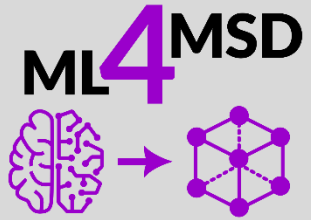


ME 5374-ST



# Machine Learning for Materials Science and Discovery

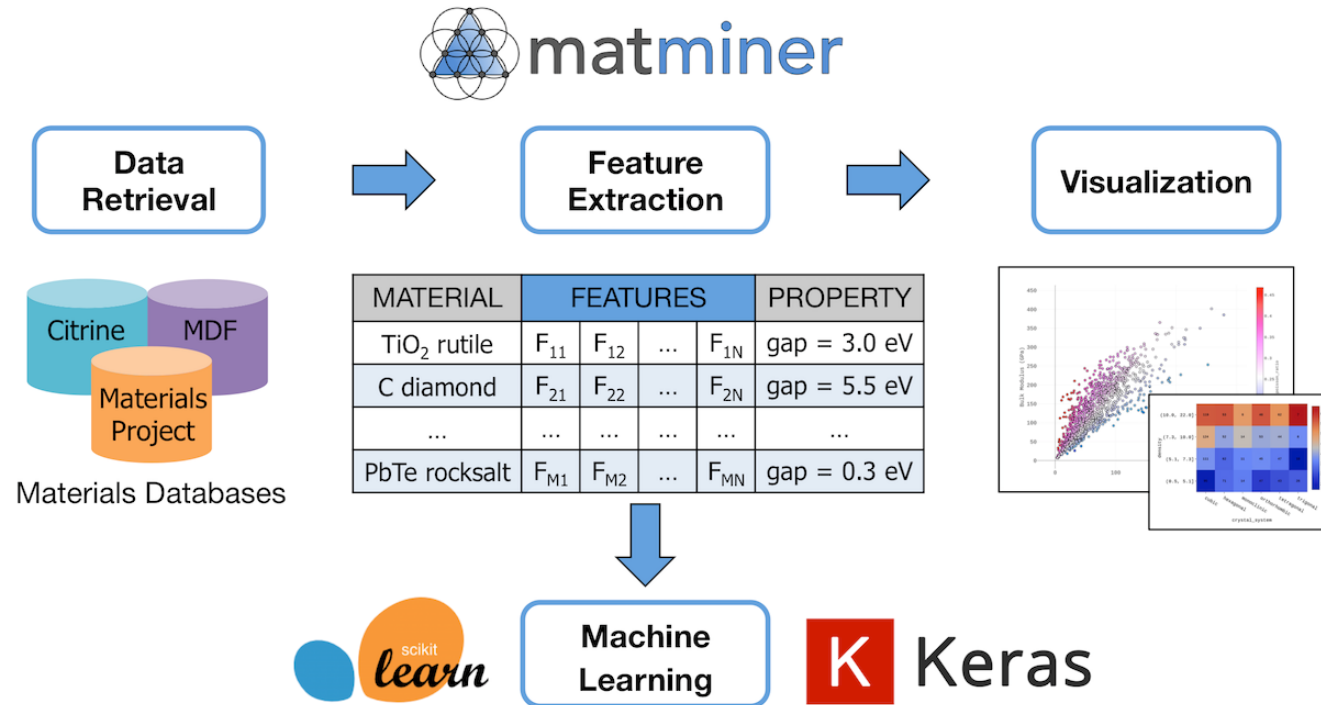
Fall 2025

Asst. Prof. Peter Schindler

Lecture 11 – Demonstration of Full ML Pipeline with a Materials Science Database

- Matminer, Matbench, and Scikit-Learn
- Overview of Materials Science ML Pipeline

