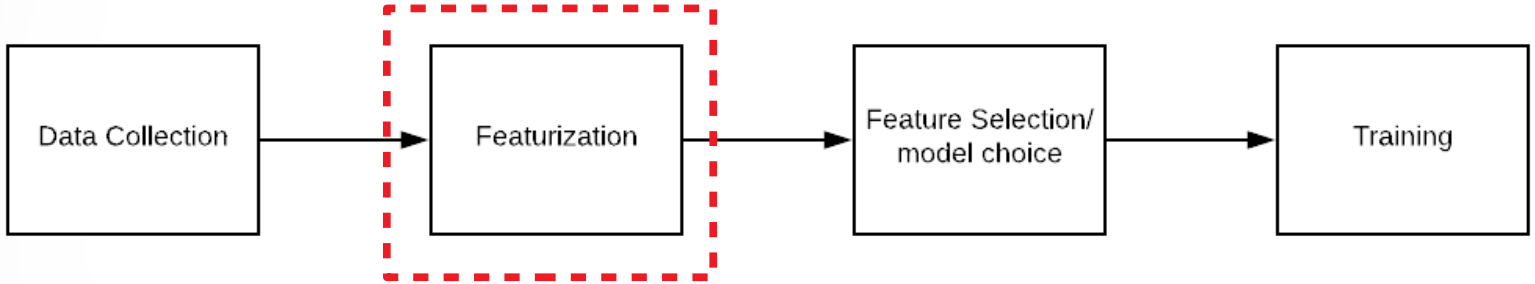
Machine Learning Workflow

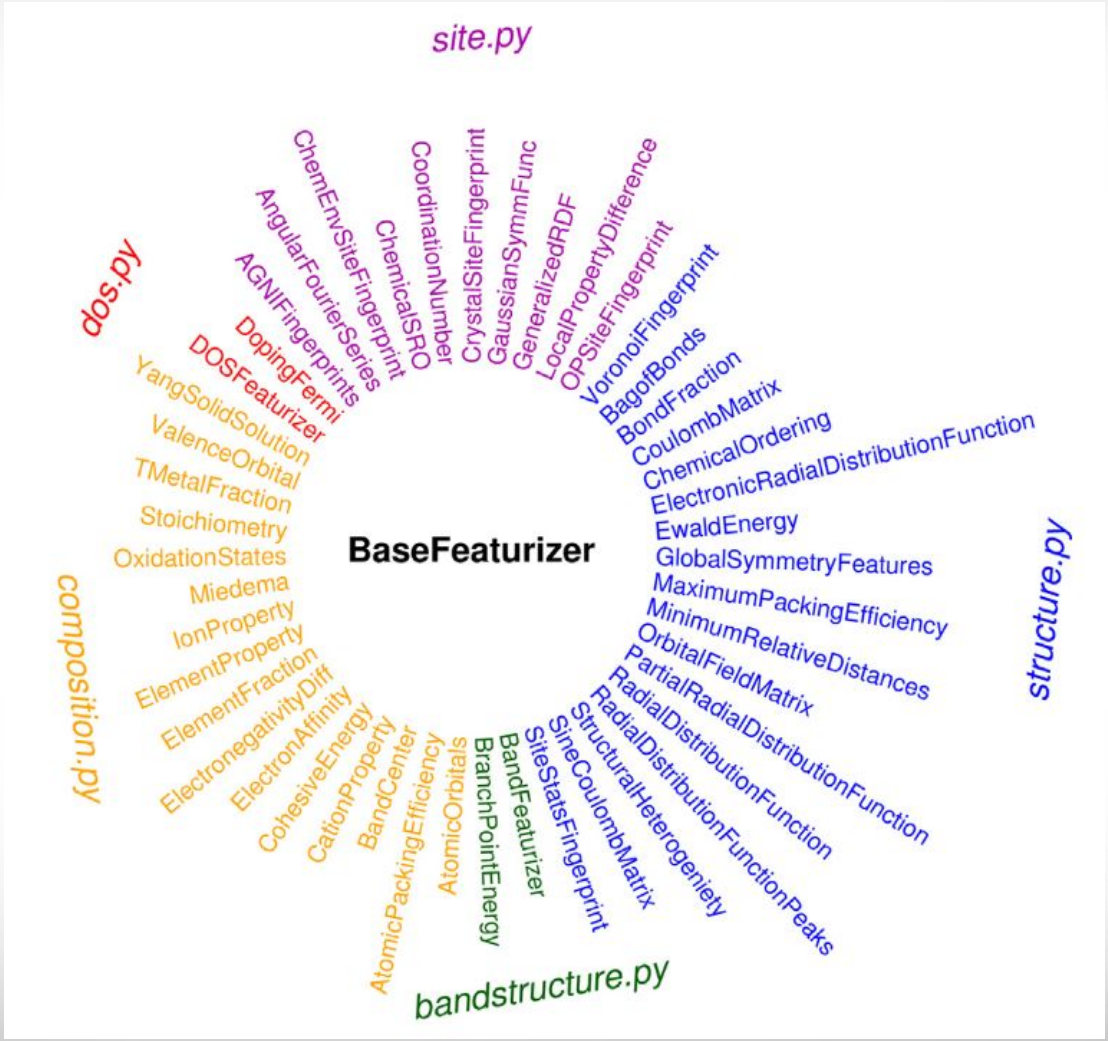


Current step

Featurization

- A feature is an individual measurable property or characteristic of a phenomenon being observed
  - In image recognition, these are the individual pixels, edges, or entire objects
  - Best image recognition neural nets take in individual pixels and can actually obtain the larger features through it.
- Features are things that should describe whatever you are looking at.
  - Informative
  - Discriminating
  - Independent

MatMiner


site.py
CoordinationNumber
ChemicalSRO
ChemEnvSiteFingerprint
AngularFourierSeries
AGNIFingerprints
CrystalSiteFingerprint
GaussianSymmFunc
GeneralizedRDF
LocalPropertyDifference
OPSiteFingerprint
VoronoiFingerprint
BagofBonds
BondFraction
CoulombMatrix
ChemicalOrdering
ElectronicRadialDistributionFunction
EwaldEnergy
GlobalSymmetryFeatures
MaximumPackingEfficiency
MinimumRelativeDistances
OrbitalFieldMatrix
PartialRadialDistributionFunction
RadialDistributionFunction
RadialDistributionFunctionPeaks
StructuralHeterogeniety
SineCoulombMatrix
SiteStatsFingerprint
BandFeaturizer
BranchPointEnergy
AtomicOrbitals
AtomicPackingEfficiency
BandCenter
CationProperty
CohesiveEnergy
ElectronAffinity
