Soil Moisture and Temperature Forecasting System

Francina Pali[1], Kishore Murugan[2] and Jeetkumar Rajendrakumar Patel[3]

Department of Computer Science, Saint Louis University

AAI-530- Data Analytics and the Internet of Things

Prof. Anna Marbut

May 8, 2023

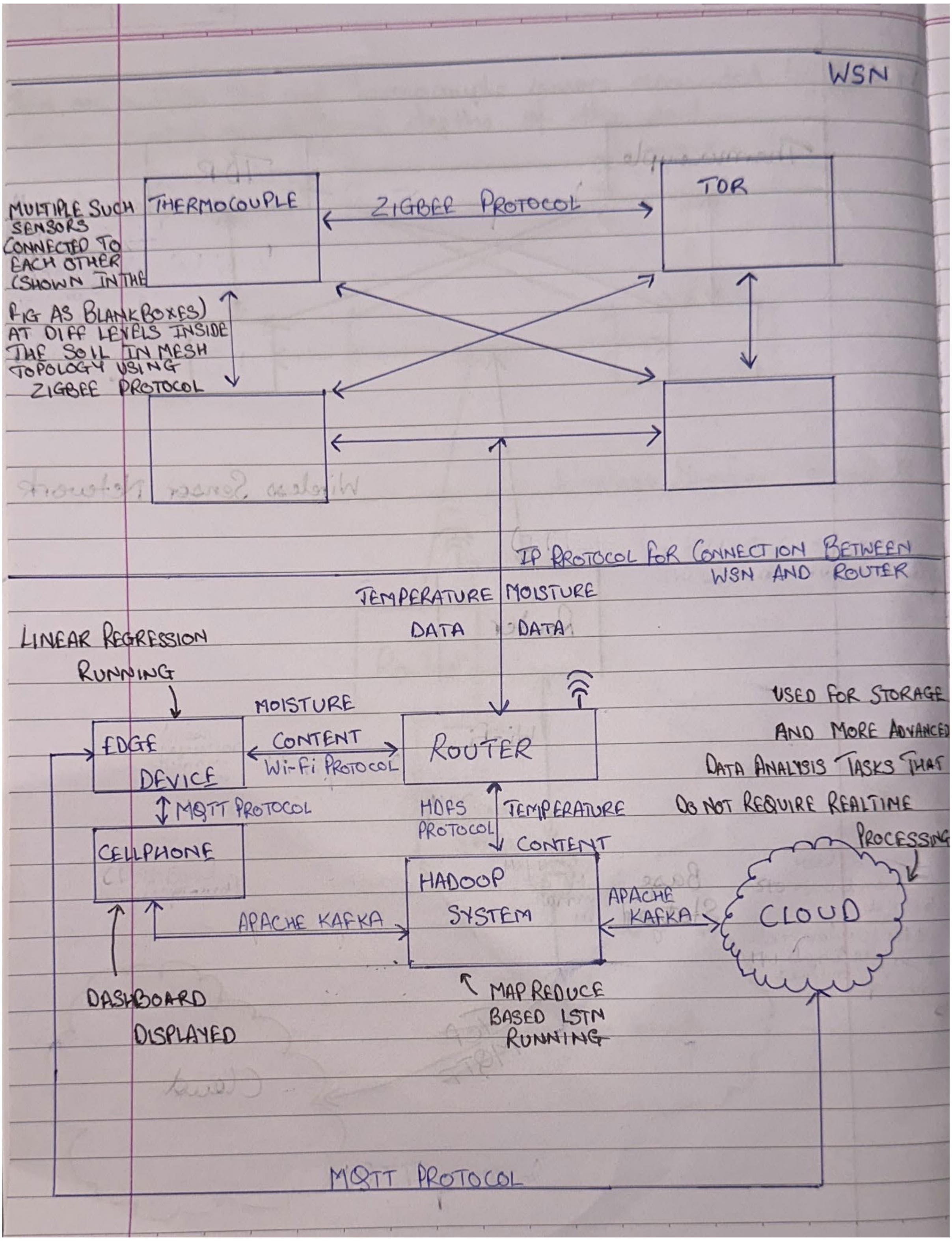
Soil Moisture and Temperature Forecasting System

Precision agriculture is rapidly evolving, with the advent of IoT technology, big data analytics, and AI algorithms. One of the most critical factors in agricultural production is the soil moisture and temperature levels. Historically, farmers have relied on experience and intuition to manage these factors, leading to inefficiencies in water usage, crop growth, and yield. However, with the advancement of technology, it is now possible to develop predictive models that can forecast future soil moisture and temperature levels at different depths of the soil. This is achieved through the analysis of historical data using AI algorithms, which can identify patterns and relationships between various factors that affect soil moisture and temperature levels. This system can help farmers and land managers to plan and adjust their irrigation and crop management practices, optimizing agricultural production, and resource management. In this way, precision agriculture can contribute to sustainable food production, reduced resource consumption, and improved profitability for farmers.