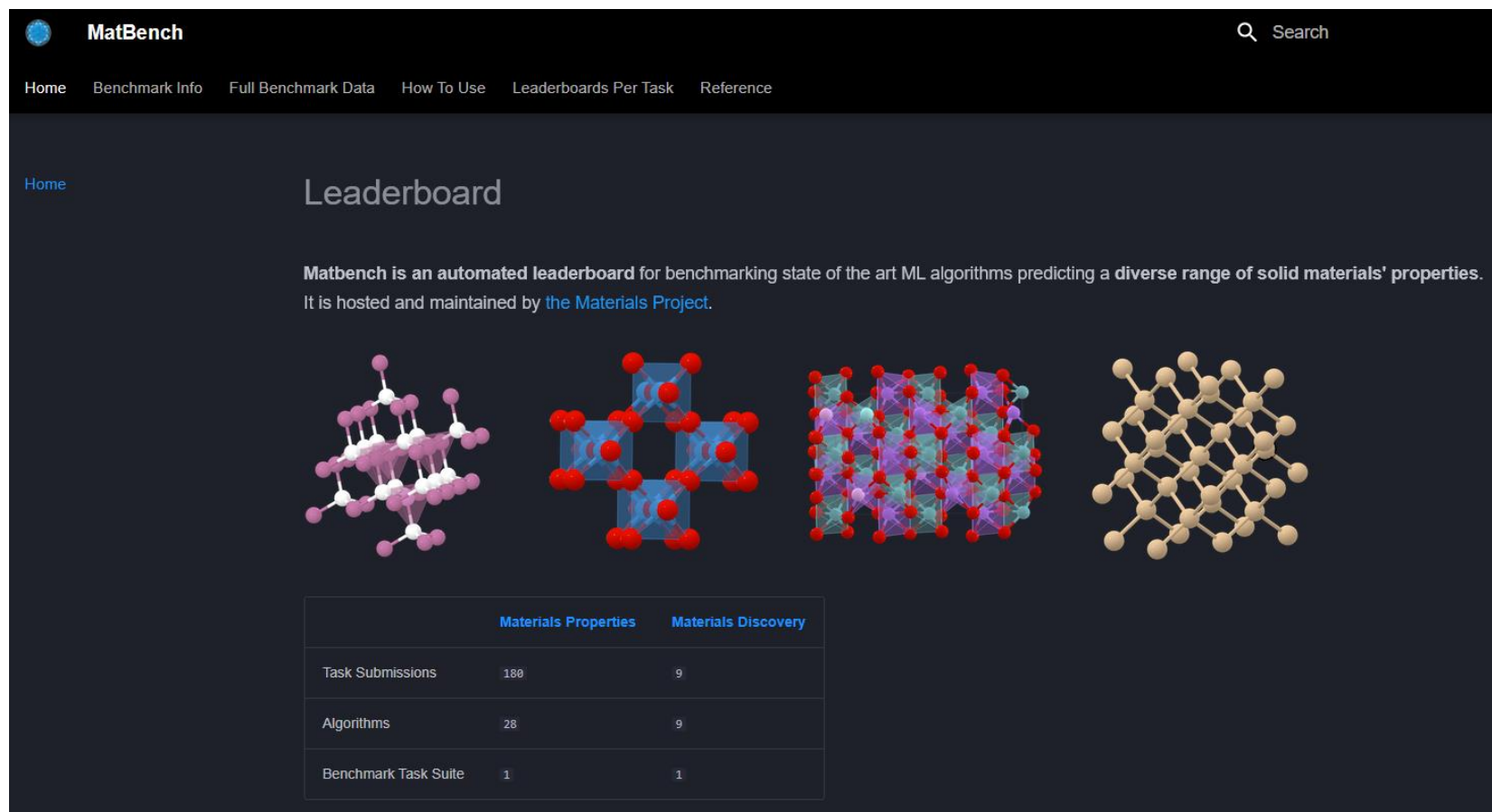
# Matminer Package



Datasets available: [https://hackingmaterials.lbl.gov/matminer/dataset\\_summary.html](https://hackingmaterials.lbl.gov/matminer/dataset_summary.html)

