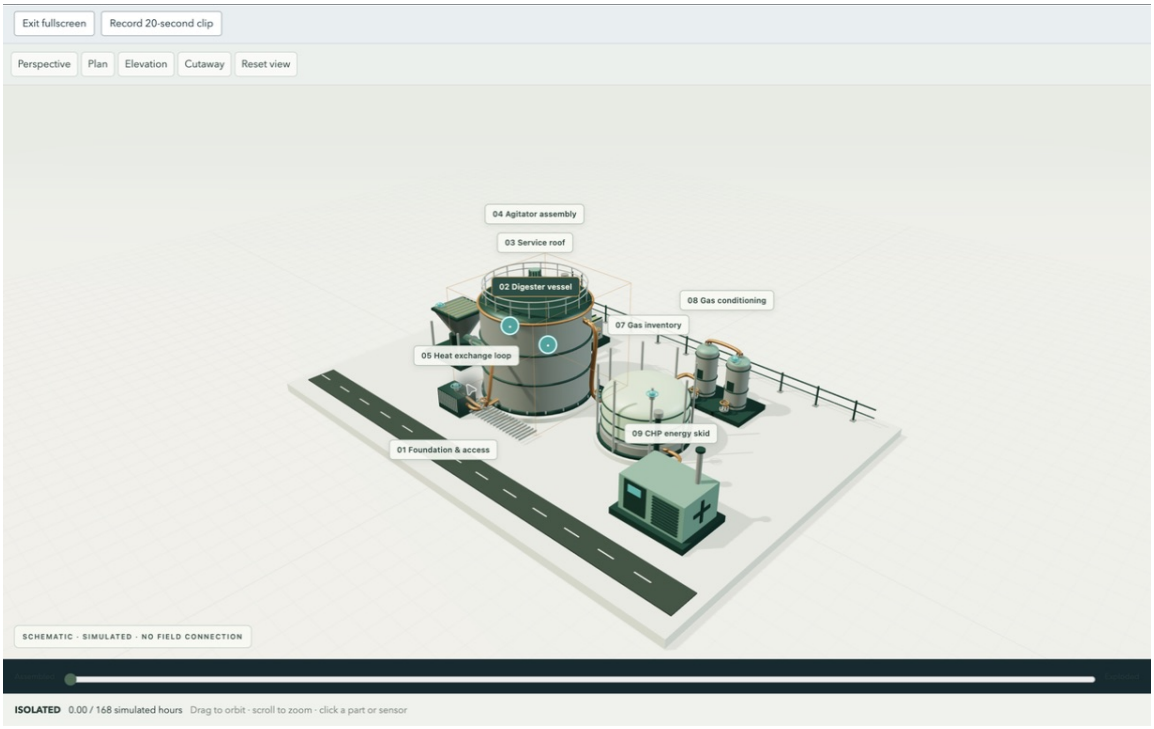
GFIS EcoTwin
Student engineering guide

SIH26200 · Renewable / Sustainable Energy

Explainable software for bioenergy plant and city-network management. Prepared locally on 9 September 2026 for the six-member student team at Shri Siddheshwar Women's Polytechnic, Solapur.



Fresh local software capture, 9 September 2026. Schematic modular plant, with no physical actuator connection.

How to use this package

Read this guide first, rehearse the separate six-slide script, then open the local demonstration in the order in section 13. The PowerPoint remains editable. The six-page presentation PDF follows the supplied slide sequence. The Marathi NotebookLM outputs are supplementary AI-generated learning aids and require a listening/viewing check before college circulation.

Before submission

The supplied SIH Guidelines, page 24, require a new idea not presented at a previous event. GFIS has prior DIPEX history. The college SPOC must resolve eligibility for the genuinely new student extension. An industry sponsorship letter does not replace the college nomination letter. Team ID remains to be entered.

01 Scope and reading map

The idea in one sentence

GFIS EcoTwin links timestamped plant observations to an explainable model-and-scenario workbench, a visible 3D state and replayable operator events, helping students evaluate renewable-bioenergy decisions before field testing.