IoT System Design

Figure 1

IoT System Design



Sensors used:

Thermocouples

A Thermocouple is a type of temperature sensor that uses two different metals to measure temperature.

The TDR sensors are position at depths of 30, 60, 90, 120 and 150 centimetres within the soil.

Limitations:

Environmental Influences

Soil Temperature Measurements can be influenced by external factors such as ambient temperature, solar radiation, and precipitation if the thermocouples sensors are not adequately shielded or protected.

Calibration and Drift

Like any temperature sensor, thermocouples may require periodic calibration to maintain accurate measurements. Additionally, over time, thermocouples can experience drift, where their output voltage changes due to aging or other factors. Regular calibration and monitoring can help mitigate these effects.

Time – Domain Reflectometry (TDR) Sensors

These sensors measure soil moisture by sending a high frequency electrical pulse along a probe inserted into the soil. The speed of the pulse is affected by the soil moisture content, and this is used to calculate the soil moisture.

Limitations:

Sensor Interference

TDR sensors rely on electromagnetic waves to measure soil moisture. Interference from nearby objects, such as rocks, roots, or other sensors, can affect the signal propagation and lead to inaccurate readings. Proper spacing between sensors and potential sources of interference should be considered.

Calibration Maintenance

TDR sensors may require periodic re-calibration to ensure accurate and reliable measurements over time. Environmental factors, sensor aging, or changes in soil properties can necessitate recalibration to maintain measurement accuracy.

Connectivity and Networking

In this system the sensors are interconnected to each other in a mesh topology. The concept of Wireless Sensor Network is used for connectivity in the following way: multiple IoT devices equipped with sensors communicate with each other using a mesh topology. So even if one of the links fails, data can still be transited through alternative paths in the network, ensuring that data loss is minimized. This could help improve the reliability of the network and reduce the chances of data loss or downtime.

The data collected by these devices in a mesh is then transmitted to a router. The router sends the moisture data to the edge device on which Linear regression is employed. The edge device then sends some of the data to the cloud for further advanced analysis and publishes the insights of the linear regression model on an app on the cell phone.

In the same way, the temperature data is sent by the router to Hadoop on which LSTM is employed. Some data is then sent by Hadoop to the cloud and the insights of the LSTM model are published to an app on the cell phone.

Edge Processing

The edge device is responsible for performing real-time data processing and low-latency analysis.

The linear regression model is trained on the edge device using historical moisture data, and once trained, it is used to make real-time predictions on the incoming sensor data. The model's weights are updated using stochastic gradient descent (SGD) algorithm. It utilizes an exponential decay factor to decrease the weights of the past. All these enable quick and low-latency decisions to be made based.

On the current moisture readings, without having to send all the sensor data to the cloud for processing. Additionally, the edge device can perform data filtering, data compression, and data aggregation to reduce the amount of data that needs to be sent to the cloud, thereby reducing the data transmission and storage costs.

Hadoop

Hadoop's Hadoop Distributed File System (HDFS) is used to store the temperature data across the cluster. HDFS provides fault tolerance and enables parallel data access, which is crucial for efficient processing. Hadoop's MapReduce framework handles the distributed processing automatically.

Cloud Computing

The cloud server is used for storage and more advanced data analysis tasks that do not require real time processing. The reason for this is that cloud storage provides a centralized and scalable platform for data management, which can facilitate long-term data analysis and decision-making.

On the cloud server, the processed data from the edge devices can be stored, analysed, and visualized to gain insights on moisture and temperature over time. the cloud server can be then used to generate trend analysis reports, anomaly detection alerts, and statistical summaries of the soil quality parameters. These reports can be accessed and visualized by authorized users through an app-based interface, which enables remote monitoring and management of the soil quality. Furthermore, the cloud server can be used to aggregate the data from multiple edge devices and Hadoop systems to gain a holistic view of the soil quality across different locations and to identify any trends or anomalies that may be occurring in the soil quality parameters.

This can be useful for making long-term decisions related to soil quality management and for identifying any potential issues before they become critical.

This division of processing between the edge device, Hadoop system and the cloud server allows for a more efficient use of computing resources, as well as a more scalable and flexible system architecture. By offloading some of the processing tasks to the Hadoop system and the cloud server, the system can leverage the additional computational power and storage capacity available in the cloud and Hadoop environment. At the same time, it can maintain low latency and high responsiveness at the edge.

Protocols Used:

- The sensors in the WSN employ the Zigbee protocol for establishing connections with other sensors in a mesh topology.
- To connect this Wireless Sensor Network (WSN) of sensors in a mesh topology to the router, the protocol used is the Internet Protocol (IP).
- The edge device supports wireless connectivity, the router connects to it using Wi-Fi protocol.
- The edge device uses the MQTT protocol for transmitting data to the cloud and the app on the cell.
- The protocol used by the router to send data to the Hadoop system is the Hadoop Distributed File System (HDFS) protocol. HDFS is a distributed file system that is part of the Hadoop ecosystem and is designed to store and process large datasets across multiple machines.