Featurizers available: [https://hackingmaterials.lbl.gov/matminer/featurizer\\_summary.html](https://hackingmaterials.lbl.gov/matminer/featurizer_summary.html)

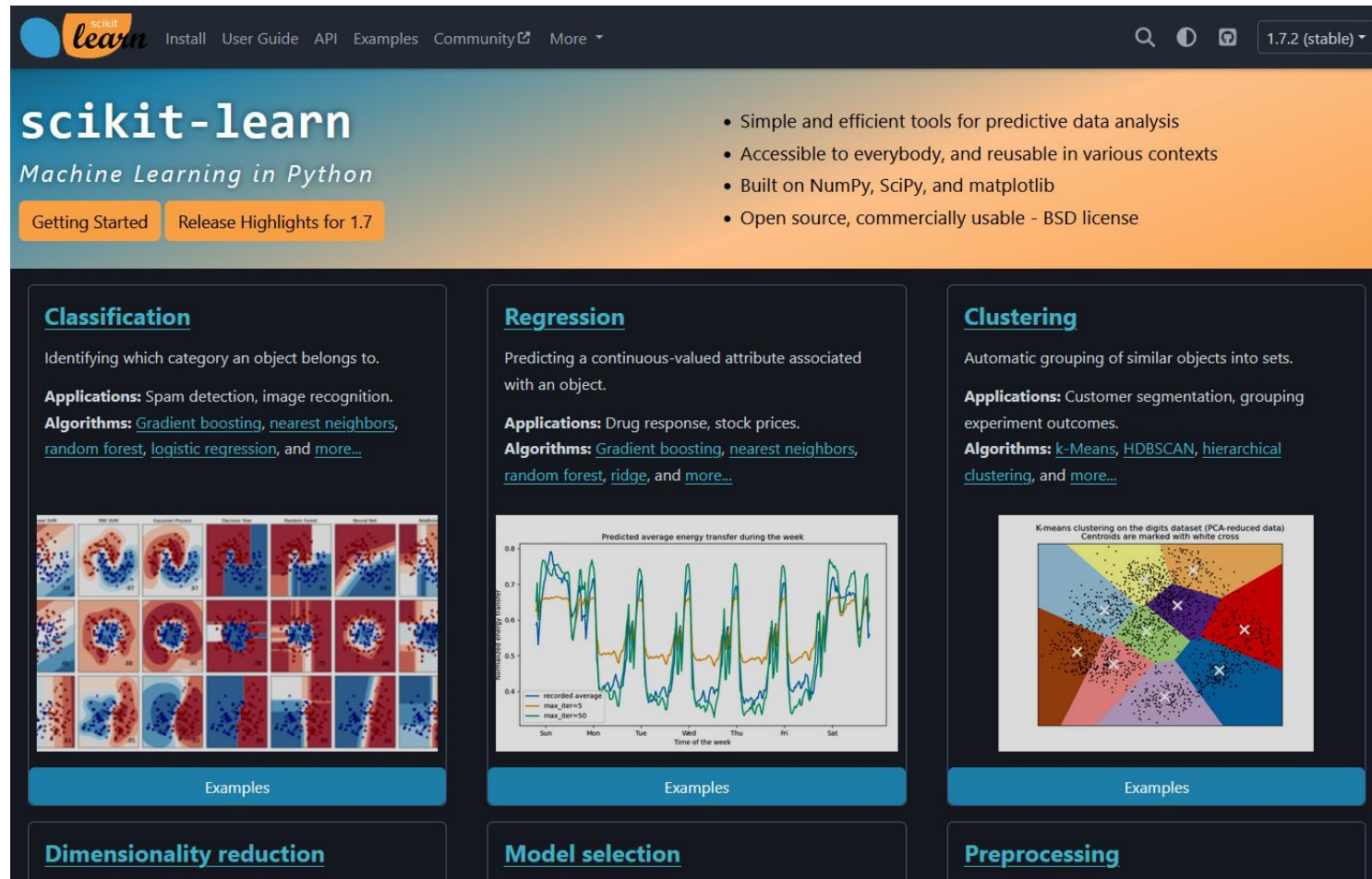
# Matbench Leaderboard



The screenshot shows the MatBench website interface. At the top is a navigation bar with links: Home, Benchmark Info, Full Benchmark Data, How To Use, Leaderboards Per Task, and Reference. A search bar is located on the right. The main content area has a 'Home' link and a 'Leaderboard' title. Below the title, a paragraph states: 'Matbench is an automated leaderboard for benchmarking state of the art ML algorithms predicting a diverse range of solid materials' properties. It is hosted and maintained by [the Materials Project](#).' Below this text are four 3D molecular models of different materials. At the bottom, there is a table with two columns: 'Materials Properties' and 'Materials Discovery'.

	Materials Properties	Materials Discovery
Task Submissions	180	9
Algorithms	28	9
Benchmark Task Suite	1	1

# Scikit-Learn: Package for Shallow ML



The screenshot shows the Scikit-Learn website homepage. At the top, there's a navigation bar with links for 'Install', 'User Guide', 'API', 'Examples', 'Community', and 'More'. The main header features the 'scikit-learn' logo and the tagline 'Machine Learning in Python'. Below this, there are buttons for 'Getting Started' and 'Release Highlights for 1.7'. A list of features is displayed: 'Simple and efficient tools