

**Stacking: Universal Dataset**

The SaratogaHouses dataset has 16 variables and 1728 records. Use “price” as target variable.

A data frame with 1728 observations on the following 16 variables.

- `price` price (1000s of US dollars)
- `lotSize` size of lot (square feet)
- `age` age of house (years)
- `landValue` value of land (1000s of US dollars)
- `livingArea` living are (square feet)
- `pctCollege` percent of neighborhood that graduated college
- `bedrooms` number of bedrooms
- `fireplaces` number of fireplaces
- `bathrooms` number of bathrooms (half bathrooms have no shower or tub)
- `rooms` number of rooms
- `heating` type of heating system
- `fuel` fuel used for heating
- `sewer` type of sewer system
- `waterfront` whether property includes waterfront
- `newConstruction` whether the property is a new construction
- `centralAir` whether the house has central air

1. Import the data into R
2. Convert the attributes to appropriate types and combine the numeric and categorical attributes.
3. Standardize the numeric data
4. Convert all categorical attributes to numeric using the dummy function, then replace the old categorical attributes with their dummied values.
5. Changing certain column names
6. Divide the data into train and test
7. Build several regression models
8. Predicting on train dataset
9. Combining the training predictions of all the models.
10. Add the original target variable to the dataset.
11. Ensemble the model with lm as Meta Learner
12. Evaluate the ensemble model on train data
13. Follow the steps from 7 to 12 on the test data and evaluate