



NEXTGEN
FEDERAL SYSTEMS

Machine Learning Platform (MLP) for Weather Applications at Scale

Unified Model (UM)/LFRic User Workshop – Data Science

June 8, 2023

Distribution Statement C. Distribution authorized to U.S. Government agencies and their contractors (Administrative or Operational Use) (August 2021). Other requests for this document shall be referred to AFLCMC/HBAW, 75 Vandenberg Dr, Hanscom AFB MA 01731.

DESTRUCTION NOTICE: Destroy by any method that will prevent disclosure of contents or reconstruction of the document.



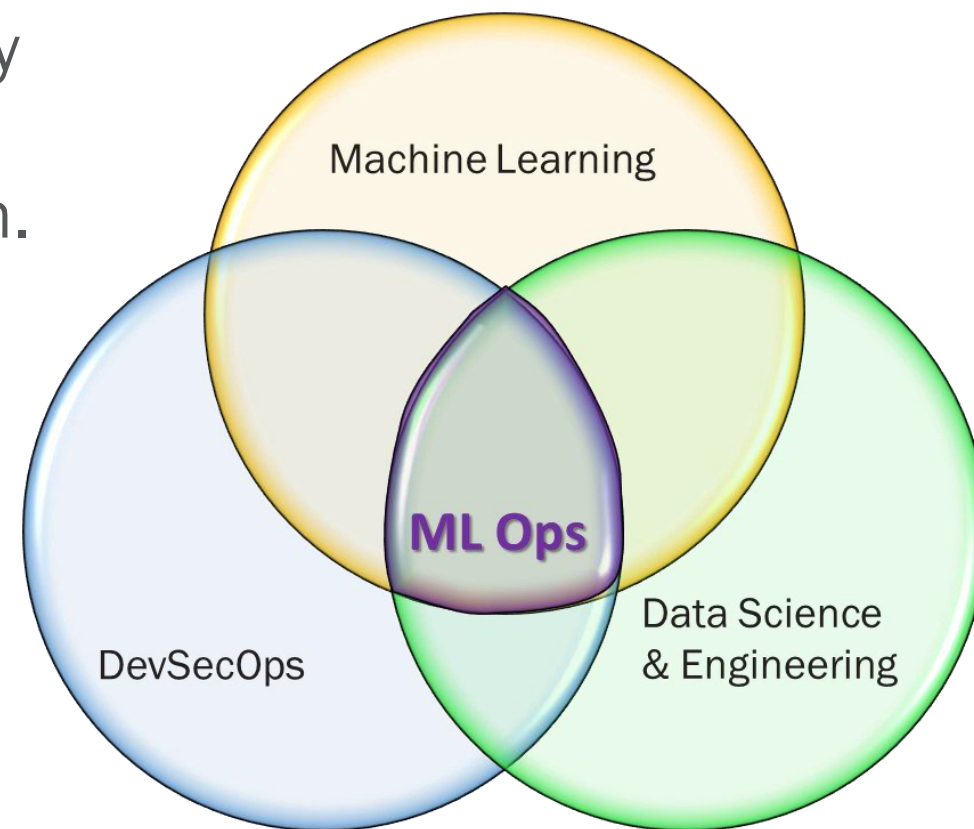
WxMLP Overview

- ML Operations
- Challenge of ML at Scale
- ML Platform
- Application of ML examples



ML Operations (MLOps)

- MLOps is the coordinated actions of a trained team applying a combination of ML, DS&E, and DevSecOps to meet mission needs. [1]
- MLOps is a set of practices used to reliably and efficiently build, test, deploy, and maintain ML models running in production.
- MLOps embodies the understanding that applying ML to meet needs is more than just training ML algorithms.
- To succeed at using ML, you need to integrate people, processes, and tools into a coherent and comprehensive operation.





MLOps Challenge

- Data-Driven machine learning approaches have been shown to outperform statistical methods
- Results have historically been difficult to reproduce
 - Runtime environment is not well documented
 - Code and datasets are not available

Reproducibility Crisis in machine learning

Just as boxers show their strength in the ring by getting up again after being knocked to the canvas, researchers test their strength in the arena of science by ensuring their work's reproducibility. If other researchers cannot replicate the research findings, the original study will draw doubters and critics. **Although reproducibility is an essential part of science, many sub-fields such as machine learning are now experiencing a reproducibility crisis.**

NEWS | SCIENTIFIC COMMUNITY

Missing data hinder replication of artificial intelligence studies

Unpublished code and sensitive training conditions aggravate reproducibility crisis in computer science

15 FEB 2018 • BY [MATTHEW HUTSON](#)



MLOps Challenge

- Data-Driven machine learning approaches have been shown to outperform statistical methods
- Results have historically been difficult to reproduce
 - Runtime environment is not well documented
 - Code and datasets are not available
- The ML community is moving towards an 'open-source' approach

ICML | 2022
Thirty-ninth International Conference on Machine Learning

Dates Calls Resources Organization

Year (2022)

Paper Writing Best Practices

Reviewing Criteria:

Accepted papers must be based on original research and must contain significant novel results of significant interest to the machine learning community. Results can be either theoretical or empirical. Results will be judged on the degree to which they have been objectively established and/or their potential for scientific and technological impact. **Reproducibility of results and easy availability of code will be taken into account in the decision-making process whenever appropriate.**

9. All source code required for conducting experiments is included in the supplementary material or the appendix

10. All source code required for conducting experiments will be made publicly available upon publication of the paper with a license that allows free usage for research purposes

NeurIPS | 2022
Thirty-sixth Conference on Neural Information Processing Systems

Dates Submit Organizers

Year (2022)

NeurIPS 2022

Code Guidelines

Your code submission should include training and evaluation code, specification of dependencies, etc. See <https://github.com/paperswithcode/releasing-research-code> for more detailed guidelines.

NeurIPS Code and Data Submission Guidelines

If any of the main contributions of your paper depends on an **experimental result**, you are strongly encouraged to submit code that produces this result. If you are using a **new dataset**, you are also encouraged to submit the dataset.