for predictive data analysis', 'Accessible to everybody, and reusable in various contexts', 'Built on NumPy, SciPy, and matplotlib', and 'Open source, commercially usable - BSD license'. The page is divided into sections for 'Classification', 'Regression', and 'Clustering', each with a brief description, applications, and algorithms. Each section includes an 'Examples' button and a corresponding figure. The 'Classification' section shows a grid of 12 scatter plots. The 'Regression' section shows a line plot of 'Predicted average energy transfer during the week'. The 'Clustering' section shows a scatter plot of 'K-means clustering on the digits dataset (PCA-reduced data)' with centroids marked by white crosses. At the bottom, there are buttons for 'Dimensionality reduction', 'Model selection', and 'Preprocessing'.

**scikit-learn**  
*Machine Learning in Python*

Getting Started Release Highlights for 1.7

- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

### Classification

Identifying which category an object belongs to.

**Applications:** Spam detection, image recognition.

**Algorithms:** [Gradient boosting](#), [nearest neighbors](#), [random forest](#), [logistic regression](#), and [more...](#)

Examples

### Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, stock prices.

**Algorithms:** [Gradient boosting](#), [nearest neighbors](#), [random forest](#), [ridge](#), and [more...](#)

Examples

### Clustering

Automatic grouping of similar objects into sets.