Reading section	What you should be able to explain
02 Biology / 03 Gasification	Wet biological conversion and dry thermochemical conversion are different processes.
04 Data / 05 Models / 06 Training	Name the source, target, units, sample count and split before quoting a score.
07 Original simulator / 08 New plant	Explain synthetic 48-hour traces and the assumptions behind the new modular engine.
09 City / 10 Architecture	Explain storage balance, transfers, telemetry and the unbuilt ROS2 bridge.
11 Prototype plan / 12 Roles	Identify the student deliverable, owner and acceptance test.
13 Demo / 14 Questions / 15 References	Rehearse navigation and answer a judge without inventing evidence.

Three evidence classes

Measured history: public laboratory and full-scale records collected by the cited researchers. Simulation: calculated trajectories from declared equations and assumptions. Proposed student work: ROS2 packet integration, supervised sensor tests and a student-owned alert workflow. These classes must stay visible in the pitch.

Existing foundation and new contribution

Akash Shivadas Chatake developed the prior GFIS dissertation, experiments and software with academic guidance. The team can build upon this foundation with permission, but must evidence its own SIH contribution. Reusing the name or adding a new interface alone does not settle the novelty rule.

02 Anaerobic digestion

Organic feed to methane

Hydrolysis breaks complex organic matter into soluble compounds. Acidogenesis produces acids and intermediates. Acetogenesis supplies acetate, hydrogen and carbon dioxide. Methanogenic archaea then generate methane. The digester excludes oxygen from the biological conversion process.



GFIS biological molecular view. This is an educational net-reaction illustration, not microscopy or experimentally measured molecular trajectories.

$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O$

The hydrogenotrophic net reaction conserves carbon, hydrogen and oxygen. Another simplified net illustration is $CH_3COOH \rightarrow CH_4 + CO_2$. Near neutral pH, acetate rather than undissociated acetic acid predominates.

What the operator needs

Feed amount and composition, temperature, pH, gas flow and laboratory measurements help interpret process state. Volatile fatty acids and alkalinity can support stability assessment, but GFIS has not established a validated VFA/ALK soft sensor. A methane fraction must be measured or explicitly assumed before converting total-biogas volume into methane volume.

03 Gasification is a separate route



Fresh local capture of the separate gasifier, cyclone and cleanup layout. Equipment geometry is illustrative.

From dry biomass to combustible gas

Preparation and drying precede pyrolysis, partial oxidation and reduction. Product gas can contain carbon monoxide and hydrogen, with other species. Cleanup is essential before end use. This pathway differs from the microbial, low-temperature digestion of wet organic material.

Representative balanced net reactions

$C + H_2O \Leftrightarrow CO + H_2$
$C + CO_2 \Leftrightarrow 2CO$
$CO + H_2O \Leftrightarrow CO_2 + H_2$
$CH_4 + H_2O \Leftrightarrow CO + 3H_2$

What the animation establishes

The display illustrates species and atom conservation on a chosen reaction basis. It does not solve gasifier kinetics, equilibrium, tar formation, pollutant removal, electron dynamics or plant-scale yield. For the SIH pitch, keep gasification as an educational extension; the primary validation story remains anaerobic digestion and renewable gas management.

04 Dataset and target register

Source / licence	Scale and target	Use / limit
Mendeley, Núñez / G-AQUA UPM 10.17632/jpnjmzp6bj.1 CC BY 4.0	Two research reactors; 69 records each, 4 Mar–7 Jul 2025. 86 eligible methane observations.	Sparse methane forecast; persistence wins. Neither 138 independent methane labels nor continuous methane validation.
Schroer & Just, University of Iowa 10.25820/data.006715 ODC Attribution 1.0	500,400 raw minute records; 491,761 processed; one Muscatine facility. Next 24-hour mean total-biogas flow, CFM.	Dense temporal benchmark. 483,121 strict timestamp-valid cases, 96,623 five-minute training issue times, 8,050 hourly issue times across eligible evaluation index.
Hafner et al., interlaboratory BMP AU-BCE-EE/Hafner-2026-IL-BMP CC BY 4.0	619 bottles; 31 held-out laboratories; 26,146 examples; mL CH4/g VS.	Laboratory transfer, not 31 full-scale plants.
Mendeley temperature data 10.17632/yrvjhj284s.1 CC BY 4.0	35 / 55 °C experimental contrast.	Retained study is inconclusive; no causal operating recommendation.