MLOps Challenge

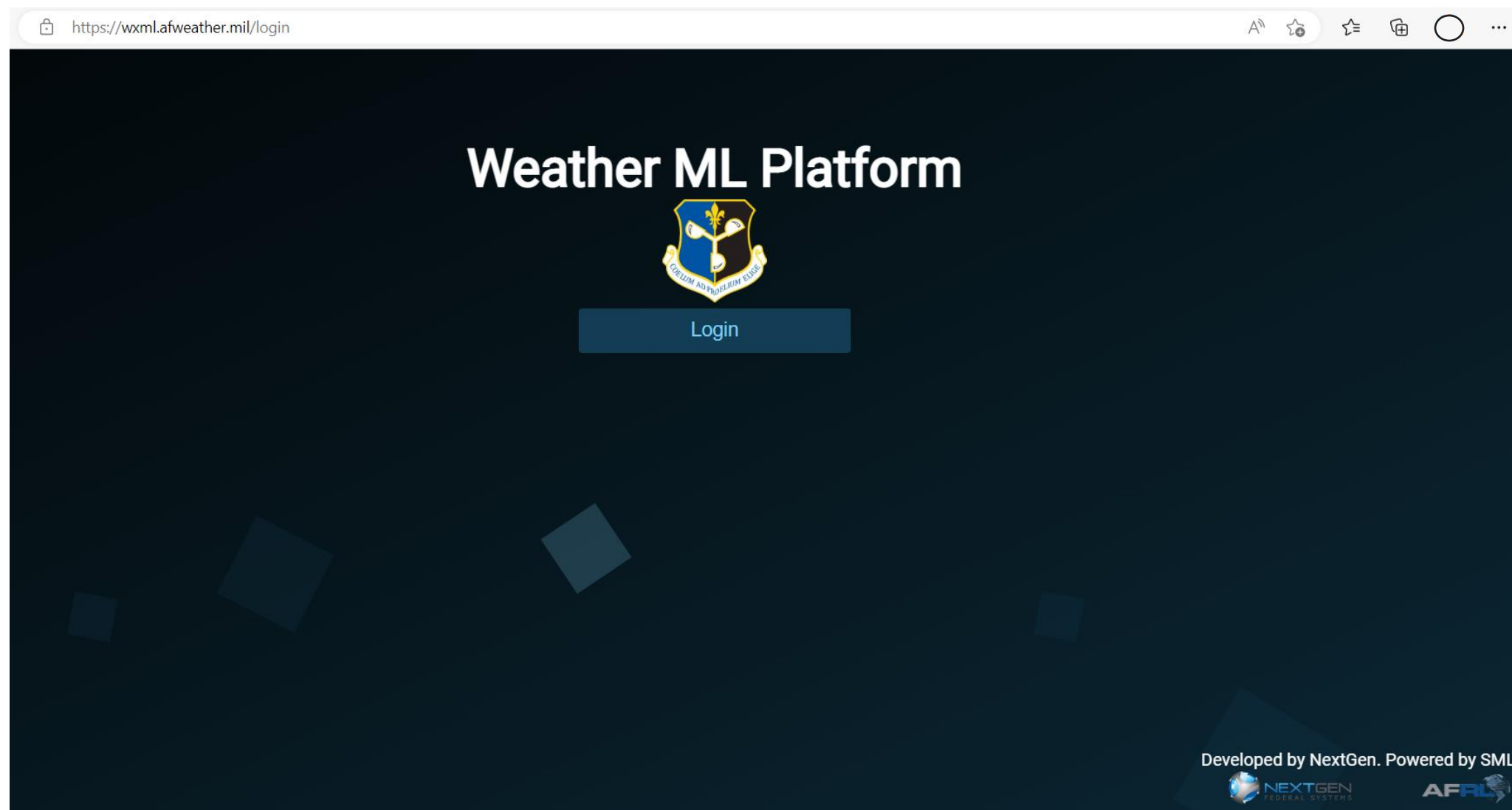
- Data-Driven machine learning approaches have been shown to outperform statistical methods
- Results have historically been difficult to reproduce
 - Runtime environment is not well documented
 - Code and datasets are not available
- The ML community is moving towards an ‘open-source’ approach
- One-off repositories of data and code scattered across various sites leads to ‘reinventing the wheel’

Repository	Data	Models
https://registry.opendata.aws/	✓	
https://datasetsearch.research.google.com/	✓	
https://msr.opendata.com/	✓	
https://github.com/awesomedata/awesome-public-datasets	✓	
https://visualdata.io/discovery	✓	
https://www.kaggle.com/	✓	✓
https://modelzoo.co/	✓	✓
https://paperswithcode.com/sota	✓	✓





MLP Login using Single Sign-On





MLP Dashboard UI

Weather ML Platform

Projects

Brianna Maze

WeatherBench

Dashboard

Overview

Tasks

Leaderboard

Repository

Models

Algorithms

Datasets

Implementations

Transforms

Activities

Transform

Label

Train

Predict

Evaluate

Members

Models

9 models

Browse Models

Register Model

Algorithms

0 algorithms

0 algorithm types

Browse Algorithms

Register Algorithm

Datasets

0 datasets

12 modalities

Browse Datasets

Register Dataset

Implementations

1 implementations

Browse Implementations

Register Implementation

Your Tasks

Job Type	Start Date	Owner	Name	Status	Age	Run Time	Details
evaluation	04-12-2022 01:59 PM	Brianna Maze	modelevaltask-16497863950...	Complete	8 months ago	45 seconds	
evaluation	01-31-2022 01:59 PM	Brianna Maze	modelevaltask-16436555821...	Complete	11 months ago	49 seconds	
data-ingest	01-31-2022 01:41 PM	Brianna Maze	WeatherBench Ground Truth ingest task 1643654437573	Complete	11 months ago	3 seconds	
prediction	01-31-2022 01:27 PM	Brianna Maze	Predict WeatherBench	Complete	11 months ago	2 minutes	
training	01-31-2022 12:46 PM	Brianna Maze	Train WeatherBench	Complete	11 months ago	16 minutes	

Items per page: 5 1 - 5 of 7

All Project Tasks

Job Type	Start Date	Owner	Name	Status	Age	Run Time	Details
evaluation	04-13-2022 02:47 PM	Steven Lack	modelevaltask-16498756234...	Complete	8 months ago	11 seconds	
prediction	04-13-2022 02:43 PM	Steven Lack	wb_results_sal	Complete	8 months ago	21 seconds	
training	04-13-2022 02:25 PM	Steven Lack	weatherbench_train_sal	Complete	8 months ago	11 minutes	
transform	04-13-2022 02:22 PM	Steven Lack	GeoRasterTemporalExtract for WeatherBench_5_degree	Complete	8 months ago	24 seconds	
transform	04-13-2022 02:18 PM	Steven Lack	GeoRasterTemporalExtract for WeatherBench_5_degree	Complete	8 months ago	1 minutes	

Items per page: 5 1 - 5 of 21



MLP Models

Tasks

Image Classification

Object Detection

Question Analysis

Part of Speech Induction

Point cloud object
classification

Image segmentation

Text classification

Text Generation

Object tracking

Image translation

Time Series Prediction

Sentiment Analysis

Weather ML Platform

Projects

WeatherBench

Dashboard

Overview

Tasks

Leaderboard

Repository

Models

Algorithms

Datasets

Implementations

Transforms

Activities

Transform

Label

Train

Predict

Evaluate

Members

Problem Type Search...

Data Type Search...

Implementations

Group By...

View...

