Advancing Organizational Capabilities in AI / ML Through Collaboration, Community Building, and a Deep Learning Academy

Brandon Smith[1,2], Jackie Kendall[1], Craig Pelissier[1,2], Shelley Stover[3]

1. Science Systems and Applications, Inc.
2. NASA Goddard Space Flight Center
3. NASA Langley Research Center

Collaborate with us!

Introduction

AI/ML is a critical focus area for the future readiness of any organization. Besides simply acquiring and fostering specialized talent in that area however, it must be recognized that successful projects invariably require some level of domain knowledge separate from AI/ML.

These two perspectives – science domain knowledge, and AI/ML aptitude – must be combined, in order to push research forward and create fruitful results.

SSAI has developed a comprehensive framework to guide workforce development in this rapidly growing field, the cornerstone of which is the **Deep Learning Academy**: a multifaceted approach aimed at augmenting existing domain knowledge with capabilities in AI/ML.

Selected applicants are sponsored by SSAI to complete intensive coursework, followed by real-world projects that are developed in-house by existing AI/ML experts, and relevant to Earth and environmental science customers.

We discuss the details of this program and focus in on the final project to examine what makes for a successful interdisciplinary and collaborative learning environment.

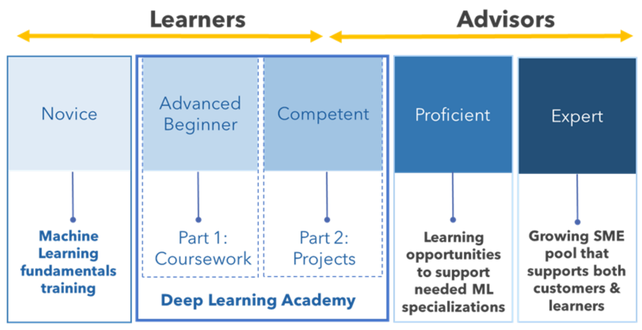
Advisors & Learners

Following the launch of SSAI's AI/ML Learning Program in October 2020, learners began a 16-week, 5-course specialization to gain a solid foundation in Deep Learning.

After the coursework learners continued to expand knowledge and skills by pairing up and implementing solutions to two applications which are under investigation by many Earth scientists: time series forecasting, and semantic segmentation.

These projects were designed by program advisors and machine learning subject matter experts Craig Pelissier and Brandon Smith, to solidify understanding of class materials by applying it to real-world data and problems in the Earth Sciences.

AI / ML Learning Program Components



Problem Statement

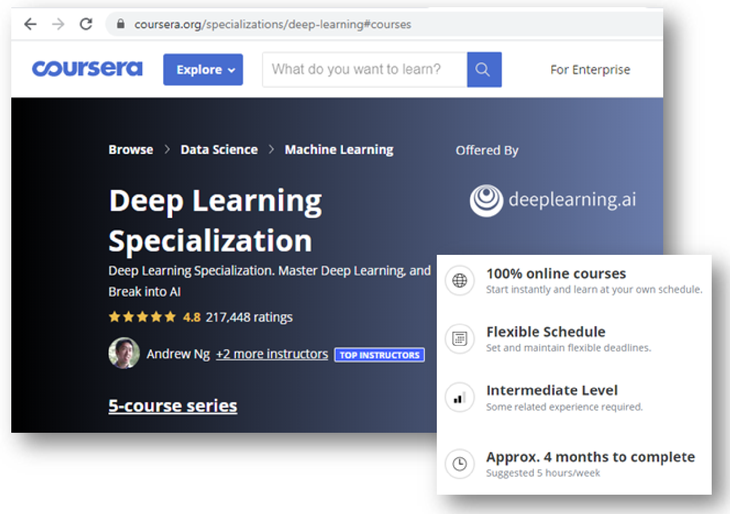
1. Many problems in science may be amenable to AI/ML solutions, but few individuals have the skill sets necessary to properly evaluate applicability.
2. Many scientists exist with domain-specific knowledge, but far fewer have the needed cross-disciplinary experience in AI/ML.
3. Acquiring this specialized talent is costly, and so organizations must find ways to grow these skills within their current workforce.