Rows are not independent experiments

Nearby minutes share operating conditions. Dense logging increases temporal resolution, not the number of independently validated plants. Every displayed metric must identify the target and unit. The headline full-scale model result concerns total biogas, not methane yield.

Provenance

Research workspaces retain source checksums, processed-data versions, split manifests, configurations, predictions and run identifiers. No dataset is combined solely to enlarge sample size.

05 Why compare three models?

Model	Mechanism	Fair use
Persistence	Repeats the defined latest relevant target value.	Strong low-cost baseline; tests whether added complexity helps.
XGBoost	Boosted trees learn from past-only lag and rolling features.	Useful challenger with tabular process history.
LSTM	A recurrent hidden state learns from an ordered sequence.	Candidate for dense temporal data; more history alone does not prove improvement.

Same target, same evaluation issue times

MAE in CFM	Persistence	XGBoost	LSTM
Rolling mean ± fold SD	16.75 ± 5.24	16.77 ± 6.34	13.94 ± 4.31
Untouched final period	18.03	14.84	15.33

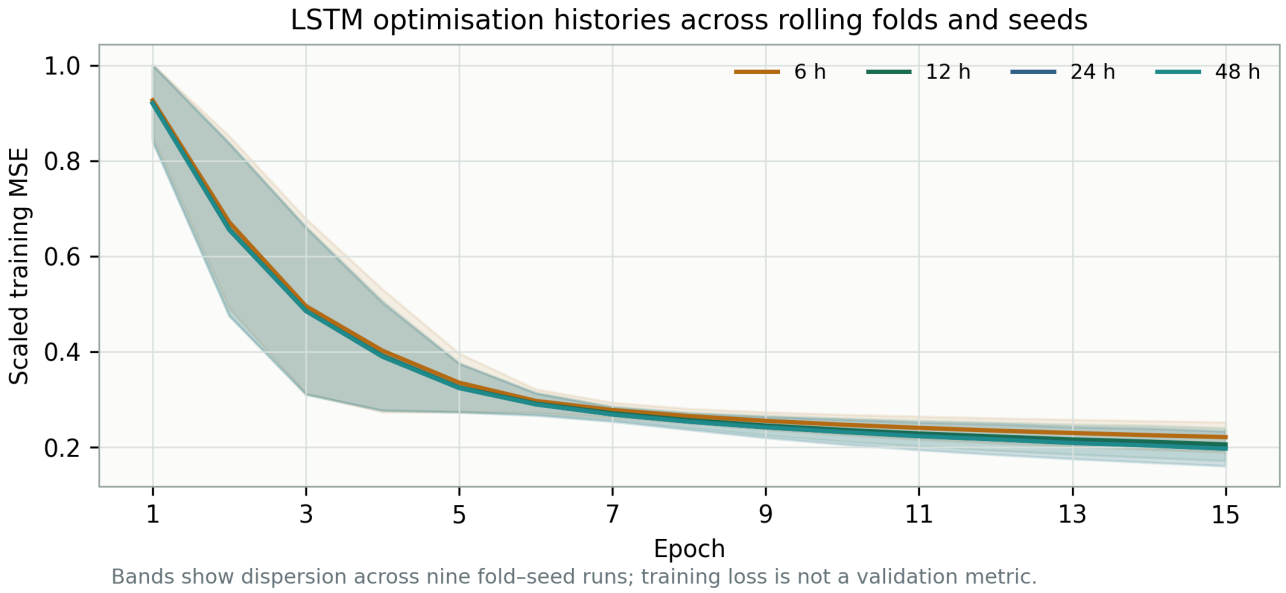
Interpretation

LSTM leads the pre-test rolling evaluation. XGBoost leads the final-period comparison. The roughly 16.8% reduction calculated from displayed rolling means is an error reduction, not measured energy savings. Fold dispersion follows seed aggregation. The separately paired estimate of improvement is 2.99 CFM with interval 1.04–4.84, and uses a different aggregation from subtracting those displayed means.

The negative finding is useful

On the 86-observation methane task, persistence has the lowest rolling MAE, 55.94 ± 39.34 mL. GFIS should retain the simpler model when the data do not justify temporal learning. On held-out BMP laboratories, XGBoost has 8.99 versus persistence 12.48 mL CH4/g VS laboratory-balanced MAE, winning in 25 of 31 labs. These targets are not interchangeable.

