NEXTGEN FEDERAL SYSTEMS

# Machine Learning Platform (MLP) for Weather Applications at Scale

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

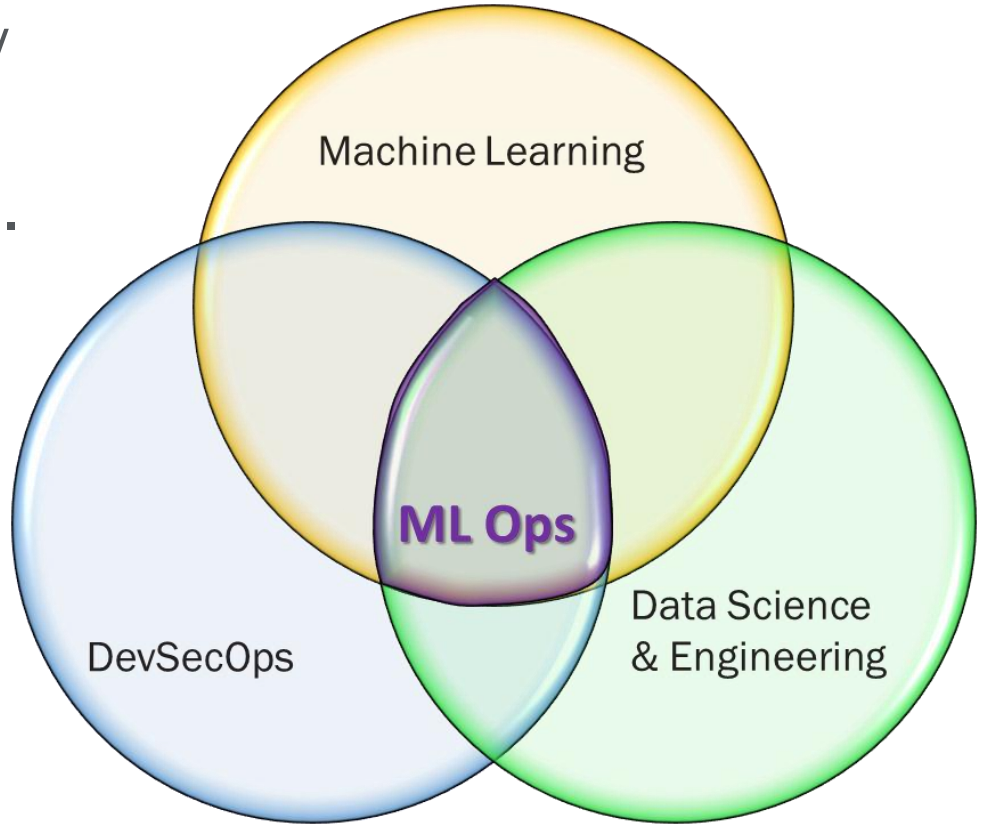
June 8, 2023

# 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.




[1] Breuel, Cristiano. "ML Ops: Machine Learning as an Engineering Discipline". https://builtin.com/machinelearning/mlops

# 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

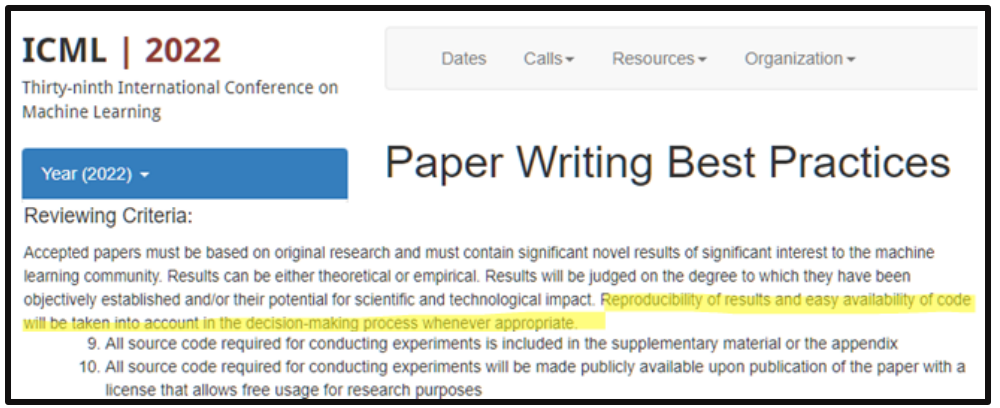
**Missing data hinder replication of artificial intelligence studies**

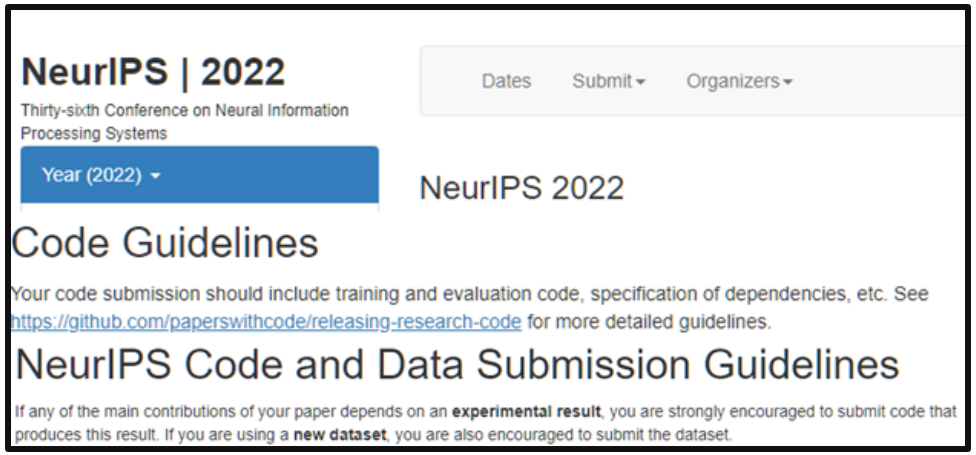
Unpublished code and sensitive training conditions aggravate reproducibility crisis in computer science

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



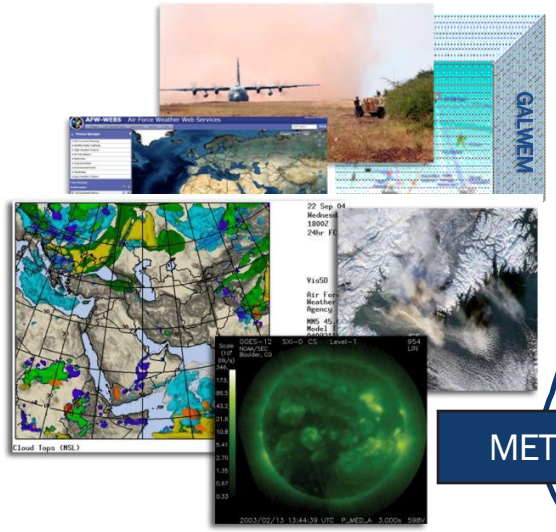



# MLOps Challenge

- 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’

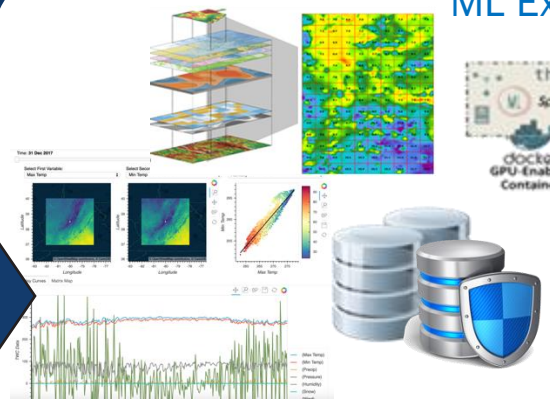
| Repository | Data | Models |
|---|---|---|
| https://registry.opendata.aws/ | ✓ | |
| https://datasetsearch.research.google.com/ | ✓ | |
| https://msropendata.com/ | ✓ | |
| https://github.com/awesomedata/awesome-public-datasets | ✓ | |
| https://visualdata.io/discovery | ✓ | |
| https://www.kaggle.com/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