**Applications:** Customer segmentation, grouping experiment outcomes.

**Algorithms:** [k-Means](#), [HDBSCAN](#), [hierarchical clustering](#), and [more...](#)

Examples

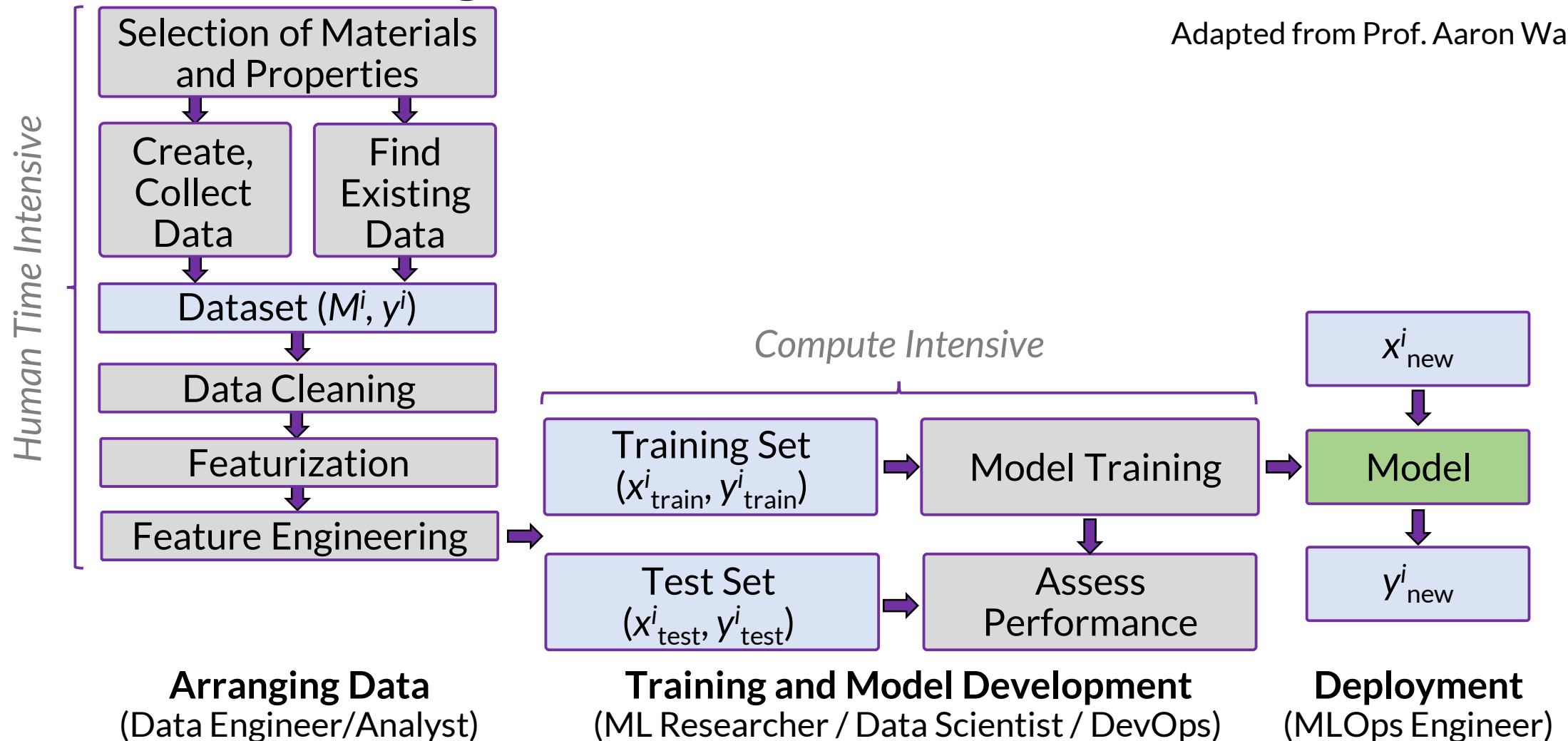
### Dimensionality reduction

### Model selection

### Preprocessing

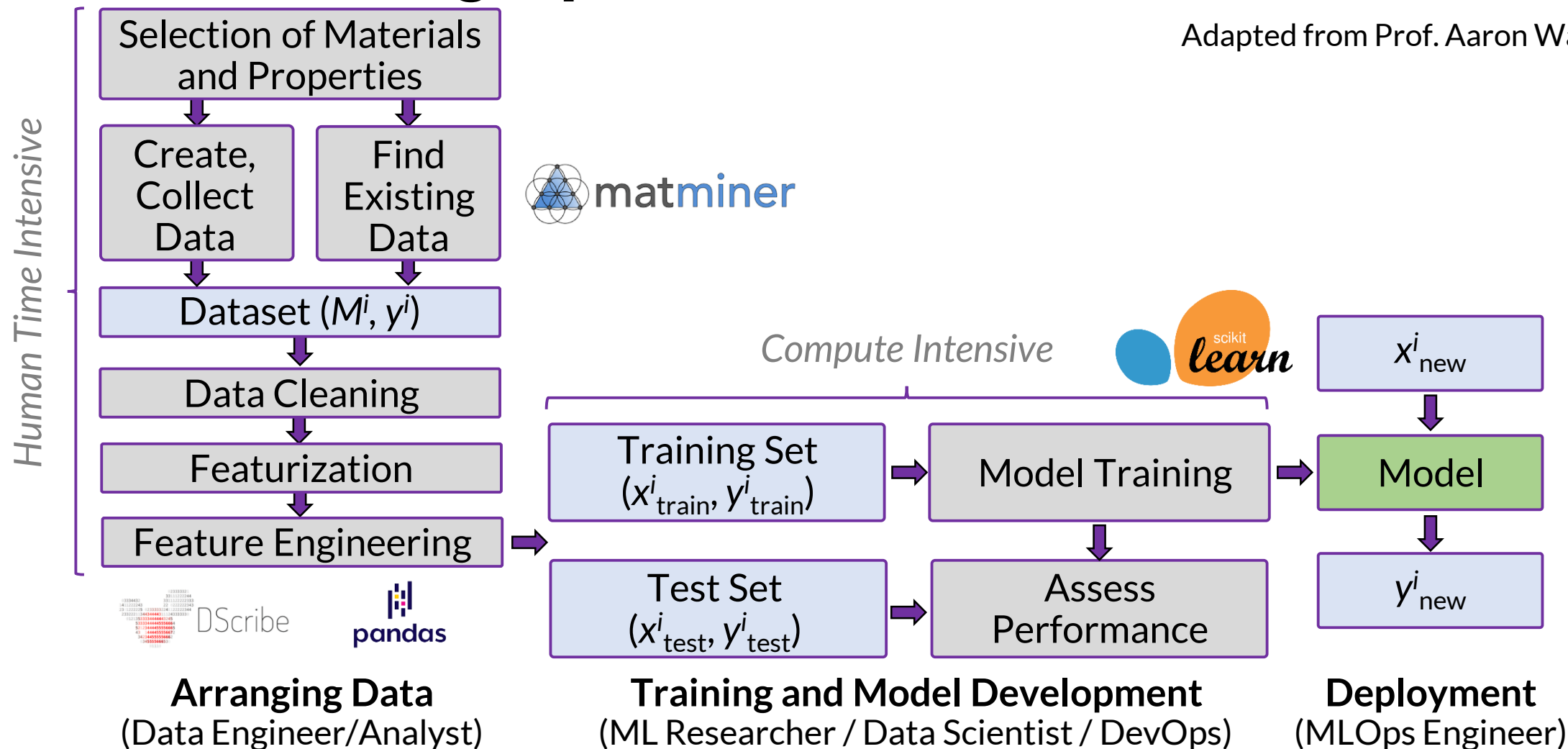
