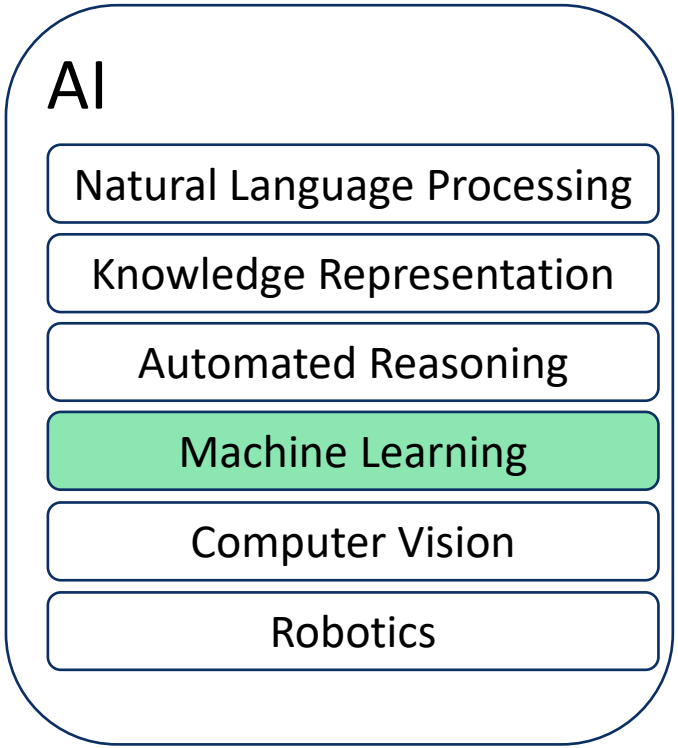
Machine Learning Overview

From Theory to Practice

Artificial Intelligence (AI) and Machine Learning (ML)

