An Introduction to Machine Learning for Materials Science: A Basic Workflow for Predicting Materials Properties

Ben Afflerbach

University of Wisconsin – Madison

5/12/2021

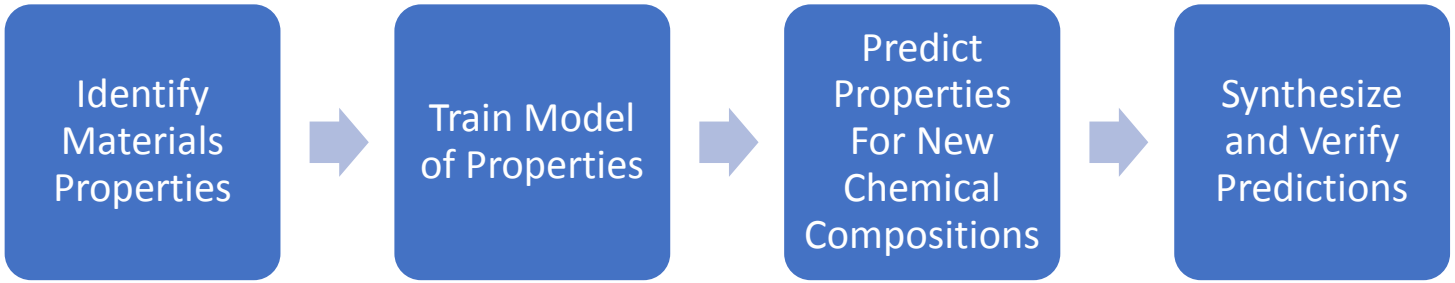
Summary

What is Machine Learning?

- Machine learning is a tool that finds patterns in large datasets that might be hard to discover otherwise

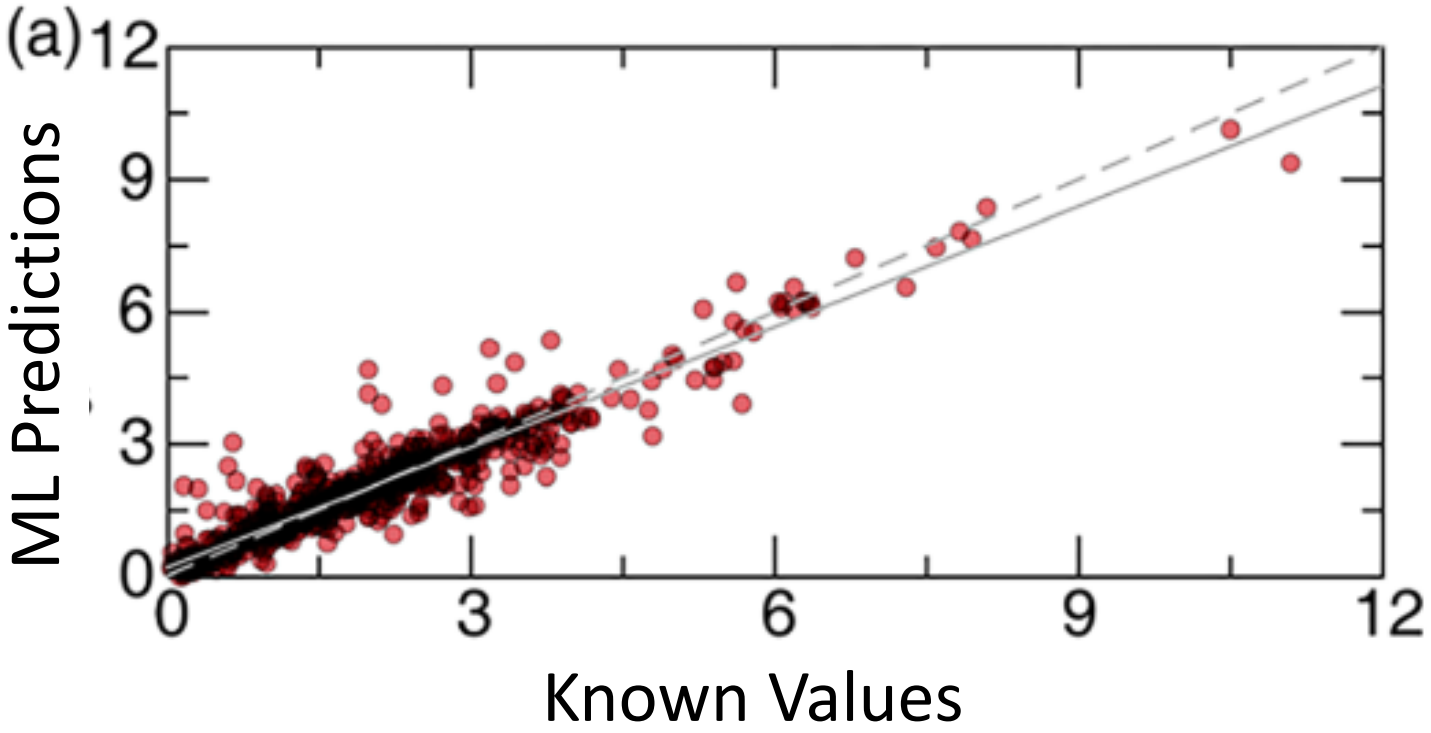
How can we use it for Materials Science?

