**Data Sources**

Existing

The Materials Project

OQMD

AFLOW
Automatic - FLOW for Materials Discovery

NOMAD
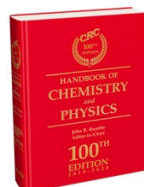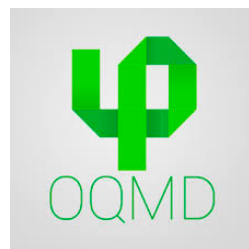
ICSD
FIZ Karlsruhe

HANDBOOK OF CHEMISTRY and PHYSICS
100TH EDITION

DIY

pymatgen

FireWorks

AiiDA

**Classification**
Decision tree
Logistic regression
...

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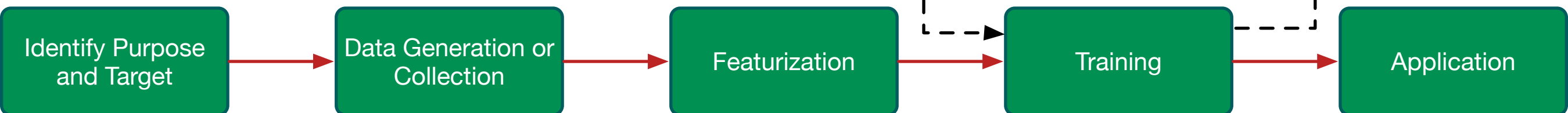Multi-linear
Random forest
SVR
Neural networks
Graph models
...

TensorFlow

scikit learn

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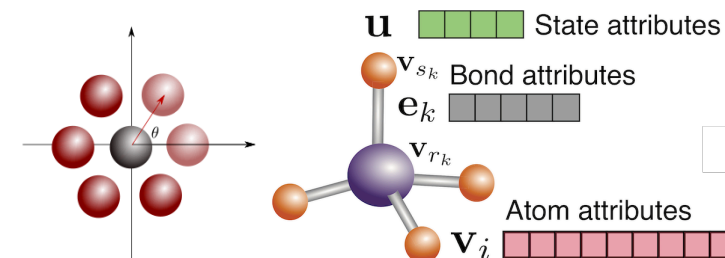Automatminer
CGCNN
DeepChem
MEGNet
PROPhet
SchnetPack
TensorMol
...

**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**

**Structural Features**

$\mathbf{u}$  State attributes
$\mathbf{v}_{s_k}$  Bond attributes
$\mathbf{e}_k$
$\mathbf{v}_{r_k}$
Atom attributes
$\mathbf{v}_i$

DScribe

matminer

**Active learning**