- Apache Kafka is used to stream data from Hadoop to the cloud The mobile app. The mobile app can subscribe to the relevant topics or streams and receive real-time or batched data updates.

Attribute Information

- VW_30cm: Volumetric water readings at 30 cm depth (m^3/m^3)
- VW_60cm: volumetric water readings at 60 cm depth (m^3/m^3)
- VW_90cm: volumetric water readings at 90 cm depth (m^3/m^3)
- VW_120cm: volumetric water readings at 120 cm depth (m^3/m^3)
- VW_150cm: volumetric water readings at 150 cm depth (m^3/m^3)
- T_30cm: temperature readings at 30 cm depth (C)
- T_60cm: temperature readings at 60 cm depth (C)
- T_90cm: temperature readings at 90 cm depth (C)
- T_120cm: temperature readings at 120 cm depth (C)
- T_150cm: temperature readings at 150 cm depth (C)
- Location
- Date Volumetric water content readings are calibrated according to: Gasch, CK, DJ Brown, ES Brooks, M Yourek, M Poggio, DR Cobos, CS Campbell. 2017. A pragmatic, automated approach for retroactive calibration of soil moisture sensors using a two-step, soil specific correction. Computers and Electronics in Agriculture, 137: 29-40. Temperature readings are factory calibrated.

Analytics

Data Collection

- There are two folders, named "Daily" and "Hourly", that contain 42 text files each. These text files contain measurements from water content and temperature sensors. There is a total of 84 files, with 42 files in each folder.
- Each text file is associated with a particular location, which is specified in the file name (such as "CAF003.txt") and in the "Location" field of the data table. The readings in each file correspond to the specific location mentioned in the file name and data table.
- The readings in the text files are sorted and arranged in a particular order. They are organized according to the 'Date' they were taken, ranging from April 20th, 2007, to June 16th, 2016. Additionally, the readings taken in hourly files are also sorted by 'Time' using a 24-hour clock format.
- The file 'CAF310.txt' has been selected for analysis because it has the highest number of valid lines. In this context, a valid line is defined as a line that does not contain 'NA' as a value.
- The file 'CAF310.txt' contains 47,567 valid lines.

Data Cleaning and Exploratory analysis

Some of the following data that can be cleaned and explored to ensure that the dataset is a good representative for a real-world problem are as follows: -

- Total no of observations recorded along with the Data type of every feature present in the dataset.
- Presence of missing values
- Visualization through plots and Correlation Analysis to investigate if there is a relation and how strong and significant it is between the features.

The data will be cleaned and pre-processed to ensure that it is ready for modelling.

Selection of Models for Analysis

Knowing the temperature of the soil is indeed important in agriculture, as it influences various biological processes, nutrient availability, and overall plant growth. However, real-time temperature data for soil may not be as critical as real-time soil moisture data in many farming scenarios.

Soil temperature typically varies slowly over time and is influenced by factors such as solar radiation, air temperature, and soil composition. While having access to historical soil temperature data can provide valuable insights for crop planning, pest management, and understanding the growth patterns of specific crops, real-time soil temperature measurements may not require constant monitoring or immediate action in the same way that soil moisture does.

In contrast, soil moisture can change rapidly due to factors such as rainfall, irrigation, evaporation, and plant water uptake. This dynamic nature of soil moisture makes real-time monitoring essential for effective irrigation management, preventing water stress or excessive water usage, and optimizing crop health and yield.

If real-time moisture content data is not available, it can have several implications for farmers and agricultural practices:

Suboptimal Irrigation: Without real-time moisture data, farmers may have to rely on scheduled or manual irrigation practices rather than adjusting irrigation based on the actual moisture needs of the soil. This can result in under- or over-irrigation, leading to water stress or waterlogging, respectively, affecting crop health and productivity.

Increased Water Usage: In the absence of real-time moisture information, farmers may tend to overcompensate by irrigating more frequently or applying higher volumes of water as a precautionary measure. This can lead to inefficient water usage and increased costs associated with water resources.

Crop Health Issues: Inaccurate or delayed knowledge of soil moisture content can impact crop health. Insufficient moisture can lead to drought stress, stunted growth, reduced yield, and susceptibility to pests and diseases. On the other hand, excessive moisture can cause root rot, nutrient leaching, and fungal diseases. Real-time moisture data allows farmers to take timely actions to maintain optimal soil moisture levels for healthy crop growth.

Resource Management Challenges: Real-time moisture data helps in efficient resource allocation. Without it, farmers may face difficulties in planning and managing resources like labour, fertilizers, and pesticides. For instance, if soil moisture is not adequately monitored, farmers may misallocate resources, resulting in inefficient use and potential economic losses.

Inefficient Nutrient Management: Soil moisture affects the availability and uptake of nutrients by plants. Without real-time moisture data, it becomes challenging to optimize nutrient application based on the current moisture conditions. This can lead to imbalances in nutrient levels and suboptimal plant nutrition.