AI Taxonomy

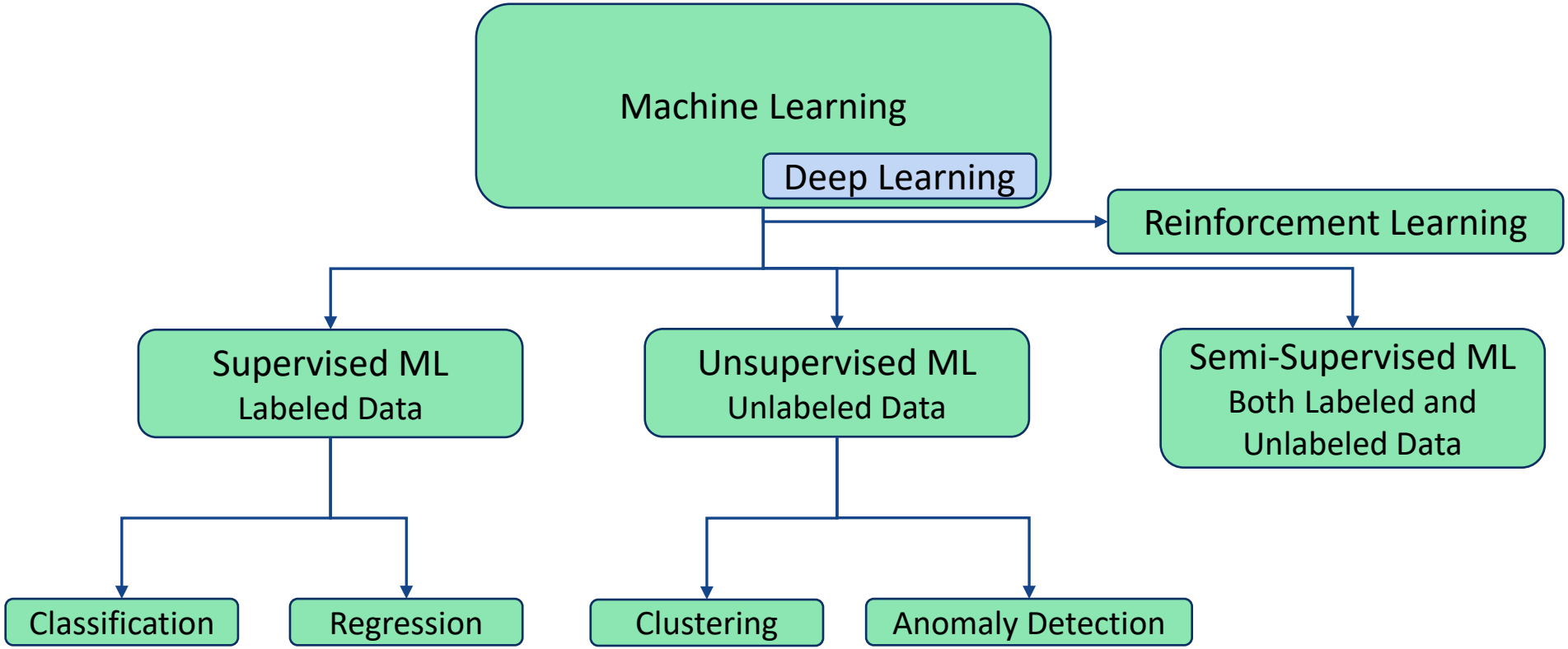


- ML as a subset of AI
- ML as the basis for other AI components
- **ML**: a subfield of artificial intelligence that gives computers the ability to learn without explicitly being programmed

https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

Russel and Norvig, 2010, *Artificial Intelligence: a Modern Approach*, https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf

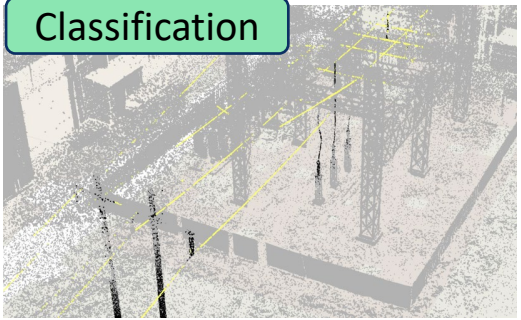
ML Ontology


Machine Learning
Deep Learning
Reinforcement Learning
Supervised ML
Labeled Data
Unsupervised ML
Unlabeled Data
Semi-Supervised ML
Both Labeled and
Unlabeled Data
Classification
Regression
Clustering
Anomaly Detection

