## Data Sources

Existing

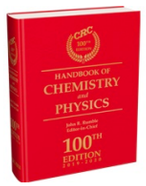
The Materials Project

OQMD

AFLOW
Automatic-FLOW for Materials Discovery

NOMAD

ICSD
FIZ Karlsruhe




DIY

pymatgen

FireWorks

AiiDA

**Classification**
Decision tree
Logistic regressio…
…

**Regression**
GPR
KRR
Multi-linear
Random forest
SVR
Neural networks
Graph models
…

TensorFlow

scikit learn

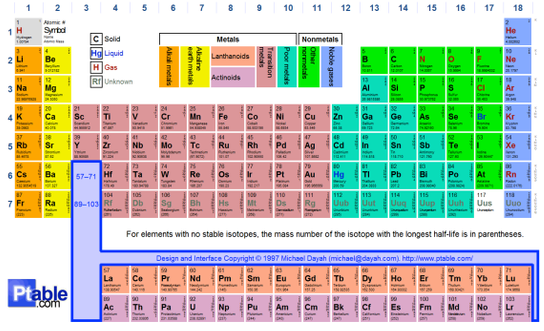
**Materials Science Specific**
ænet
Automatminer
CGCNN
DeepChem
MEGNet
PROPhet
SchnetPack
TensorMol
…

**Supervised**
- Cross-validation
- Hyper-parameter optimization

Identify Purpose and Target → Data Generation or Collection → Featurization → Training → Application

**Domain knowledge**
- Is target learnable?
- Is target ambiguous?

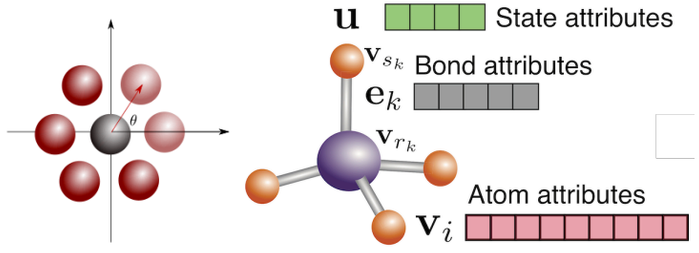
**Elemental Features**




DScribe

matminer

**Structural Features**




**Active learning**