In summary, while both soil temperature and soil moisture are important parameters in agriculture, real-time monitoring of soil moisture is typically considered more critical for making timely decisions related to irrigation management, crop health, and resource allocation. Soil temperature, on the other hand, may be monitored less frequently or assessed through periodic measurements to gain a broader understanding of the soil environment and its impact on plant growth. Therefore, it was decided to compute the predictions for moisture on the edge device itself. As edge computing is better suited for applications where there is a need for real-time processing and decision-making.

Linear Regression

Linear Regression was implemented on the edge device. It is a lightweight method that does not require a lot of computational power as Linear regression involves a simple mathematical formula for predicting the target variable making it suitable for applications where low latency is important. A linear regression model assumes a linear relationship between the input features and the target variable. Therefore, we can make use of mu while using the linear regression model. mu is the forgetting factor or exponential decay. Since the readings are not going to be linear, we don't want our historical data to have much influence over the current prediction. The goal of mu is to assign more weight to more recent data points and to reduce the weight of older data points in the forecast.

LSTM (Long Short-Term Memory)

It is a type of neural network architecture that is particularly effective in handling sequential data such as speech, text, and time series data. It is a type of recurrent neural network (RNN) that has additional memory cells in an LSTM model, the output of a layer can be passed not only to the next layer but also to the previous layer. This allows the network to remember information from previous time steps and use it to inform its current output. This is achieved using a memory cell and gates, which allow the LSTM to selectively forget or remember information. Therefore, LSTM model is well-suited to capturing the temporal dependencies and patterns in the data that are likely to be important for accurate predictions over time. Previous readings of temperature can provide important information about how the soil environment has changed over time, and how it is likely to continue changing in the future. For example, if the soil has been consistently dried for several days, this can indicate that the temperature may be increasing due to reduced evaporative cooling. Conversely, if the soil has been consistently wet for several days, this can indicate that the temperature may be decreasing due to increased evaporative cooling.

Let's say the temperature has been steadily increasing for the past few hours with LSTM layers, the previous readings would also be considered, and the network would learn that the temperature is likely to continue increasing.

LSTM are also flexible in the kind of output that they produce. In the case of a Soil Moisture and Temperature Forecasting System monitoring system, we might want to predict a single value (e.g., the temperature in the next hour) if we are interested in short-term forecasting, or we might want to predict a sequence of values (e.g., the temperature content for the next 24 hours) if we are interested in longer-term forecasting.

Evaluation and Result

Two LSTM based models were trained.

One was a standard LSTM model and the other was an LSTM model with several convolutional layers.

Figure 2

Visualization of the performance of the standard LSTM model on the validation set.

```
50/50 - 0s - loss: 0.7191 - mse: 0.7191 - 221ms/epoch - 4ms/step

MSE: 0.7191160321235657
50/50 [==============================] - 1s 5ms/step
```