# MLP Streamlines ML Model Production



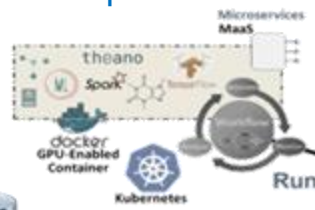

Archived Historical data in any format

METOC Data



Data analysis, exploration,& management

ML Experimentation

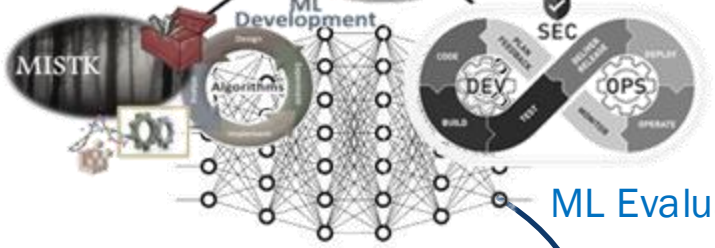



Configuration Management

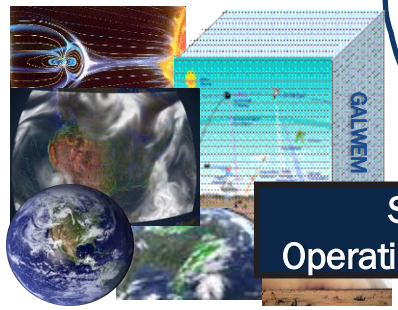
ML Training




MLOps DevSecOps




ML Evaluation




Streaming Operational Data

AFW Enterprise Services

Data Processors

Data Pre-Processor

Accepted ML Models

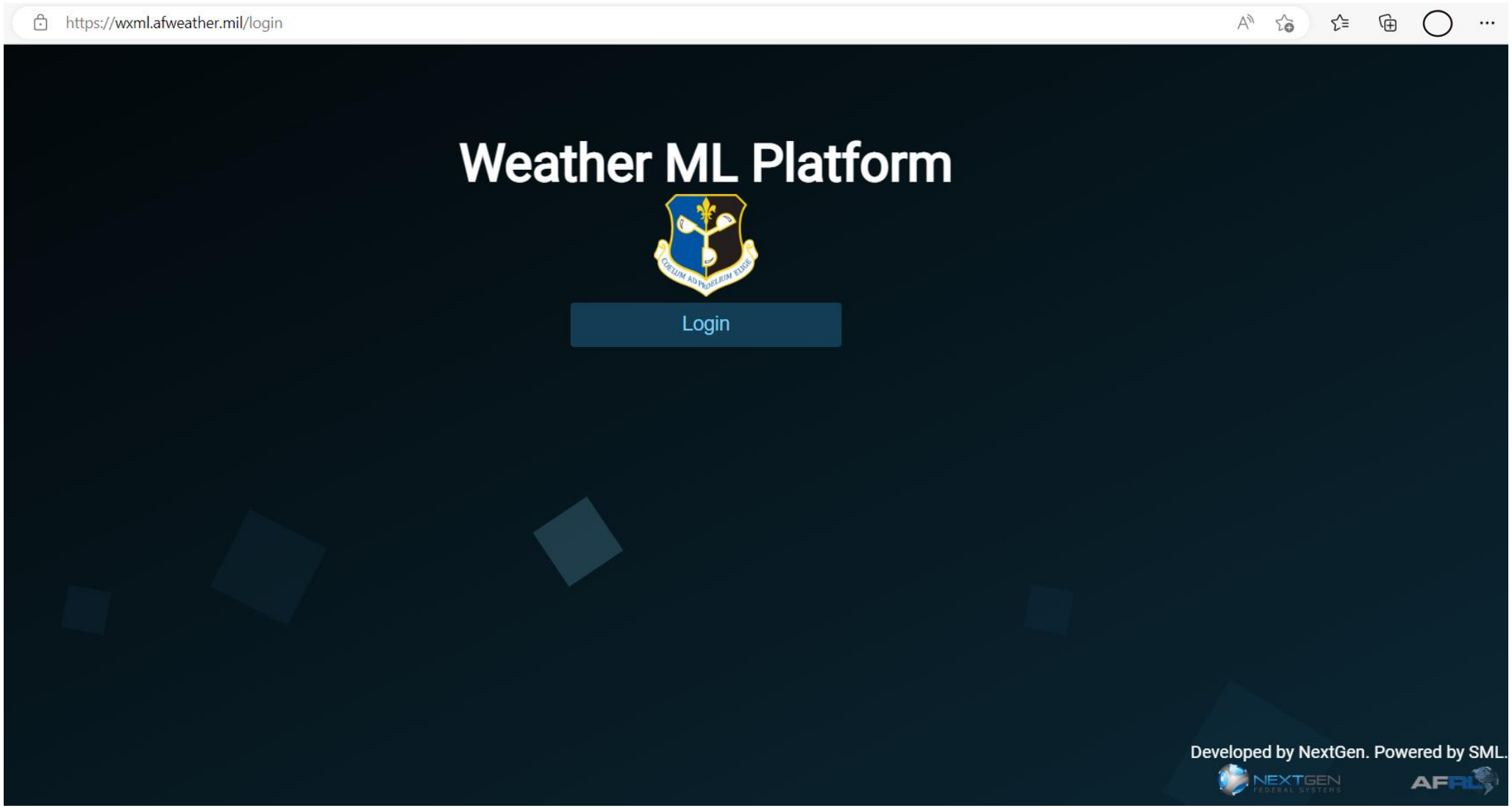
ML Ops Engine

Output

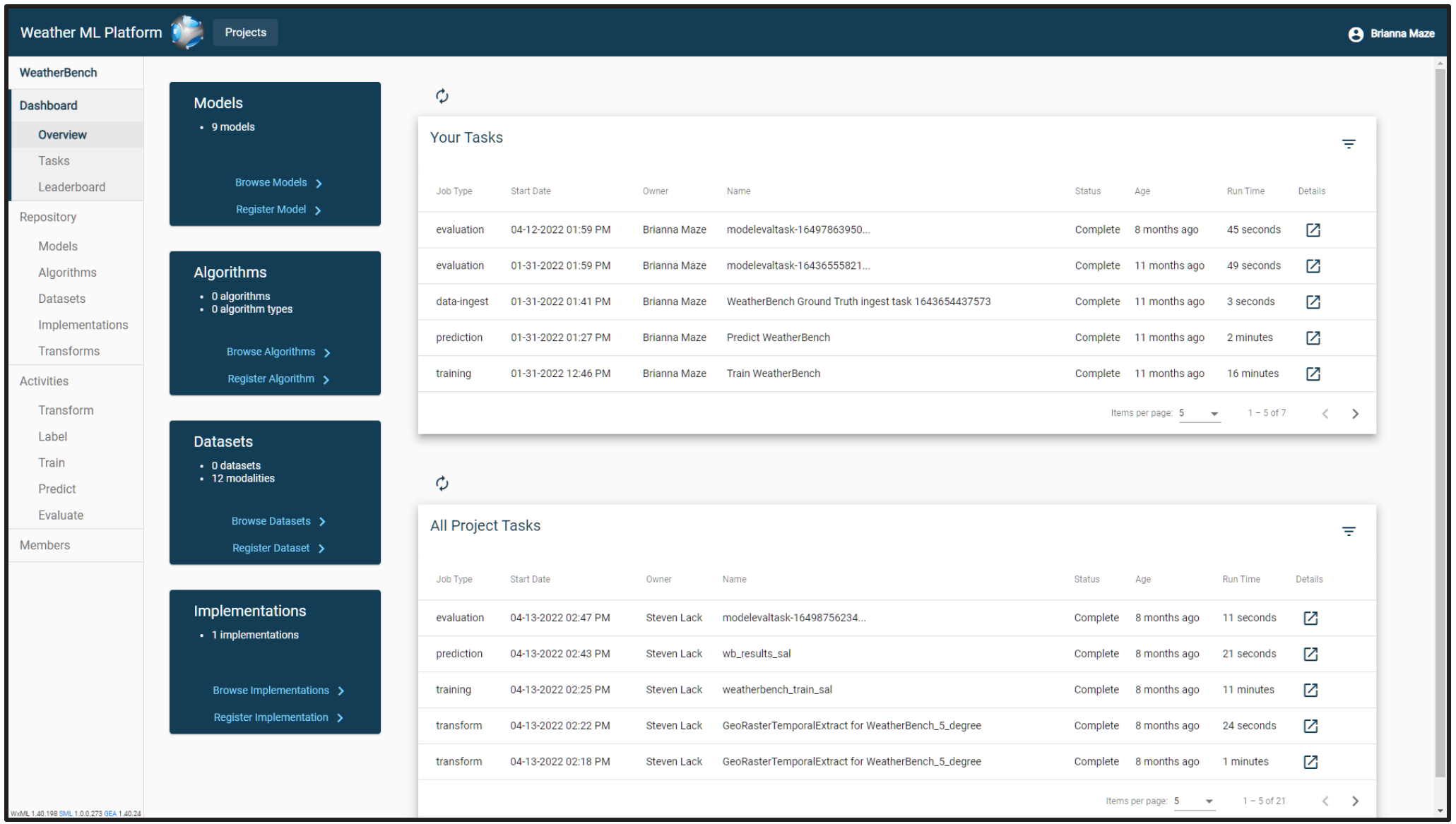
AFWE GovCloud

A runtime ecosystem to quickly train, test, and validate & verify machine learning algorithms

