واد مصر الرقمية

SMART FARMING SYSTEM

Where Innovation Cultivates Growth

AGRISENSE

Our team

Yasmeen Abdallah

Yasmine Abdelgawad

Faress Farrag

Mohamed Sherif

Omar Mostafa

Omar Ahmed

Table of contents

01

Overview

Problem statement and solution

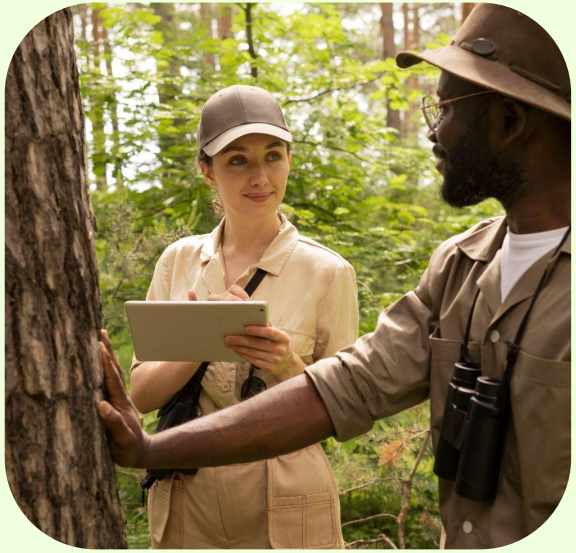
The problem

- “Farmers lose up to 50% of irrigation water due to poor monitoring and delayed decisions.”
- Farmers can’t monitor soil & climate continuously
- Irrigation decisions are delayed or inaccurate
- Data is scattered, not analyzed, not actionable

Our solution

AGRISENSE:

- provides a smarter, data-driven way to monitor farms
- Simulated farms using live data, ready for real production deployment.
- Gives real-time insights, alerts, and trend analysis
- Supports better irrigation and soil-health decisions



Data Simulation & Real-Time Streaming

02

- Real farms normally use IoT sensors to capture soil & climate conditions.
- For this project, we created a simulation to test the system end-to-end.
- We use two data flows:
  - Historical Data: full 2024 simulation (minute-by-minute)
  - Live Stream: new reading generated every 3 seconds

Simulated Farms & Regions

10 farms created with different climate and crop conditions

Nile Delta

humid & moderate climate

Upper Egypt

Hot & Dry

Sinai

Cooler with Microclimate

Sensor Variables & How They Affect Each Other

☀️ Sunlight	🌡️ Temp ↑	💧 Moisture ↓	
🌧️ Rain	💧 Moisture ↑	💨 Humidity ↑	🧪 pH ↓
🌡️ Temp	💨 Humidity ↓		
💧 Moisture	🧪 pH ↑		

Live Data Generation + Kafka Producer

➔ Every ~3 seconds, we generate a new sensor reading
Calculated using:

- time of day
- farm region
- current weather patterns
- last moisture/pH state

1. Reading is turned into JSON
2. Sent instantly to Kafka topic smart_farming_data
3. Spark consumes the stream in real time

03

Setting up the pipeline

End-to-End Data Pipeline

Our pipeline processes real-time agricultural sensor data through four key stages:

- **Data Ingestion →** Apache Kafka
- **Stream Processing →** PySpark Structured Streaming
- **Dual Storage →** PostgreSQL + HDFS
- **Analytics & ML →** Streamlit Dashboard + Predictive Models

kafka

Apache Kafka - Real-Time Data Ingestion

Distributed streaming platform for real-time data pipelines

Our Implementation

- Topic: **smart_farming_data**
- Producer: Python script reads sensor data and publishes to Kafka Message
- Format: JSON with sensor readings (soil moisture, temperature, pH, rainfall, humidity, sunlight, pesticide usage)
- Frequency: Continuous streaming.

PySpark - Connecting Kafka to Storage

Acts as the middle layer that reads from Kafka and writes to both PostgreSQL and HDFS simultaneously

Our Implementation

- Consume from Kafka: Continuously reads messages from **smart_farming_data** topic
- Parse & Transform: Converts JSON to structured format
- Dual Write Operation: Sends processed data to two destinations in parallel: PostgreSQL → For real-time queries HDFS → For historical storage