- It can be included in existing materials science workflows to accelerate research, materials design, and materials discovery



An Application: Predict a Materials Property

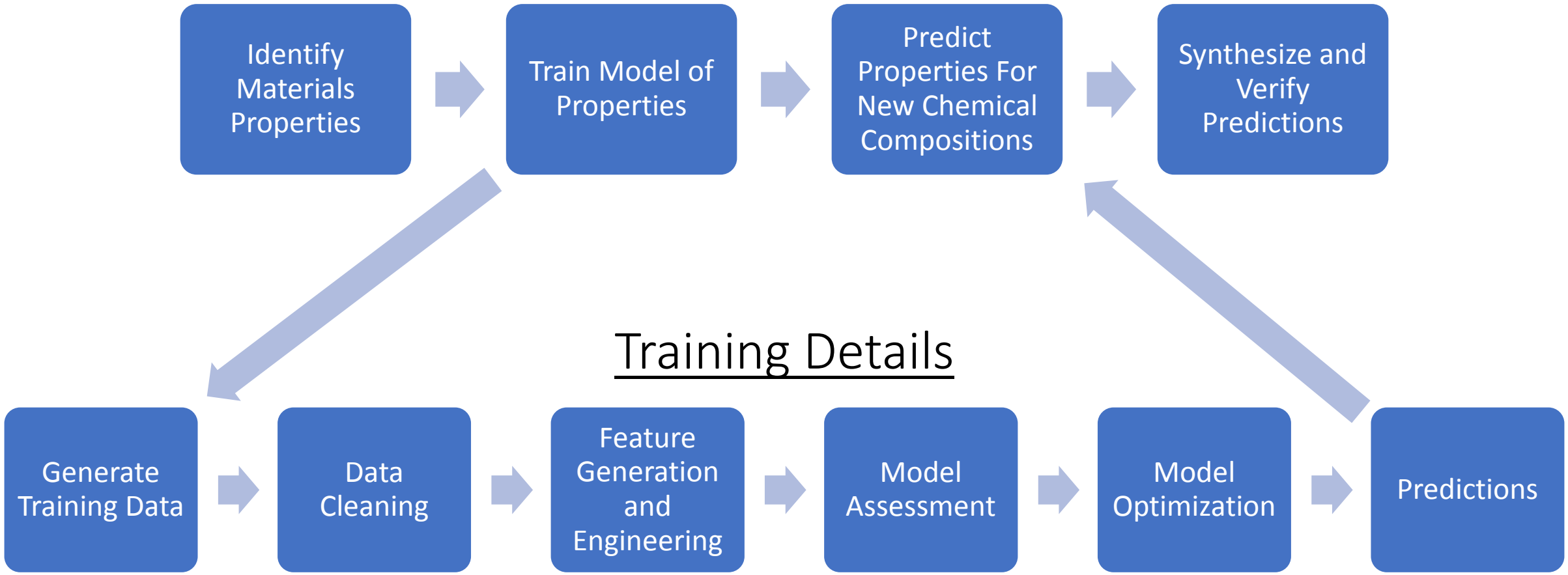
Bandgaps in Semiconductors



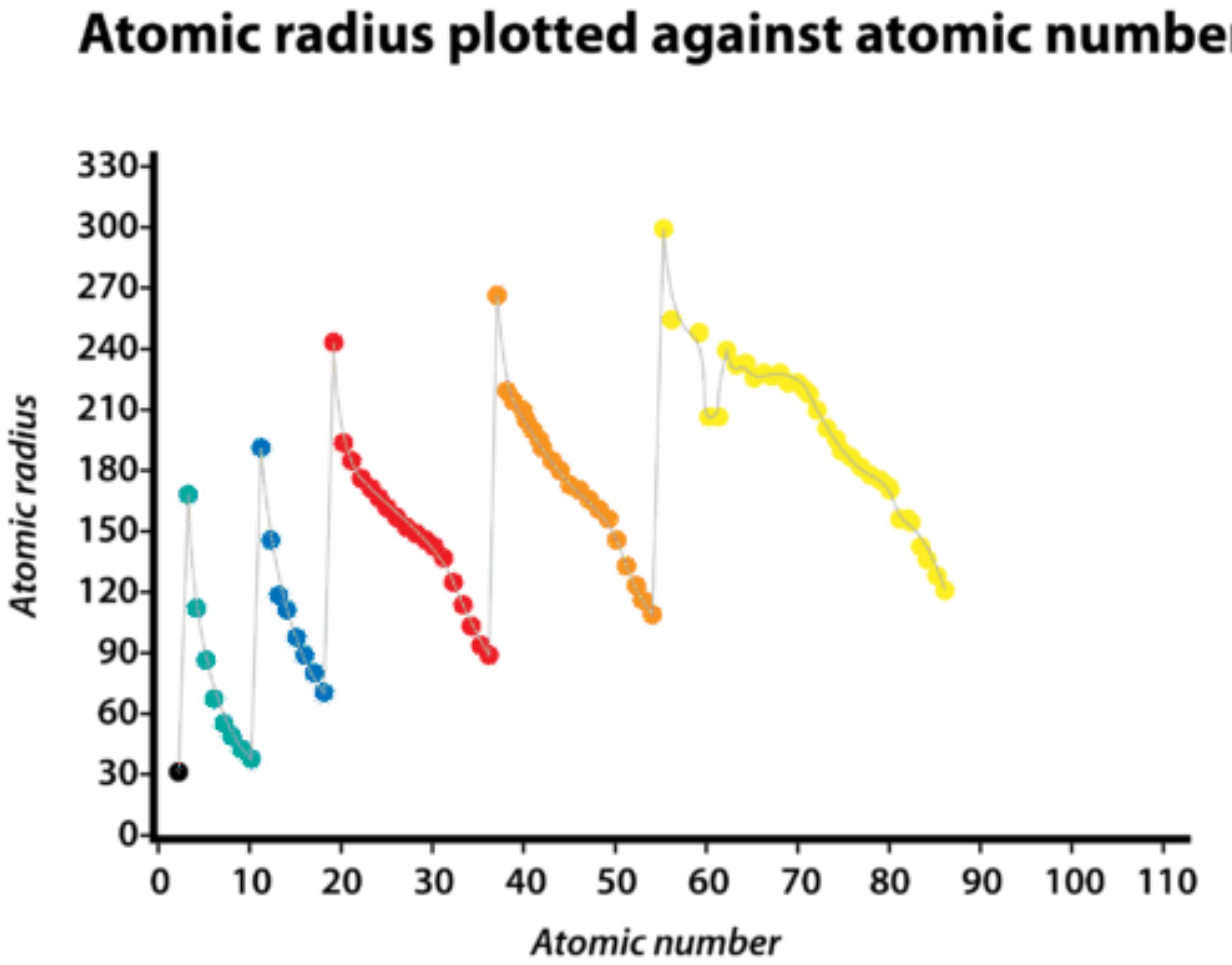
Machine learning prediction here is obtained from only properties of the elements in the material!

