Precision Irrigation with Cost-effective and Autonomic IoT Devices using Artificial Intelligence at the Edge

D3.2.

Report on the development of the OSIRRIS light-weight AI irrigation model

Responsible Editor: Innotec21

Contributors: Waziup e.V.

Document Reference: D3.2

Distribution: Public

Version: 1.0

Date: Septembre, 2023

Table of Contents

1. Introduction

The "Osirris AI irrigation model" is a pivotal element within the comprehensive Osirris Irrigation System, offering precise and dependable predictions of soil moisture. This model empowers informed decision-making by farmers, trained using data collected during the data collection phase and external agricultural data. Following its integration into the Osirris Irrigation System, it will undergo real-world testing.

The irrigation model forecasts near-future soil moisture content, facilitating proactive water management. It goes beyond simple moisture prediction, incorporating advanced features like projecting days until critical dryness for a specific crop. Additionally, it estimates necessary water quantities to maintain optimal soil moisture levels.

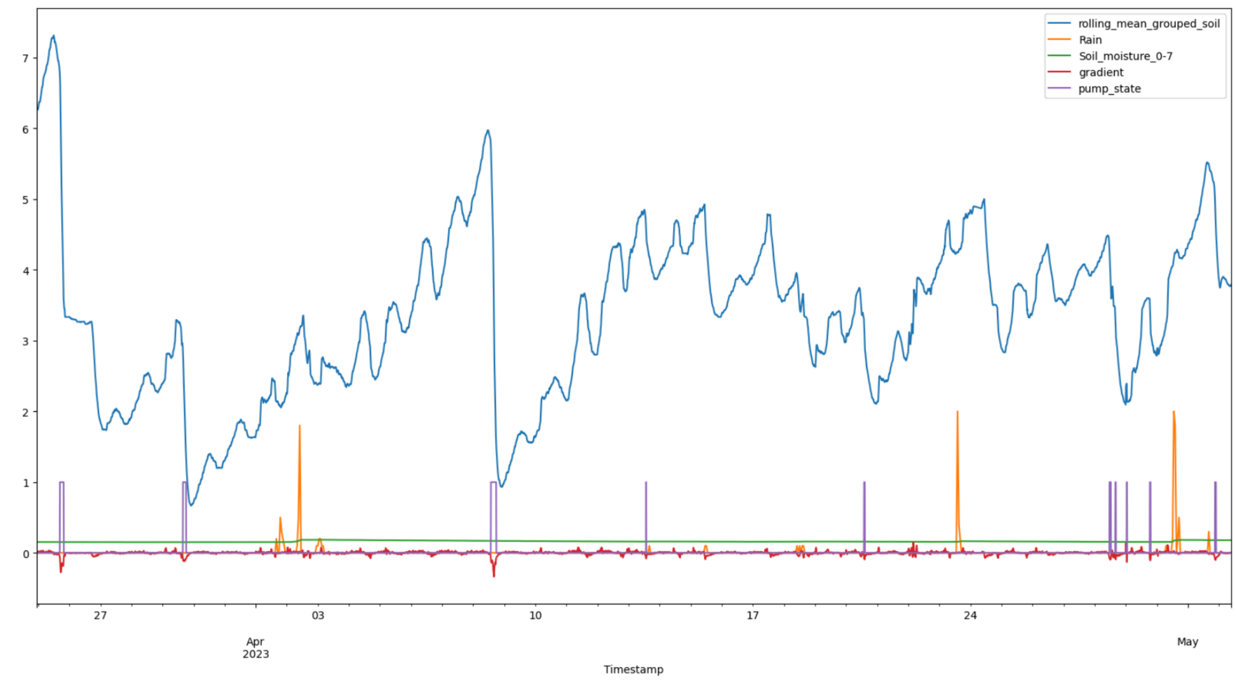

Figure 1: The graph illustrates soil water tension changes over time for an apple tree in Sbeïtla (Sandy soil). The upward trend and daily fluctuations signify soil drying, while irrigation events are shown by downward "slopes," replenishing soil moisture. Gradient(s) represent the change in soil moisture tension, coupled with precipitation (in the legend mentioned as “Rain”), assumptions can be made on when there was an irrigation period.

