



Data Science Upskilling Workshop

Oren Livne and Jiangang Hao
Educational Testing Service

NCME 2023 Training Workshop – 4/12/2023

About Us



Oren Livne (olivne@ets.org)

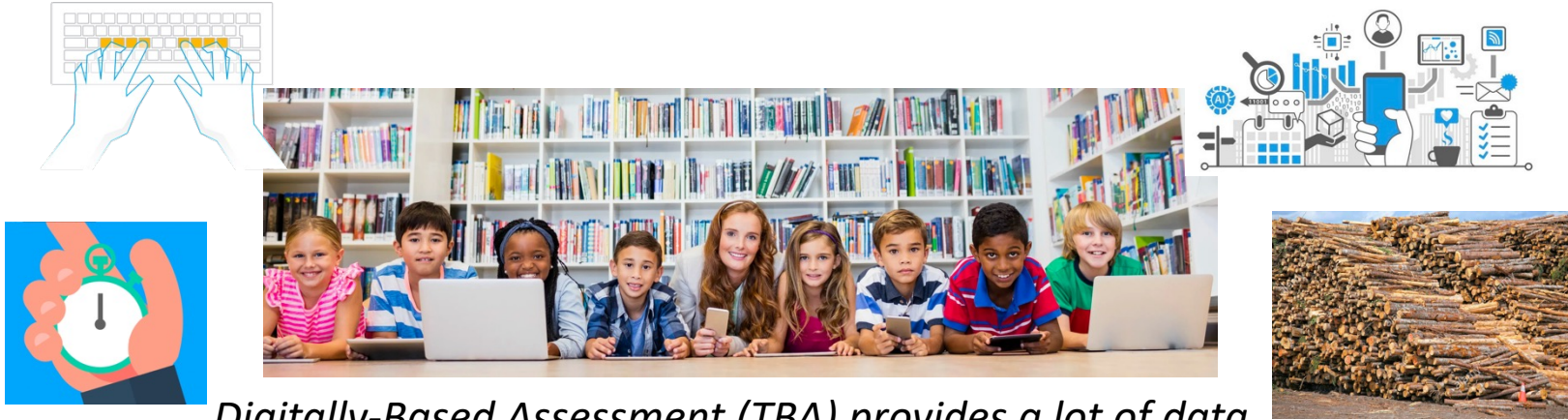


Jiangang Hao (jhao@ets.org)

Welcome to the Data Science Upskilling Workshop!

- We assume you performed the setup beforehand (we will also briefly go through the setup)...
- Materials is presented on slides and jupyter notebooks.
- Follow along, and *try the code yourself*. We added optional homework exercises you can do later.
- Please ask any questions; reserve them to the end of each section to save time.
- Have fun!

An Urgent Challenge in Assessment



Digitally-Based Assessment (TBA) provides a lot of data about learners, but how do we make sense of it?

Train Psychometricians & Data Scientists



Engineering Challenges



Statistical & Computational Challenges



Increase measurement precision
Identify response patterns
Identify new constructs

ETS Data Science Academy

Syllabus

IN A NUTSHELL

PYTHON ECOSYSTEM

Anaconda, Jupyter, Python

Week 1

DATA WRANGLING

Json, Xml, Pandas, Validation

Week 2

SPEEDING UP COMPUTATION

Jit, parallelization, GPU

Week 3

INTERACTIVE VISUALIZATION

Matplotlib, plotly, ipywidgets

Week 3, 4

DATA MODEL & PROCESSING

Relational, document, batch, stream

Week 5

VERSION CONTROL

Hash, Git, Github

Week 5

MACHINE LEARNING

Supervised, Unsupervised

Week 6

AI AND DEEP LEARNING

DNN, RNN, CNN

Week 7

CLOUD COMPUTING

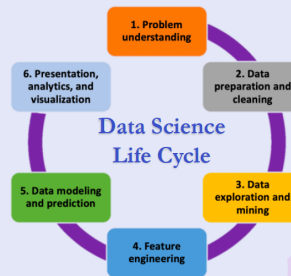
AWS, SageMaker, Lambda, S3

Week 8

Capstone Projects

Week 9 – 10

Make sense of the process data
from remote testing



UPSKILLING

Learning Goals

We aim at introducing core data science techniques and machine learning/AI methods in the context of digitally based learning and assessments to help psychometrics practitioners better prepared for the increasingly big and complex data!

Instructors

Oren Livne: olivne@ets.org

Jiangang Hao: jhao@ets.org

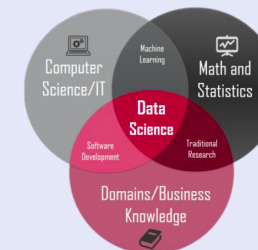
Other guest speakers

Data Science Academy

FOR PSYCHOMETRICIANS & DATA ANALYSTS

SPONSORED BY: PSDS

2 - 4 HOURS/WEEK FOR 10 WEEKS



LECTURES & HANDS-ON EXERCISES

CAPSTONE PROJECTS

BASED ON REAL DATA EXAMPLES AT ETS

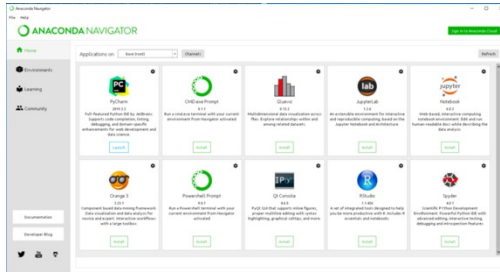


Season 1: July - October 2020

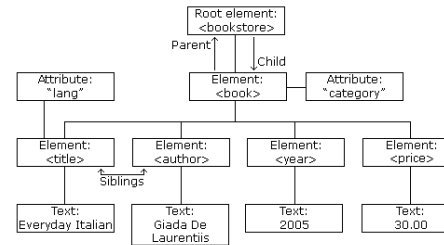


Learning Goals

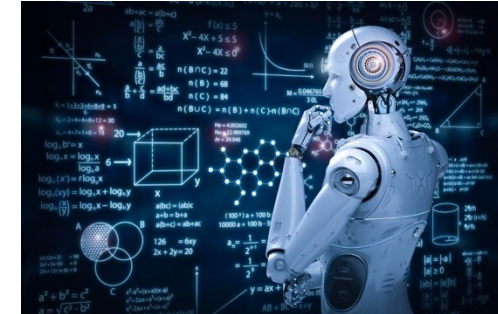
Develop participants' basic working skills for manipulating complex processed data.



Work within a development environment



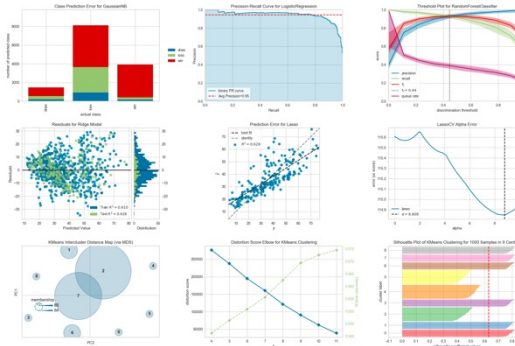
Wrangle XML & JSON data



Understand machine learning basics and apply them to solve real problems



Write high-quality Python code



Create data visualizations



Environment Setup

Development Environment is there to make your life easier

1. Install anaconda

<https://www.anaconda.com/products/individual>



NOTE: Conda's default package distribution is no longer free for commercial use. You should use the conda-forge channel for free use.

Development Environment is there to make your life easier

1. Install anaconda.
2. Create notebook.
3. Manage virtual environments *within anaconda*.
4. Install and uninstall packages.