# MLP Login using Single Sign-On

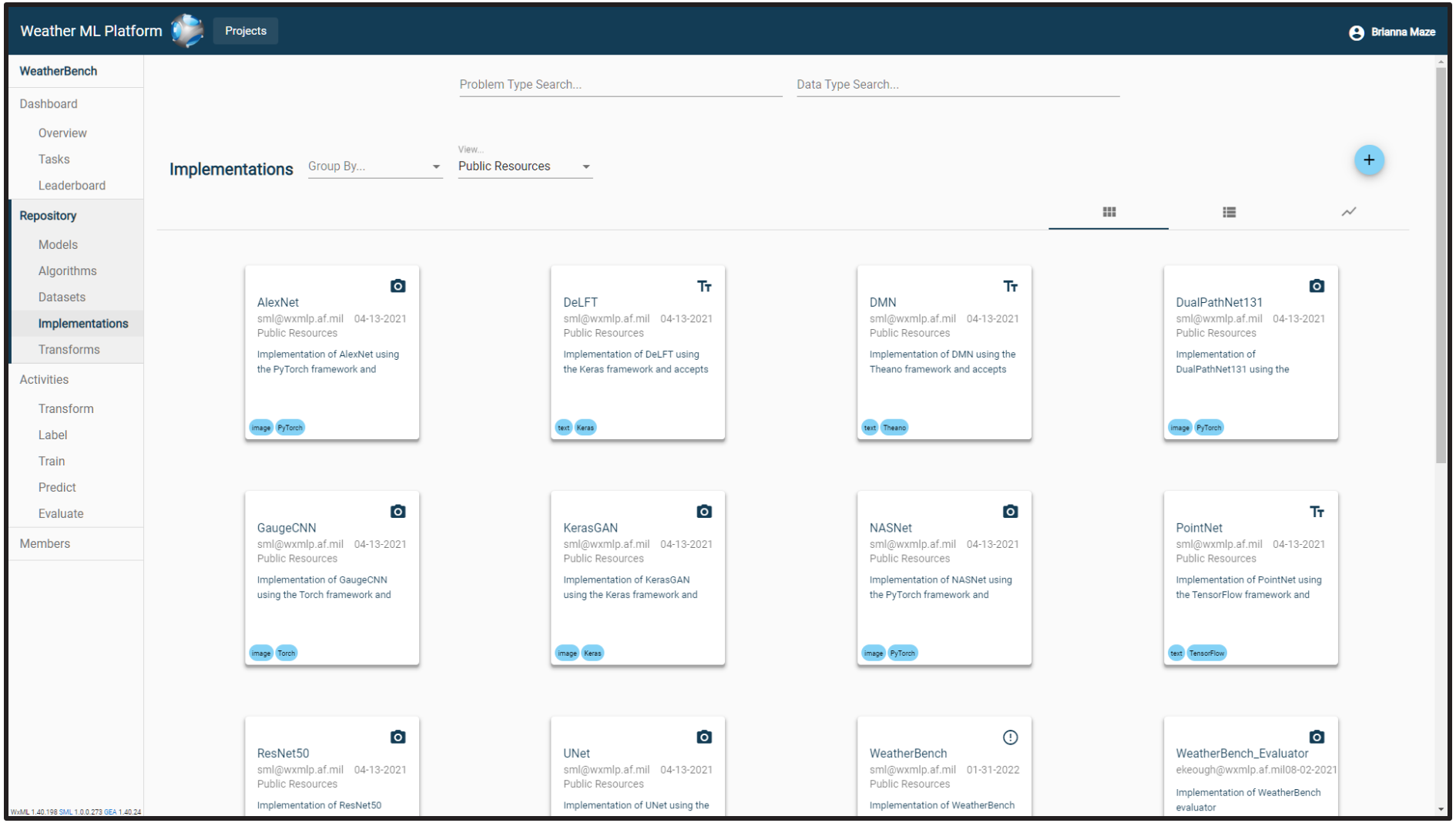


# MLP Dashboard UI



# 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 |

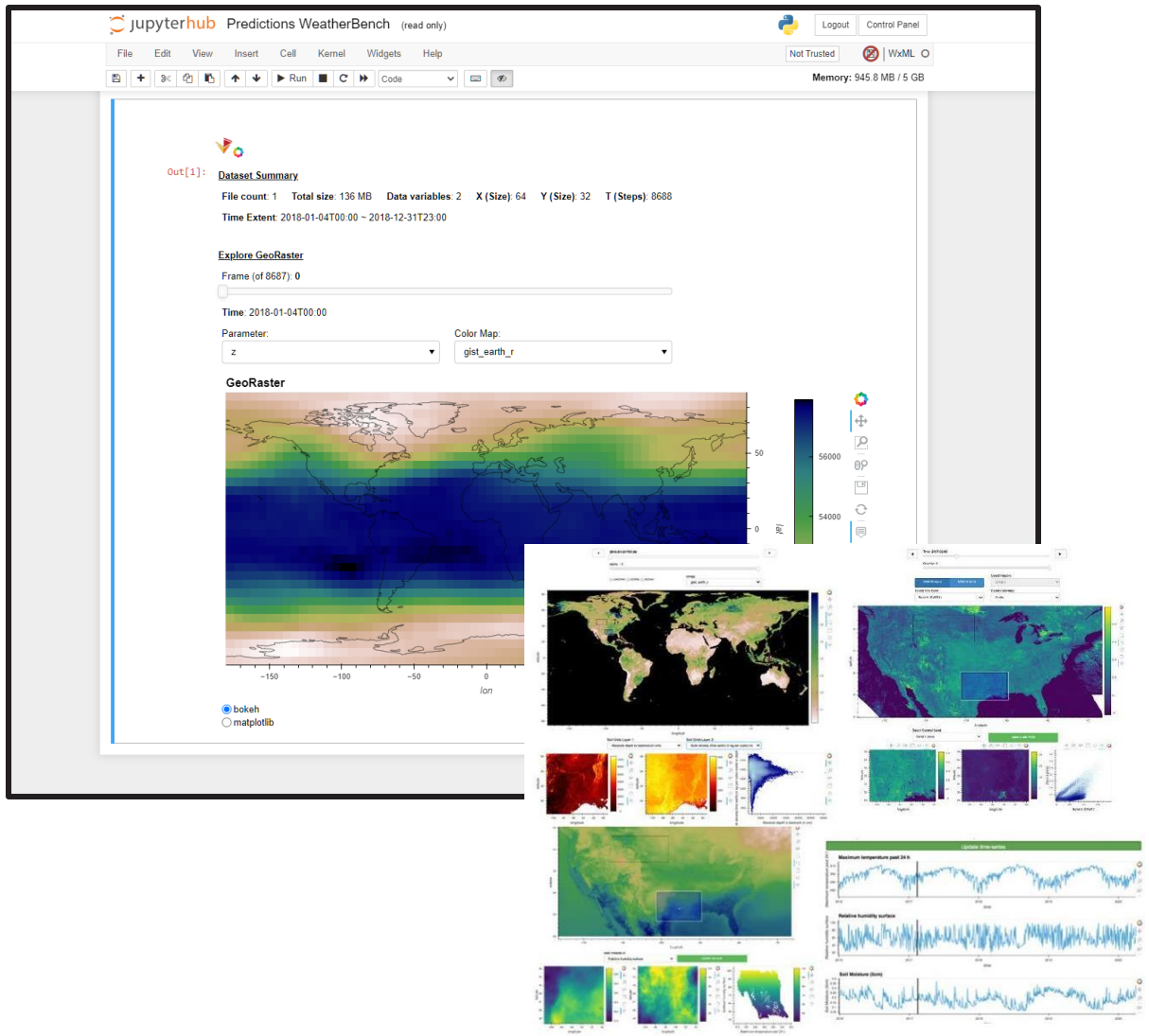



Totals: 27 algorithms, 12 task types

New algorithm: ~9 days

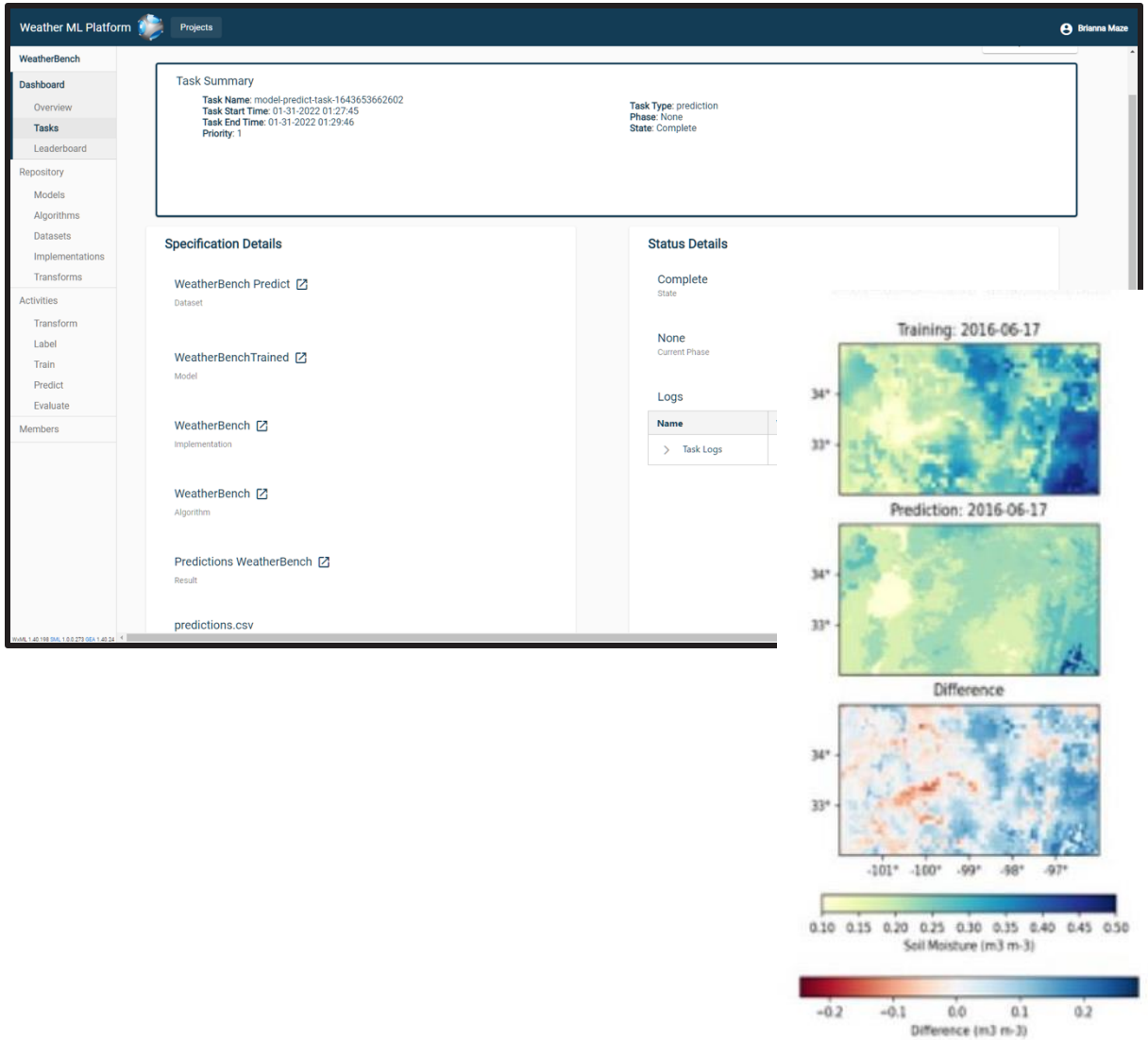
# MLP Data Science & Engineering

- 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