The model's capacity to estimate water needs enhances its utility. By accounting for weather, evapotranspiration, and soil attributes, it guides precise irrigation amounts. This optimizes water management, encourages sustainable agriculture, and curbs wastage.

Our AI model development follows an Agile approach, releasing iterative versions to enhance predictive capabilities. Each iteration integrates additional data sources. Validation and analysis inform adjustments for improved accuracy, robustness, and generalization.

2. Data Collection

The initial prototype models were trained using data collected from an on-site weather station, which measured a range of meteorological attributes such as:

- Température extérieure(°C)
- Humidité extérieure(%)
- Pression absolue(hpa)
- Vitesse du vent(km/h)
- Direction du vent
- Facteur vent(°C)
- Précipitations dernière heure(mm)
- Précipitations dernières 24 h(mm)
- Précipitations de la semaine(mm)
- Précipitations du mois(mm)
- Précipitations totales(mm)
- Lumière(lux)

The weather station used saves measurement results to a CSV file. These measurements can be manually downloaded from the SD card. However, due to the iterative model training process, we needed a solution to access the most recent weather data.

While this approach provided localized data, we recognized the need for a more versatile and comprehensive solution. Consequently, we shifted to utilizing the weather-meteo.com API, a valuable resource that allows us to input geographical coordinates and retrieve a wealth of historical weather data and also forecasts according to different weather forecast models. Forecasts will help the model produce more reliable results, because it can also leverage sophisticated meteorological predictions into account.

The weather-meteo.com API encompasses an extensive array of data points, including:

- Temperature (2 m)
- Relative Humidity (2 m)
- Dewpoint (2 m)
- Apparent Temperature
- Precipitation Probability
- Precipitation (rain + showers + snow)
- Rain
- Showers
- Snowfall
- Snow Depth
- Weathercode
- Sealevel Pressure
- Surface Pressure
- Cloudcover (Total, Low, Mid, High)
- Visibility
- Evapotranspiration
- Reference Evapotranspiration (ET_0)
- Vapor Pressure Deficit
- Wind Speed (10 m)
- Wind Speed (80, 120, 180 m)
- Wind Direction (10, 80, 120, 180 m)
- Wind Gusts (10 m)
- Temperature (80, 120, 180 m)
- Soil Temperature (0, 6, 18, 54 cm)
- Soil Moisture (0-1, 1-3, 3-9, 9-27, 27-81 cm)

By incorporating this wide-ranging data, our AI model gains a more holistic understanding of environmental conditions that impact soil moisture. To ensure our AI model remains efficient and avoids overfitting, we strategically selected a subset of these features for training:

- Temperature
- Humidity
- Precipitation (rain + showers + snow)
- Cloudcover
- Shortwave_Radiation
- Windspeed

- Winddirection
- Soil_temperature_7-28
- Soil_moisture_0-7
- Et0_evapotranspiration

This thoughtfully curated feature set strikes a balance between predictive accuracy and model complexity.