06 Training, epochs and leakage



Retained research figure: training loss across the 6/12/24/48-hour memory study. Bands represent nine fold-seed runs. Falling training loss is not validation accuracy.

Primary full-scale LSTM configuration

24 hourly states, 14 input variables, 32 recurrent units and a 16-unit dense layer. Adam learning rate 0.001, batch size 1024, 15 epochs per fit, seeds 17/42/73. Twelve primary fits retain 180 epoch rows. The four-window study retains 720 epoch rows. An epoch is one pass through a fit's training examples.

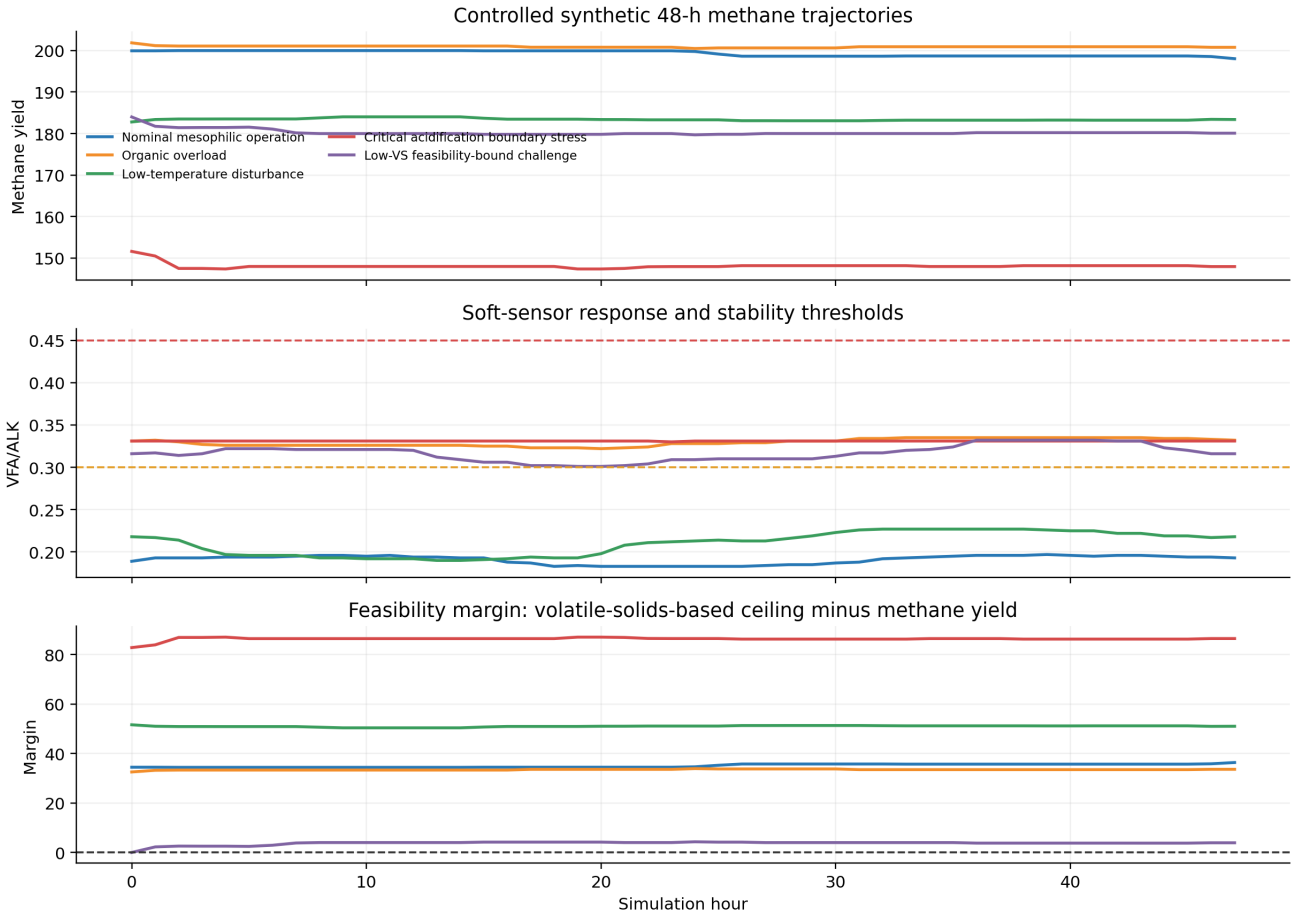
Chronological design

Train before validation; keep the final period untouched during model selection. Three rolling validation periods start 1 June, 15 August and 1 November 2022. A 24-hour embargo stops forward labels from crossing the next evaluation boundary. Imputation and scaling fit only training observations. The final evaluation period is 30 December 2022–27 February 2023.