ElectronegativityDiff
ElementFraction
ElementProperty
IonProperty
Miedema
OxidationStates
Stoichiometry
TMetalFraction
ValenceOrbital
YangSolidSolution
DOSFeaturizer
DopingFermi
dos.py
composition.py
BaseFeaturizer
structure.py
bandstructure.py

MatMiner

Composition.py

- Creates features based on composition
  - Composition is the atoms and ratio of them
  - i.e. NaCl, H20, BaTiO3, and SiO2
- Cohesion energy, max electronegativity difference, etc.

Structure.py

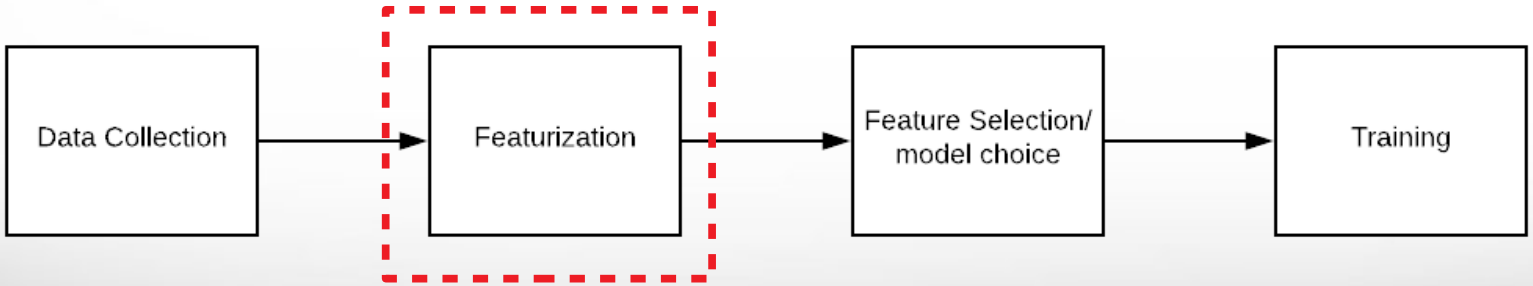
- Creates features based on structure
  - Structure is the position of the atoms
  - Similar to sites
- Radial distribution function(statistical mechanics method), Symmetry information, etc

Site.py

- Creates features based on sites
  - Sites are the where you can put an atom
  - Similar to structure
- Coordination number, local chemical environment
- Might be restricted to same number of site structures

Current Goal for Matminer

- Create a method to input CIF, POSCAR, JSON, etc. (some file format) into Matminer
  - Discussed in Sections 3.1
  - “The final data format will be compatible with the subsequent data featurization tools of matminer”
- Use matminer to generate all features.
  - Use Composition.py, Structure.py, and Site.py



Questions?