The data from the API is returned in a JSON formatted response, this is being saved in a data frame. Afterwards the dates are converted to datetimeindex in ISO format, set as index and

```python
import subprocess
import json
from datetime import date

# Example data
lattitude, longitude = 35.222866, 9.090245
start_date, end_date = '2022-12-08', '2023-07-04'
today = date.today()

def get_historical_weather_api(lattitude, longitude, start_date, end_date):
dct                                = subprocess.check_output(['curl',
f'https://archive-api.open-meteo.com/v1/era5?latitude={lattitude}&longitude={longitude}&start_date={start_date}&end_date={today.strftime("%Y-%m-%d")}&hourly=temperature_2m,relativehumidity_2m,rain,cloudcover,shortwave_radiation,windspeed_10m,winddirection_10m,soil_temperature_7_to_28cm,soil_moisture_0_to_7cm,et0_fao_evapotranspiration']).decode()
  dct = json.loads(dct)
  # I also convert it to a pandas dataframe

  data = (pd.DataFrame([dct['hourly']['temperature_2m'],
           dct['hourly']['relativehumidity_2m'],
           dct['hourly']['rain'],
           dct['hourly']['cloudcover'],
           dct['hourly']['shortwave_radiation'],
           dct['hourly']['windspeed_10m'],
           dct['hourly']['winddirection_10m'],
           dct['hourly']['soil_temperature_7_to_28cm'],
           dct['hourly']['soil_moisture_0_to_7cm'],
           dct['hourly']['et0_fao_evapotranspiration'],
           dct['hourly']['time']],
              index = ['Temperature', 'Humidity', 'Rain', 'Cloudcover', 'Shortwave_Radiation', 'Windspeed',
'Winddirection', 'Soil_temperature_7-28', 'Soil_moisture_0-7', 'Et0_evapotranspiration', 'date'])
     .T
     .assign(date = lambda x : pd.to_datetime(x.date, format='%Y-%m-%dT%H:%M'))
     .set_index(['date']).dropna())

  return data

data_weather_api = get_historical_weather_api(lattitude, longitude, start_date, end_date)
```

An important factor in accurate soil moisture predictions is accounting for soil type variability. To address this, we calculated volumetric soil water content using the Van Genuchten approach. We have also integrated predefined soil water retention curves corresponding to major soil types. Moreover, recognizing the diversity of soil compositions across agricultural fields, we offer farmers the flexibility to input their custom soil water retention curve. This adaptation ensures that the AI model's predictions align closely with the unique characteristics of the cultivated soil.

3. Data Preparation

The heart of the Osirris AI irrigation model lies in its ability to process diverse and dynamic datasets from various environmental conditions, soil types, weather data, and crop variations. To ensure its versatility and reliability, the data preparation pipeline has been meticulously designed and tested.

The project is written in python programming language. For easy access and collaboration, we have created a GitHub repository where you can find the code using the following link:

GitHub repository

This section outlines the comprehensive process that transforms raw data into a structured and informative format for accurate predictions. This process includes the following steps:

1) Interpolation and Outlier Handling:

Upon retrieval of sensor data through the Wazigates API, an initial step involves linear interpolation. This method tackles missing data points and mitigates the impact of outliers, thus fostering a more continuous and coherent dataset.

2) Weather Data Integration:

Weather-related data is fetched from the API, typically available on an hourly basis. To align it with the desired sampling rate, a linear resampling technique is applied. This ensures synchronization between sensor and weather data timestamps, crucial for meaningful correlation.

3) Timestamp-Based Merging:

Merging the two dataframes is a pivotal step, requiring accurate alignment of timestamps. This integrated dataset forms the foundation for subsequent analysis and modeling.

4) Temporal Alignment:

To prevent spurious learning from non-existent data, the integrated dataframe is trimmed to match the shortest time series duration, fostering uniformity and accuracy in predictions.

5) Evaluation Metric and Data Quality:

A prior evaluation methodology computed deviation from ground truth data, evaluating preprocessing steps including interpolation and imputation methods using diverse metrics. This quality assurance step ensures robust and reliable data.

6) Data Type Standardization:

Aligning data types is critical for consistent analysis. For instance, recalculating wind directions into degrees ensures uniformity across features.

7) Feature Engineering:

The model's performance is augmented by engineered features. Grouped soil metrics, rolling means, and temporal indicators like hour, minute, date, and month provide contextual information to the model, enhancing its predictive capability.

8) Soil Water Content Calculation:

Volumetric water content (VWC) is pivotal in soil moisture predictions. The calculation depends on the soil type, and predefined soil water retention curves are provided. Farmers can contribute by inputting custom curves, refining predictions for their specific conditions.

