

# Final Exam – DA6813

Due Date: August 12, 2018 at 11:59 pm

## Problem

Your task in this final exam is to build a model to predict property values (prices) which have been taken from a small region of Bexar County. You have been given a dataset with the attributes for five properties which do not have a listed a value (price). You will use a dataset of 608 properties with known values to build a set of models of any form (linear, lasso, SVM, etc.). After you build your models, you will have to make a choice of which model to use to determine your final results – the values of the five unpriced properties.

## Data

There are two datasets for this exam. The data for training and testing your models is called `final_exam_train.csv`. The target/response variable is named 'value.' The dataset with the 5 houses with unknown values is in the file `final_exam_test.csv`. The features in the data are related to attributes of a house. Most of the attributes are related to the size of features that make up a house. To give you a sense of what the features are: "AG" stands for "Attached Garage", "LA" stands for "Living Area," "PAC" stands for "Patio Covered", etc. These features are in square feet.

## Evaluation

Your submission will be evaluated in two ways. 25% of the grade will be based on the accuracy of your model results.

The other 75% of your grade will be based on the approach that you take to get the answer. You are expected to:

1. Preprocess the data
2. Build several models.
3. Select the best performing model by clearly estimating its accuracy.
4. Refine your approach by considering different input feature combinations and preprocessing techniques.

You will be given an \*.Rmd file that will help guide you in organizing your submission. Please comment your code where you think an extra explanation of your approach is necessary.

## Submission of Final Exam

Your final exam will be submitted through Blackboard. This will be due on August 12, 2018 at 11:59 pm. Late submissions will NOT be accepted. You will have almost 3 weeks to put together your solutions for this exam. Make sure to use the extra time wisely.

Three (3) files are expected in the submission of your final exam. Do not zip them. Attach them together on Blackboard. The files to be submitted are:

1. Your predictions of the house values in the following format in a word document (\*.docx):

property_id	value
2001	<b>100,000</b>
2002	<b>100,000</b>
2003	<b>100,000</b>
2004	<b>100,000</b>
2005	<b>100,000</b>

The values delivered by your selected model of choice will replace bolded text above.

2. Your final .Rmd file.
3. The HTML Output of your .Rmd File.

### Expected Conduct

You are expected to work on this take-home final exam on your own. Do not discuss your approach or answers with other students. Your answers will be evaluated using plagiarism checking software on Blackboard to ensure that your responses are your own. If there is any evidence of collaboration on your final exams, it will be handled according to UTSA's policies on scholastic dishonesty.