# Machine Learning Pipeline

Adapted from Prof. Aaron Walsh

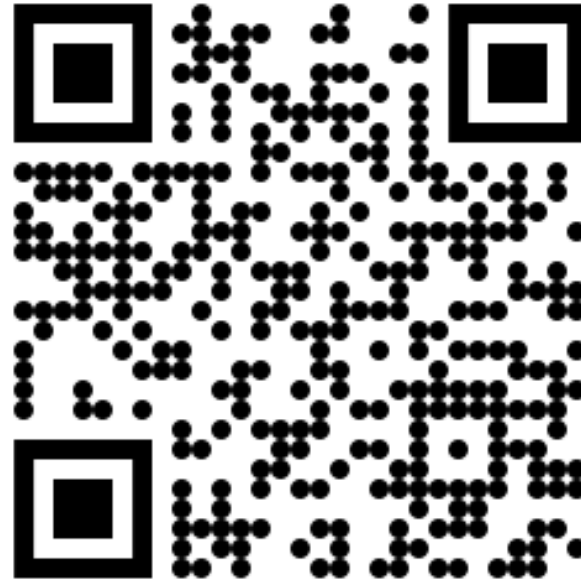


# Machine Learning Pipeline

Adapted from Prof. Aaron Walsh



# Lecture Feedback



Please, scan the QR code and take a minute to let me know how the lecture was and mention any **feedback/questions**

This form is **anonymous!**