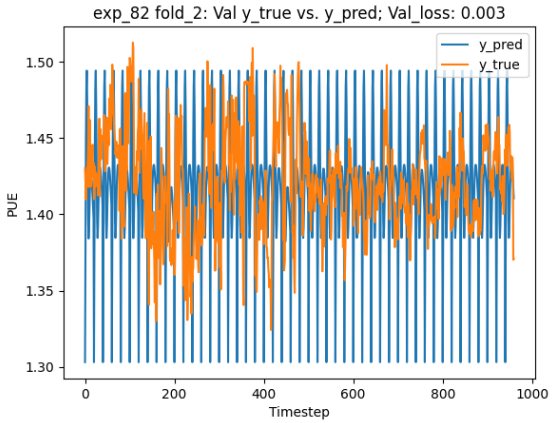
Supervised ML
Labeled Data

Unsupervised ML
Unlabeled Data

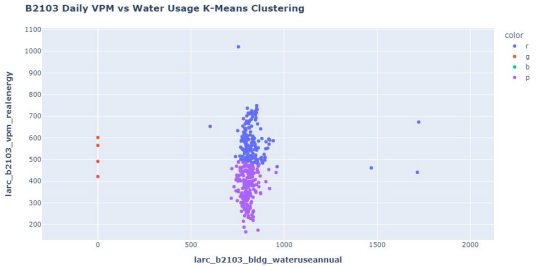
Classification



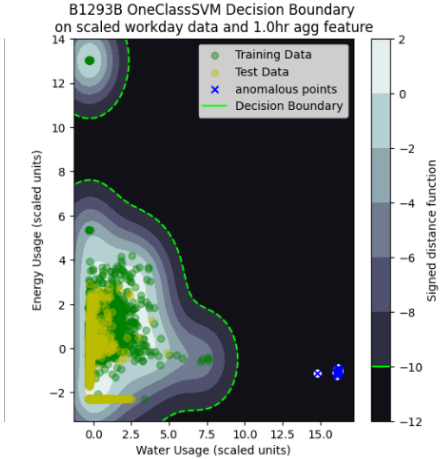
Regression



Clustering



Anomaly Detection



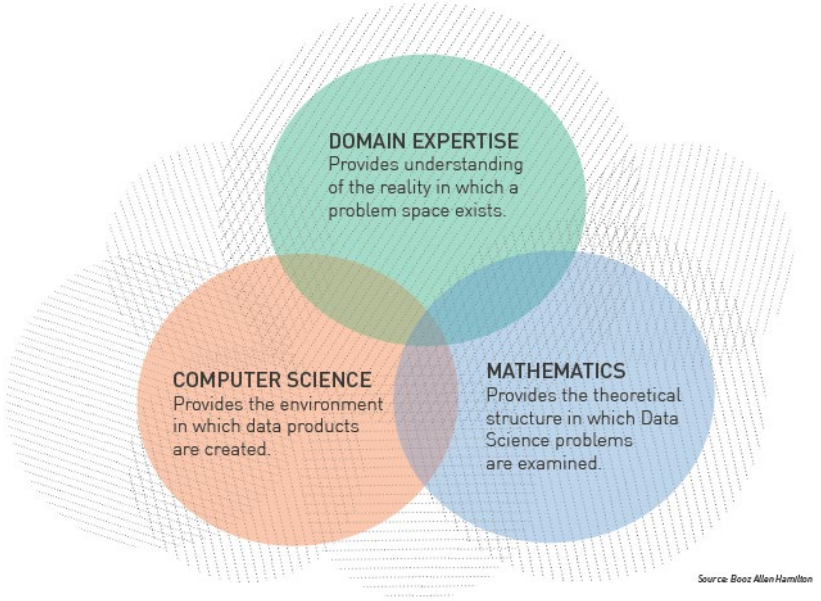
Big Data

- Volume: size of data
- Velocity: speed of transmission, reception
- Variety: different types of data, different sources
- Variability: changes in dataset over time, non-stationarity
- Veracity: data accuracy / quality

https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf
https://ntrs.nasa.gov/api/citations/20170000676/downloads/20170000676.pdf

Getting Started

- Define clear objectives for ML
  - Much easier for supervised ML
  - Performance metrics
- Engage key stakeholders
  - Energy management
  - Operations and maintenance
  - Procurement
  - IT
  - Data owners
- Assemble data science team
  - Domain / subject matter expert
  - Computer Scientist
  - Mathematician / Scientist
- Establish team battle rhythm



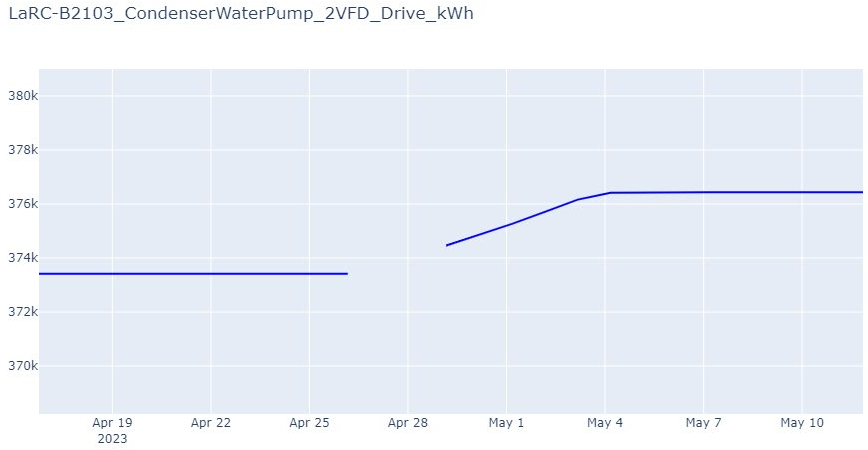
https://www.boozallen.com/s/insight/publication/field-guide-to-data-science.html

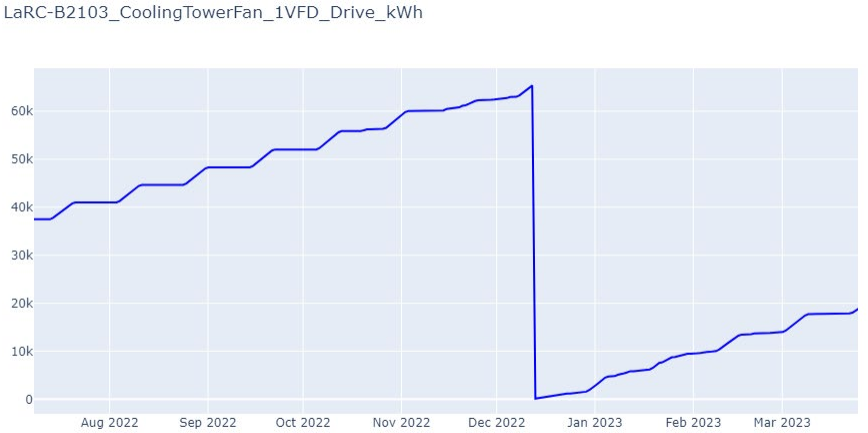
Example Project: Datacenter Power Usage Effectiveness (PUE)