9) Feature Normalization:

Normalization harmonizes feature ranges, a crucial step for AI models. The linear max-min normalization method is employed, ensuring equal importance for features regardless of their value range.

10) Gradient Calculation and Pump State:

Detecting rapid soil tension declines indicates precipitation or irrigation. The gradient feature signifies this change. The pump state, denoting irrigation periods, is integrated, enhancing prediction accuracy.

11) Cleaning and Cutting:

Data cleanliness is paramount. The data frame is refined by renaming columns, eliminating missing values, and aligning with rolling mean window sizes, thereby optimizing input quality.

12) Training Data Preparation:

To facilitate training, the data frame is segmented into separate subsets, each encompassing periods between irrigation events. This strategic partitioning enhances model learning and predictive accuracy.

By embracing this comprehensive data preparation pipeline, the Osirris AI irrigation model not only accommodates a multitude of scenarios, but also ensures the robustness and generalization required for reliable predictions across diverse environmental settings, soil conditions, and crop types.

4. Model Training

The journey of training and refining the Osirris AI irrigation model is characterized by meticulous steps, each contributing to the model's accuracy and reliability within the specified irrigation time frame. By delving into the intricacies of each phase, the following elucidation sheds light on the iterative process that transforms raw data into informed predictions.

1) Initial Setup: Data Split and Feature Pruning

At the outset, the already segmented dataframe is split into separate training and testing sets. To streamline the model's learning process, redundant or unnecessary features like the previously engineered "gradient" are discarded, refining the input.

2) Model Training and Comparison: Evaluating Regression Models

The exploration continues with a rigorous comparison of various regression models. From this evaluation, the top-performing models are earmarked for subsequent phases. This preliminary selection lays the foundation for an optimized model. In the following there is a list of the models that are being compared:

- Linear Regression
- Extra Trees Regressor
- Random Forest Regressor
- Gradient Boosting Regressor
- Decision Tree Regressor
- Light Gradient Boosting Machine
- CatBoost Regressor
- Extreme Gradient Boosting
- AdaBoost Regressor
- Ridge Regression
- Bayesian Ridge
- Huber Regressor
- Elastic Net
- K Neighbors Regressor
- Lasso Regression
- Lasso Least Angle Regression

We also evaluated other models, but due to bad performance metrics we omitted them and did not waste resources on training them. In the following there is a list of the models that are not used:

- Least Angle Regression
- Dummy Regressor
- Orthogonal Matching Pursuit
- Passive Aggressive Regressor

In the development of the Osirris AI irrigation model, thorough model comparison and testing have emerged as critical steps. The rationale behind this lies in the fact that different models perform differently across various datasets and time periods. Each dataset carries its unique characteristics, which can significantly influence a model's effectiveness.

To validate this, we conducted testing across five distinct time periods, representing diverse environmental conditions and crop growth stages. This extensive evaluation served as a benchmark to assess model performance. Models consistently displaying poor performance across all test periods were deliberately excluded from further consideration.

The core message here is the acknowledgment that no single model universally excels in all scenarios. Variability in soil conditions, weather patterns, and crop attributes demands a comprehensive exploration of multiple modeling approaches. By systematically comparing models across diverse datasets and temporal contexts, we not only identify the best-suited candidates but also gain insights into the strengths and limitations of each model.

In summary, this iterative process of model comparison and testing underscores the need for adaptability and versatility in AI systems. It ensures that the final Osirris AI irrigation model not only performs well on training data but also exhibits robust adaptability to the dynamic challenges posed by varying environmental conditions, soil compositions, and crop characteristics.

3) Hyperparameter Tuning: Precision Enhancement through Parameter Optimization

Hyperparameter tuning follows, a critical step where model parameters are adjusted. This fine-tuning can occur automatically or guided by a predetermined parameter grid. By conducting both methods and comparing the outcomes, any improvements are meticulously tracked.

