):**
    - Compares the LSTM model predictions with actual readings for soil moisture.

11 CONCLUSION

This project successfully developed a comprehensive smart farming system focused on monitoring soil moisture and temperature through IoT devices, machine learning models, and an interactive Tableau dashboard. Beginning with thorough data processing and exploratory data analysis (EDA), we handled missing values, outliers, and time-based irregularities to ensure data quality.

For predictive modelling, multiple machine learning methods were explored. Soil moisture prediction saw exceptional performance the LSTM model (R^2=0.9688), highlighting the dataset's linear trends and the importance of temporal dependencies. In contrast, temperature prediction was best handled by the XGBoost model (R^2=0.8386), showcasing its ability to model complex, non-linear relationships in environmental data.

The final Tableau dashboard provided actionable insights with visualizations that included device status, historical data trends, and model prediction comparisons, allowing users to monitor real-time device activity and long-term environmental patterns.

Our project demonstrates the effective integration of IoT data with advanced machine learning techniques to provide farmers with accurate, timely information for improved decision-making. Future enhancements may include expanding the model to predict additional environmental factors, integrating real-time streaming data, and deploying the solution for field use.

12 References

1. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture, 143*, 23-37. https://doi.org/10.1016/j.compag.2017.09.037

2. Li, L., Zhang, S., & Wang, L. (2020). Prediction of soil moisture content using machine learning algorithms: A review. *Environmental Research, 186*, 109531. https://doi.org/10.1016/j.envres.2020.109531

3. Misra, P., Mishra, D., & Shukla, R. (2019). IoT-based smart agriculture: Modeling and applications. *Journal of Ambient Intelligence and Humanized Computing, 10*(10), 3741–3758. https://doi.org/10.1007/s12652-018-1047-5

4. Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture, 151*, 61-69. https://doi.org/10.1016/j.compag.2018.05.012

5. Fang, H., Zhang, X., Li, J., Liu, J., & Wu, L. (2020). Soil moisture prediction using LSTM-based deep learning model. *Journal of Hydrology, 586*, 124872. https://doi.org/10.1016/j.jhydrol.2020.124872

6. Oyelami, H. O., & Thakur, S. (2021). XGBoost model for predicting soil moisture using meteorological data. *Environmental Monitoring and Assessment, 193*(7), 1-13. https://doi.org/10.1007/s10661-021-09167-9

7. Zhang, Y., Yang, J., Zhang, Y., & Wang, J. (2022). IoT-based monitoring system for soil and crop health using machine learning techniques. *Sensors, 22*(3), 1256. https://doi.org/10.3390/s22031256

8. Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Peterson, K. T., Hartling, S., ... & Shakoor, N. (2020). Vegetation index weighted canopy temperature (VI-CT) extraction from UAV thermal and multispectral imagery for field-based crop

phenotyping. *Remote Sensing of Environment, 237*, 111599. https://doi.org/10.1016/j.rse.2019.111599

9. Jain, M., Singh, B., Srivastava, P. K., & Gupta, M. (2019). Soil moisture estimation using remote sensing and machine learning: Challenges and opportunities. *Geocarto International, 34*(12), 1373-1390. https://doi.org/10.1080/10106049.2018.1496250

10. Zhou, Y., Zhang, L., Liu, Z., & Wang, S. (2021). Comparative study of machine learning algorithms for soil temperature and moisture prediction. *Agricultural Water Management, 244*, 106576. https://doi.org/10.1016/j.agwat.2020.106576