Y. Zhuo, et al. *The Journal of Physical Chemistry Letters* 2018 *9* (7), 1668-1673. DOI: 10.1021/acs.jpclett.8b00124

A Basic Materials Design Workflow



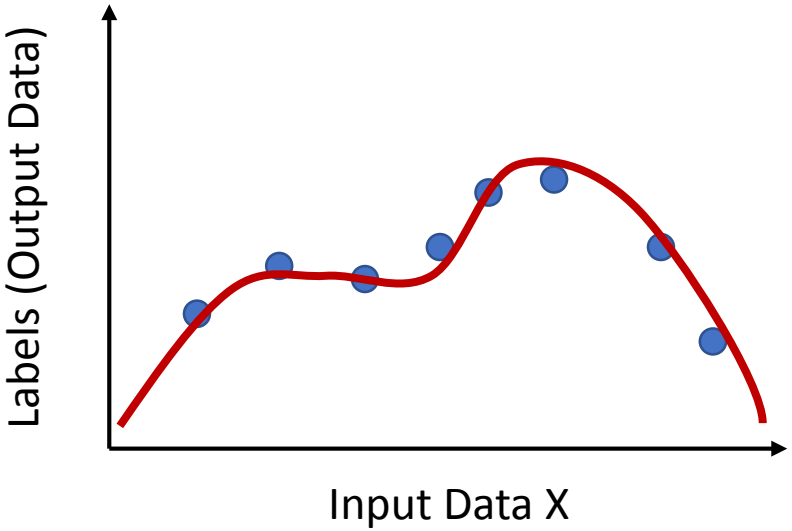
Machine Learning is Pattern Matching



https://chem.libretexts.org/Bookshelves/Introductory_Chemistry (accessed May, 2020)

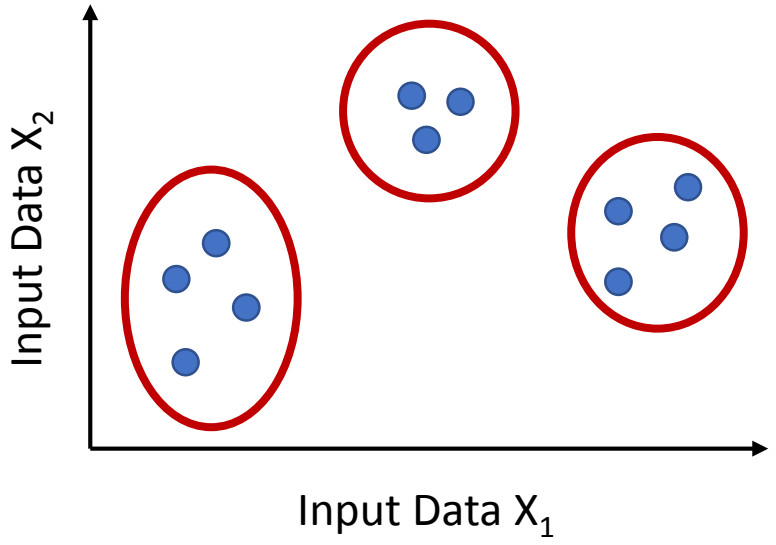
Key Distinction in ML

Supervised Learning



Find a **Function** that represents the data

Unsupervised Learning



Find **Structure** in the data
No Labels

Key Distinction in ML

Regression

- Bandgap
- Fatigue Strength
- Transformation Temp.

Classification

- Insulator / Conductor
- Brittle Failure / Ductile Failure
- Is Shape Memory Alloy / Not Shape Memory Alloy