Public Resources

AlexNet

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of AlexNet using the PyTorch framework and

DeLFT

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of DeLFT using the Keras framework and accepts

DMN

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of DMN using the Theano framework and accepts

DualPathNet131

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of DualPathNet131 using the

GaugeCNN

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of GaugeCNN using the Torch framework and

KerasGAN

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of KerasGAN using the Keras framework and

NASNet

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of NASNet using the PyTorch framework and

PointNet

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of PointNet using the TensorFlow framework and

ResNet50

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of ResNet50

UNet

sml@wxmip.af.mil 04-13-2021

Public Resources

Implementation of UNet using the

WeatherBench

sml@wxmip.af.mil 01-31-2022

Public Resources

Implementation of WeatherBench

WeatherBench_Evaluator

ekeough@wxmip.af.mil 08-02-2021

Implementation of WeatherBench evaluator

WxML 1.40.198 SML 1.0.0.273 GEA 1.40.24

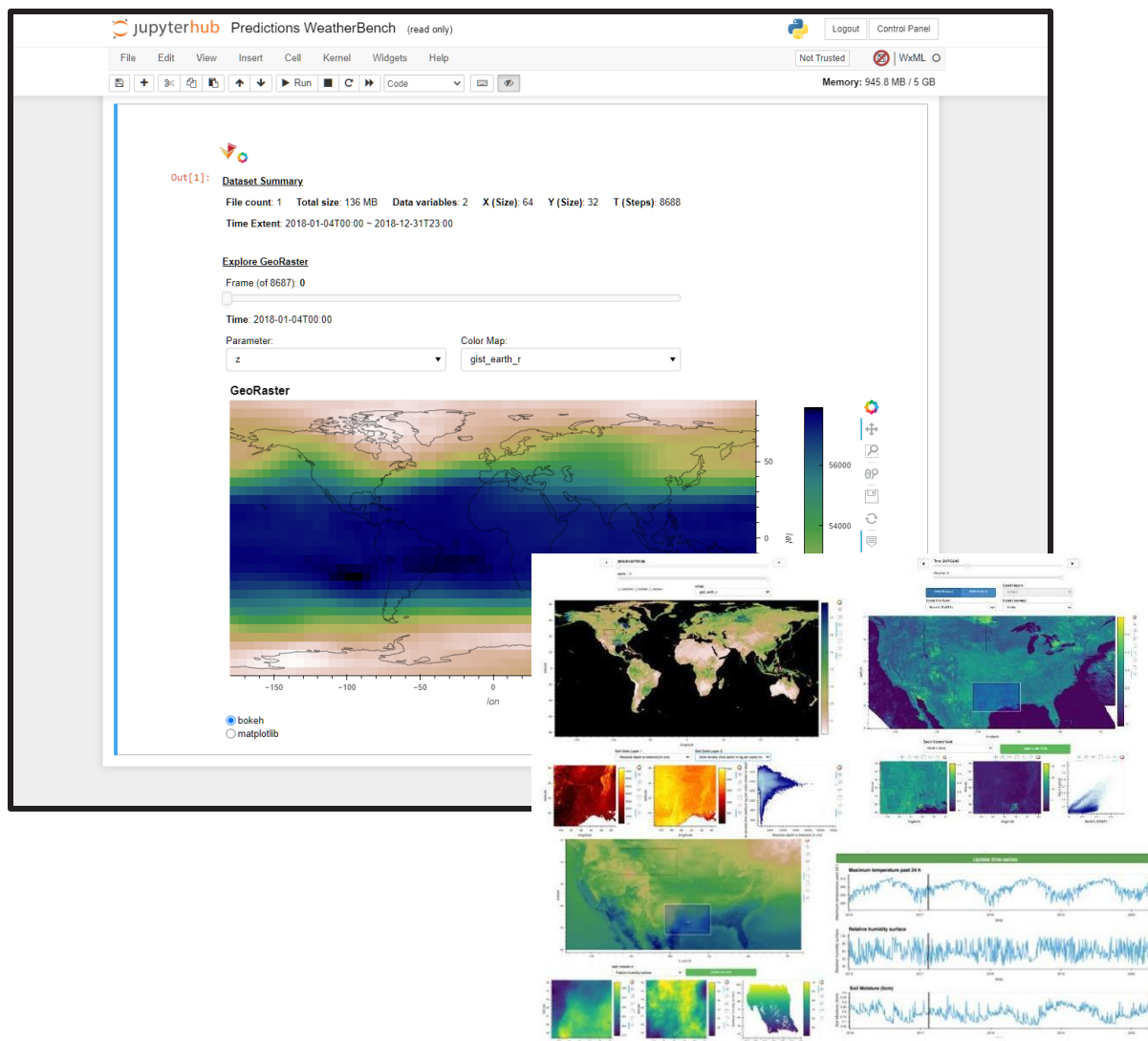
Totals: 27 algorithms, 12 task types

New algorithm: ~9 days



MLP Data Science & Engineering

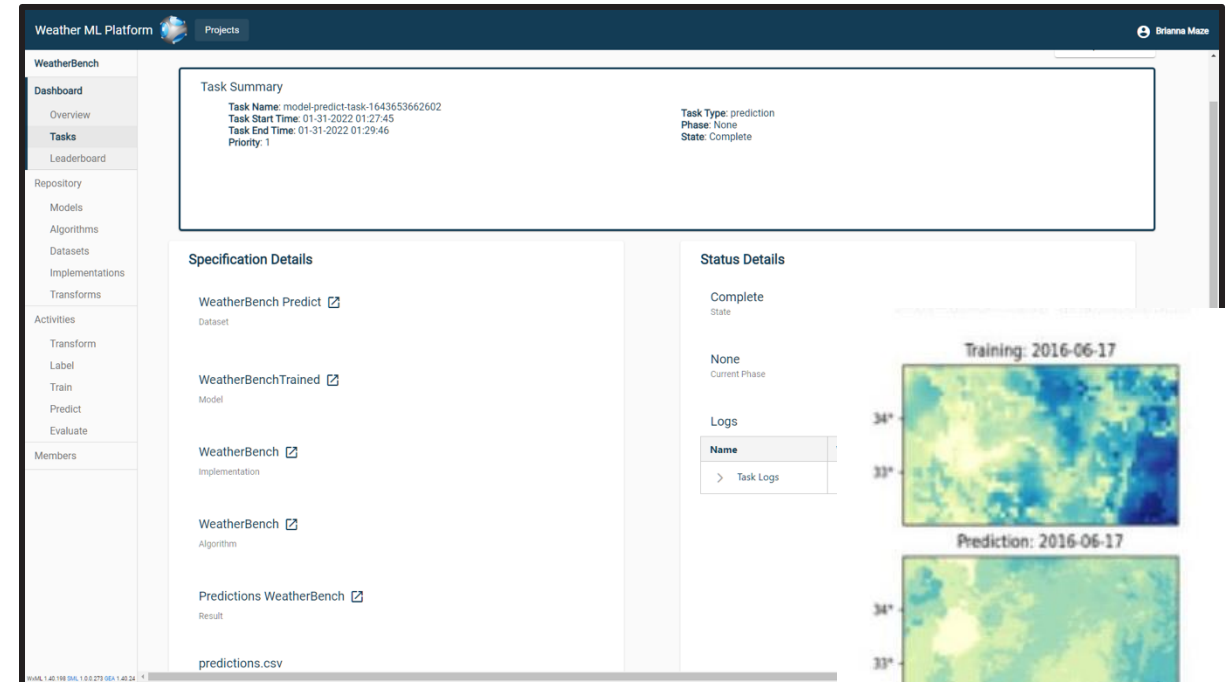
- Supports the ingest of multiple weather data formats
- Users can explore and analyze the data in Jupyter notebooks directly from the client UI
- Use built-in data transforms to create canonical, ML-friendly datasets