PostgreSQL - Real-Time Operational Database

Stores latest sensor readings for immediate access and real-time dashboards

	sensor_id [PK] uuid	timestamp timestamp with time zone	soil_moisture double precision	soil_ph double precision	temperature double precision	rainfall double precision	humidity double precision	sunlight_intensity double precision	pesticide_usage_ml double precision	farm_id character varying (50)	region character varying (100)	crop_type character varying (100)
1	1b072f70-3c4e-4b91-b9f1-0144ae3df4b7	2025-11-26 22:10:07.006285+00	40.09	6.43	20.43	0	55.03	0	0	farm_1	NileDelta	Wheat
2	11ee3ff9-a8cb-4b63-886c-9097a8c2d48a	2025-11-26 22:10:08.662812+00	40.06	6.36	22.36	0	51.54	0	0	farm_2	NileDelta	Rice
3	18235bee-d115-40f4-af65-d69cd82be18b	2025-11-26 22:10:09.681076+00	35.11	6.74	19.77	0	56.82	0	0	farm_3	NileDelta	Onion
4	f2942b4c-ab44-47e7-8d09-9c3cce0484ba	2025-11-26 22:10:10.71585+00	39.81	6.13	23.2	0	54.77	0	0	farm_4	UpperEgypt	Tomato
5	fc08fdb5-d0f9-45a3-bdc4-a24ede2e5381	2025-11-26 22:10:11.763671+00	36.46	6.28	24.73	0	50.87	0	0	farm_5	UpperEgypt	Dates
6	1f065f96-5531-41e1-ba26-b2e3dbddd96d	2025-11-26 22:10:12.78212+00	36.76	6.36	25.4	0	50.84	0	0	farm_6	UpperEgypt	Peanuts
7	3cb435dc-f105-49ec-8896-da49408e30...	2025-11-26 22:10:13.811801+00	30.96	6.67	23.39	0	51.78	0	0	farm_7	Sinai	Corn
8	51e44d30-6ee2-4474-8131-5fc9b19a1d...	2025-11-26 22:10:14.841818+00	37.74	6.3	19.16	0	56.14	0	0	farm_8	Sinai	Olive
9	fcd6b5ab-08e4-4411-90a1-2e6a2512d8...	2025-11-26 22:10:15.851954+00	36.53	6.08	20.63	0	54.17	0	0	farm_9	Sinai	Barley
10	55c48152-9092-4bd0-9525-d9c1f595c7...	2025-11-26 22:10:16.863276+00	39.64	6.8	21.52	0	52.84	0	0	farm_10	Sinai	Potato
11	66fc7034-0927-4e37-b0bb-6f5c18cea424	2025-11-26 22:10:17.917938+00	40.03	6.44	18.93	0	57.59	0	0	farm_1	NileDelta	Wheat
12	4348f609-ebbd-44e8-aadd-8ef8395f1382	2025-11-26 22:10:18.927769+00	39.79	6.38	20.11	0	54.08	0	0	farm_2	NileDelta	Rice
13	bfe4d53b-a878-468d-a425-38b5035a3d...	2025-11-26 22:10:19.944158+00	35.21	6.73	18.16	0	56.84	0	0	farm_3	NileDelta	Onion
14	ece81923-bb92-4fcd-8e04-ca84cf41ffc5	2025-11-26 22:10:20.956343+00	39.66	6.13	20.54	0	57.06	0	0	farm_4	UpperEgypt	Tomato
15	5aedf060-e837-4e8e-a417-4ed152be87...	2025-11-26 22:10:21.977674+00	36.41	6.29	21.76	0	55.39	0	0	farm_5	UpperEgypt	Dates
16	e9c6e8e1-68ff-4939-9909-2d84e227d333	2025-11-26 22:10:22.991787+00	36.66	6.36	22.13	0	50.45	0	0	farm_6	UpperEgypt	Peanuts
17	c6404e22-9da3-4f19-8dfb-a1878e461ed7	2025-11-26 22:10:24.004761+00	30.63	6.66	20.88	0	53.73	0	0	farm_7	Sinai	Corn
18	3a01428c-189d-4899-bc55-708f4665f298	2025-11-26 22:10:25.015775+00	37.9	6.32	18.11	0	54.13	0	0	farm_8	Sinai	Olive
19	a7d1c495-6ace-449c-b155-37724f1436...	2025-11-26 22:10:26.033794+00	71.52	5.99	18.94	58.94	73.08	0	0	farm_9	Sinai	Barley
20	21682945-0859-41e2-bf86-d80972bdd1...	2025-11-26 22:10:27.057828+00	39.55	6.82	19.69	0	56.17	0	0	farm_10	Sinai	Potato
21	cfa63b86-96ae-4dde-91e8-a3186769214f	2025-11-26 22:10:28.144578+00	39.98	6.44	17.67	0	57.36	0	0	farm_1	NileDelta	Wheat