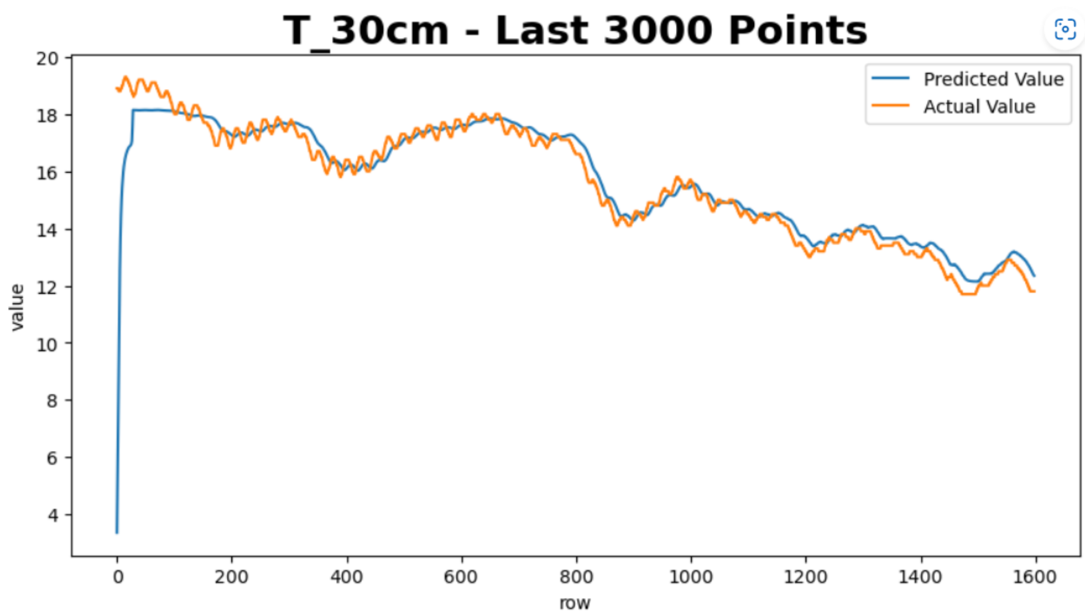
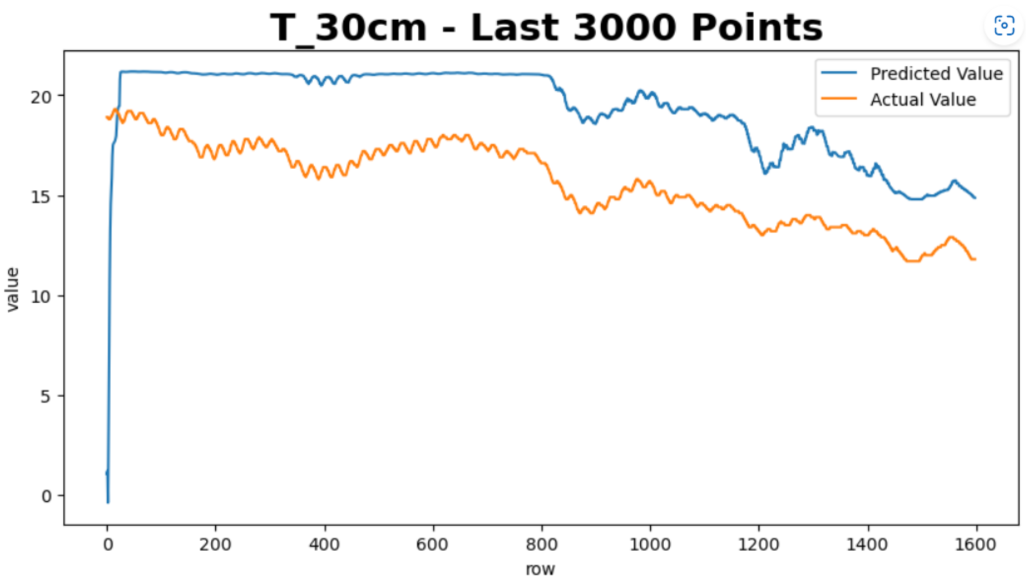


Figure 3

Visualization of the performance of the LSTM model with convolutional layers on the validation set.

```
50/50 - 0s - loss: 15.4859 - mse: 15.4859 - 490ms/epoch - 10ms/step

MSE: 15.48589038848877
50/50 [==============================] - 1s 11ms/step
```



- After evaluating the above two models, we decided to choose the standard LSTM model without convolutional layers. This decision is based on several factors. Firstly, when applying the standard LSTM model on the validation set, we observed that the predicted values closely align with the actual data. Additionally, the mean squared error (MSE) value of the standard LSTM model is significantly lower at 0.719 compared to the LSTM model with convolutional layers, which had an MSE value of 15.485. These findings support the conclusion that the standard LSTM model is a better fit for the temperature prediction task.
- Five Linear Regression based models were trained.
- One with mu=0.9, second with mu=1 and third with mu=0.01
- In The fourth model we added an extra feature “T_30cm” for predicting “V_30cm”.

- In the fifth model we Used the moving average for both the moisture and temperature data.