MLP Model Training and Prediction

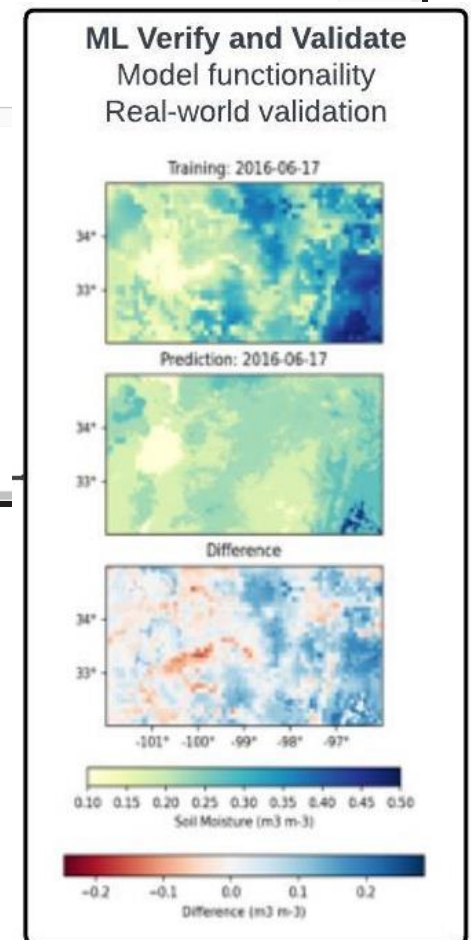
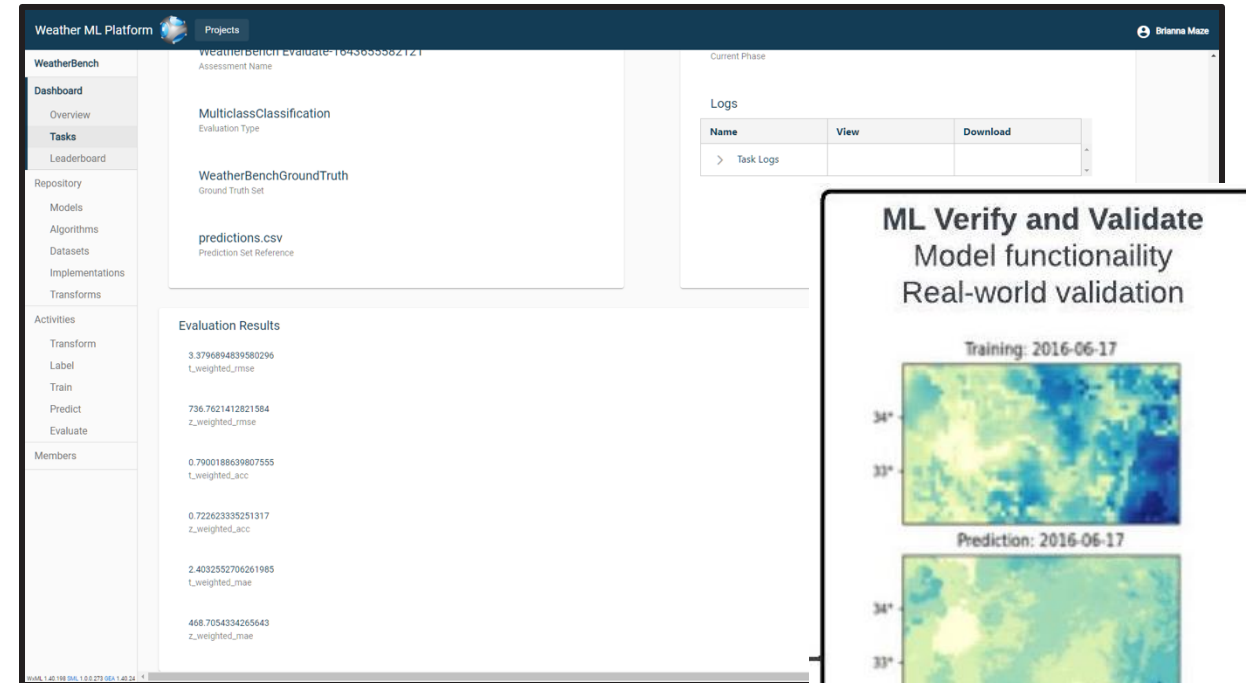
- Perform model training and prediction in a code-less environment
- All model metadata is tracked through its lifecycle
- Monitor the training and prediction tasks through the Tasks page





MLP Evaluation

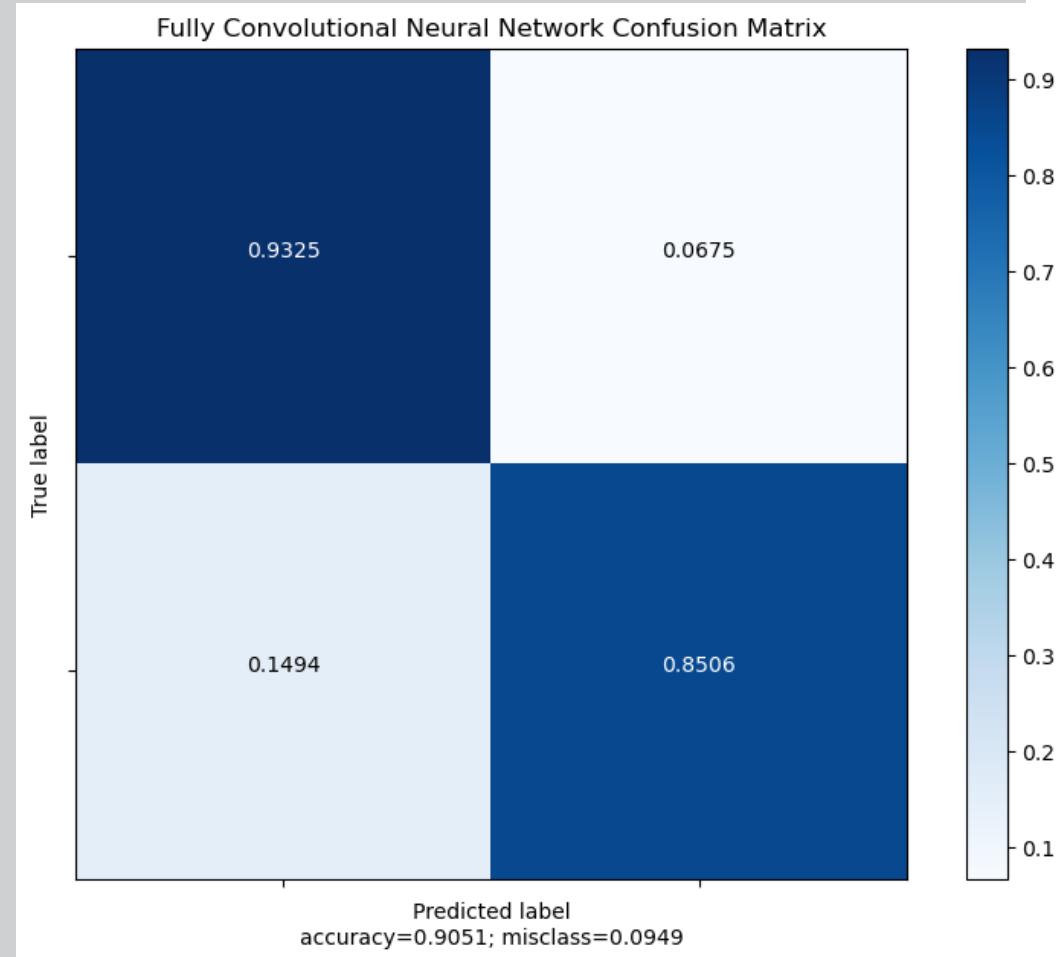
- Qualitatively evaluate predictions through dashboard view
- Quantitatively evaluate predictions with standard evaluation metrics
- Easily compare model performance through the leaderboard view