Hadoop HDFS - Batch Data

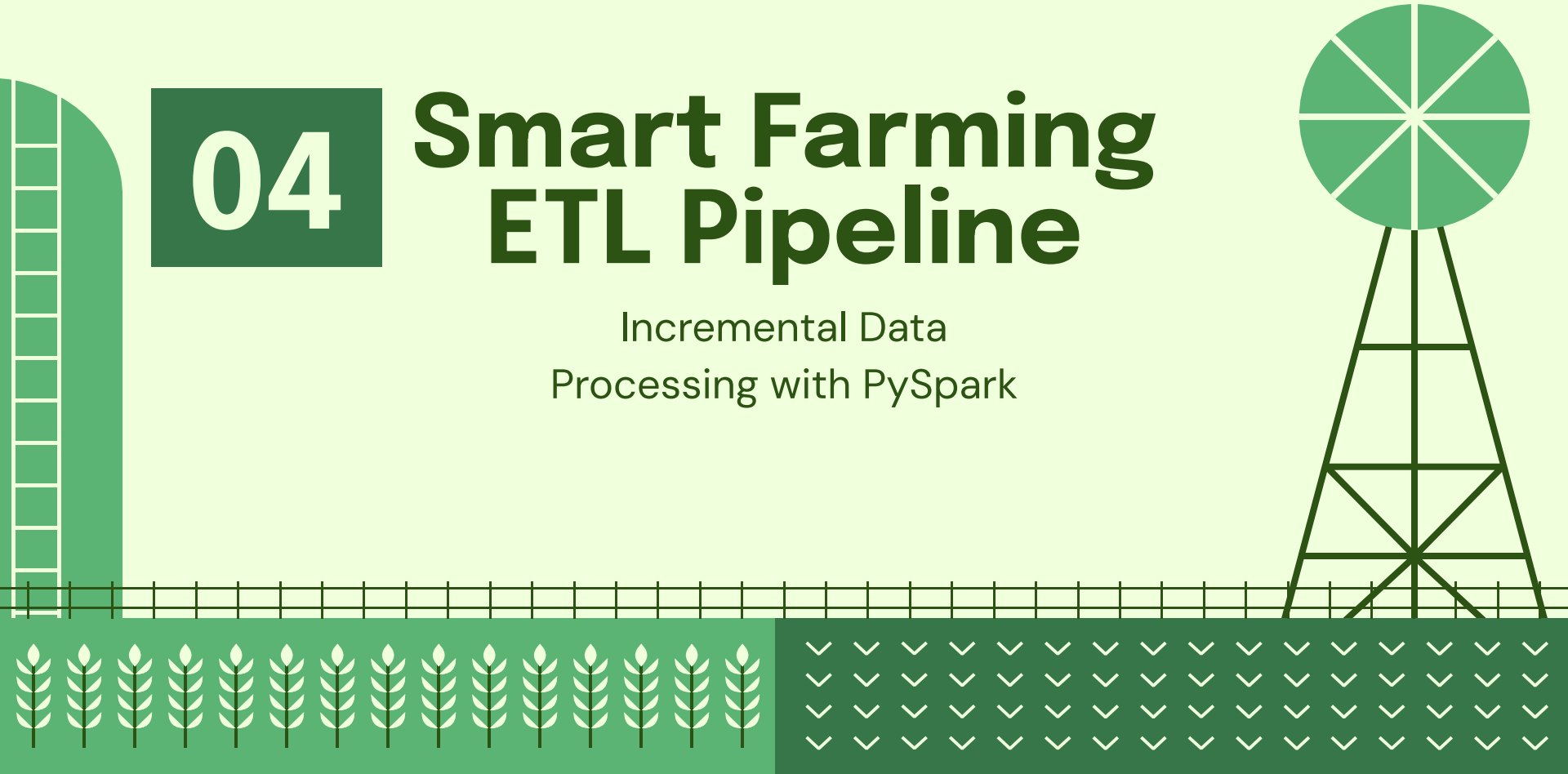
Stores all historical sensor data for batch analytics, machine learning, and data warehousing

Why HDFS?