Figure 4

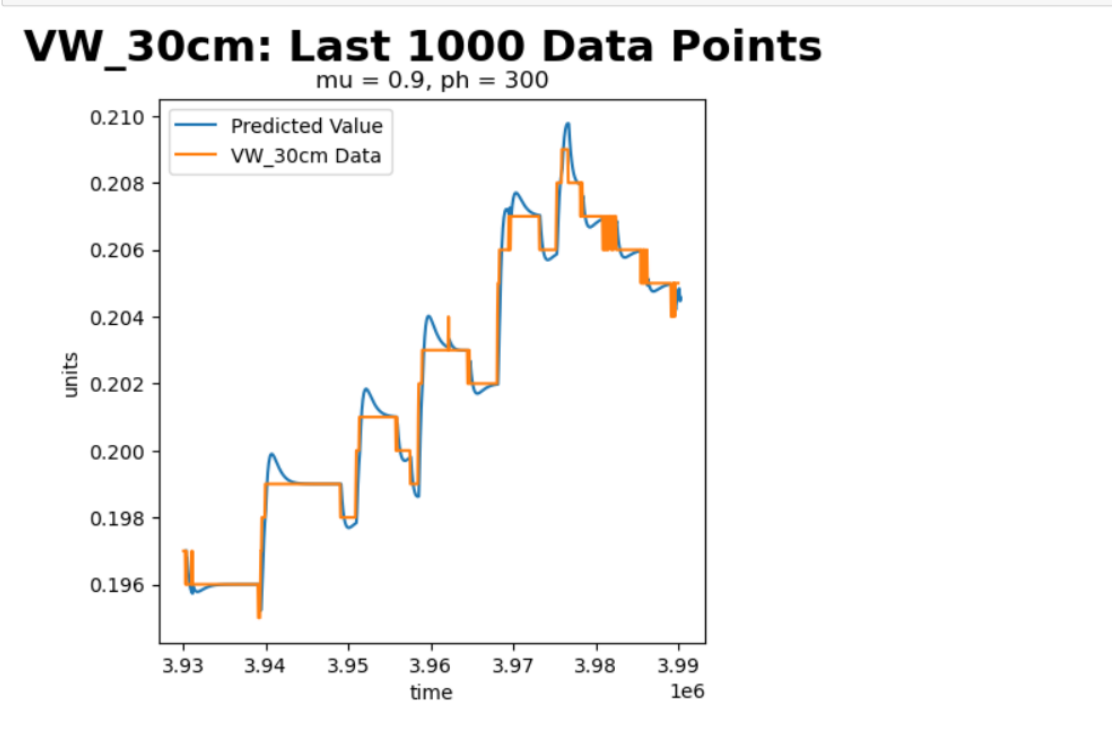


Figure 5

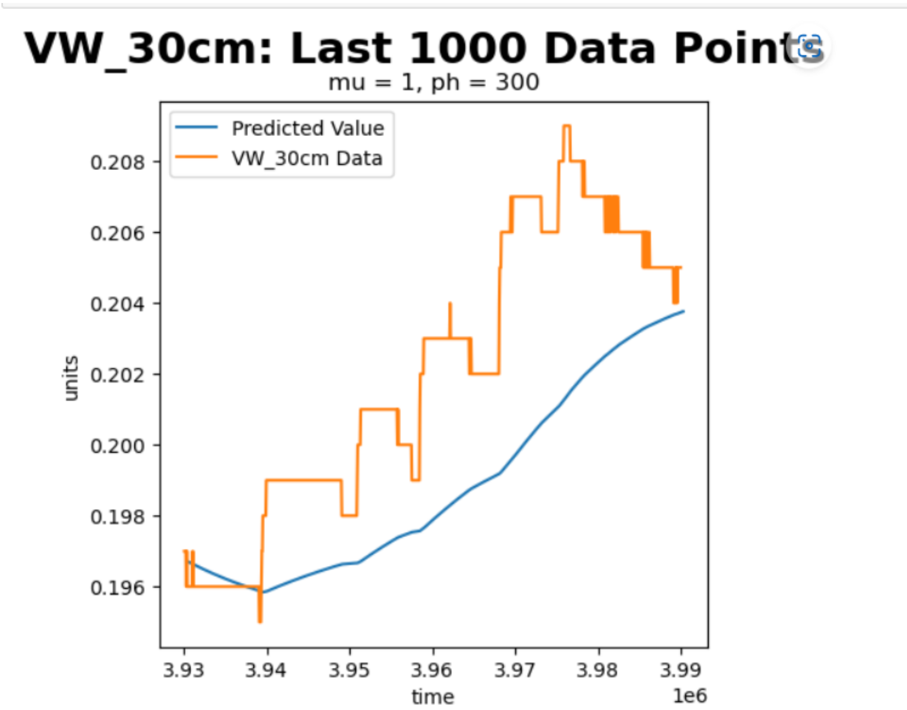


Figure 6

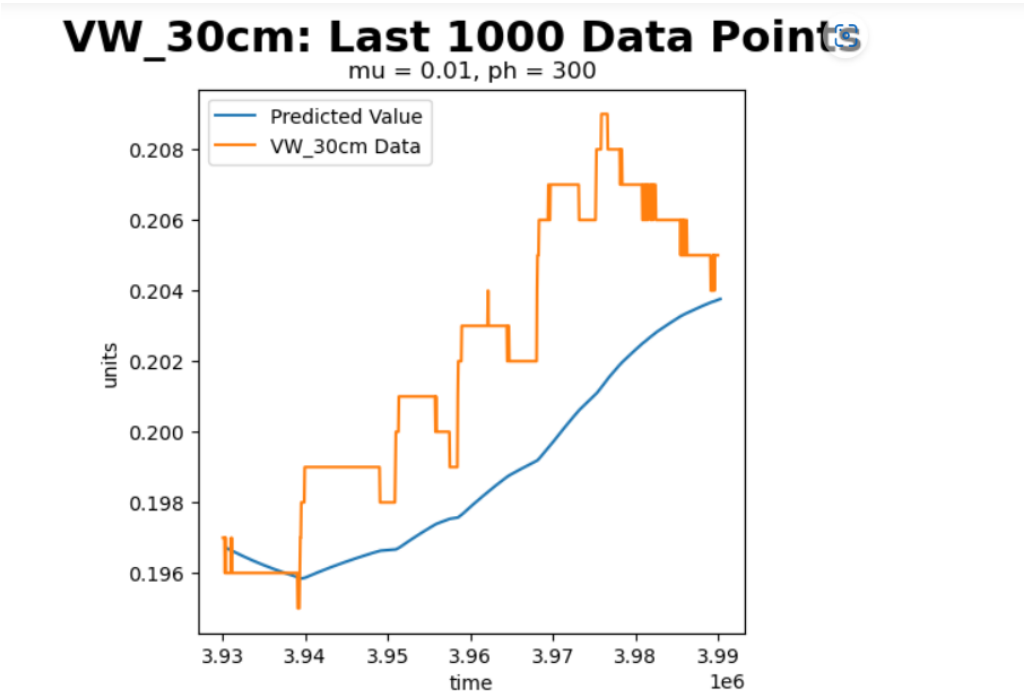


Figure 7

Visualization after Adding temperature data at a soil depth of 30 cm as a second variable to our model.

