Frost Risk Forecast Data Challenge

Transforming California's Central Valley into a global ag-tech innovation hub.

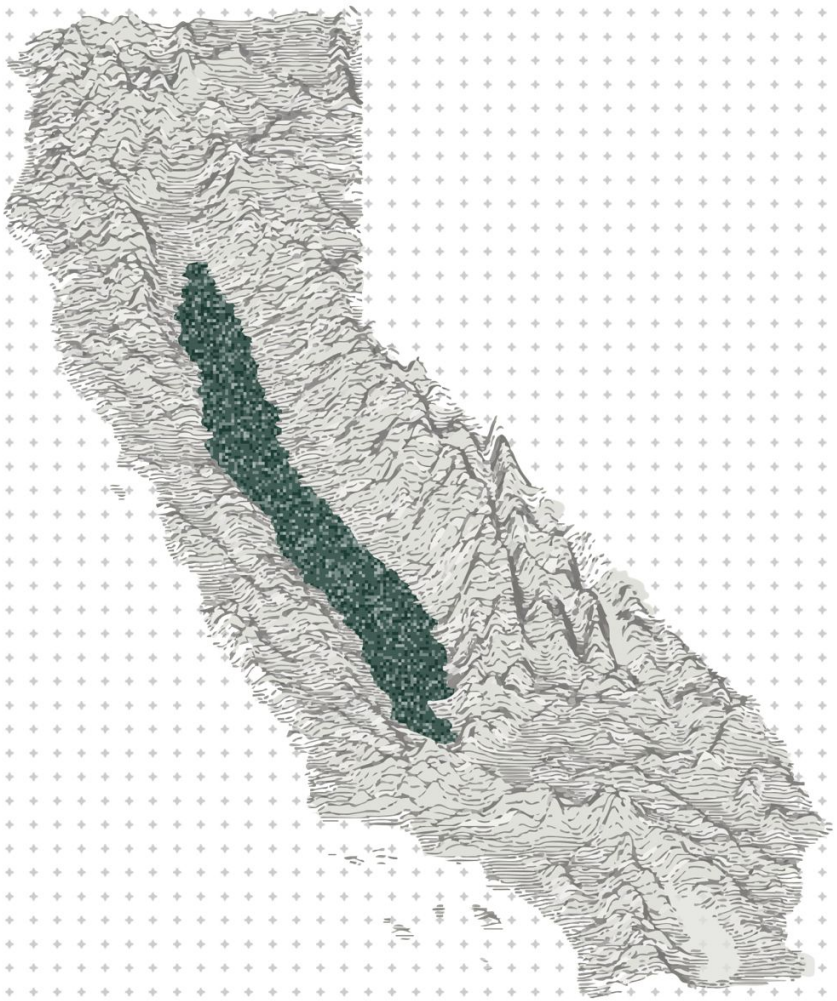
In partnership with UC San Diego & the National Data Platform (NDP)

11/10/2025

SAN DIEGO SUPERCOMPUTER CENTER

NATIONAL DATA PLATFORM

F3i



A single stretch of farmland produces one quarter of the nation's food.

F3 Innovate is a nonprofit founded in California's Central Valley to develop the region into the global epicenter for innovation in agriculture and food security. Where the scale of production helps build and test solutions that strengthen our most vital supply chain.

5M+	>350	13.6k	$70B
Acres in Production	Commodities Grown	Farming Enterprises	Annual Revenue

Why Frost Risk Forecasting?

- Frost events are among the most damaging weather risks for California agriculture — and the country. In the United States, the economic losses due to frost damage exceed all other weather-related phenomena.
- Frost protection has relied on manual readings from orchard thermometers and growers' local knowledge of "frost pockets," slopes, and microclimates
- Emerging research using machine learning and spatial modeling suggests that data-driven approaches could deliver earlier, more localized, and more reliable frost warnings.