What longer memory showed

The common-target 6/12/24/48-hour study did not establish a reliable benefit from longer history. Six hours remains competitive and 12 hours is a compact candidate. Do not describe window performance as proven equivalence or the longest window as optimal. Playback of a 48-hour plant scenario does not retrain these models.

07 The original 48-hour study



Retained dissertation-era figure. All trajectories and soft-sensor outputs here are synthetic scenario responses. The source figure's methane-yield label lacks a unit, so it must not be used for an absolute quantitative comparison.

What to show

Compare nominal operation, organic overload, low temperature, acidification-boundary stress and low-VS challenge on the same simulation clock. Explain how outputs, warning thresholds and feasibility margins change under each declared scenario.

What this does not validate

The smooth curves do not establish 48 hours of industrial operation, measured VFA/ALK soft sensing or an experimentally trained physics loss. The source chart is retained as historical software evidence. The student team should show configuration, input changes and saved events rather than interpreting the plotted values as plant observations.

08 New modular plant engine



Schematic modular assembly. Instrument hotspots, roof, mixing, heating, gas storage and CHP are visible in the local product.

Transparent reduced-order model

$$r = 0.0025\, M \exp[-((T - 37)/12)^2]\, a(R)$$

$$a(R) = 0.65 + 0.35 \min(R/12, 1)$$

$$V_{CH4} = Y \times \text{destroyed organic mass}$$

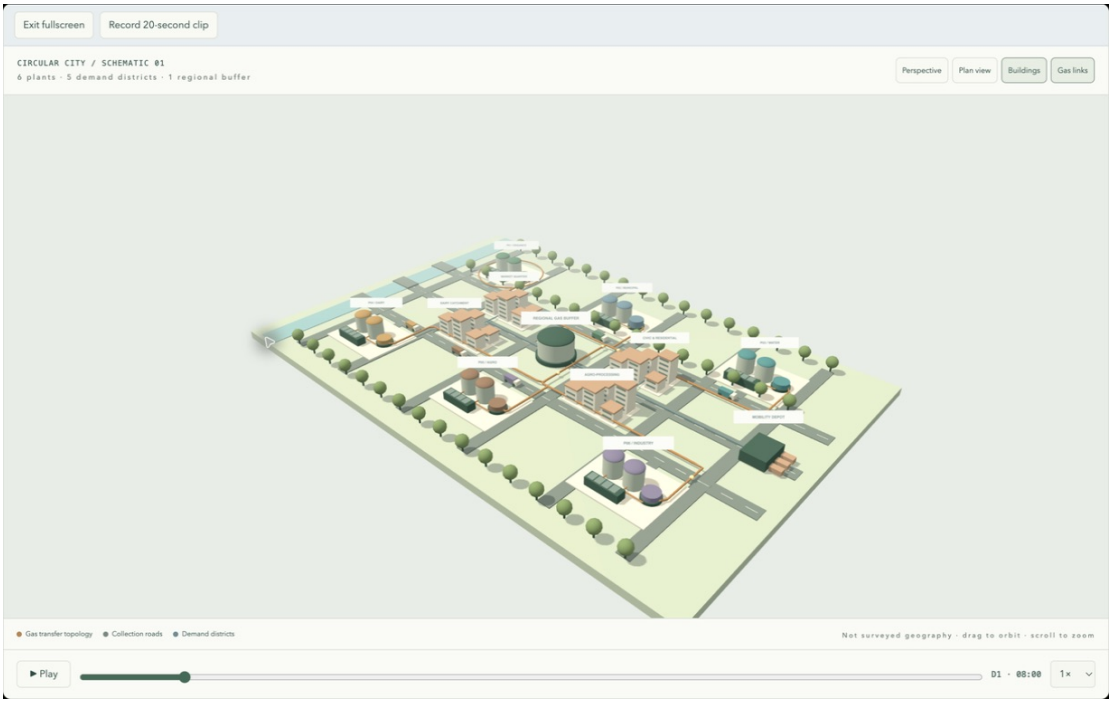
$$V_{biogas} = V_{CH4} / x_{CH4}$$

M denotes organic inventory, T is temperature in °C, R is mixing speed and a is an assumed mixing response. The model applies its rate on a declared timestep. Default methane yield Y = 0.35 Nm³/kg and fraction x = 0.60 are scenario assumptions, not universal learned coefficients.

Testable behavior

Configuration drives vessel geometry and scenario inputs. A temperature fault can trip the plant; reset returns to isolation and restart is explicit. Cutaway, exploded assembly, time replay and CSV/JSON export support inspection. Gas and organic balance residuals check numerical bookkeeping, not plant calibration. This reduced engine is not a full ADM1 implementation.

09 City-network reasoning



Fresh local city-network capture. Six campuses, five demand districts and one regional buffer. This is schematic, not surveyed geography.

An inventory model across a network

Each campus uses a declared supply profile. Storage, local demand, hub transfer limits and outages determine the 72-hour scenario. Internal transfers cancel in the city-wide conservation sum. The campus profiles are not automatically coupled to the custom plant's biological kinetics.

$G_{next} = G_{current}$ + production − dispatch − flare

Use the same standard-volume basis and timestep for all terms. A link outage can reduce served demand even when gas exists elsewhere, depending on topology, capacity and storage. Export the scenario before and after a change to explain the difference.

Energy example, with assumptions

