ORIE 4741 Project Proposal

Zillow Prize: Zillow’s Home Value Prediction (Zestimate)

Wei Zou (wz299), Jiahe Xu (jx266)

Sep 22,2017

1 Background

Zillow?s Zestimate home valuation has influenced the U.S. real estate industry since first released 11 years ago. A house is always the most important purchase a person makes in his lifetime. In this case, it is incredibly important to ensure homeowners have a trusted way to monitor this asset. The Zestimate was created to give consumers as much information as possible about homes and the housing market, marking the first-time consumers had access to this type of home value information at no cost.

2 Object

In our project, we are going to build a model to improve Zestimate residual error using linear regression, logistic regression and random forest methods. Specifically, it is to predict the log-error between their Zestimate and the actual sale price, given all the features of a home. The log error is defined as:

logerror=log(Zestimate)-log(SalePrice)

3 Partial Data

’airconditioningtypeid’	’architecturalstyletypeid’	’basementsqft’
’buildingqualitytypeid’	’buildingclasstypeid’	’calculatedbathnbr’
’finishedfloor1squarefeet’	’calculatedfinishedsquarefeet’	’finishedsquarefeet6’
’finishedsquarefeet15’	’finishedsquarefeet50’	’fips’

4 Approach

The data consist of a full list of real estate properties in three counties (Los Angeles, Orange and Ventura, California) data in 2016. (https://www.kaggle.com/c/zillow-prize-1/data) The train data has all the transactions before October 15, 2016, plus some of the transactions after October 15, 2016.