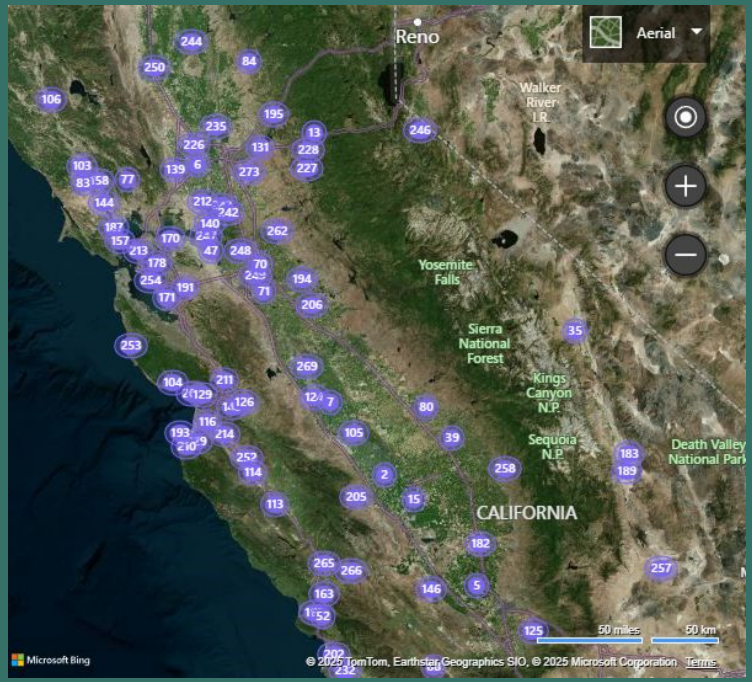
Sources: UC Davis LAWR; California Citrus Mutual/ Growing Produce (2014); Venner (1995);

The Data

CIMIS weather stations were established to help farmers, researchers, and water managers improve irrigation efficiency by providing accurate, localized weather data.

Each CIMIS station continuously collects several meteorological variables, including but not limited to:

- Relative humidity
- Solar radiation
- Soil temperature
- Vapor Pressure
- Precipitation

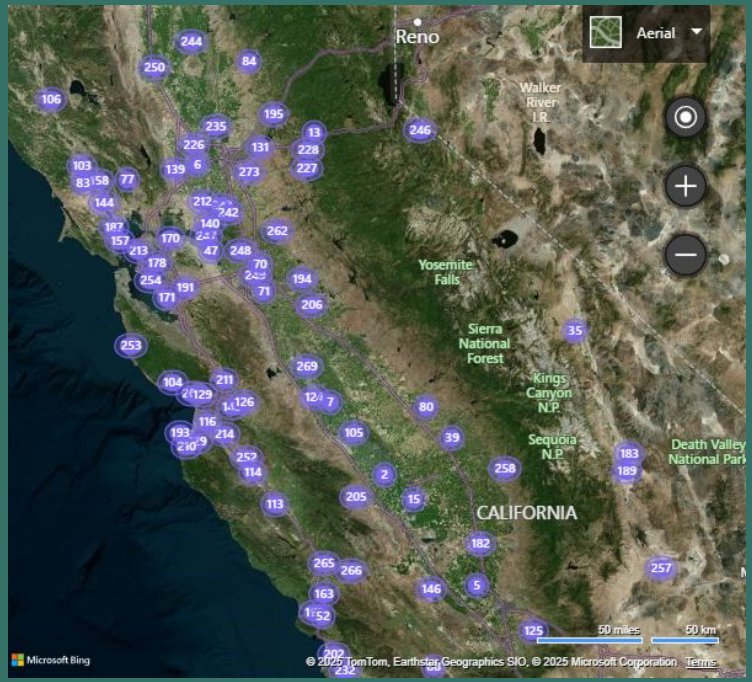


Your Mission

Build and evaluate machine learning models that can forecast frost risk and intensity at various lead times using weather station data.