- Scalable: Handles petabytes of data across distributed nodes
- Cost-Effective: Cheap storage for historical data
- Fault Tolerant: Data replicated across multiple nodes
- Analytics Ready: Optimized for large-scale batch processing

04 Smart Farming ETL Pipeline

Incremental Data Processing with PySpark



ETL Execution Flow

Step 1
Extract from HDFS

Step 2
Use Timestamp to only get new rows

Step 3
Handle NULL values

Step 4
Remove Outliers (IQR method)

Step 5
Feature Engineering: Temporal dimensions

Step 6
Star Schema design

Step 7
Loading to MySQL: 11 tables

Our ETL Logic with pyspark

Incremental Loading:

- Stored last processed timestamp
- Only processed new rows using timestamp and checkpoints
- Prevented duplication
- Exits if no new data is found

Missing Value Handling:

- Filled numeric NULLs using **mean imputation** for soil_moisture, soil_pH, temperature, humidity, sunlight_intensity

Outlier Detection & Removal:

- IQR method on numerical columns
- Removed noisy/physically impossible sensor readings

Text Cleaning:

- Trimmed whitespace in region & crop names

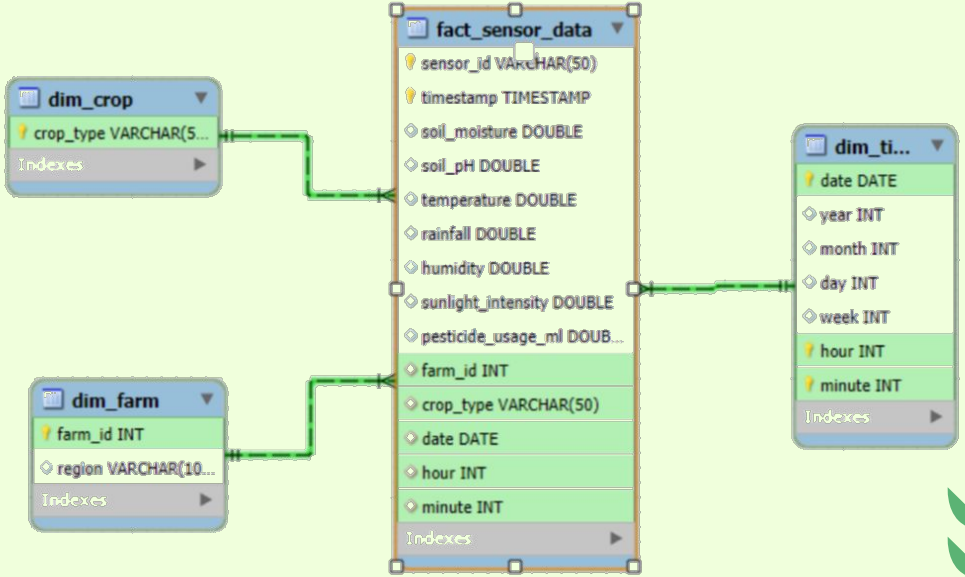
Feature Engineering

- Extracted date hierarchy: year, month, week, day, hour, minute.
- Derived moisture trend, rainfall, moisture interaction, climate effects, pH trend, pesticide usage patterns, sunlight exposure.
- Built multiple trend tables with grouped statistics.
- Prepared data for dashboard-ready insights.
- Created 7 analytical aggregation tables that cover moisture trends, rainfall correlations, climate impacts, pH monitoring, pesticide patterns, and sunlight analysis.

Star Schema Topology

Data Warehouse Model

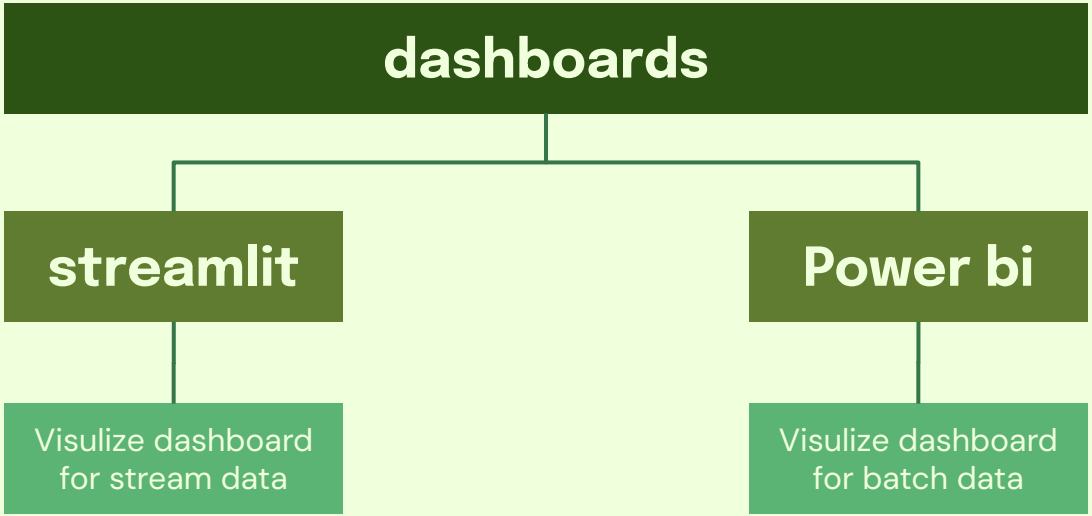
- 4 core tables:
  - 3 dimensions & 1 fact
- 7 analytical aggregations
- 11 total tables loaded incrementally to MySQL data warehouse with optimized column types and append mode



