To-do **Important points** General text

1. Introduction
   1. Description of stance toward failure and learning
   2. Inverse problem: People have used physics to build models, might make sense to learn in the opposite way
   3. Interpretability? Move to make ML for MSE more interpretable….
2. Description of methods
   1. Dataset
   2. Features
   3. Models
3. Typographical errors
   1. Outliers can skew some models ability to fit. RF should be rather agnostic.
4. Physics/modeling errors
   1. Features can be examined and learned from
      1. Feature importances are “gini importance”
         1. Purity at each node.
      2. Probably need to remove highly correlated variables first
      3. Feature removal…? Probably not what we want to do when using features to determine
   2. **Predicting difference between a simple model and real data can be useful**
      1. Can improve model significantly
      2. Can identify features that are different 🡪 point to physical reasons why we might deviate from a naïve model
   3. How does MP DFT predict/perform

Literature, machine learning

1. G. Pilania, R. Ramprasad, Accelerating materials property predictions using machine learning, Scientific Reports 2013
   1. Material -> Fingerprint vector -> Chemical differences -> Machine learning
   2. Fingerprint vector similar to matminer features, possibly.
   3. Kernel Ridge regression (KRR)
2. Fdsf1