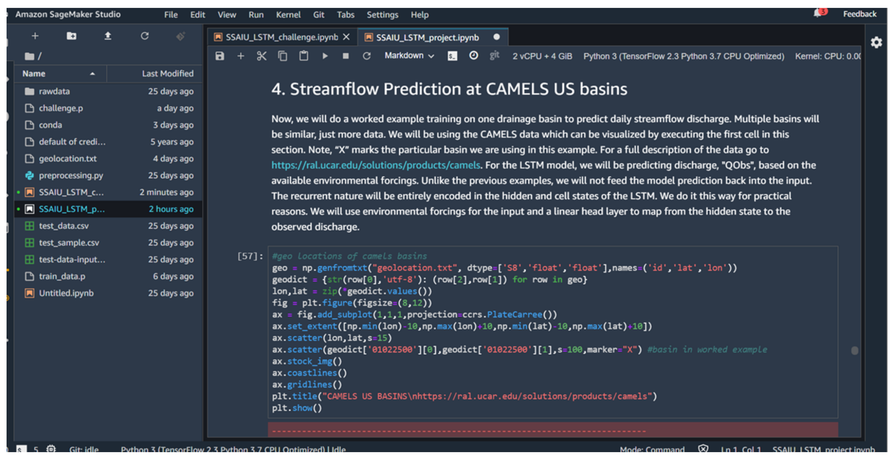
Deep Learning Academy

Part 1 of the program involves learners earning the certificate from Andrew Ng's highly regarded *Deep Learning Specialization* on Coursera, which includes 5 courses:

1. Neural Networks and Deep Learning
2. Improving Deep Neural Networks
3. Structuring Machine Learning Projects
4. Convolutional Neural Networks
5. Sequence Models

Each course takes place over several weeks, having learners watch lecture videos and complete graded quizzes, assignments, and labs. Advisors are available for one-on-one mentoring, and learners engage with each other for peer support via dedicated channels in *Microsoft Teams*.



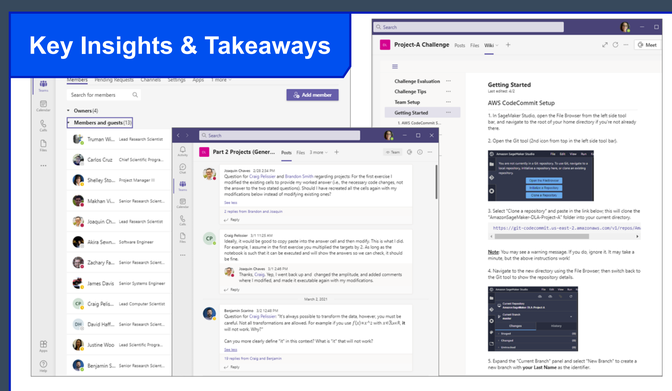


Part 2 involves learners developing solutions to two projects designed by our advisors. In addition to paid time for project work, SSAI provides computing resources via *Amazon SageMaker* for model development and training.

The projects include Jupyter notebooks and code templates that conceal the complexities of interacting with the underlying infrastructure. With this framework learners can hit the ground running, immediately applying what they've learned to the problem at hand rather than struggling with the details of *AWS* and *SageMaker*.

Both projects are also framed as a competition between teams, with *Kaggle* allowing learners to submit and track progress on key metrics.

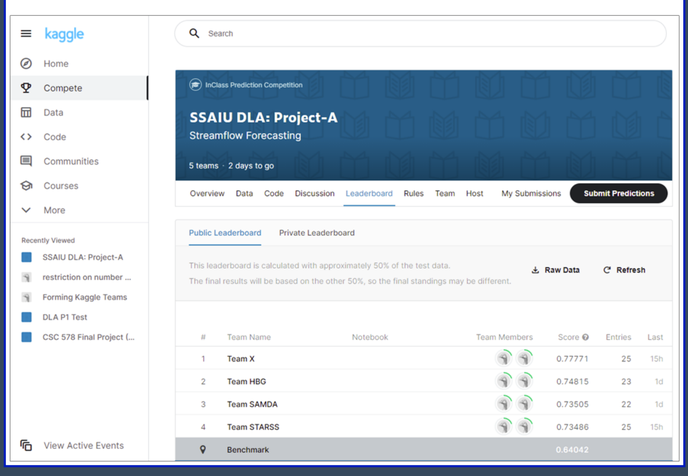
Key Insights & Takeaways



Community creates motivation. Allowing learners to discuss their ideas, questions, successes, struggles, and much more; we used Microsoft Teams to facilitate connections between learners, and grew a community of practitioners focused on the advancement of their own understanding.

Competition fuels growth. We used Kaggle competitions to track each teams' performance on relevant metrics. Friendly competition on the leaderboard provided a fun and motivating environment, with rewards for top scoring teams.

Applications are essential. Having real-world projects for learners to apply their course knowledge makes everything come together - driving comprehension of the material, insights into their own work, and experience in addressing the challenges of applying machine learning in Earth Science.