100 Nm³/day biogas × 0.60 methane × 9.94 kWh/Nm³ CH4 × 0.35 electrical efficiency = 208.74 kWh/day gross. Deduct auxiliary loads before reporting net electricity. No rupee saving is asserted without measured loads, tariff, operating costs and capital assumptions. Pipeline pressure, gas quality compatibility and city rollout cost remain further engineering work.

10 System and student interface

Layer	Input	Output
Data contract	Timestamp, plant ID, variable, value, unit, quality, source mode	Validated observation or explicit missing/stale flag
Research inference	Past-only sequence / lag features	Forecast with model, target, unit and evaluation context
Scenario engine	Configuration, demand, mixing, temperature, events	Synthetic state, balance terms and failure states
3D scene	Shared timestamped state	Geometry, instrument hotspots, flow/state indicators
Operator memory	Input snapshot, alert, explanation, acknowledgement	Replayable event with export and source context
Proposed ROS2 adapter	Bench/replay message with clock and units	Mapped observations; no actuator command

Technology

Three.js renders the new scenes; JavaScript implements their reduced state engine. Python research pipelines use PyTorch and XGBoost. APIs expose retained evidence and scenario results. These are distinct components: a live API connection is not proof of a live physical instrument.

ADM1 in the architecture

Anaerobic Digestion Model No. 1 represents biological and physicochemical states through coupled differential equations. GFIS retains a PyADM1 280-day numerical ring test. An experimentally calibrated neural ADM1 residual is not established. VFA/ALK learning needs paired laboratory labels and held-out evaluation. Do not substitute visual chemistry for those experiments.

11 Fourteen-day prototype plan

Nomination is an immediate, separate step

Confirm the novelty rule, Team ID and college authorisation now, before the portal deadline. The following fourteen-day engineering sprint is a proposed development schedule, not permission to delay registration or a claim that the new student integration is complete.

Days	Deliverable / owner	Acceptance evidence
1–3	Packet schema, source mode and timeline Sunaina + Shweta	Unit checks, null/stale packets, monotonic timestamps and plant-ID isolation
4–7	ROS2 replay adapter and twin mapping Sakshi + Sneha	Same packet timestamp in scene and log; replay reproduces selected states
8–10	Alert lifecycle and memory Samina + Vijayalaxmi	Missing sensor never silently becomes zero; acknowledgement persists; export/reload preserves event
11–12	Supervised bench and scenario tests All pairs	Safe low-voltage sensor test, baseline versus fault trace, latency and missing-packet measures
13–14	Rehearsal and evidence Vijayalaxmi + team	Six-slide PDF, confirmed Team ID, SPOC novelty decision, offline fallback and clear speaking handoffs

Safety and practicality

Use recorded data and a low-voltage classroom bench. Gas handling, pressure equipment and methane generation are outside this student demo without suitable facilities and competent supervision. Do not connect autonomous actuators to a real plant. ROS2 experience transfers to the adapter and visualization work, but its implementation must be demonstrated.