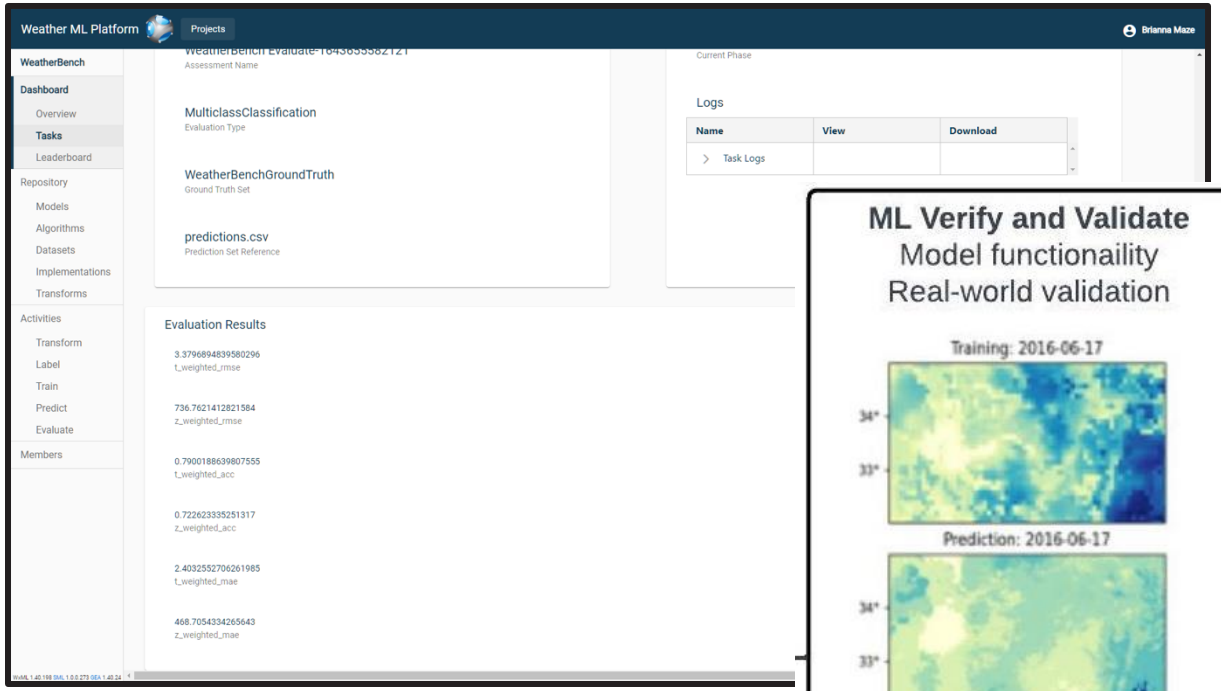
# MLP Model Training and Prediction

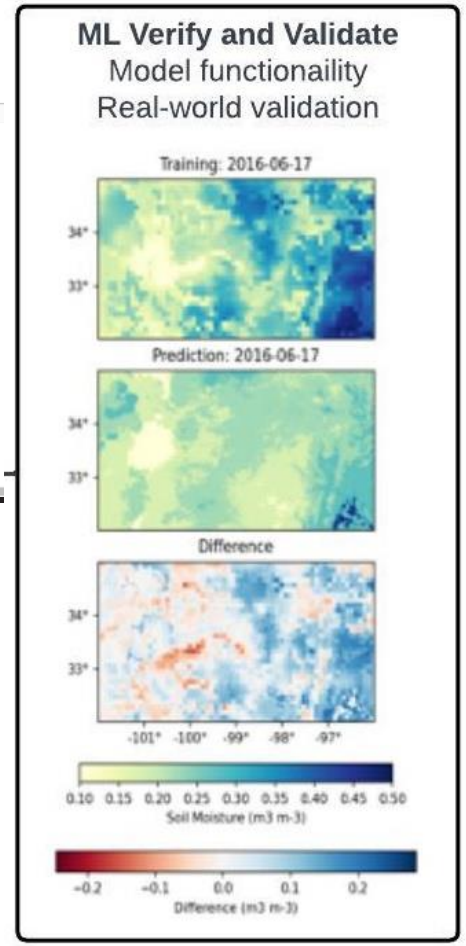
- Perform model training and prediction in a code-less environment
- All model metadata is tracked through its lifecycle
- Monitor the training and prediction tasks through the Tasks page