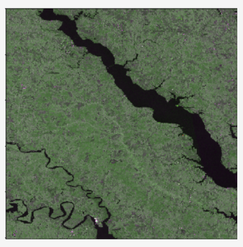


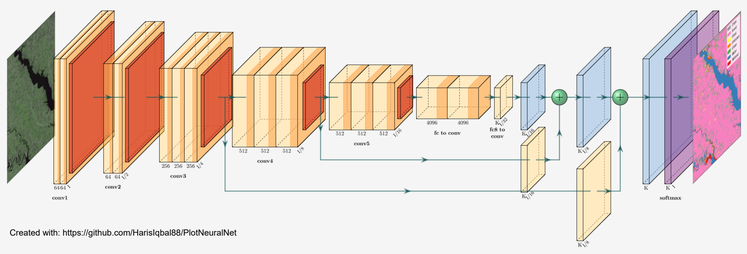
The final project in the program focuses on the problem of land surface classification, using *Sentinel-1* SAR imagery and *Copernicus PROBA-V* land classes via CEOS Open Data Cube (courtesy of Brian Killough, NASA LaRC).

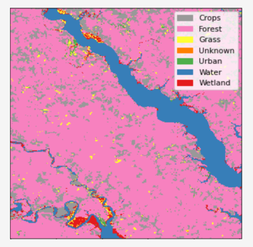
Solutions to this problem focus on applying convolutional neural networks (CNNs) to generate surface class predictions (e.g. Urban, Forest, Crop, etc.) for each pixel in a given image.

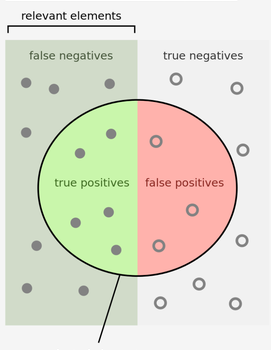
Learners are provided with a notebook walking through all concepts and ideas necessary to develop a successful model, as well as a library providing a baseline solution for the *Kaggle* competition.

One of the primary difficulties in this problem is dealing with issue of class imbalance (e.g. a scene might be 95% forest pixels). Without prior experience in this type of environment, many practitioners fail to consider the impact.









Learners must address this imbalance through weighted loss functions, biased data sampling, or some other combinations of approaches.

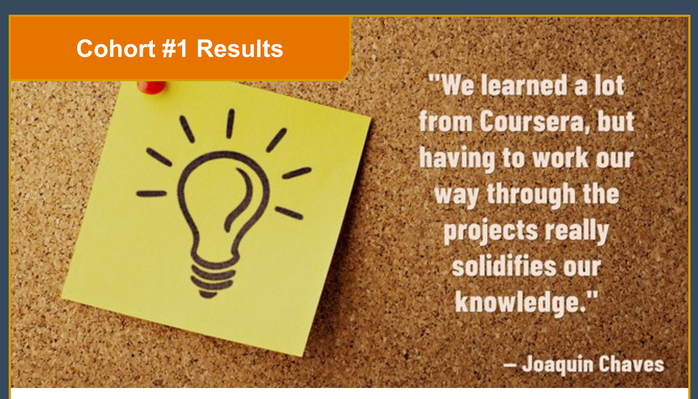
Accordingly, the *Kaggle* competition uses a held-out dataset to rank teams via their macro F1-score - a metric based on precision and recall, which penalizes models that only perform well on one or a few classes.

Other requirements for a successful model include standard machine learning considerations (data augmentation and scaling, model design and optimization, etc.); as well as domain-specific issues such as speckle correction.

These hurdles are fundamental to the learning process, and participants agreed that the greatest learning benefit came when they hit a wall and needed to re-evaluate their approach.

What drives progress the most however, are the learners themselves: partnering up to work on projects provides accountability and motivation to reach for improvement in the competition. This in turn leads to significant improvement in AI/ML skills for everyone in the program.

Cohort #1 Results



Participants were highly motivated. All learners enthusiastically completed the program from beginning to end, earning their certificates and developing competitive models.

Developed models surpassed expectations. Solutions crafted by learners in most cases were near or equal to expert-developed model performance in the competitions.

Leadership praised improvements. Learners applied the gained knowledge to their own work, resulting in promising new research directions and high satisfaction from managers.

Our Roadmap Ahead

Quantifying Results

- Systematically assess AI/ML capability and learning needs across the organization
- Use 3rd-party assessment tools to measure initial knowledge and demonstrate growth

Expanding Topics

- Add new tracks for Data Engineering & MLOps to the Deep Learning Academy
- Pair prospective Data Scientists with Data and ML Engineers to work on the projects

Interested in AI/ML?
Have ideas for a project?
Collaborate with us!

Community Partnerships

Actively seeking external collaborators to:

- Broaden our DL Academy to include participation by external Learners and Advisors
- Engage with external partners to identify high priority Earth and space science use cases
- Host curated community competitions to promote collaboration on these use cases
- Promote open science and collaboration by sharing reusable artifacts for learning and research
- Evolve and expand our capacity building efforts based on community needs and feedback

For more information, contact:
Brandon.Smith@ssaihq.com

Our Vision