

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

Brandon Smith^{1,2}, Jackie Kendall¹, Craig Pelissier^{1,2}, Shelley Stover³

NASA Goddard Space Flight Center 3. NASA Langlev 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

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



Deep Learning Academy Advisors





Craig Pelissier Brandon Smith

Shelley Stover

Academy Cohort #1







Joaquin Chaves

Carlos Cruz Zach Fasnacht Ben Scarino

For Enterprise

🔘 deeplearning.ai

100% online courses

Flexible Schedule

Intermediate Level

Suggested 5 hours/week

Set and maintain flexible deadlines.

Some related experience required.

Approx. 4 months to complete

Start instantly and learn at your own schedule.



What do you want to learn?

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.

Machine Learning fundamentals training	Part 1: Coursework	Part 2: Projects	Learning opportunities to support needed ML	Growing SME pool that supports both customers &

Makhan Virdi Akira Sewnath

coursera

C @ coursera.org/specializations/deep-learning#courses

Explore 🗸

Deep Learning

Specialization

★ ★ ★ ★ ★ 4.8 217,448 ratings

5-course series

Break into Al

Browse > Data Science > Machine Learning

Deep Learning Specialization. Master Deep Learning, and

Andrew Ng +2 more instructors TOP INSTRUCTORS

Truman Wilson Justine Woo

Offered By

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.

				Q. Search			🚯 – I
Ke	y Insights	8	Takeaways	Project-A Chall	enge Posts	Files Wiki ~ +	∠7 C … @ Me
				=			
(ji)	Members Pending Requests Channels	settings Apps	1 more ~	Challenge Evaluation	n	Getting Started	
Teams	Search for members Q		🖧 Add member	Challenge Tips		Last edited: 4/2	
EI* Calendar	 Owners (4) 			Getting Started		Aws CodeCommit Setup	
ß	 Members and guests (13) 			1. AWS CodeCommit	S	 In SageMaker Studio, open the File Brows bar, and navigate to the root of your home 	er from the left side tool directory if you're not already
caño (Pa	候 Truman Wi Lead Research Scientist	$\langle \rangle$, Search	👔 – 🛛 🕅		there. 2. Open the Git tool (2nd icon from top in th	e left side tool bar).
Files	Carlos Cruz Chief Scientific Progra	Activity C	Part 2 Projects (Gener Posts Files 3 more ~ +	⑦ Team Q ¹	-	Amazon SageMaker Studio File Edit View You are not currently in a Git repository. To use Git, navig	Run Ke
	Shelley Sto Project Manager III	Chat Literans	Joaquin Chaves 2/28 2/34 PM Question for Craig Pelissier and Brandon Smith regarding projects: If modified the existing cells to provide my worked answer (i.e., the ne the answer to the two stated questions). Should I have recreated all modifications below instead of modifying existion one?	For the first exercise I cressary code changes, not the cells again with my	ise I inges, not ith my	locar repetitory, initialize a repeatory here, or clore an existing repeatory. Cpen the Androne or Hildler a Repetitory Creat the Androne or Creating Androne or	
	Makhan Vi Senior Research Scient	Et: Calendar	See less			Cone a reporter?	a liak balaw this will slave the
	Joaquin Ch Lead Research Scientist	S.	2 repres from srandon and Joaquin ¢/ Reply			"AmazonSageMaker-DLA-Project-A" folder	into your current directory.
		0	CP Graig Pelissier 3/1 11:25 AM	e medific This is what I did		<pre>https://git-codecommit.us-east-2.</pre>	amazonaws.com/vl/repos/Am

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

Deep Learning Academy

preprocessing.py

SSAIU_LSTM_c

SSAIU_LSTM_p

test_data.csv

test_sample.cs

test-data-input

🗅 train_data.p

Untitled.ipynb

25 days ago

2 minutes ago

2 hours ago

25 days ago

25 days ago

25 days ago

6 days ago

25 days ago

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*.

Amazon SageMaker Studio	File Edit	View Run Kernel Git Tabs Settings Help	
+ 🗈 🛨	G 🍕	SSAIU_LSTM_challenge.ipynb × SSAIU_LSTM_project.ipynb •	
I /		🖬 + 🛠 🗇 🖺 🕨 🔳 C Markdown 🗸 💁 🕺 🖉 git 2 vCPU + 4 GiB Python 3 (TensorFlow 2	2.3 Pyt
Name 🔺	Last Modified		
🖿 rawdata	25 days ago	4. Streamflow Prediction at CAMELS US basins	
🗅 challenge.p	a day ago		
🗅 conda	3 days ago	Now, we will do a worked example training on one drainage basin to predict daily streamfl	.ow di
default of credi	5 years ago	be similar, just more data. We will be using the CAMELS data which can be visualized by ex	ecution
geolocation.txt	4 days ago	better (/ml user edu/relutions/products/comple. For the LCTM model, we will be predicting	

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

framework learners can hit the ground running, immediately applying what

Both projects are also framed as a competition between teams, with *Kaggle*

complexities of interacting with the underlying infrastructure. With this

they've learned to the problem at hand rather than struggling with the

allowing learners to submit and track progress on key metrics.



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.



reasons. We will use environmental forcings for the input and a linear head laver to map from the hidden state to the observed discharge xt("geolocation.txt", dtype=['S8','float','float'],names=('id','lat','lon'))

r(row[0],'utf-8'): (row[2],row[1]) for row in geo] (1.1.1.projection=ccrs.PlateCarree()) :le("CAMELS US BASINS\nhttps://ral.ucar.edu/solutions/products/camels")

relevant elements

false negative

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.

details of AWS and SageMaker.



0

selected elements

Accordingly, the *Kaggle* competition uses a held-out dataset to rank teams via their macro F1score - 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 domainspecific 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.

n 3.7 CPU Optimized) Kernel: CPU: 0.00

harge. Multiple basins will: the first cell in this

f the data go to

	DLA P1 Test								
	CSC 578 Final Project (#	Team Name	Notebook	Team Members	Score 😧	Entries	Last	
		1	Team X		99	0.77771	25	15h	
		2	Team HBG		99	0.74815	23	1d	
		3	Team SAMDA		99	0.73505	22	1d	
		4	Team STARSS		99	0.73486	25	15h	
ሮ	View Active Events	Q	Benchmark			0.64042			



Credit: Wikipedia

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.



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 expertdeveloped 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



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