Pilot scorecard

Measure packet loss and latency, correct missing-data alerts, forecast MAE against persistence, repeatable replay, gas-balance residual and operator task completion. Field energy gain, leakage reduction and economic payback remain hypotheses to measure, not pitch percentages.

Budget

Use existing computers and licensed open-source software first. Quote any sensor, interface and safe bench purchases before commitment. This package adds no deployment or cloud service. NotebookLM generation uses the existing signed-in service without a new plan purchase.

12 People and presentation order

Pair / student	Department	Work and speaking role
1 Vijayalaxmi Kalidas Sundalam	CO	Team leader; idea, integration, nomination and closing ownership.
1 Sakshi Ajay Malwadkar	CO	Opening technical flow; proposed ROS2 replay bridge and 3D tests.
2 Sunaina Shashikant Gaikwad	IT	Dataset provenance, model comparisons and leakage-safe explanation.
2 Sneha Mallinath Wale	IT	3D scene, city replay and ROS2 state mapping.
3 Samina Mahemood Shaikh	CO	Faults, alert lifecycle, event explanation and memory.
3 Shweta Laxmikant Vibhute	CO	Telemetry schema, supervised bench and validation evidence.

Mentorship and attribution

Industry sponsor: Chatake Innoworks Private Limited.
Industry mentor: Akash Shivadas Chatake.
College guide: Prof. N. K. Mane, HOD, Computer Engineering Department.
Institution: Shri Siddheshwar Women's Polytechnic, Solapur.
Research academic guidance: Dr. Dipti Jadhav, attributed to the prior GFIS research. She is not being appointed as this SIH team's mentor by this document.

Before college circulation

Confirm roster spelling against student registration, portal Team ID and nominated institution. The college SIH coordinator/SPOC is a separate role; do not relabel the HOD as coordinator. Preserve the company sponsorship letter as a separate document; college nomination needs the prescribed signatory and seal.

13 Demonstration navigation

Keep one controlled starting point

Open the local product at http://127.0.0.1:8390/. Use the in-product menu, not a chain of unfamiliar portal tabs. The local server must remain running. These localhost URLs cannot be opened from a student's phone without an explicitly configured network setup.

Stop	Local path / action	Narration
1 Plant	/#assembly Fullscreen, cutaway, hotspot	Geometry and sensors describe a schematic plant. Inputs and physics assumptions stay visible.
2 Scenario	/#plant Compare nominal and fault inputs	These are 48 simulated hours. Saved configurations make the comparison repeatable.
3 Data	/#industrial and /#models	One full-scale biogas source; sparse methane source stays separate.
4 Training	/#training	15 epochs per fit. Loss decreases; evaluation decides whether the model helps.
5 City	/#network Inspect campus and link outage	Declared supply profiles, demand and storage drive the scenario.
6 Processes	/#biology then /#gasification	Biological methane and thermochemical syngas are different pathways.
7 Memory	/#memory Inspect retained run / export	Explain timestamp, source mode, settings and replay. No live field claim.

Existing public fallback

gfis.energy/upgrade/ is the prior deployed preview. gfis.energy/sih/ is the prior team portal. This new deck, guide and media have not been deployed. Keep the local PDF open if connectivity or WebGL fails. Never describe an unavailable route as verified during judging.

14 Questions judges may ask

Why use AI instead of a rule dashboard?

Rules expose units, missingness and basic constraints. Data-driven models test whether past operating history improves a defined forecast. A persistence comparison prevents adding AI without demonstrated benefit.

Is the whole system physics-informed ML?

It combines evaluated ML with explicit physics screening and transparent scenario equations. It does not establish a fully calibrated ADM1-residual neural model. Constrained-learning experiments and mechanistic numerical tests must be described separately.

Do half a million rows prove generalisation?

