Workshop Summary

Empowering the Statistician with Spark, Machine Learning and Deep Learning

ASA Hawai‘i Chapter

Workshop sponsors

- *American Statistical Association (ASA) and its Hawai‘i Chapter*
- *University of Hawai‘i at Mānoa Office of the Vice Chancellor for Research*
- *UHM College of Tropical Agriculture & Human Resources (CTAHR)*

Background on the Workshop

As part of the Continuing Education program of the ASA, chapters may annually apply for a Traveling Course (workshop) to be held in their jurisdiction. In 2019 and 2020 the Hawai‘i Chapter applied for and was awarded **Empowering the Statistician with Spark, Machine Learning and Deep Learning**. With sponsorship from the UHM Office of the Vice Chancellor for Research and CTAHR, the workshop was held November 19-20, 2020 as an ASA-hosted Zoom Webinar.

November 19, 2020

- *Data Science*
- *Cloud Platform & Big Data*
- ***Hands-on** (Databricks Community Edition, R, Load Spark Dataframe, Python)*
- *Decision Tree*
- ***Hands-on** (R, Python)*

November 20, 2020

- *Feedforward Neural Network (FFNN)*
- *FFNN **Hands-on** (R, Python)*
- *Convolutional Neural Network (CNN)*
- *CNN **Hands-on** (R, Python)*
- *Recurrent Neural Network (RNN) (Introduction, Embedding)*
- *RNN **Hands-on** (R, Python, Tokenize and Pad)*

Workshop Instructors

Ming Li

Ming Li is a Research Scientist at Amazon and previously worked at Walmart and General Electric. With his background and experience in statistics, data science and machine learning, he has trained and mentored many individuals from a variety of backgrounds. Ming holds a Ph.D. in Statistics from Iowa State University.

Hui Lin

Hui Lin is a Quant Researcher at Google and previously worked at Netlify and DuPont. She is the co-founder of Central Iowa R User Group and blogger of https://scientistcafe.com/. She enjoys making analytics accessible to a broad audience and teaches data science tutorials and workshops for practitioners. Hui holds a Ph.D. in Statistics from Iowa State University.

Spark, Machine and Deep Learning Workshop Overview

Who was the workshop for?

The workshop was designed for audiences with a statistics education background to bridge the gap between traditional statisticians and data scientists. No software downloads or installations were needed; all hands-on sessions were done through an internet browser (Chrome or Firefox) within the Databricks free cloud environment.

What was the workshop about?

Apache Spark is an open-source distributed engine for querying, processing and modeling big data. In the workshop participants learned how to leverage Spark and R/Python to process and model big data with a common machine learning algorithm. The workshop was a combination of lecture and hands-on sessions where participants set up community accounts in Databricks and loaded content prepared by the instructors. This allows participants to revisit their Databricks environment after the workshop, for exploration and in depth analysis of their own data.

The workshop learning objectives were to:

- Get familiar with Spark, a cloud-based big data platforms, for data preprocessing and machine learning model development
- Understand what data scientists “in the wild” are actually doing to better prepare statisticians to be successful data scientists in the future
- Learn how to leverage Spark to process and model big data with a common machine learning algorithm.
- Get familiar with basic deep learning methods and do hands-on exercises in R/Python.

The course syllabus and resources are online at https://course2020.scientistcafe.com

Who came to the workshop?

44 individuals participated in the workshop. Attendees participated from 15 different Hawai‘i zip codes as well as from overseas. 23 individuals (52%) participated on both days; 17 on DAY 1 only (39%); and 4 on DAY 2 only (9%). The majority of participants (86%) had UH/UHM/RCUH ties, 23% had CTAHR ties, and 23% had ASA ties. Feedback was obtained each day via an online Mentimeter form resulting in 29 responses over the 2 days of the course. 50% of participants provided feedback on DAY 1 and 33% on DAY 2. A summary of feedback received follows.