Fully Convolutional Network

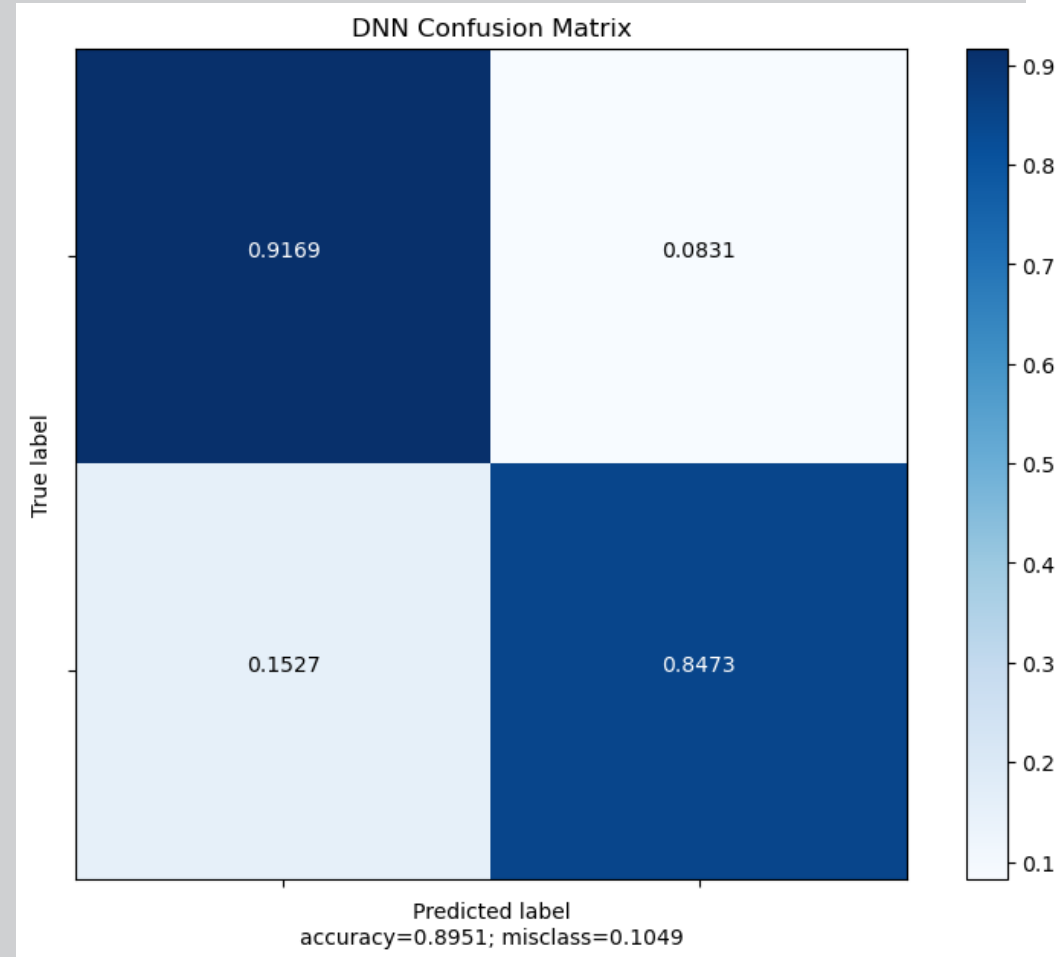
- Fully Convolutional Network
 - Raw Accuracy: 90.51
 - Probability of Detection: 85.06
 - False Alarm Rate: 13.62
 - F1: 85.71
- Model was evaluated on randomized held-out data split from MSL based master dataset





Deep Neural Network

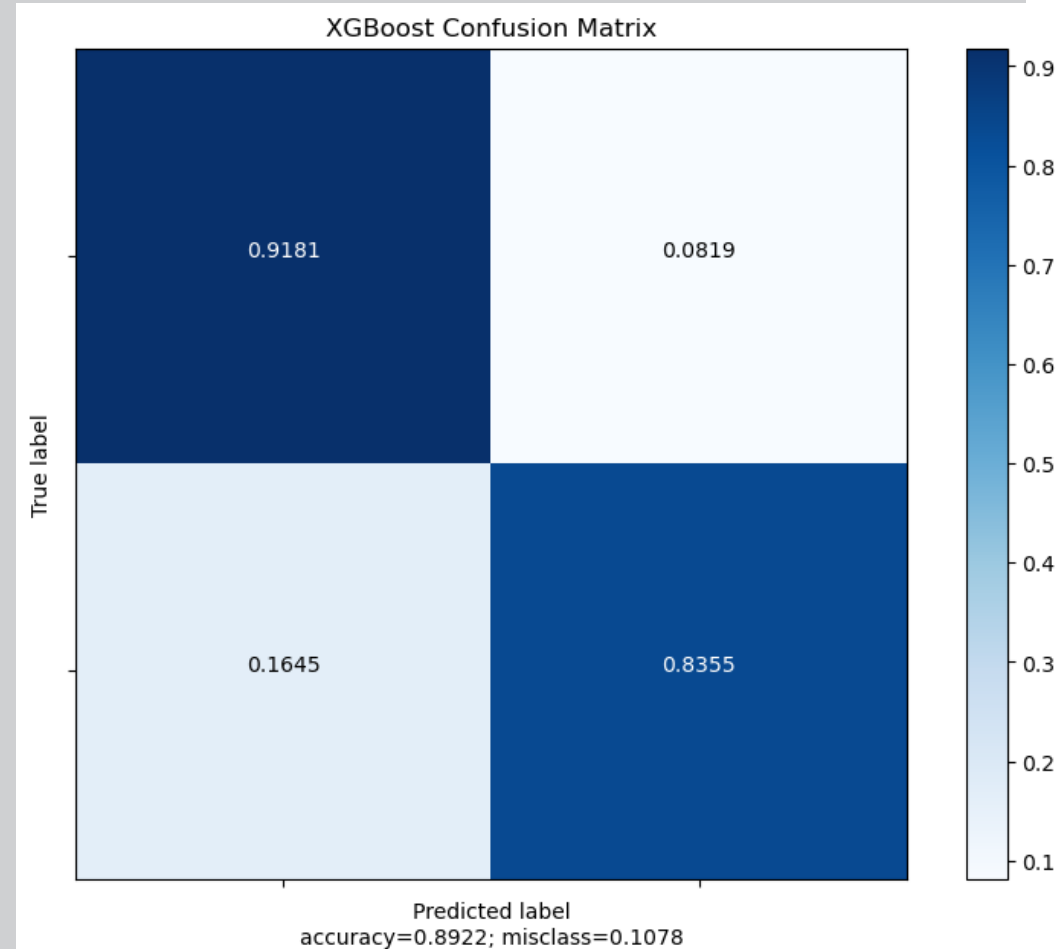
- Deep Neural Network
 - Raw Accuracy: 89.51
 - Probability of Detection: 84.73
 - False Alarm Rate: 17.74
 - F1: 83.47
- Model was evaluated on randomized held-out data split from MSL based master dataset