No. The dense data come from one full-scale facility and are temporally dependent. Cross-plant external validation remains necessary. Held-out batch laboratories answer a different transfer question.

What is new for this student team?

The proposed deliverable is a reproducible ROS2 telemetry-to-twin workflow with data-quality failures, asset mapping, alert acknowledgement and replay. Existing founder research remains explicitly attributed. The SPOC must determine whether this new work satisfies SIH's prior-event rule.

Can we say the prototype improved energy efficiency?

Not yet. Forecast-error reduction and simulated balance consistency are not measured energy gains. A supervised pilot must measure useful gas, flare, unserved demand, parasitic load and operating outcomes against a baseline.

Can it control an industrial digester?

This package demonstrates software/replay and scenario reasoning. It does not command physical equipment. Industrial use requires calibrated instruments, safety interlocks, secure integration, plant-specific validation and operator approval.

15 Sources and submission checks

Scientific sources

Schroer, H. W., and Just, C. L. (2024). Laboratory, operational, and SCADA datasets for anaerobic co-digestion at the Muscatine, IA Water Resource Recovery Facility. University of Iowa. doi:10.25820/data.006715. ODC Attribution 1.0.

Núñez / G-AQUA UPM. Research-reactor data. doi:10.17632/jpnjmzp6bj.1. CC BY 4.0.

Interlaboratory BMP sources: AU-BCE-EE/Hafner-2026-IL-BMP, CC BY 4.0. Temperature contrast: doi:10.17632/yrvjhj284s.1. PyADM1: CaptainFerMag/PyADM1, MIT.

Retained local evidence

DS06-Q1-20260827T155346Z (primary fits); DS06-MEMORY-20260827T155839Z (window study); v11 scientific figures and v12 editorial audit. Original dissertation screenshots and synthetic 48-hour figure remain unchanged in their source workspaces. New assembly, city, gasification and biology screenshots were captured locally on 9 September 2026.

Supplied SIH documents

SIH 2026 Guidelines: p11 team composition; p14 college letter; p17 presentation PDF; p20 evaluation criteria; p21 mentors; p24 originality and IP provisions. The supplied presentation-format PDF is a completed six-slide example. Its slide order, white background, blue headers/footer and official SIH logo informed this deck. NavaKrishi was a visual reference only; its technical claims and savings were not imported.

Release checklist

Resolve originality with SPOC; confirm Team ID and roster; verify PS26200 selection in the portal; check the latest deadline and upload size there; review PDF after any edit; rehearse all six speakers; listen to generated Marathi media for errors; keep prior research attribution. No signed declaration, portal submission or production deployment was performed for this package.

16 Mathematical and technical glossary

Forecast errors

For N evaluated timestamps, observation y_i and prediction p_i:

$MAE = (1/N) \Sigma |y_i - p_i|$

$RMSE = \mathrm{sqrt}[(1/N) \Sigma (y_i - p_i)^2]$

$R^2 = 1 - \Sigma(y_i - p_i)^2 / \Sigma(y_i - \mathrm{mean}(y))^2$

MAE and RMSE retain target units. RMSE gives greater weight to large errors. R-squared is undefined when the observed target has zero variance and can be negative on poor forecasts. Scores must use matching issue times, horizon and target definition.

Term	Meaning for this project
CFM	Cubic feet per minute: the dense full-scale total-biogas target unit.
Nm³	Gas volume on a declared normal reference basis; compare only compatible reference conditions.
VS / VFA / ALK	Volatile solids / volatile fatty acids / alkalinity. VS-based methane targets differ from bulk gas flow.
SCADA	Supervisory control and data acquisition; the public dataset is historical SCADA, not a current GFIS instrument feed.
LSTM / epoch / seed	Long short-term memory network / one training pass / random initialisation identifier.
ROS2	Robot Operating System 2: proposed message/replay adapter here, not existing plant control.
CHP	Combined heat and power. The illustrative electricity estimate excludes auxiliary loads.
Digital twin	Here, an inspectable software state and scenario/replay environment. Physical operational synchronisation still needs validation.

Explain a prediction in this order

State the source and plant; target and units; forecast horizon; past input window; split; baseline; chosen model; error; missing-data behavior; and the action the operator may inspect. This sequence is more useful than saying that the platform uses AI.