05 Analysis

Batch and Streaming dashboards

Analysis

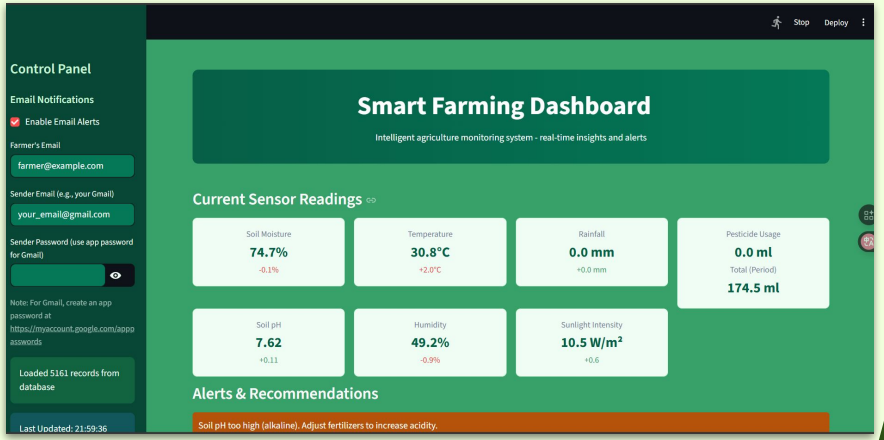

dashboards
streamlit
Power bi
Visulize dashboard for stream data
Visulize dashboard for batch data

Streamlit

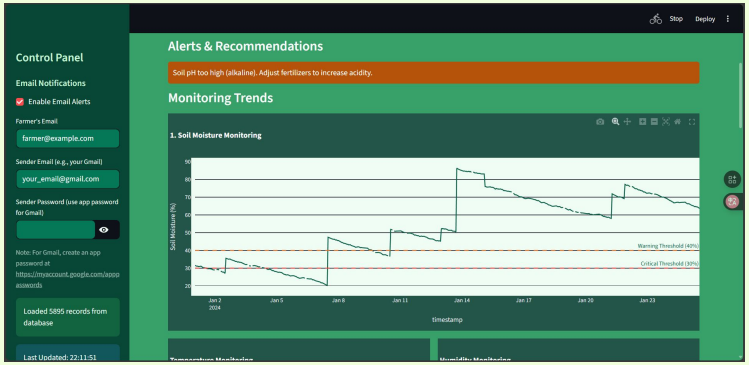
provides real-time monitoring, alerts, and interactive visualizations of sensor data.