- Katherine Johnson Computational Research Facility
- Well-instrumented for sensor collection
- Goal: build an automated framework for PUE measurement and prediction

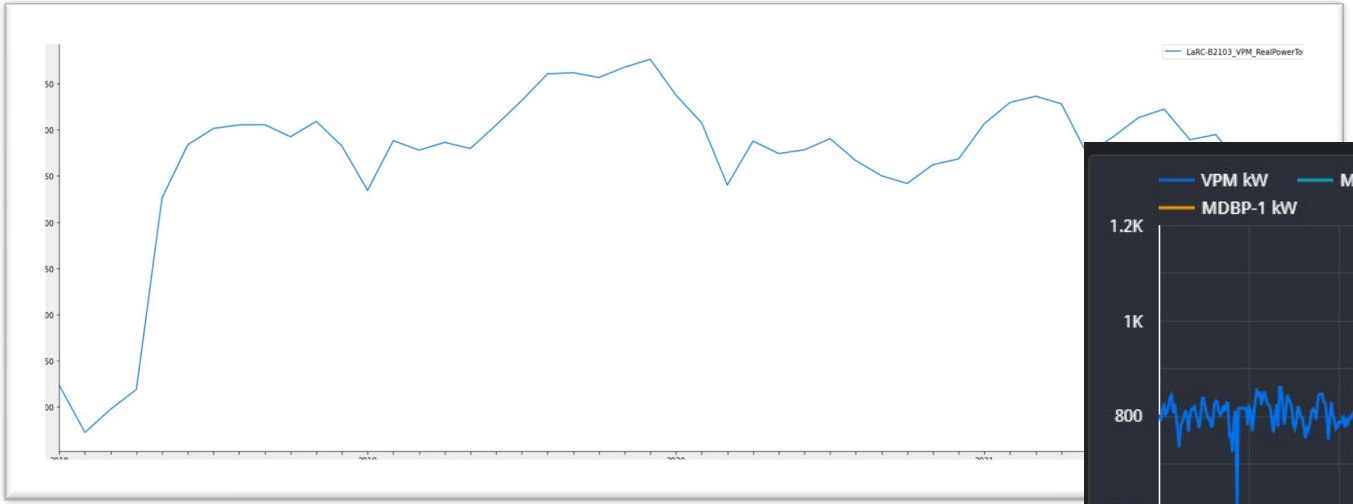
Exploratory Data Analysis (EDA)

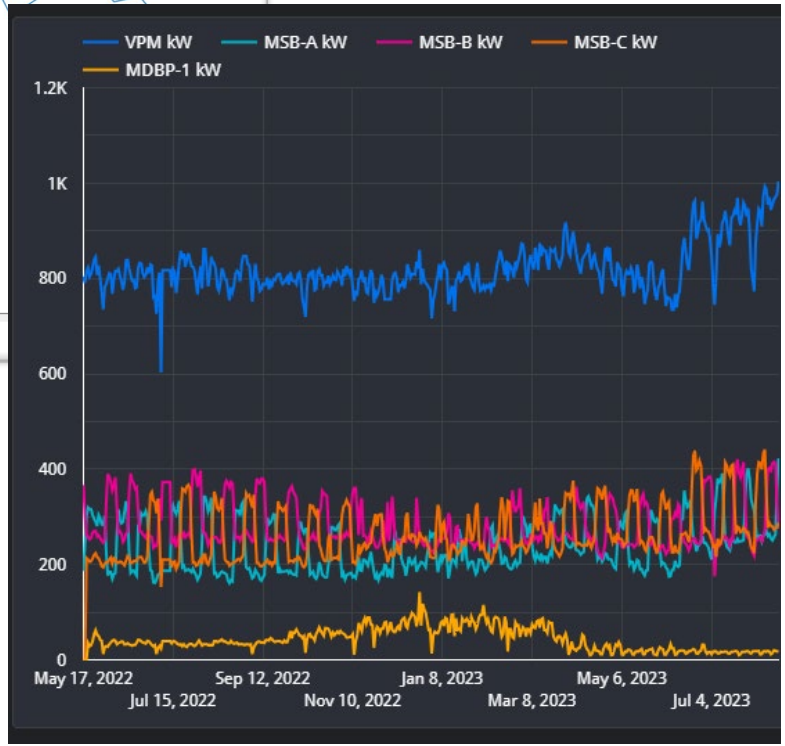
- Descriptive statistics
- Data visualization
- Data Discrepancies
  - NaN values (missing)
  - Sensor dropout
  - Meter rollover
- Current vs desired state of data