Data

- 15 years of hourly data from 18 CIMIS stations
- 2.3 million observations across different topographies and climates

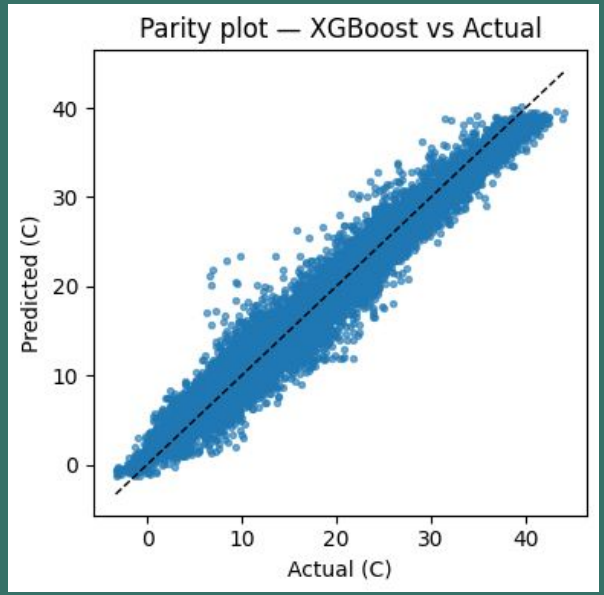


Core Tasks

- Train models to predict the probability that a frost event **(T < 0°C)** occurs as well as the expected temperature within the next 3, 6, 12, and 24 hours, using only information available up to the forecast time.

  Your model should output calibrated probabilities equivalent to: **“There is a 30% chance of frost in the next 3 hours, predicted temperature: 4.50 °C”**

- Test how well your model transfers to unseen sites



Total frost events in test set: 276

timestamp	p_frost	actual	pred_temp6h
2024-01-08 00:00:00	0.982456	1	-1.095667
2024-01-08 01:00:00	0.982456	1	-1.148667
2024-01-12 01:00:00	0.962117	1	-0.942333

Deliverables

Reproducible Pipeline (Code + Documentation)

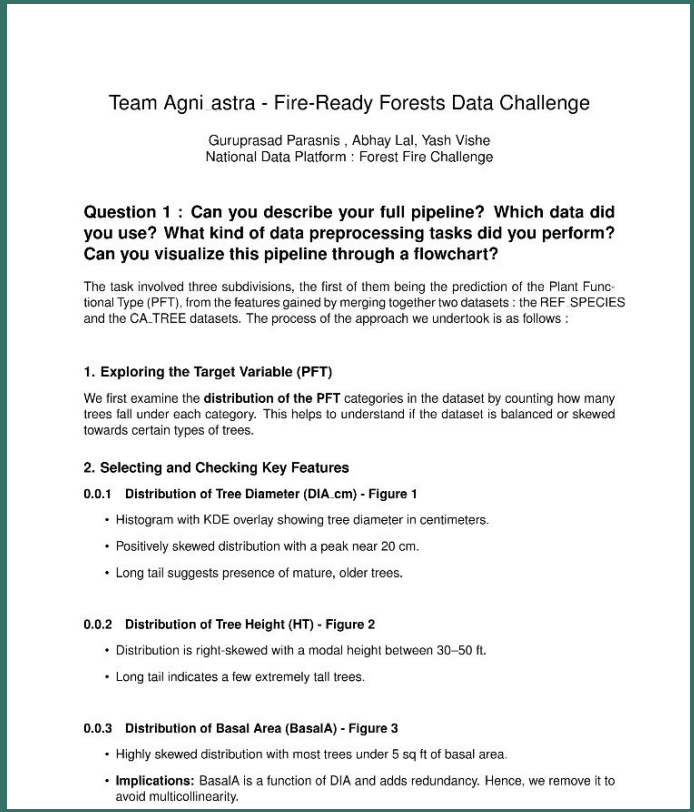
Submit either:

- A Jupyter Notebook that executes end-to-end
- Or, a custom workflow with clear run instructions in a README.md.
- Include links to your public **GitHub repo.**

PDF Report:

- Submit a report in PDF format, addressing the four questions outlined in the challenge brief

See **Section 10** of the Challenge Brief for Example Deliverables



Team Agni astra - Fire-Ready Forests Data Challenge

Guruprasad Parasnis , Abhay Lal, Yash Vishe
National Data Platform : Forest Fire Challenge

Question 1 : Can you describe your full pipeline? Which data did you use? What kind of data preprocessing tasks did you perform? Can you visualize this pipeline through a flowchart?