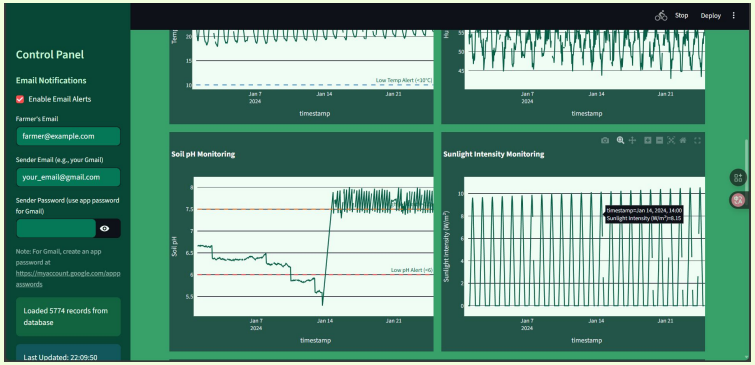
Dashboard Features

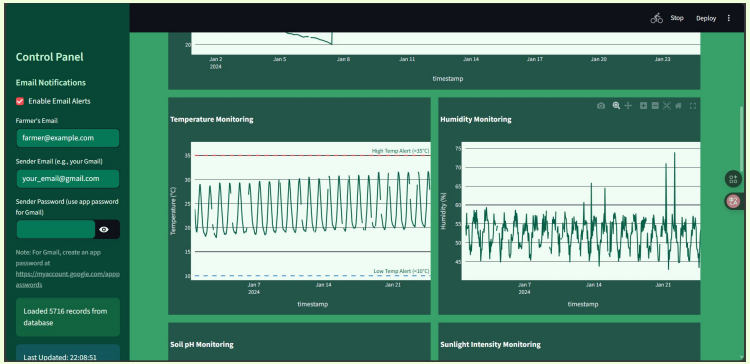
-Real-time sensor readings (soil moisture, temperature, pH, etc.)
-Interactive trend charts with threshold alerts
-Critical/Warning alerts based on sensor thresholds
-Email notifications for alerts (configurable in sidebar)
-Farm & Region filtering
-Historical data table for the last
-Auto-refresh every 1 second

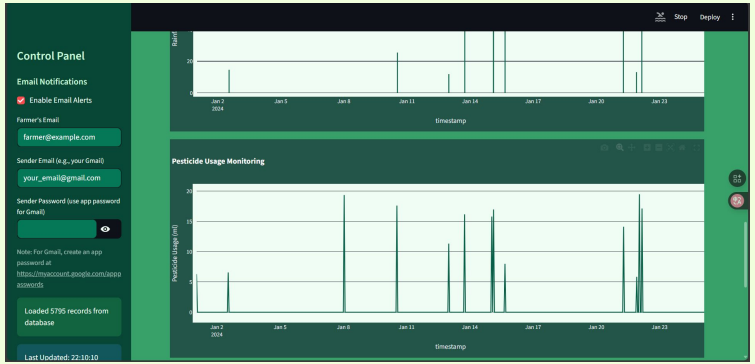


Charts


Stop
Deploy
Control Panel
Email Notifications
Enable Email Alerts
Farmer's Email
farmer@example.com
Sender Email (e.g., your Gmail)
your_email@gmail.com
Sender Password (use app password for Gmail)
Note: For Gmail, create an app password at https://myaccount.google.com/apppasswords
Loaded 5895 records from database
Last Updated: 22:11:51
Alerts & Recommendations
Soil pH too high (alkaline). Adjust fertilizers to increase acidity.
Monitoring Trends
1. Soil Moisture Monitoring
Soil Moisture (%)
Warning Threshold (40%)
Critical Threshold (30%)
timestamp


Stop
Deploy
Control Panel
Email Notifications
Enable Email Alerts
Farmer's Email
farmer@example.com
Sender Email (e.g., your Gmail)
your_email@gmail.com
Sender Password (use app password for Gmail)
Note: For Gmail, create an app password at https://myaccount.google.com/apppasswords
Loaded 5774 records from database
Last Updated: 22:09:50
Soil pH Monitoring
Soil pH
Low pH Alert (<6)
Sunlight Intensity Monitoring
Sunlight Intensity (W/m²)
timestamp


Stop
Deploy
Control Panel
Email Notifications
Enable Email Alerts
Farmer's Email
farmer@example.com
Sender Email (e.g., your Gmail)
your_email@gmail.com
Sender Password (use app password for Gmail)
Note: For Gmail, create an app password at https://myaccount.google.com/apppasswords
Loaded 5716 records from database
Last Updated: 22:08:51
Temperature Monitoring
High Temp Alert (>35°C)
Low Temp Alert (<10°C)
Temperature (°C)
Humidity Monitoring
Humidity (%)
timestamp
Soil pH Monitoring
Sunlight Intensity Monitoring