XGBoost

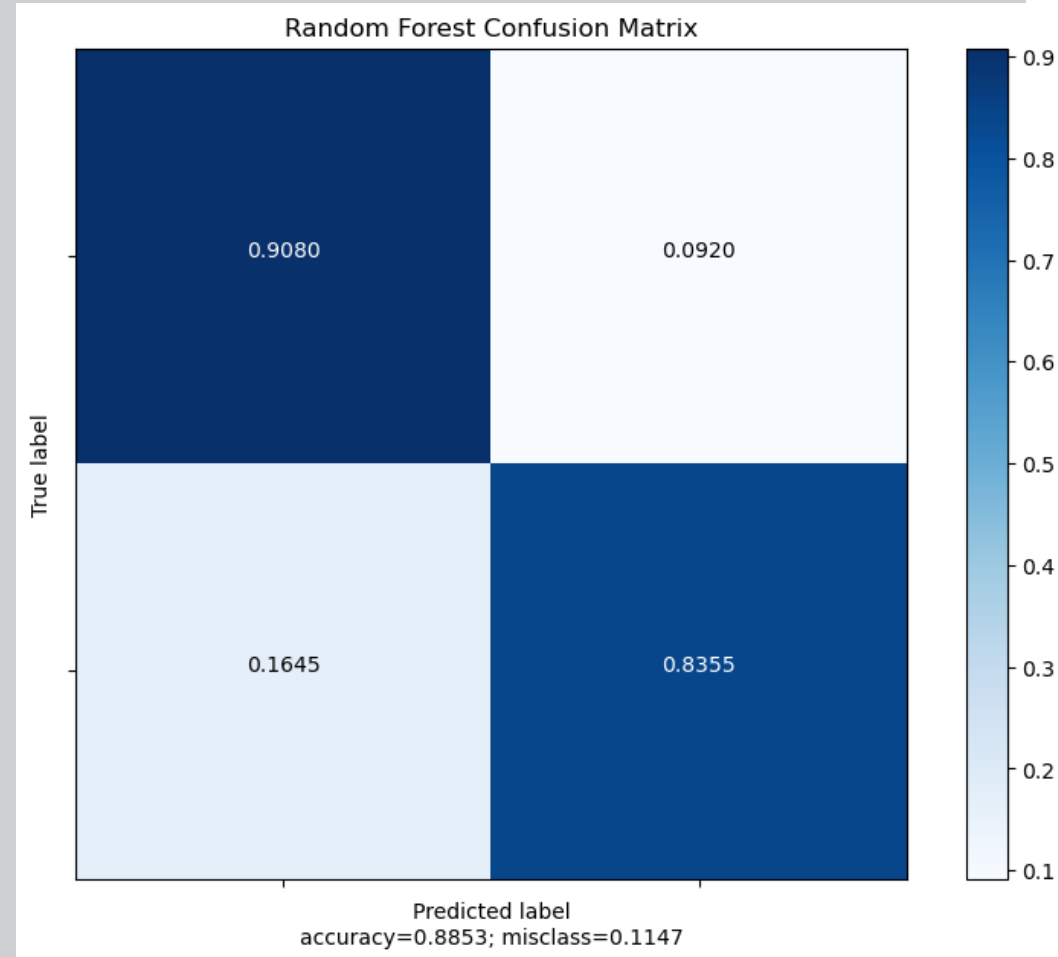
- XGBoost
 - Raw Accuracy: 89.22
 - Probability of Detection: 83.55
 - False Alarm Rate: 17.74
 - F1: 82.90
- Model was evaluated on randomized held-out data split from MSL based master dataset





Random Forest

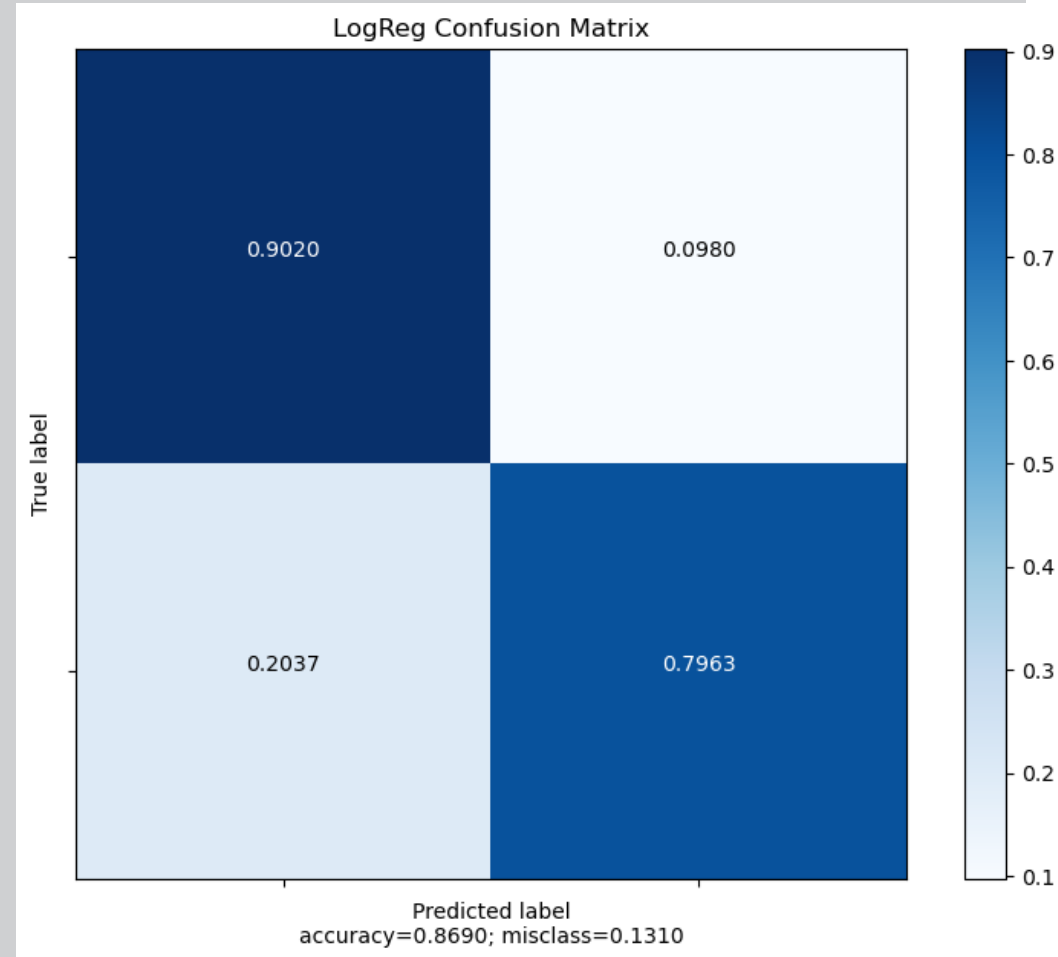
- Random Forest
 - Raw Accuracy: 88.53
 - Probability of Detection: 83.55
 - False Alarm Rate: 19.49
 - F1: 81.99
- Model was evaluated on randomized held-out data split from MSL based master dataset