The task involved three subdivisions, the first of them being the prediction of the Plant Functional Type (PFT), from the features gained by merging together two datasets : the REF_SPECIES and the CA_TREE datasets. The process of the approach we undertook is as follows :

1. Exploring the Target Variable (PFT)

We first examine the **distribution of the PFT** categories in the dataset by counting how many trees fall under each category. This helps to understand if the dataset is balanced or skewed towards certain types of trees.

2. Selecting and Checking Key Features

0.0.1 Distribution of Tree Diameter (DIA_cm) - Figure 1

- Histogram with KDE overlay showing tree diameter in centimeters.
- Positively skewed distribution with a peak near 20 cm.
- Long tail suggests presence of mature, older trees.

0.0.2 Distribution of Tree Height (HT) - Figure 2

- Distribution is right-skewed with a modal height between 30–50 ft.
- Long tail indicates a few extremely tall trees.

0.0.3 Distribution of Basal Area (BasalA) - Figure 3

- Highly skewed distribution with most trees under 5 sq ft of basal area.
- **Implications:** BasalA is a function of DIA and adds redundancy. Hence, we remove it to avoid multicollinearity.

*See **Section 9** of Challenge Brief

Data Access & Platform

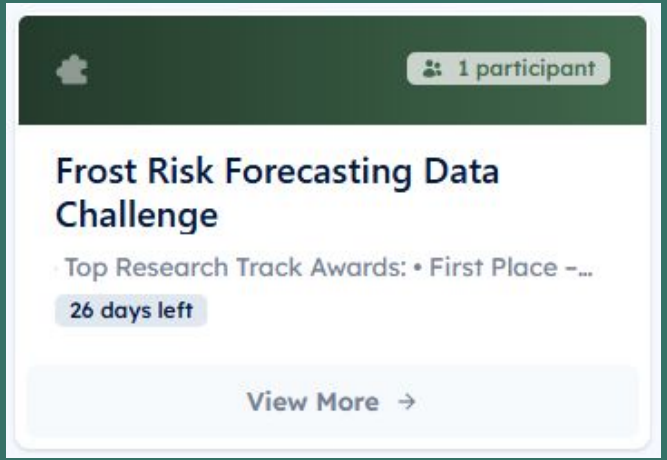
Built on the National Data Platform (NDP)

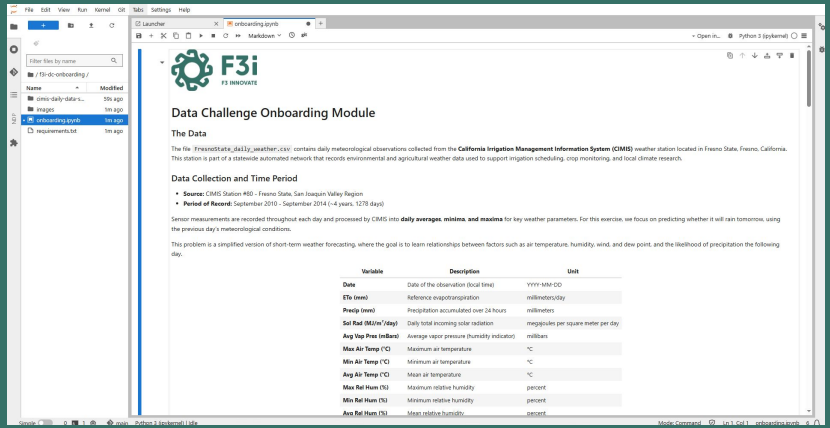
nationaldataplatform.org

- Cloud-based research environment from UC San Diego
- Cataloged datasets, compute resources, and collaboration tools
- Free for anyone with an **institutional email**