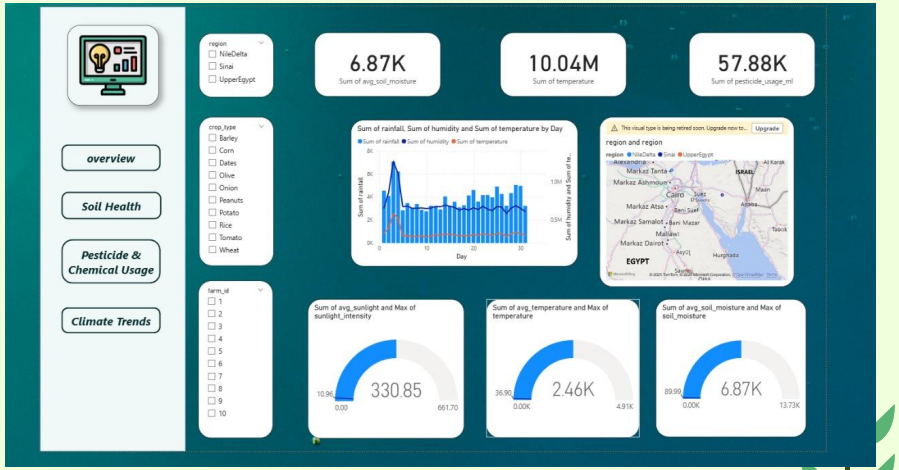

Stop
Deploy
Control Panel
Email Notifications
Enable Email Alerts
Farmer's Email
farmer@example.com
Sender Email (e.g., your Gmail)
your_email@gmail.com
Sender Password (use app password for Gmail)
Note: For Gmail, create an app password at https://myaccount.google.com/apppasswords
Loaded 5795 records from database
Last Updated: 22:10:10
Pesticide Usage Monitoring
Pesticide Usage (ml)
timestamp

Power BI

Dashboard provides real-time insights into farm operations, soil health, pesticide & chemical usage, moisture levels, and climate trends for farms in Egypt (Nile Delta, Sinai, Upper Egypt, and surrounding regions).

Overveiw

- Shows overall farm conditions: soil moisture, temperature, rainfall, pesticides.
- Map shows farm regions in Egypt (Nile Delta, Sinai, Upper Egypt).
- The main line chart shows the daily changes in rainfall, humidity, and temperature, giving insights into climate patterns and their effect on crops.
- Gauges show key ranges for sunlight, temperature, and moisture.
- The main line chart shows the daily



Soil health

- This dashboard focuses on overall **soil health and fertility.**
- pH vs. moisture chart shows how soil conditions change.
- Compares moisture levels for each **crop type**.
- Donut chart shows moisture distribution across farms.

Pesticide & Chemical Usage

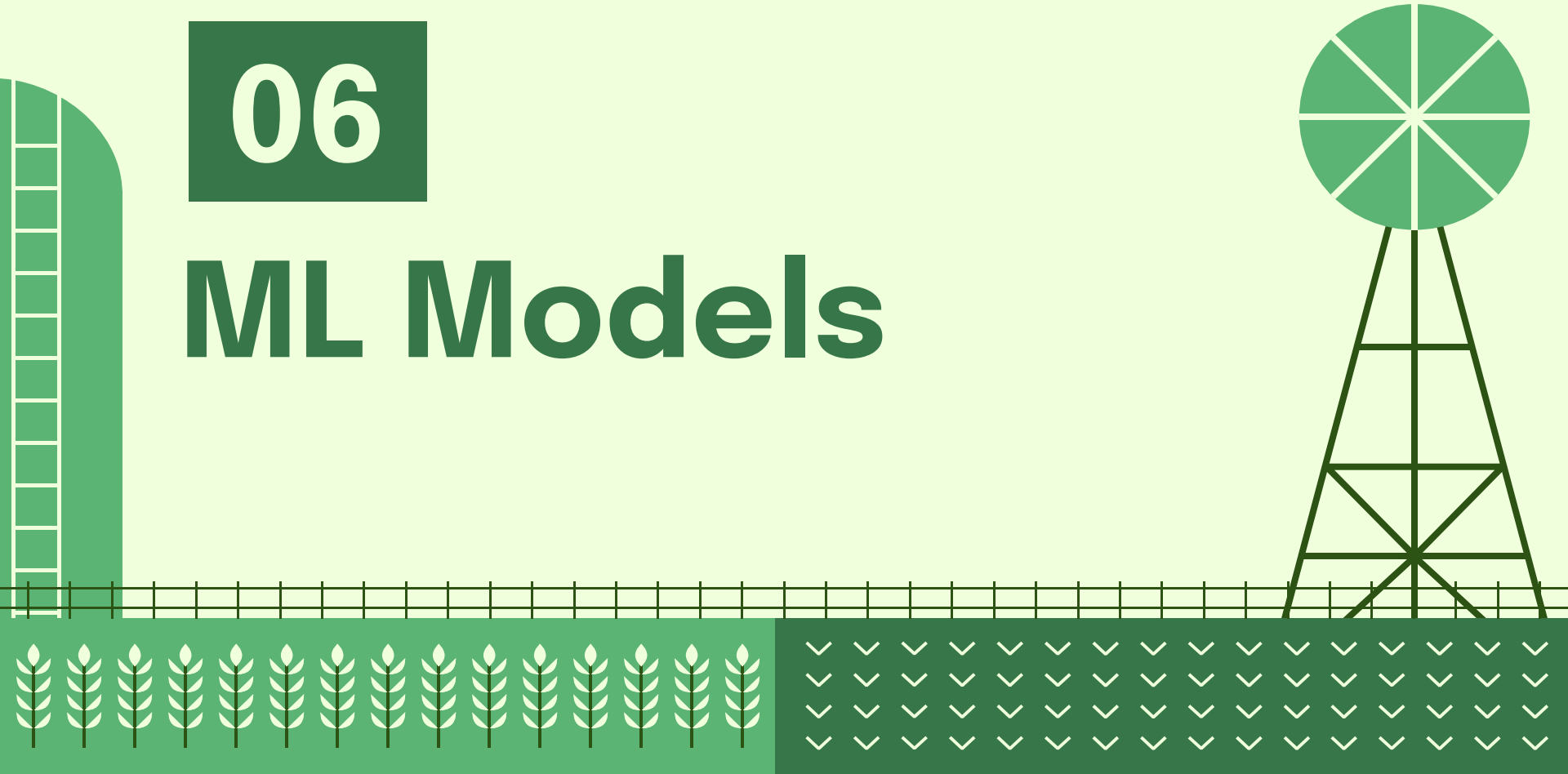
- Displays pesticide usage by **farm and week**.
- Highlights total pesticide consumption.
- Shows trends between **rainfall and pesticide use**.
- Compares pesticide usage across **different crop types**.