4) Ensemble Model Creation: Augmenting Predictive Power

A pivotal step encompasses the creation of ensemble models, to further increase the quality of predictions. This is done in various ways:

- **Blended Models and Timespan Comparison:** These blended models are generated from the same and different timespans, necessitating data alignment through precise model cutting. The ensuing evaluation gauges the performance of these ensemble models, identifying optimal combinations.
- **Model Stacking: Combining Predictive Abilities:** Further advancing ensemble strategies, model stacking is deployed. The performance of these stacked models is scrutinized, leading to the selection of the most proficient option. This approach augments the model's predictive prowess through careful combinations.

5) Evaluation on Unseen Data: Robustness Assessment

Evaluating the model's reliability on unseen data is pivotal. Fresh time spans are extracted from the "ground truth" dataset, uninvolved in the training or testing sets. Comparing these time spans against the model's predictions furnishes insights into performance, ensuring the model's robustness.

6) Data Enrichment for Enhanced Predictions: Integration of Forecasted Data

To enhance prediction accuracy, known or forecasted data is injected into the model. Weather forecast data from the www.open-meteo.com API becomes a cornerstone. Timestamps are converted into time-related data columns, while Volumetric Water Content (VWC) and soil temperature are acquired from the forecast data, replacing previously sensor-derived calculations. The predictive influence of the pump state feature is also set to zero to estimate potential soil dryness.

7) Validation and Metrics Computation: Gaining Insights into Performance:

To validate the model's refined predictions, a meticulous comparison is established between the "ground truth" sensor data and the model's projections. A spectrum of metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Percentage Error (MPE), and R-squared Score (R2-Score), are calculated. These metrics serve as compasses, directing the assessment of the model's proficiency in forecasting independent data.

Embarking on this intricate process, the Osirris AI irrigation model evolves through stages of assessment, comparison, and refinement. The amalgamation of careful evaluation, predictive enrichment, and meticulous model comparisons culminate in an AI system that not only thrives on training data but also demonstrates robust adaptability to unforeseen environmental dynamics, soil conditions, and crop characteristics.

5. Soil Water Tension Forecasting

5.1. Visual Analysis

The visualization process commences by rendering the best performing models through a series of insightful plots. Each plot serves as a visual snapshot of the model's predictions against the actual "ground truth" data.

