## Kaggle House Prices – Algorithm Pipeline

1. Importing
   1. Import packages
   2. Import practice\_sample and evaluation\_sample data
2. Preprocessing
   1. Append practice\_sample and evaluation\_sample data
   2. Deal with NAs
      1. Numeric Columns: Replace NAs with mean value for entire column
      2. Categorical Columns: Replace NAs with “None”
   3. Split practice\_sample and evaluation\_sample data
   4. Split practice\_sample in X\_train, X\_test, y\_train, y\_test
   5. Feature Selection
      1. Numeric Columns:
         1. Check Correlation Matrix
         2. Select k most important features based on correlation with target variable
      2. Categorical Columns:
         1. Encode values to integers
         2. Select j best features based on Chi squared/Mutual information statistic test
   6. Categorize selected columns
   7. Transform columns into dummy variables
3. Machine Learning pipeline
   1. Hyperparameter tunning:
      1. Initialize dictionary with possible model parameters to try
      2. Initialize model
      3. Initialize best parameters search algorithm with previous initialized model
      4. Fit X\_train, y\_train to the algorithm
      5. Check best algorithm score and select best params to fit model
   2. Initialize model with best parameters found
      1. Fit model to X\_train, y\_train
      2. Predict output for X\_test
      3. Compute evaluation metric Root Mean Squared Logarithmic error
4. Fit ML model to evaluation\_sample
5. Export results csv