Simple Logistic Model (Reference)

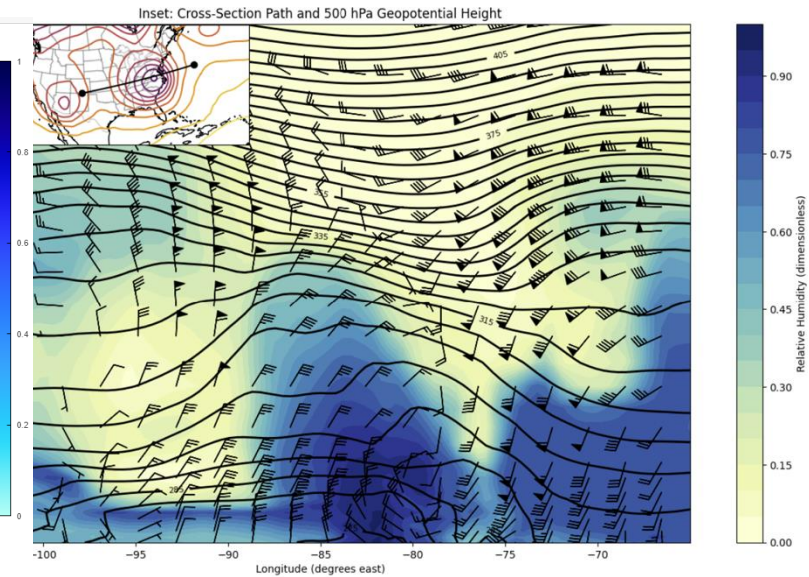
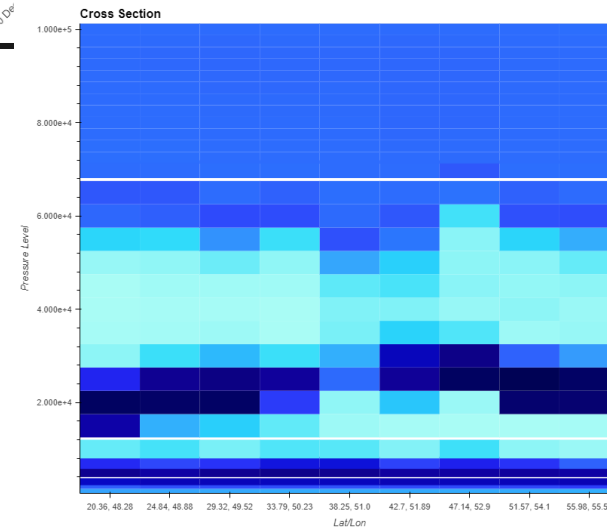
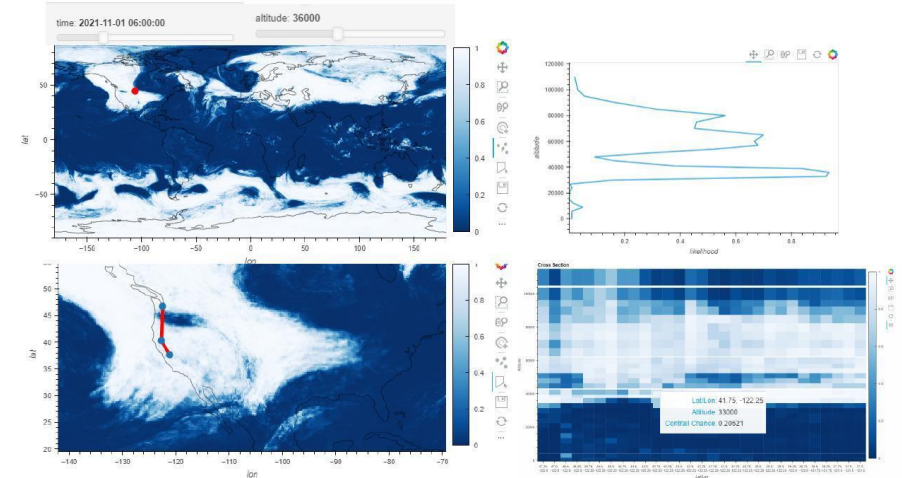
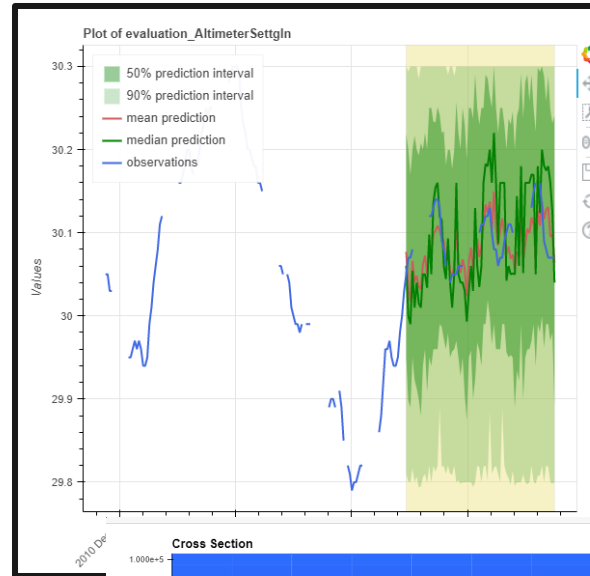
- Simple Logistic Model (Reference)
 - Raw Accuracy: 86.89
 - Probability of Detection: 79.63
 - False Alarm Rate: 21.29
 - F1: 79.17
- Model was evaluated on randomized held-out data split from MSL based master dataset