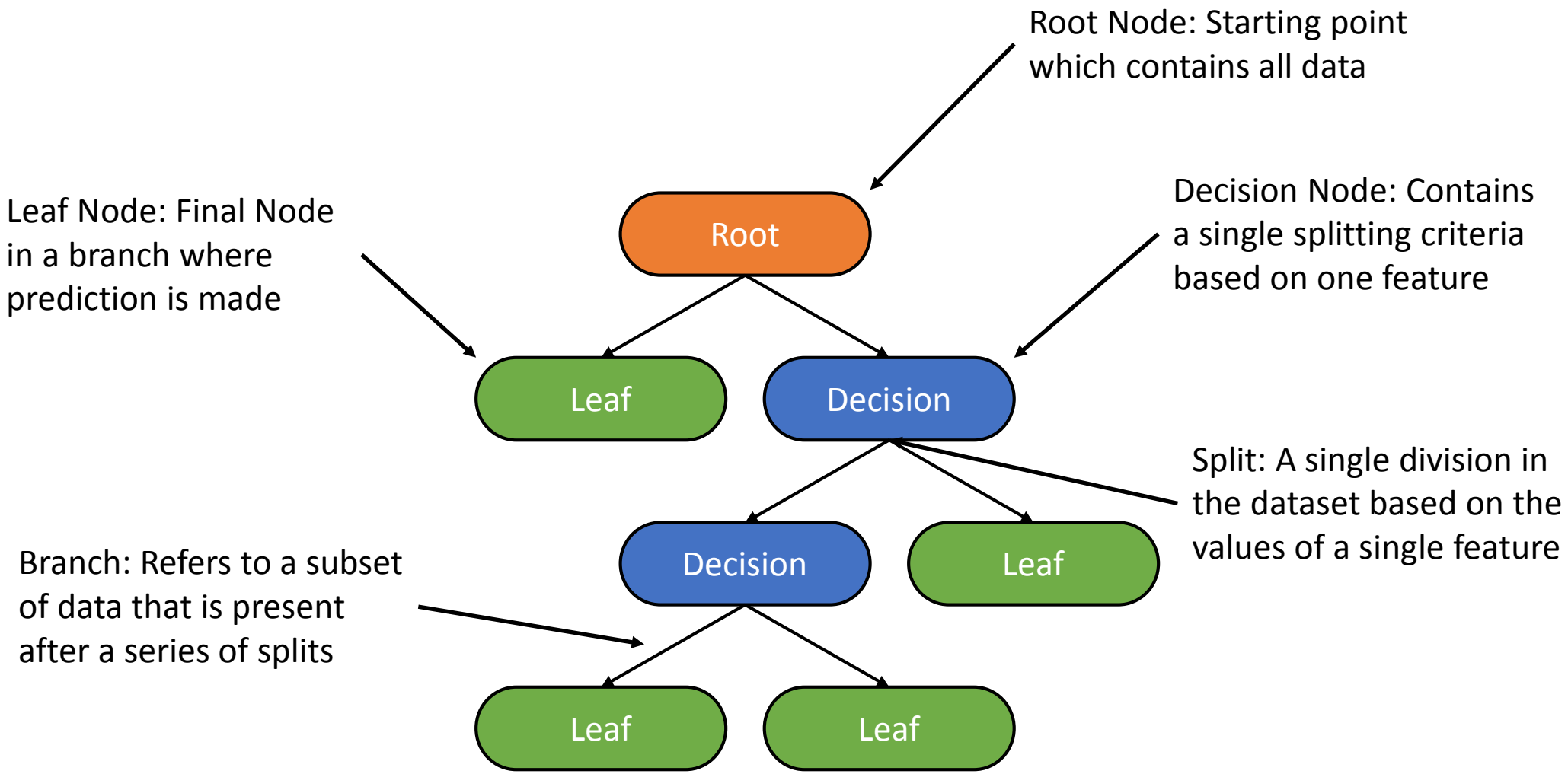
Model Types

- Linear Models
- Kernel Ridge
- Support Vector Machines
- Nearest Neighbors
- Gaussian Processes
- **Decision Trees**
- **Random Forests**
- Neural Networks

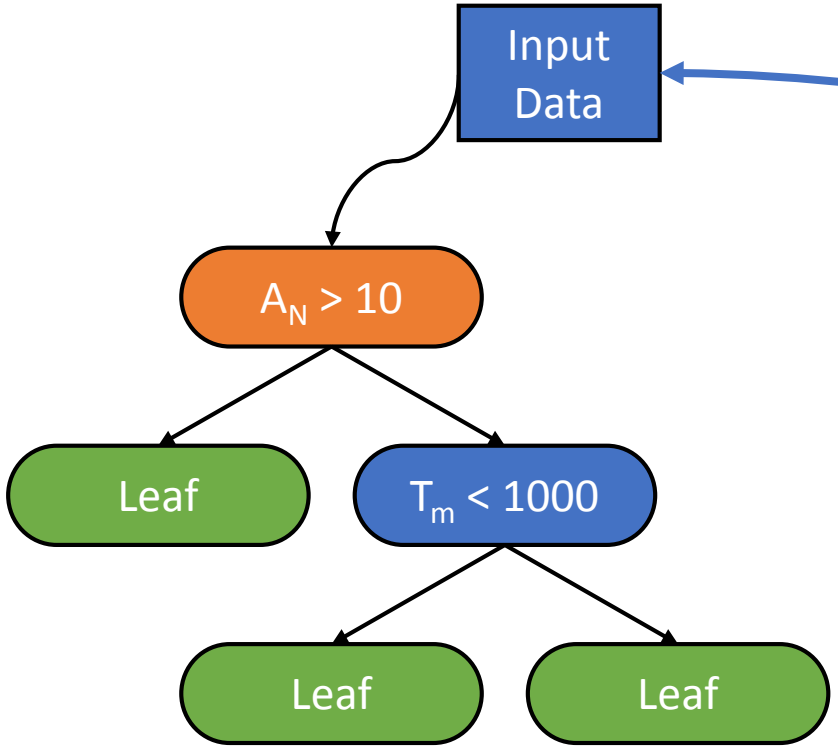
We'll focus on just one type that is easier to understand conceptually and doesn't require advanced math