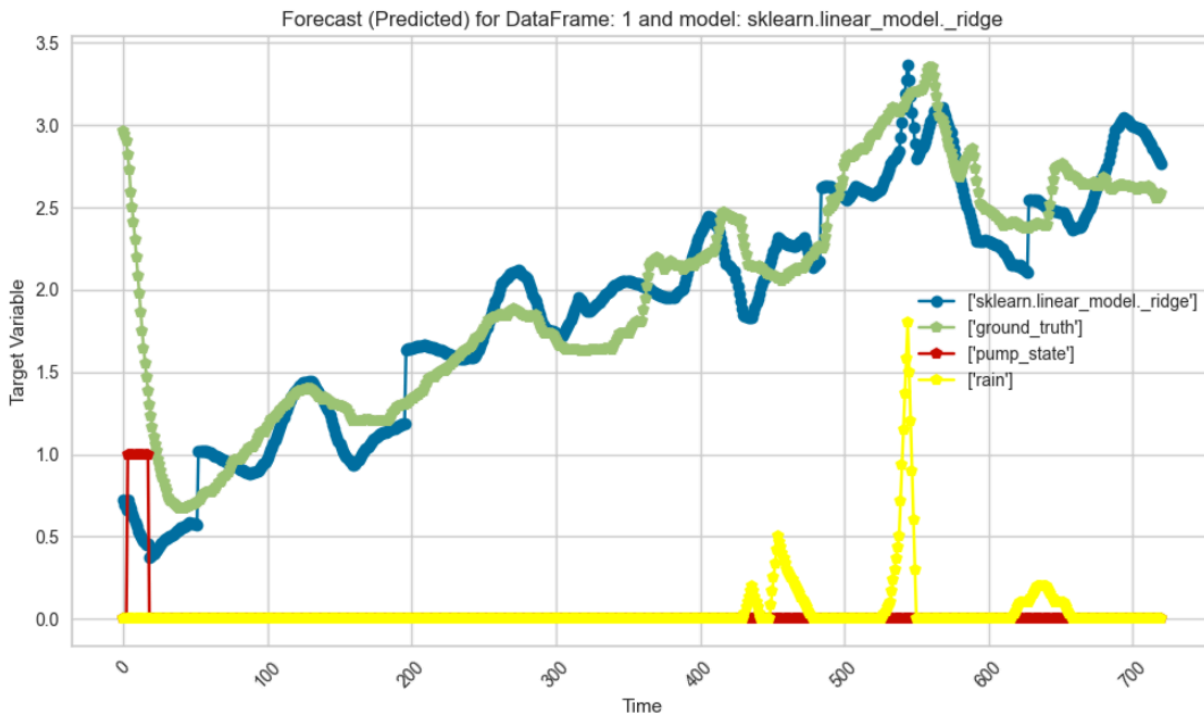

Figure: 5 day forecast for Soil tension values on unseen data

This visual juxtaposition provides an immediate understanding of the model's accuracy across diverse scenarios.

- **Spotting Potential Problems:** Addressing Outliers and Anomalies
- **Diverse Scenarios:** Comprehensive Insights into Model Behavior
- **Fine-Tuning through Visual Insight:** In instances where problematic areas emerge, visual validation acts as a stepping stone for iterative improvement. These insights guide the refinement process, empowering data scientists to fine-tune the models further, potentially mitigating anomalies and outliers.

In the journey of the Osirris AI irrigation model from raw data to predictive capability, visual validation stands as a crucial checkpoint. This phase presents the model's intricate dynamics in a comprehensible format. It serves a dual purpose: ensuring the model's reliability and bridging the gap between data science intricacies and practical decision-making.

5.2. Validation on Unseen Test Dataset

When assessing the performance of our models, it's crucial to focus on metrics generated from unseen data rather than metrics produced during training. To achieve this, we adopt a two-step validation process.

Validation Set: During training, we set aside the most recent part of the data frame, ensuring it is not used for training or even as a validation set. This reserved portion becomes our test set. The validation set, comprising the remaining data, serves as the basis for cross-validation during model training.

Test Set Evaluation: After training, we rigorously evaluate our model's performance on the test set, which it has never encountered before. This step is paramount in selecting the best-performing model for the task. It guards against the risk of overfitting, where models excel in training data but perform poorly on new, unseen data.

Model Refinement: The top-performing model is further refined by retraining it individually, incorporating the previously withheld test set. This ensures the model has the most up-to-date data points for optimal accuracy.

Performance Metrics

For our Subaytilah test site in Tunisia, we obtained the following performance metrics for a 5-day forecast horizon using unseen data:

Metric	**Result**
MAE	0.23
RMSE	0.35
MPE	12.52 %
R2	0.73

In the field of statistics, the Mean Percentage Error (MPE) represents the calculated average of percentage errors, indicating how forecasts made by a model deviate from the actual values of the forecasted quantity. These metrics offer a promising outlook, particularly in the context of agricultural applications, where a 12.5% MPE signifies a reliable forecasting performance.

5.3. Predicting Optimal Irrigation Timing

The final stride of the Osirris AI irrigation model's journey brings its culmination in empowering farmers with actionable insights—specifically, predicting the optimal timing for irrigation. This pivotal step serves as the bridge between sophisticated machine learning and real-world agricultural practices, enabling farmers to make informed decisions aligned with their crops' needs.

Predictive Guidance for Optimal Irrigation

Harnessing the predictive prowess of the model, the system computes and projects when the pre-established soil tension threshold will be reached. This projection, grounded in data-driven insights, offers farmers a tangible timeframe within which their plants will require irrigation. By predicting the impending soil tension threshold crossing, the model provides an invaluable cue to initiate irrigation practices.