# MLP Evaluation

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

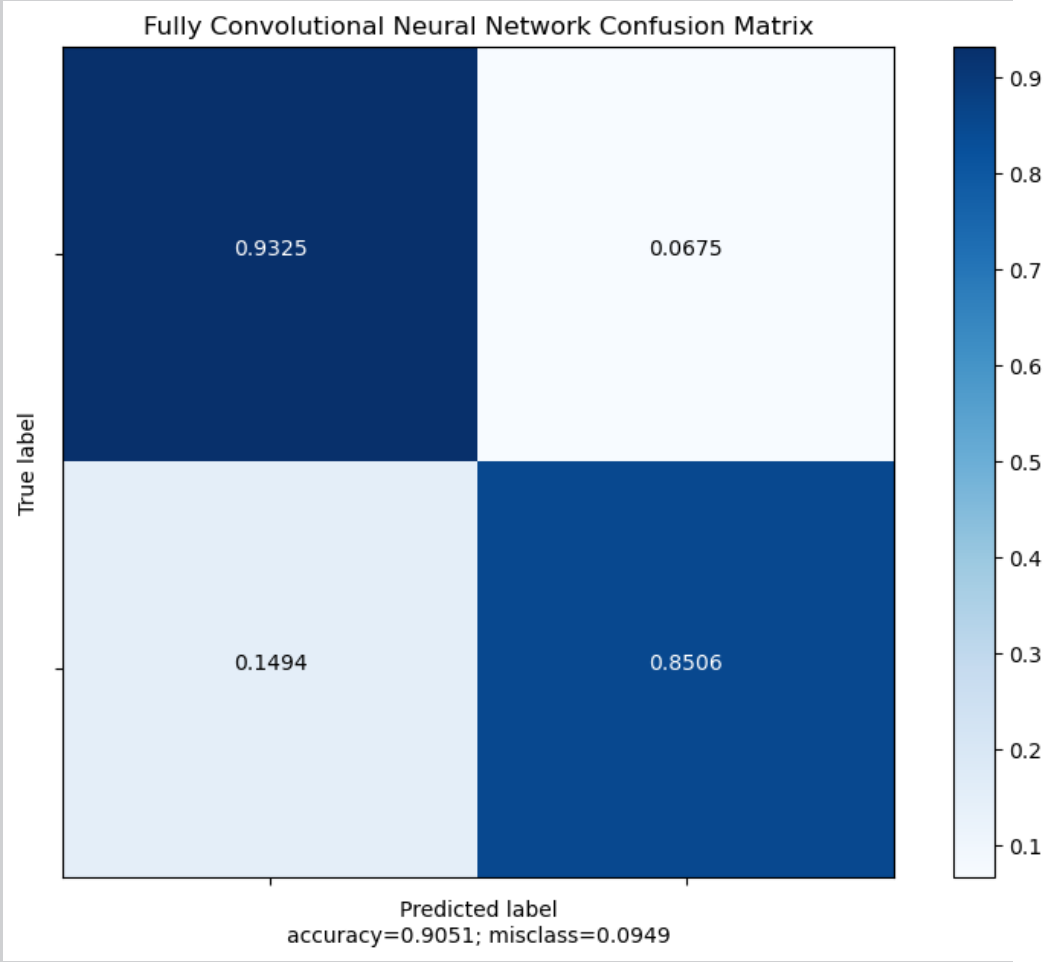


# Fully Convolutional Network

- Fully Convolutional Network
  - Raw Accuracy: 90.51
  - Probability of Detection: 85.06
  - False Alarm Rate: 13.62
  - F1: 85.71

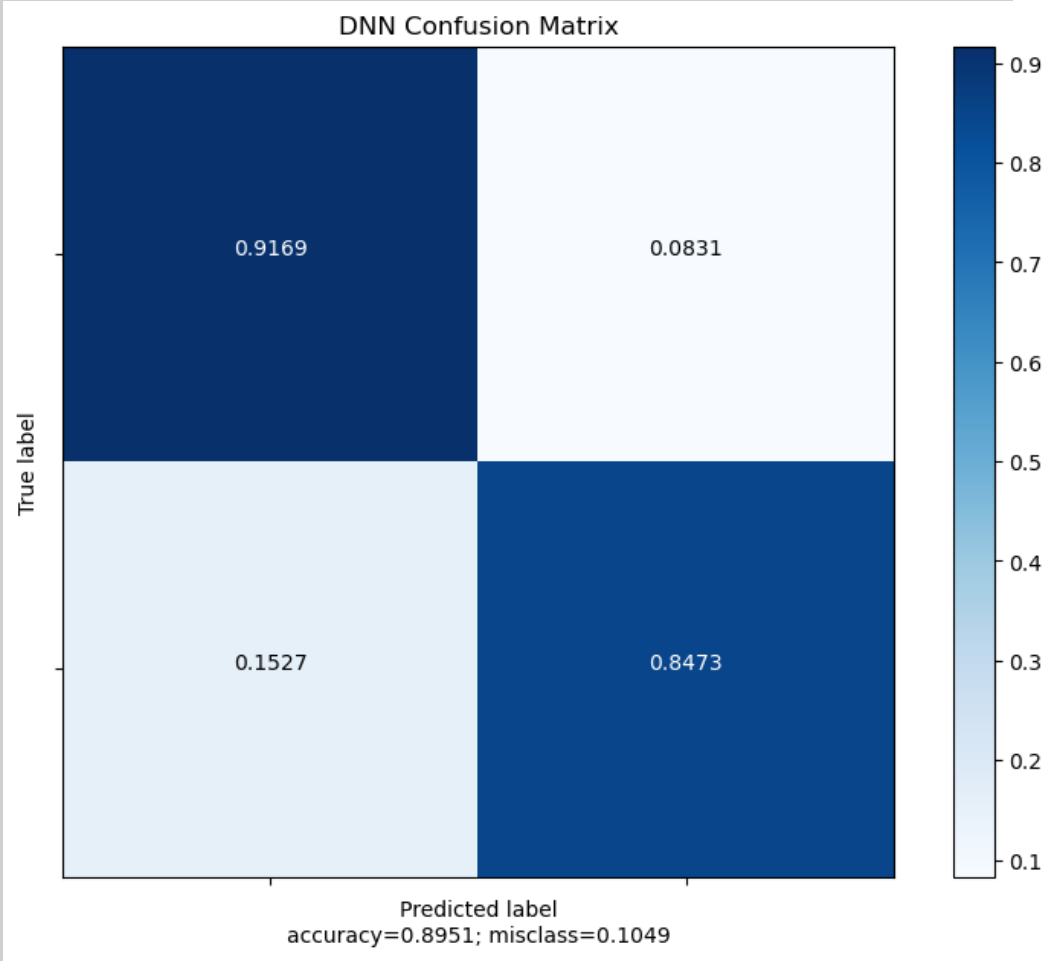
- Model was evaluated on randomized held-out data split from MSL based master dataset



# Deep Neural Network

- Deep Neural Network
  - Raw Accuracy: 89.51
  - Probability of Detection: 84.73
  - False Alarm Rate: 17.74
  - F1: 83.47