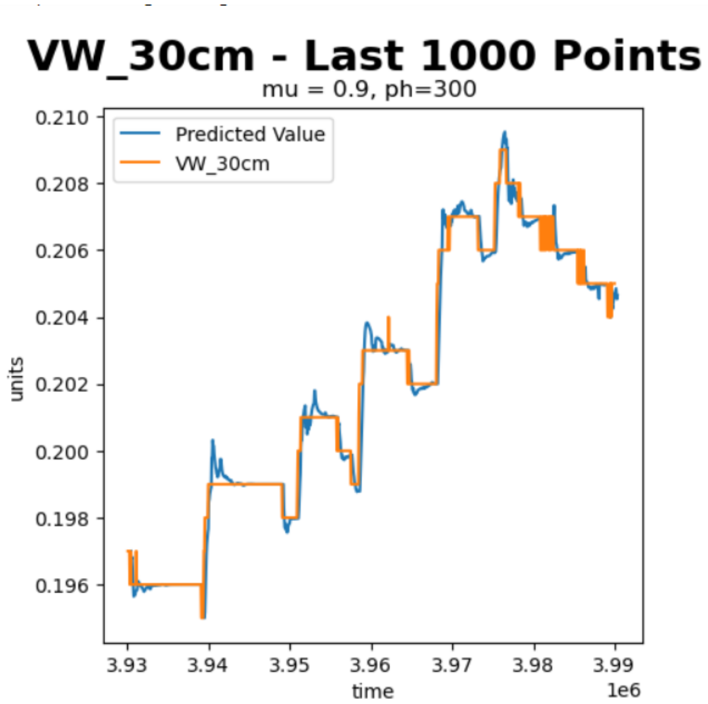
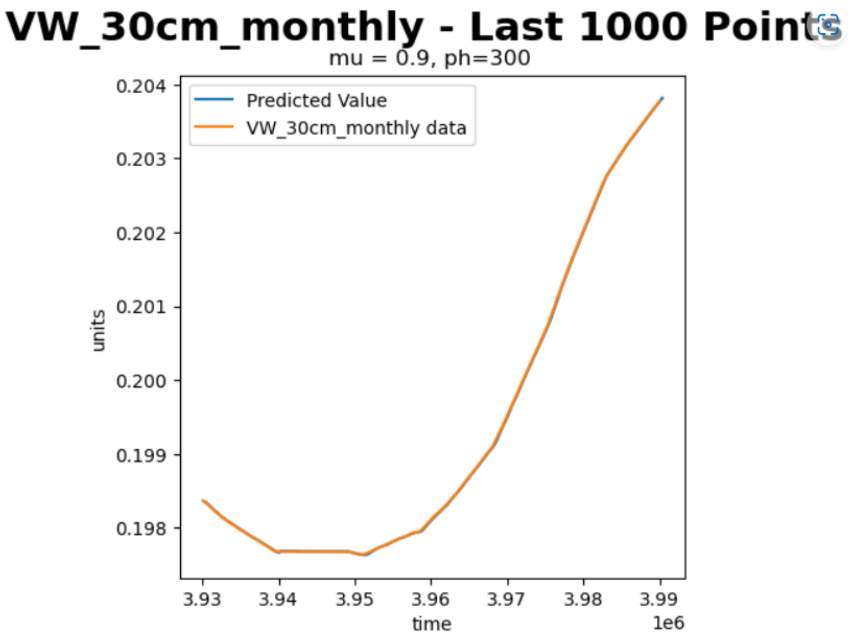