For a more complete list of models
https://scikit-learn.org/stable/supervised_learning.html

Decision Trees: Structure



Decision Trees: Inputs

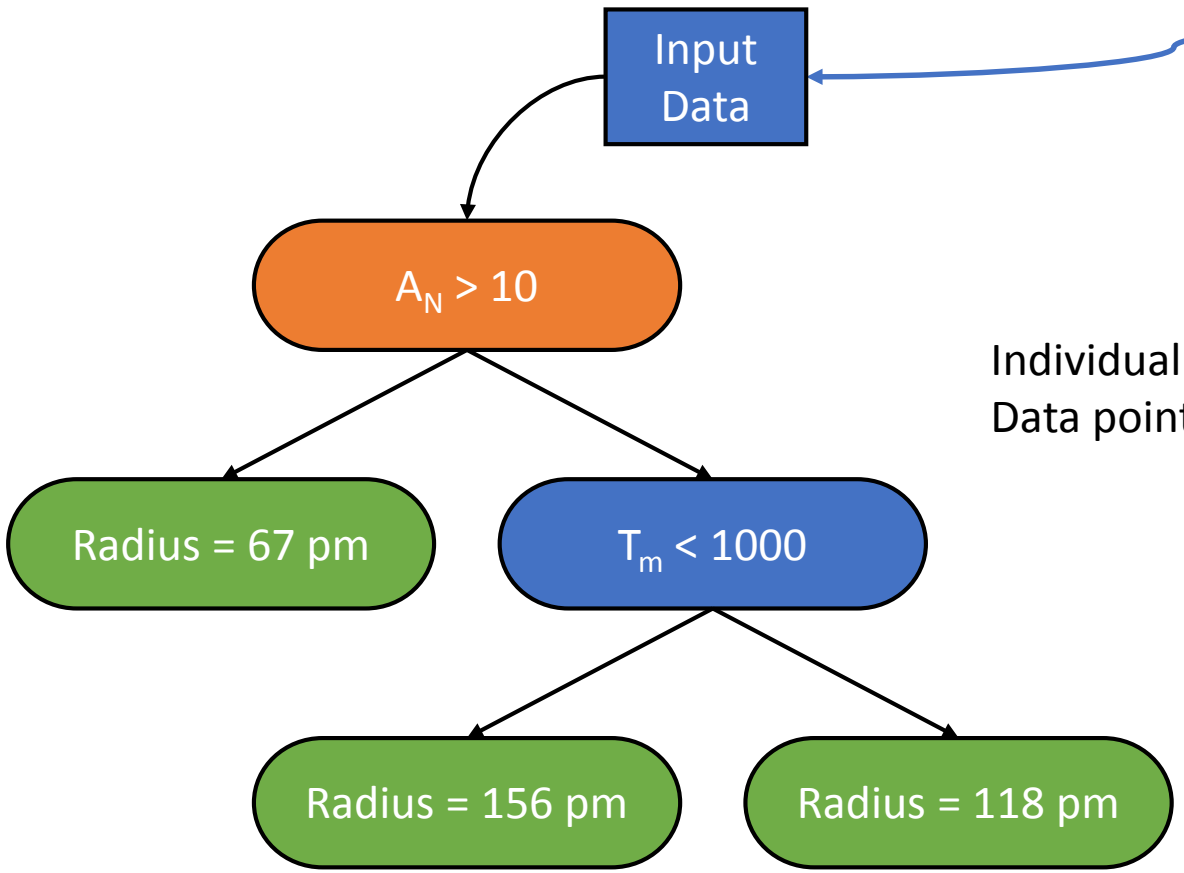


Features (descriptors)

Individual Data points

Index	A_N	T_m (K)
Fe	26	1800
Al	13	930
C	6	3800

Decision Trees: Outputs



Features (descriptors)

Individual Data points

Index	A_N	T_m (K)
Fe	26	1800
Al	13	930
C	6	3800

Summary

What is Machine Learning?

- Machine learning is a tool that finds patterns in large datasets that might be hard to discover otherwise

How can we use it for Materials Science?

- It can be included in existing materials science workflows to accelerate research, materials design, and materials discovery