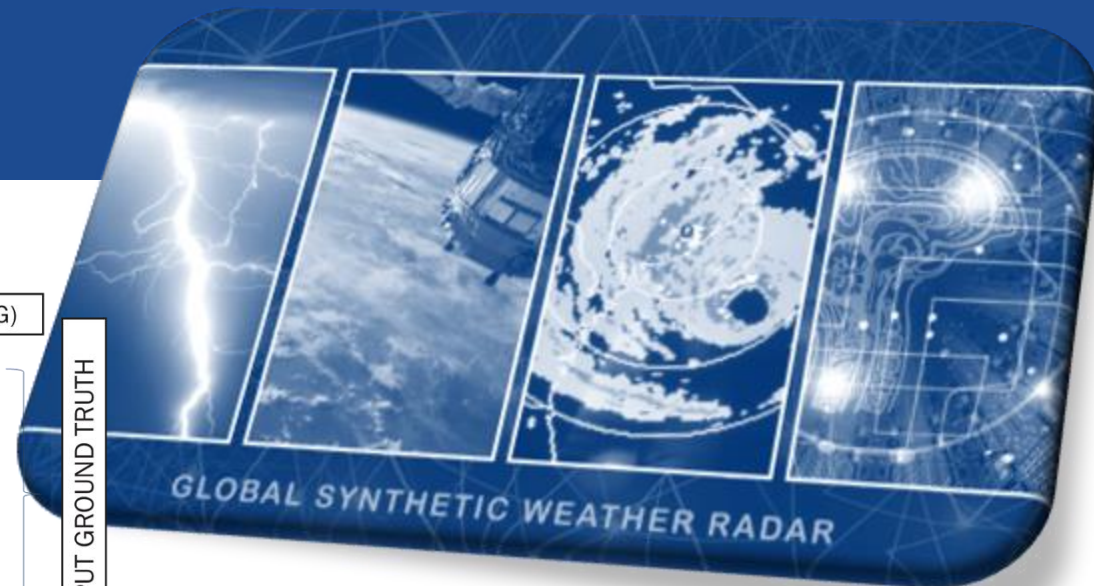
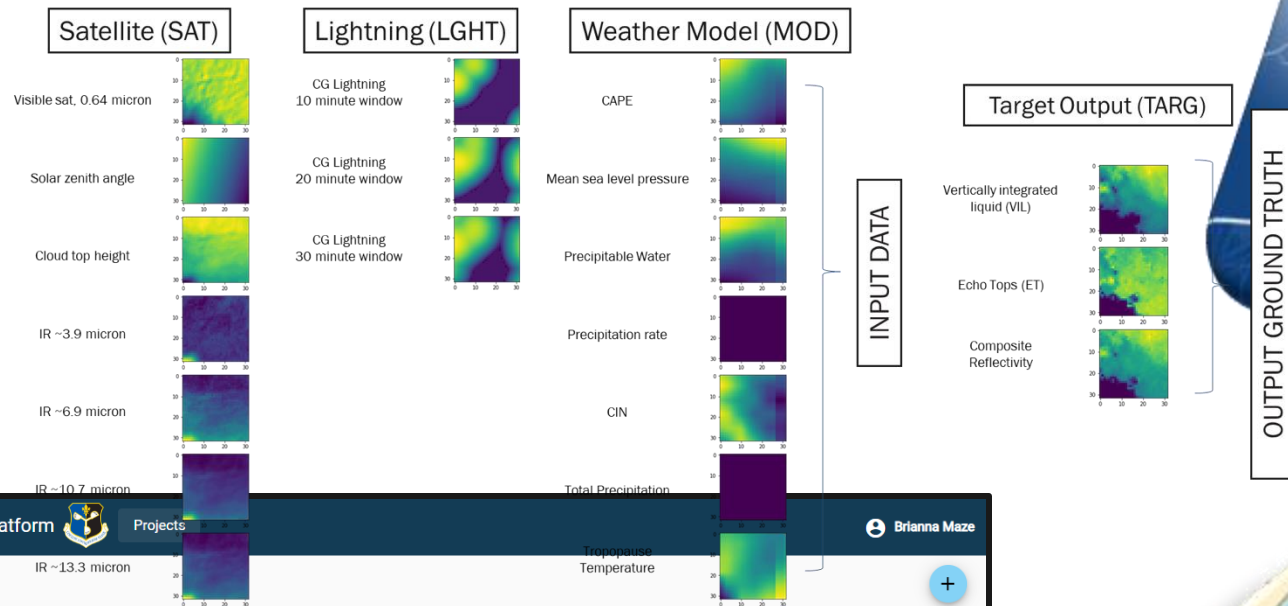
Forecasting Applications

- Weather
 - TAF
 - Precip Type
- Contrails
- Turbulence
- Lightning
- Synthetic Weather Radar





GSWR



Weather ML Platform

Projects

Brianna Maze

Projects

Your Projects Public Projects

Project Name

16th WxS

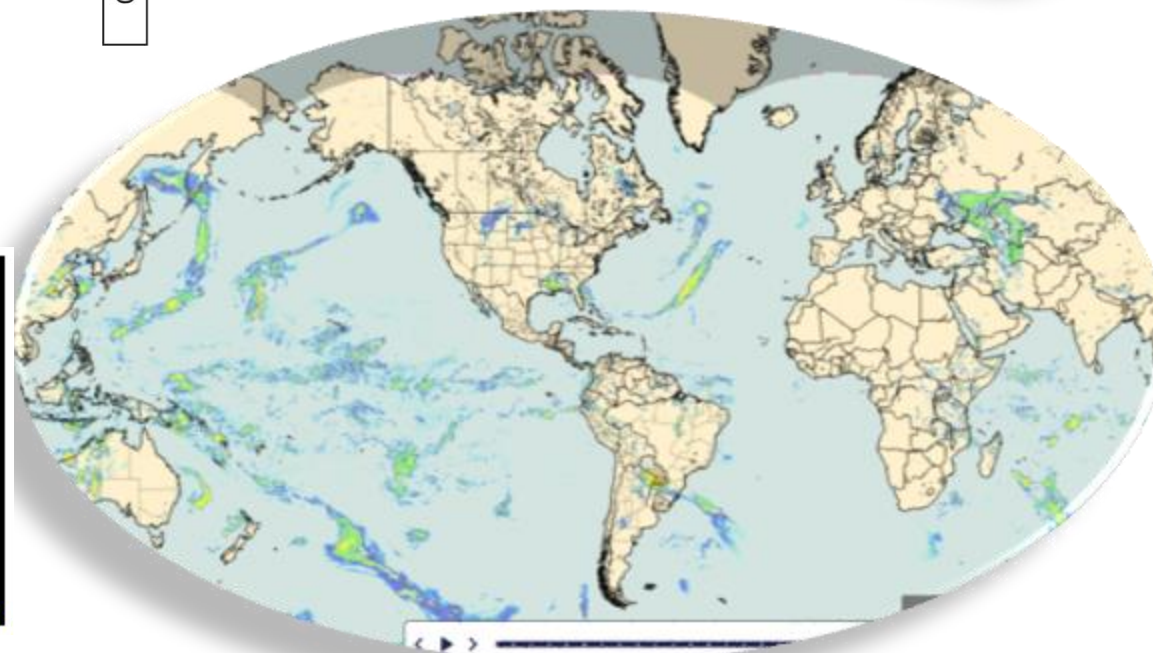
Contrails project-owner

GSWR project-owner

Public Resources

WeatherBench project-owner

Radar observations vs. ML improvement





MLP Summary

- MLP provides a unified platform to rapidly prototype ML models for use across a variety of task types including Weather applications
- Reduces the cost and time associated with rapid ML prototyping
- No-code/low-code environment presented in an intuitive interface
- Easy sharing and re-use of models through a centralized catalog of containerized models, data transformations, and evaluators
- Deployed operationally in the USAF Weather Enterprise Cloud
 - Adheres to US DoD cybersecurity standards
 - Supporting several MLOps studies

Questions? rfarrell@nextgenfed.com