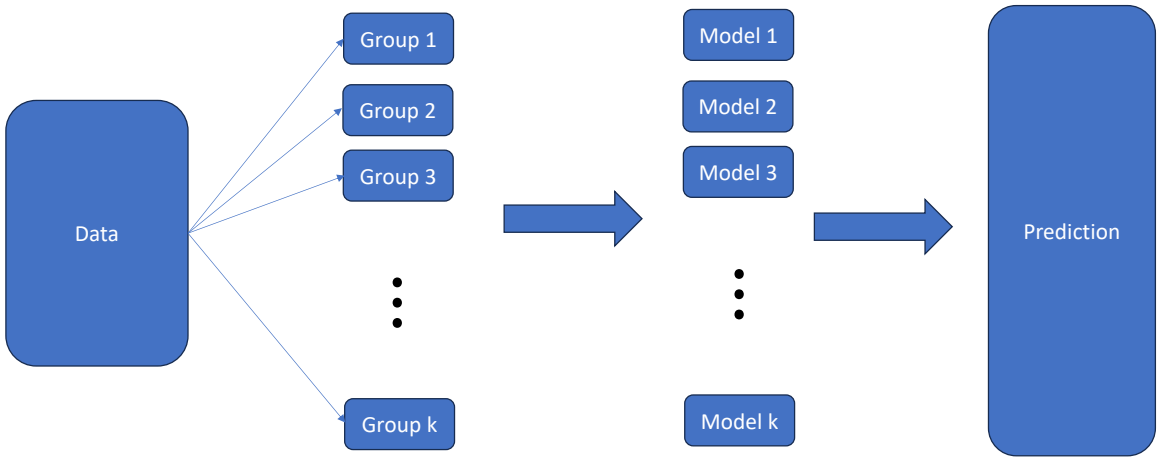
CS559: Machine Learning
Final Project: House Price Prediction Competition

Our second project is to predict the house price in Ames, Iowa. In the given train and test data sets, there are 79 features. The project aims to make a model that predicts the individual house price in the test set.

Features: Be creative to reduce the number of features to **35 features** via feature engineering, feature selection, and dimensionality reduction. Using LDA or PCA to reduce the number of features without feature engineering or feature selection to measure the number of features is prohibited.

Train Model Process:



1. Subgroup the train data into k many subgroups and build a classifier that predicts the group. The trained classifier must be validated using grid-search cv.
2. Make a model that predicts the price for each group.
    a. **The group model must be validated with at least four machine-learning algorithms**.
    b. Validate each algorithm to have its own best model.
    c. Then, use a **stacking method** to predict individual houses in each group.
3. Evaluate the model using MSE, $E = \frac{1}{2N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$. The target must not be standardized or normalized.

Test data prediction:

1. Write a function, test_data_prediction(train data, test data), that transforms the test data into the train data format, e.g., the same features, scales, etc, and predicts the house price using the trained best models.

Submission:

1. Write a report on workflow and results.
    a. The training data process must be explicitly explained.

   b. Evaluate the train models.
   c. Evaluate the test predictions.
2. Record a brief video presentation for about 10 – 15 minutes.
3. Report the trained MSE the alpha-model by 12/1/2023 Friday. The alpha-model rank will be shared with the class.

Grade Scheme:

1. Machine Learning Modeling: 65 pts
2. Report and Video Presentation: 25 pts
3. RMSE Rank: 10 pts