Spark, Machine and Deep Learning Workshop Summary

Participant Demographics

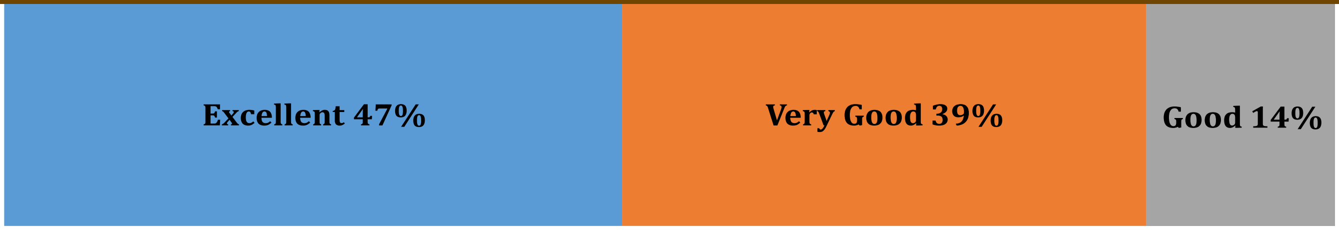
Gender	%
Female	55
Male	45

Age	%
Under 30	31
30-39	21
40-49	17
50 and over	31

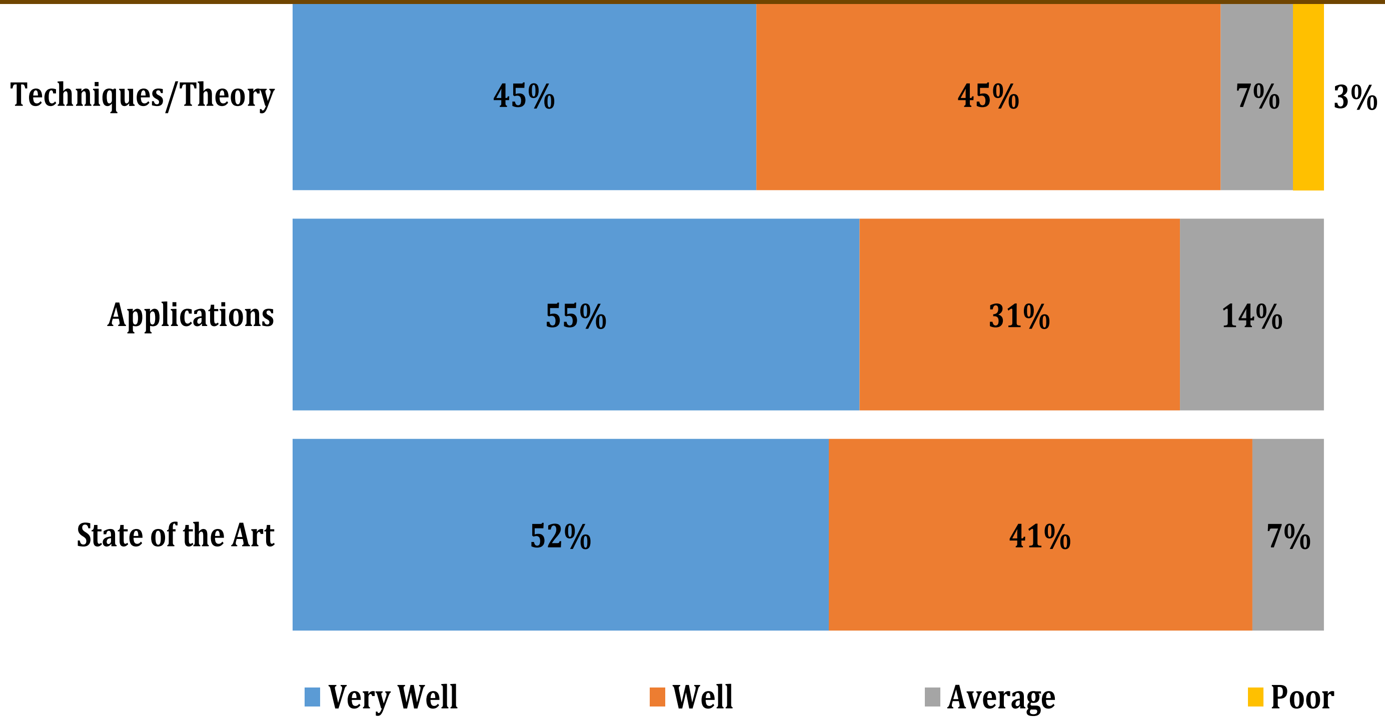
Education	%
< Bachelor’s	0
Bachelor’s	17
Master’s	31
Doctorate	52

Employer	%
Academic	70
Government	7
Private	13
Other	10

Overall, how would you rate the course?



How well were the following items covered?



Would you attend a course in the future?

Yes	Probably	Maybe	Probably Not	No
75%	14%	11%	0%	0%

Spark, Machine and Deep Learning Workshop Summary

Additional Participant Feedback

Excellent teaching and course materials - Full of rich information	19
Great opportunity to learn about a complicated topic and big data platforms - Eye opening!	14
More time needed for hands-on portions of the workshop	10
Include more practical use cases with real datasets	7

This was a really good introduction to many terms and concepts I've often only wondered about.

A lot of us don't have the luxury of taking years and years to concentrate our attention on learning this content, so I really appreciate this offering and particularly the experience and backgrounds of the presenters.

I appreciate the way instructors are explaining in very simple way with details.

Very good explanations making everything clear and simple, but detailed at the same time.

I enjoyed the Spark hands-on modules and wish that was longer with more variations.

Maybe spend less time on theory and more time discussing applications.

MAHALO TO WORKSHOP SPONSORS!

44 participated in the workshop

UHM / UH / RCUH	CTAHR	ASA	Other
38 (86%)	10 (23%)	10 (23%)	2 (5%)

83 registered & were emailed a link to resources

UHM / UH / RCUH	CTAHR	ASA	Other
67 (81%)	25 (30%)	14 (17%)	3 (4%)

Link to Resources from the Workshop

Nov 19 and Nov 20 workshop materials & edited video recordings will be available in this folder (starting in February 2021)

American Statistical Association Hawai‘i Chapter Email List

Contact us at amstatassochi@gmail.com and we'll add you to our email list :-)