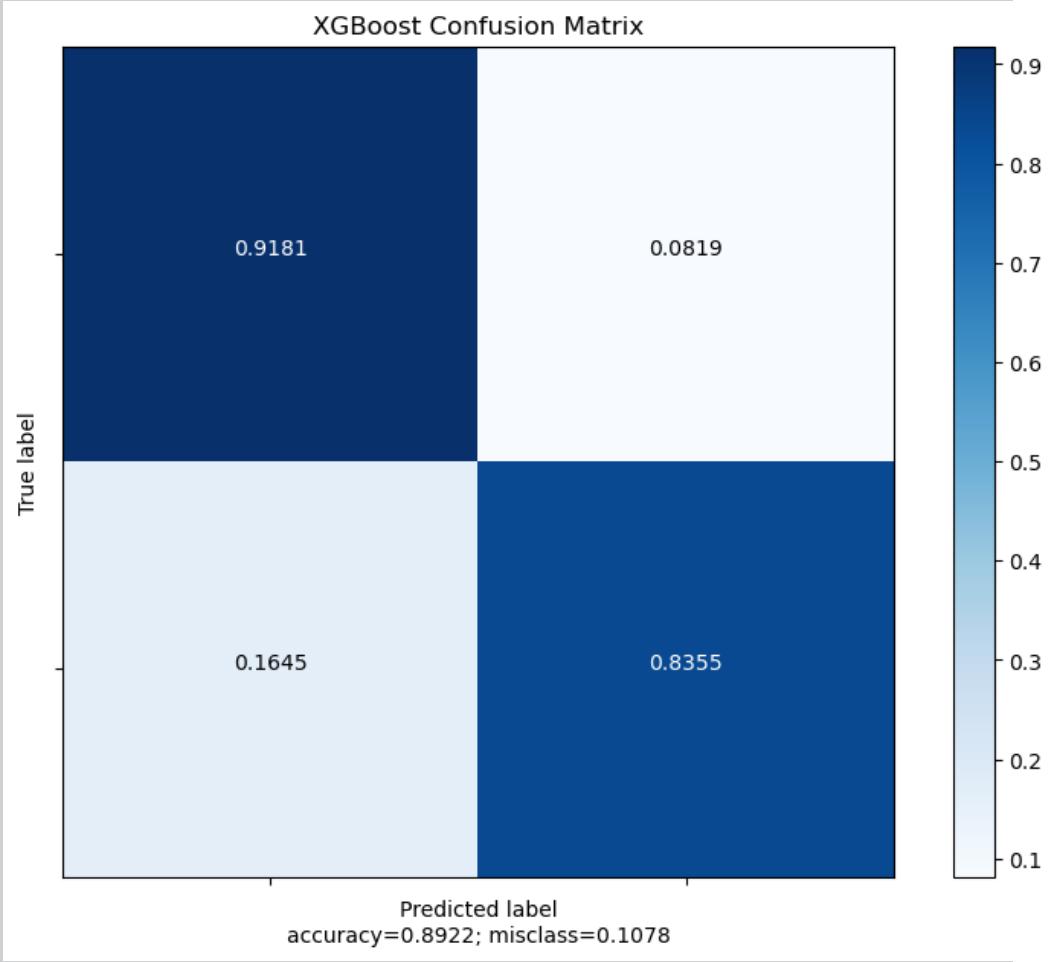
- Model was evaluated on randomized held-out data split from MSL based master dataset



# XGBoost

- XGBoost
  - Raw Accuracy: 89.22
  - Probability of Detection: 83.55
  - False Alarm Rate: 17.74
  - F1: 82.90

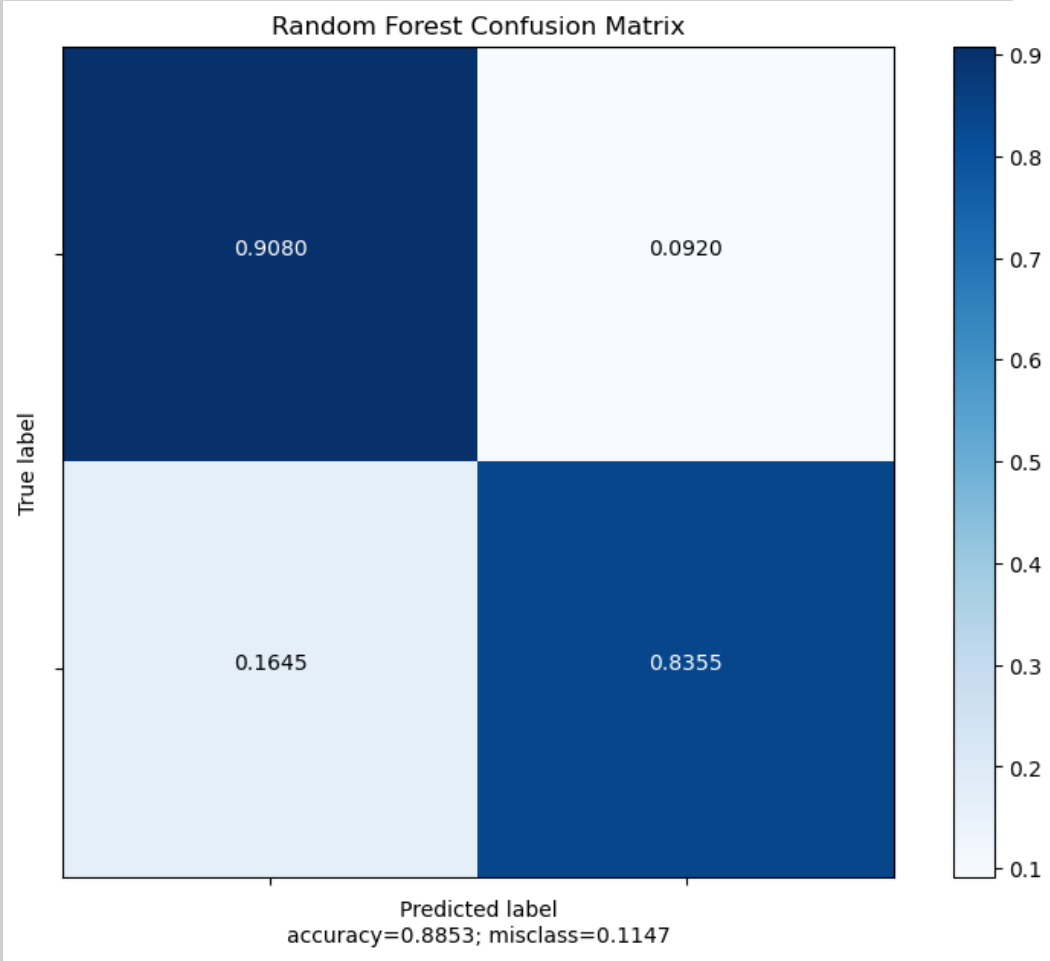
- Model was evaluated on randomized held-out data split from MSL based master dataset



# Random Forest

- Random Forest
  - Raw Accuracy: 88.53
  - Probability of Detection: 83.55
  - False Alarm Rate: 19.49
  - F1: 81.99

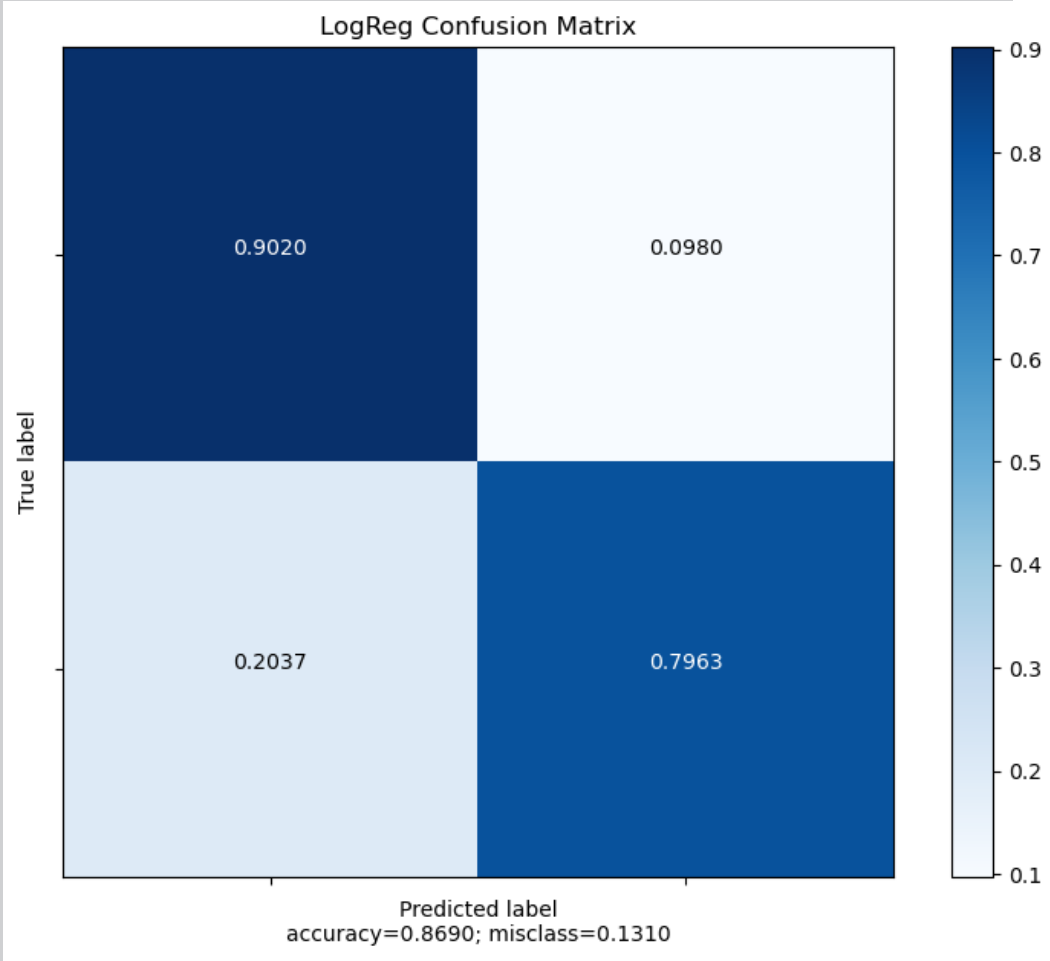
- Model was evaluated on randomized held-out data split from MSL based master dataset