Climate Trends Dashboard

- This dashboard shows overall climate patterns including **total rainfall, average humidity, and average temperature**.
- The gauges highlight the **minimum, maximum, and average** levels for both humidity and temperature.
- The line chart shows how **temperature, humidity, and rainfall change across weeks**.
- The bar chart displays the **rainfall distribution per week and month**, helping compare rainfall trends .

06

ML Models



AI-Powered Predictive Agriculture

Model 1: Irrigation Prediction

An intelligent machine learning system that predicts irrigation needs for agricultural fields based on real-time environmental sensor data and crop characteristics.

- **Algorithm:** XGBoost Classifier
- **Accuracy:** 95%
- **Input:** Soil moisture, temperature, humidity, rainfall, sunlight, pH

  Output: Irrigate (1) or Not (0)
- **Purpose:** Optimize water usage and prevent over/under watering

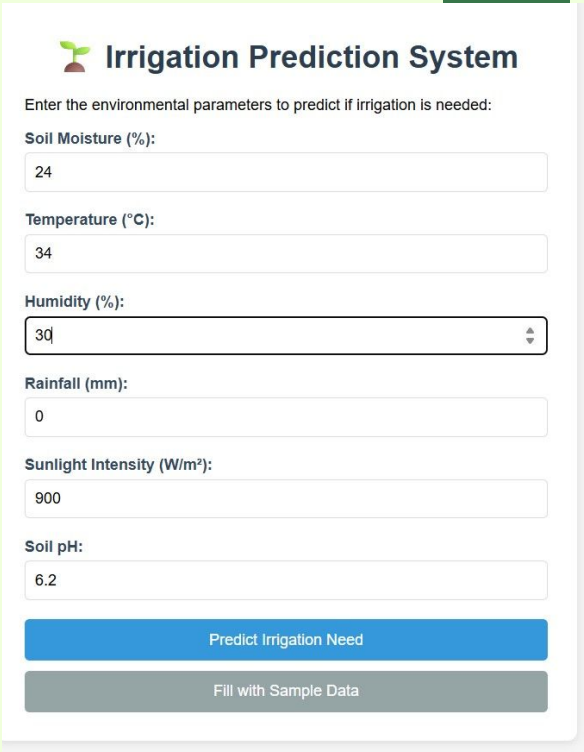
ML Training & Deployment Pipeline

01

Deployment of Irrigation Prediction System

API: Flask

02



AI-Powered Predictive Agriculture

Model 2: Soil Health Index

A machine learning regression model that predicts soil health status based on environmental conditions, agricultural practices, and soil characteristics.
The model outputs a continuous Soil Health Index (0-1 scale).

- **Algorithm:** Random Forest Regressor
- **Accuracy:** $R^2 = 0.9987$
- **Input:** Soil pH, moisture, climate, pesticide usage, crop type, region
- **Output:** Health score 0.0-1.0

  (🔴 Poor → 🟡 Moderate → 🟢 Healthy)

Thank YOU

Special Thanks to Eng. Mohamed Hammed for his invaluable guidance, support, and mentorship throughout this project