Precision and Customization: Adapting to Individual Needs

Recognizing the diversity in soil types, crop characteristics, and local conditions, the predictive guidance takes into account the specific settings of each agricultural plot. The Osirris AI irrigation model's adaptability ensures that the forecasted irrigation timing is tailored to individual needs, enhancing its precision and practicality.

Enabling Proactive Decision-Making

The predictive irrigation timing empowers farmers to proactively plan their irrigation routines. This proactive approach not only ensures that plants receive the right amount of water at the right time but also optimizes water resource management. By aligning irrigation practices with plants' actual needs, water wastage is minimized, and crop health is maximized.

Holistic Integration into Agricultural Practices

The predictive guidance seamlessly integrates into existing agricultural routines, bridging the gap between cutting-edge technology and traditional farming wisdom. Farmers can now rely on data-backed predictions to guide their actions, making the irrigation process more efficient, effective, and responsive to real-time conditions.

A Glimpse into the Future

As the Osirris AI irrigation model evolves, its predictive capabilities are poised to become increasingly refined and accurate. Through continuous learning and iterative enhancement, the model's ability to anticipate optimal irrigation timing will only improve, solidifying its role as a trusted companion in modern agriculture.

In this final step, the Osirris AI irrigation model transcends data algorithms to deliver tangible benefits to farmers—offering the gift of foresight and the power of timely, informed decision-making.

6. Conclusion

In the realm of modern agriculture, the fusion of technology and cultivation practices has yielded transformative results. This report has explored the evolution and implementation of the "Osirris AI irrigation model," a pivotal component within the comprehensive Osirris Irrigation System. As we conclude this discourse, we recognize the overarching objective: to revolutionize irrigation practices by harnessing the power of artificial intelligence, offering pragmatic solutions to age-old challenges.

The journey embarked with a vision to usher in a new era of precision agriculture, where water management goes beyond tradition. The inception of the Osirris Irrigation System unveiled a transformative paradigm, deploying sensor devices armed with real-time data collection capabilities. This transition from conventional methodologies to data-driven insights served as the foundation for informed decision-making, curbing water wastage, and augmenting crop health.

The Osirris AI irrigation model represents a significant innovation achieved through iterative prototyping. Its primary function is predicting soil moisture content, leveraging a combination of meteorological data. This predictive capability empowers farmers by providing valuable insights for precise and timely irrigation. In a world where sustainability is a pressing concern, this predictive power aids in optimizing resource usage, a significant step towards resource-efficient agriculture.

The model's architecture is designed with meticulous attention to detail. It incorporates diverse sensor data, weather forecasts, and soil attributes, making it adaptable to various environmental conditions. By utilizing machine learning techniques, the model handles complexity effectively, capturing the nuances of irrigation dynamics with high accuracy.

A cornerstone of the Osirris AI irrigation model's evolution was its inherent modularity. The flexibility to accommodate different hardware setups, sensor types, and crop specifications is a testament to its versatility. Furthermore, the system's capacity to integrate seamlessly with the cloud infrastructure and remote access amplifies its practicality.

As the Osirris AI irrigation model journeys forward, continual refinements and enhancements remain a paramount endeavor. The emphasis on predictive precision, data-driven optimization, and real-world applicability sets the stage for ongoing innovation. The integration of predictive irrigation timing bestows farmers with the invaluable ability to make well-informed decisions, underpinned by advanced technological insights.

In closing, the Osirris AI irrigation model serves as a testament to the remarkable symbiosis between technology and agriculture. This report has illuminated its evolution, from conception to realization, encapsulating its intricate components, methodologies, and aspirations. As the agricultural landscape continues to evolve, the Osirris AI irrigation model stands as a beacon of progress, a harmonious blend of innovation and tradition, poised to redefine the contours of modern farming.