# Simple Logistic Model (Reference)

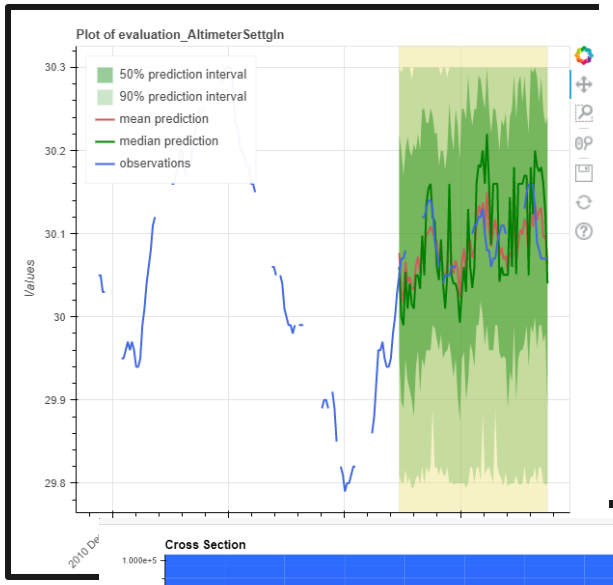
- Simple Logistic Model (Reference)
  - Raw Accuracy: 86.89
  - Probability of Detection: 79.63
  - False Alarm Rate: 21.29
  - F1: 79.17

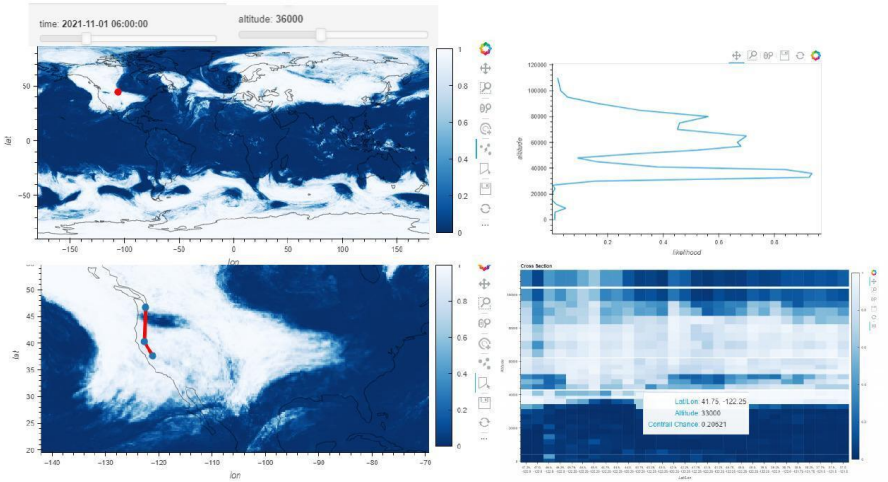
- Model was evaluated on randomized held-out data split from MSL based master dataset

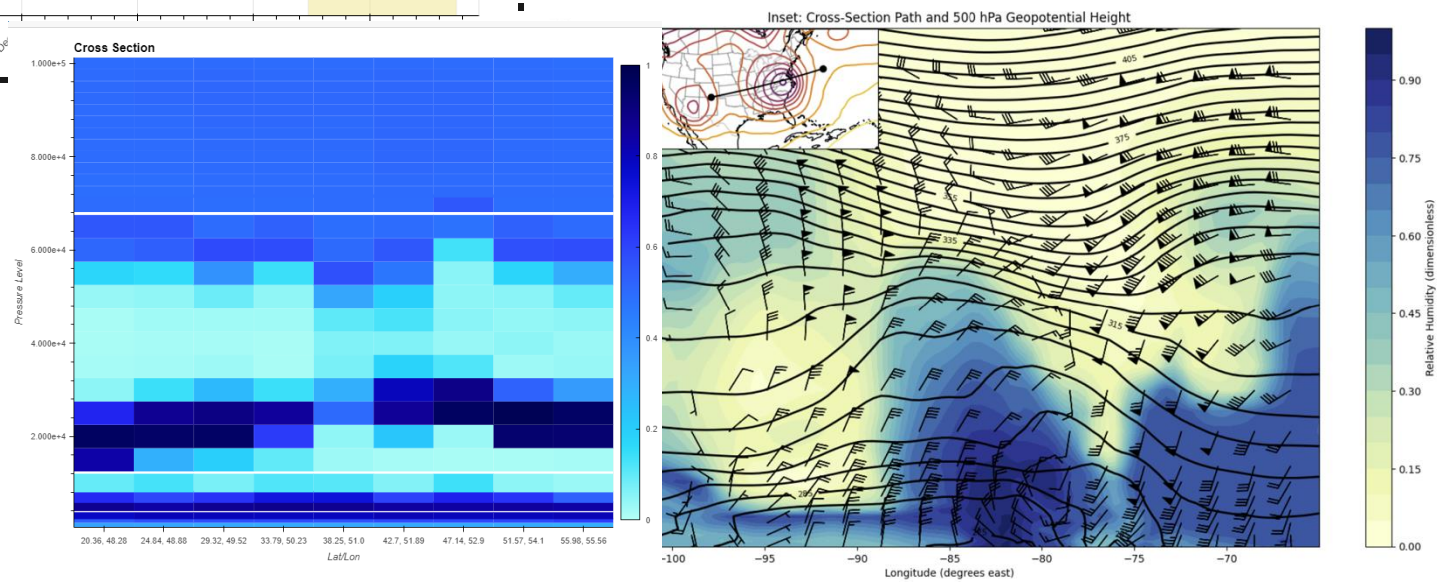


# Forecasting Applications

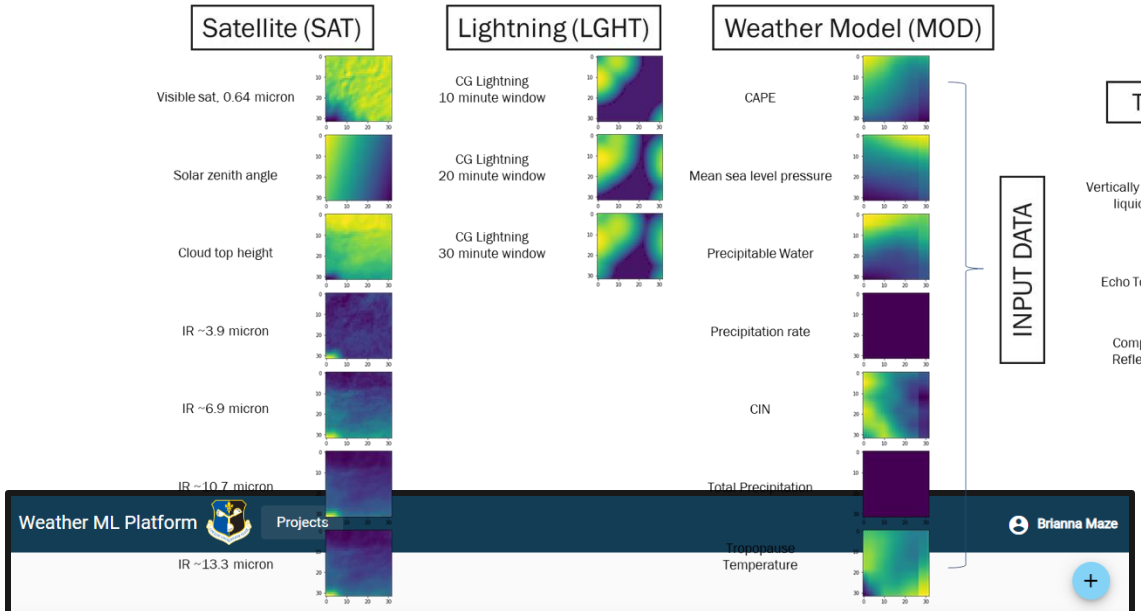
- Weather
  - TAF
  - Precip Type
- Contrails
- Turbulence
- Lightning
- Synthetic Weather Radar