EDA to Dashboarding

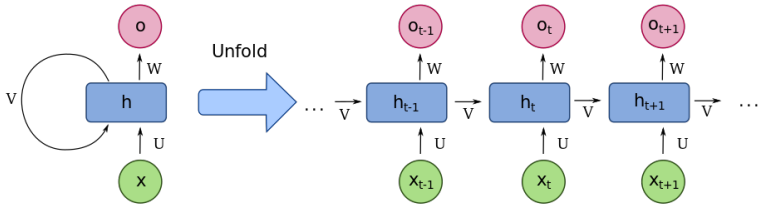




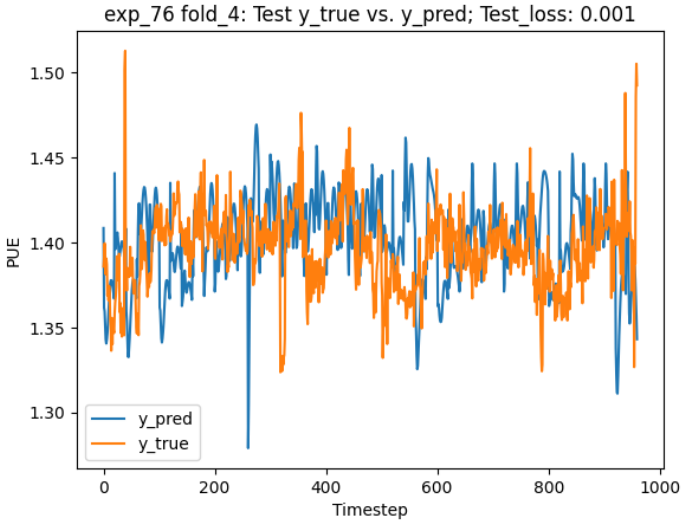
- Data acquisition
- Data historization: historize 60 additional submeters
- Data engineering: handle abnormalities
- Dashboarding

Machine Learning Approach

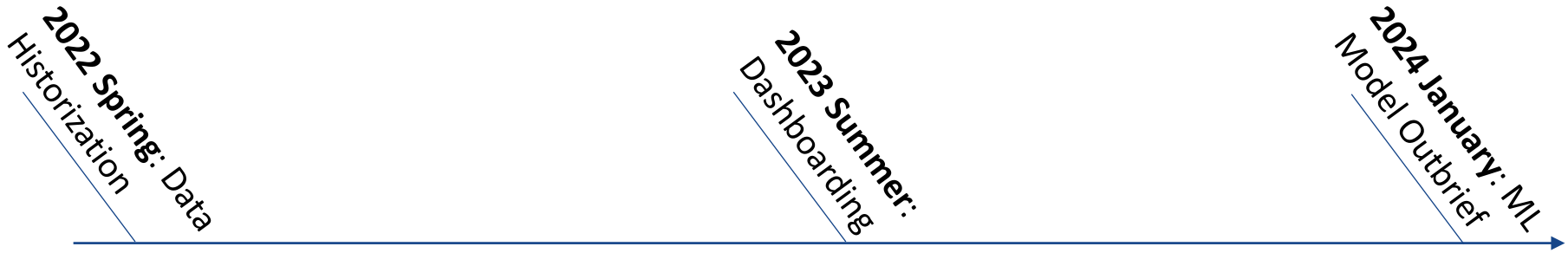
- Data engineer open-source weather data as ML model another input
- Utilize recurrent neural networks for time series prediction
- Visualize results
- Share



https://upload.wikimedia.org/wikipedia/commons/b/b5/Recurrent_neural_network_unfold.svg



Timeline and Summary



- Successful ML requires discipline and patience
- Honest evaluation of model performance
- Key stakeholder engagement

Thanks

- Mark Pugh, OCIO Cloud and Computing Services (CCS)
- NASA Langley Integrated Operations Center Team
- NASA LaRC Energy Efficiency Team
- NASA GIS Team
- NASA Interns
  - Heather Anderson
  - Julieanna Bacon
  - Samuel Briggs