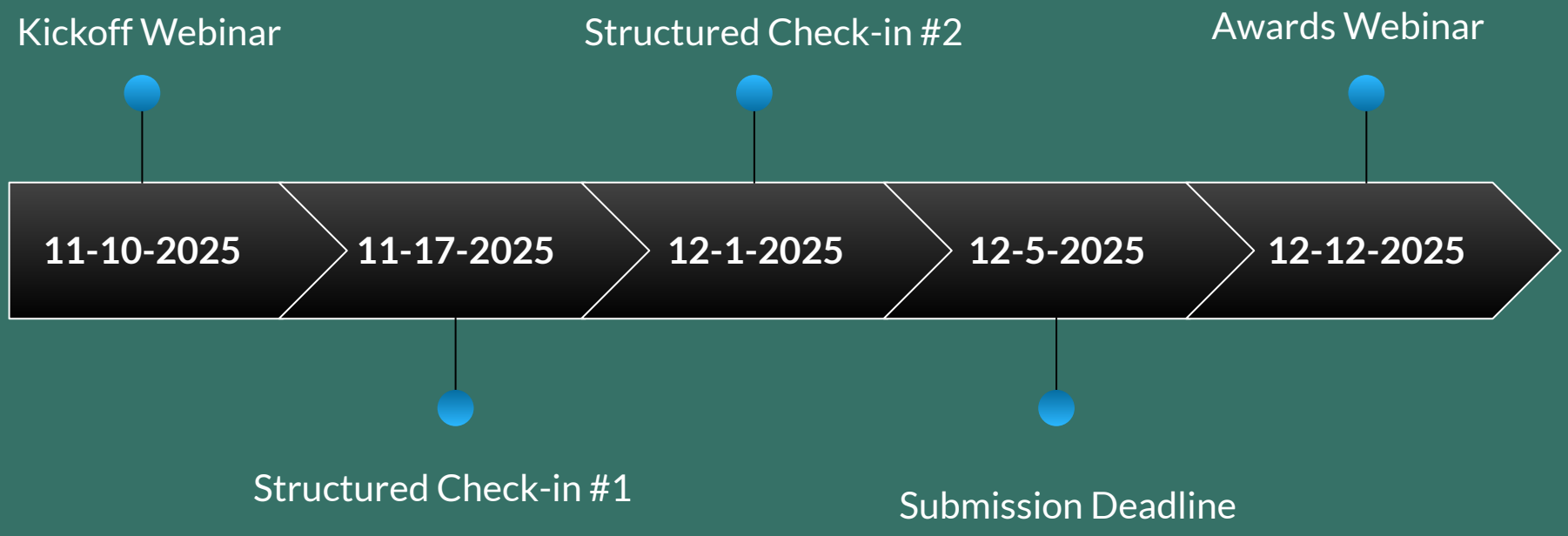
You can:

- Access CIMIS and challenge data directly
- Run analyses and train models in the cloud
- Collaborate via shared workspaces





Challenge Timeline



Office hours available upon request throughout the duration of the challenge

Challenge Workshops

Forming or Joining a Team

- November 12th, 17th
  - 10 am - 10:30 am
  - 2 pm - 2:30 pm

- Thursday, November 13th
  - 9am - 9:30am
  - 3pm - 3:30 pm

NDP Access & Usage

- November 12th
  - 10:30am - 11am
- November 13th
  - 10 am - 10:30 am
  - 2 pm - 2:30 pm
- November 17th
  - 9:30am -10am
  - 2:30pm - 3pm

Office hours available upon request throughout the duration of the challenge

All participants are encouraged to request office hours with NDP representative ***Pedro Ramonetti*** for technical assistance or onboarding support throughout the duration of the challenge. When booking, please note in your reservation message that you are participating in the F3i Frost Risk Forecasting Challenge.

Rewards & Opportunities

Top Research Track Awards

- **First Place:** $1500
- **Second Place:** $750
- **Third Place:** $400

Presentation & Recognition

The Top 3 teams will be invited to present their methods and findings to F3 Innovate's industry and grower partners.

Mentorship & Collaboration Opportunities

Selected participants may be invited to join future sprint design teams and commercialization cohorts supported by F3 Innovate and partner institutions.

Certificates & Digital Badges

All teams submitting valid, reproducible entries will receive a Certificate of Participation.

The **Top 3** teams will earn an official F3 Innovate Digital Credential, verifiable through our open badging system and suitable for inclusion on professional profiles or CVs.

Questions?

Ryan Dinubilo
ryan@f3innovate.org

This challenge is the first of many.

We want to create a thriving, open, data-driven ecosystem where ag-tech innovation can flourish.

Thank you for joining us on this journey. Let's get started!