Figure 8

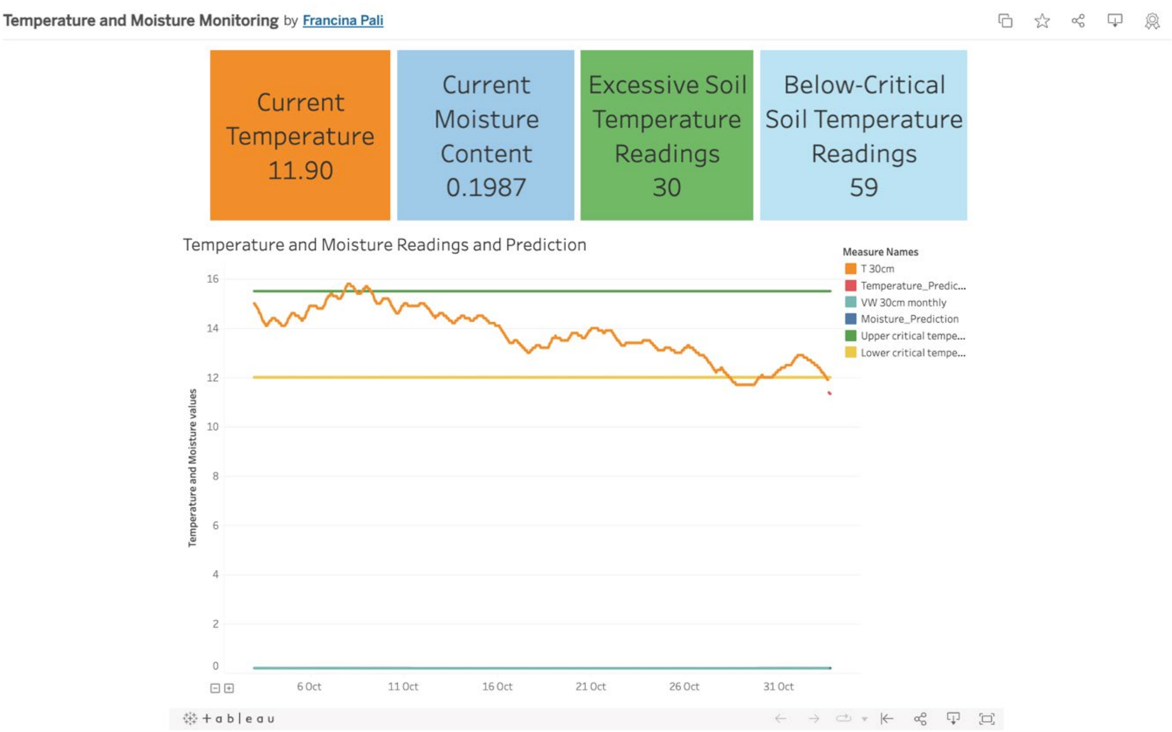
Visualization after computing the 30-day moving average for both the moisture and temperature data.



- The inclusion of temperature data further enhanced the performance of the model. This improvement is logical because we are providing the model with additional information beyond solely the past moisture content values. Our exploratory data analysis also did not indicate a correlation between moisture content and temperature, so it is expected that incorporating temperature data would lead to improved predictions.
- Using the moving average for both the moisture and temperature data improved the model even further. Therefore, we selected the fifth model for the moisture prediction task.
- The MSE value also for the fifth model is the lowest i.e., 7.654973094557386e-11 compared to all the other linear regression models trained.

Dashboard

Figure 9



- we created a time series visualization.
- we made a line chart of the temperature and the moisture readings of the last 30 days over time and show the prediction at the end of that period.
- We also created two summary tiles named “Excessive soil temperature readings" and “Below-Critical soil temperature readings. They display the number of times that the soils temperature went above and below the critical thresholds.
- we used the green and yellow colour for these two summary tiles as they were the colours used for the lines on the time series chart.
- Assumptions made: -
- Upper critical temperature threshold=15.5.
- Lower critical temperature threshold= 12.

- We also created a status tile which reports the most recent moisture and temperature reading.
- We decided to use orange colour for the status tile that displayed the most recent temperature reading and blue colour for the status tile that displayed the most recent moisture reading as they were the colours used for the lines on the time series chart for temperature and moisture.
- We then created a dashboard by adding in all our 5 sheets.

Conclusion

In conclusion, the use of AI algorithms and IoT technology to predict future soil moisture and temperature levels offers immense benefits for farmers, agriculture researchers, and land managers.

Farmers could use this system to optimize their irrigation schedules and crop management practices by receiving predictions on future soil moisture and temperature levels. This can help them reduce water usage and minimize crop stress, leading to higher yields and more efficient resource utilization. Agriculture researchers could also use this system to gather data on soil moisture and temperature levels to study the effects of different management practices on crop growth and health.

Land managers could use this system to plan and monitor land-use activities, such as reforestation or restoration projects, to ensure that soil moisture and temperature levels are within the optimal range for successful growth and development.

In terms of industry, this IoT application/system could fit into the agriculture industry, specifically in the subcategory of precision agriculture, which aims to optimize agricultural production and resource management using technology and data analytics.

References

Kim, D., Lee, H., & Kim, H. (2020). Soil moisture and temperature forecasting system using linear regression and LSTM for agricultural applications. Computers and Electronics in Agriculture, 176, 105614.

Liu, Z., Xu, X., Zhang, Q., & Li, J. (2021). A soil moisture and temperature forecasting system based on linear regression and LSTM networks for precision irrigation. Agricultural Water Management, 245, 106601.

Li, Y., Li, J., & Wang, X. (2022). A deep learning-based soil moisture and temperature forecasting system using linear regression and LSTM models. Journal of Environmental Management, 313, 114287.