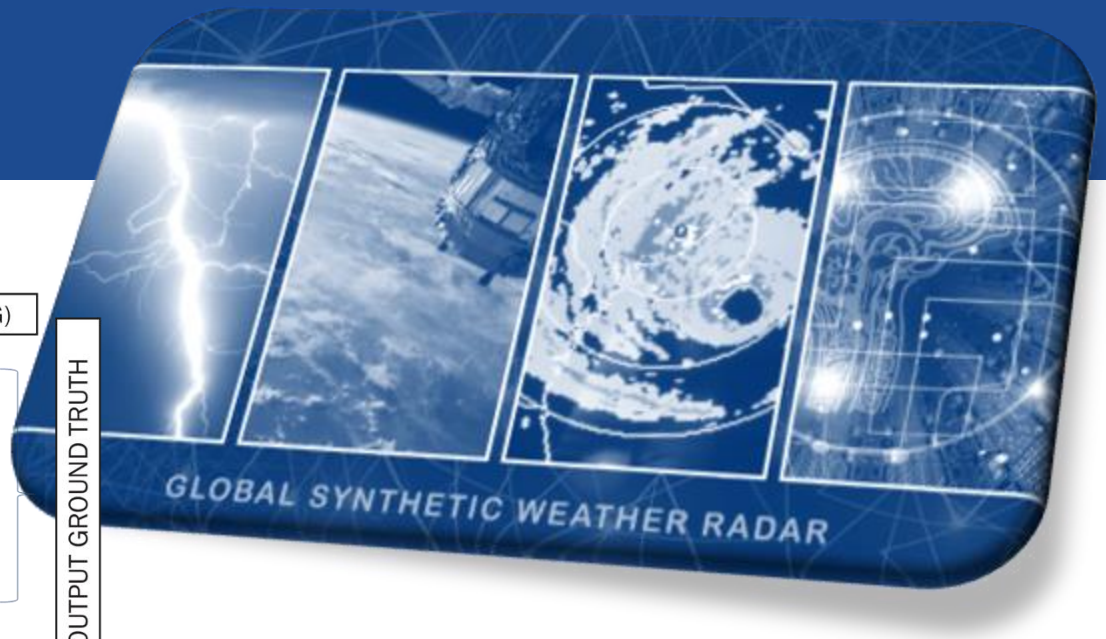


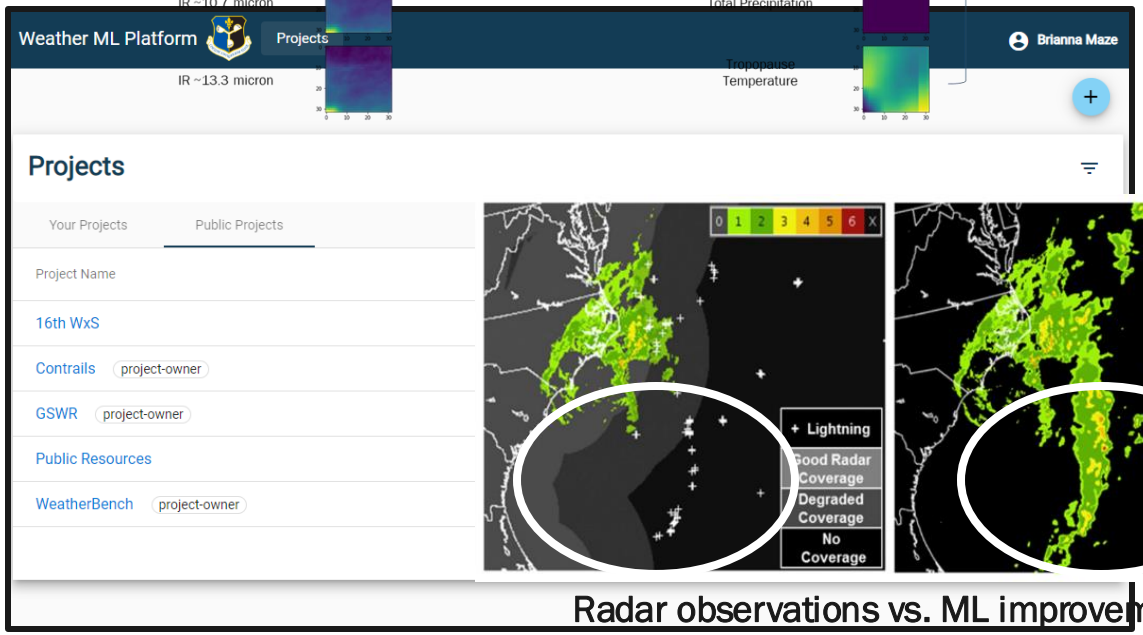



# GSWR

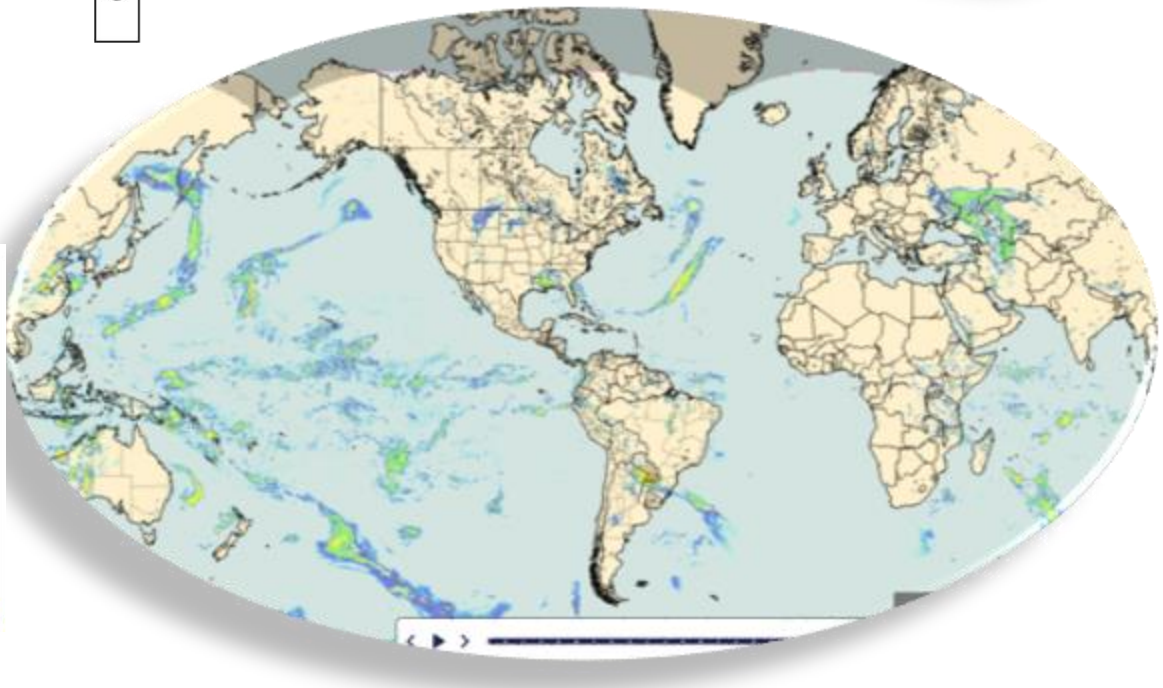








Radar observations vs. ML improvement

# 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