

House Price Prediction in Los Angeles

Team 110
Chen Zhang(czhang613)
Shaojuan Liao(sliao033)
Kai Ni(knimr3)
Shengchen Liu(sliu651)
Zheng Kuang(zkuang30)
Jinjun Liu(jliu788)

Project Overview

- House purchase: a BIG decision
- Real estate websites provide detailed information
- Absence of information: school district and community quality
- This project aims at develop an accurate evaluation model.
 - Physical, Community, school and environment variables
 - Advanced machine learning techniques
 - Interactive visualization

Existing Solutions

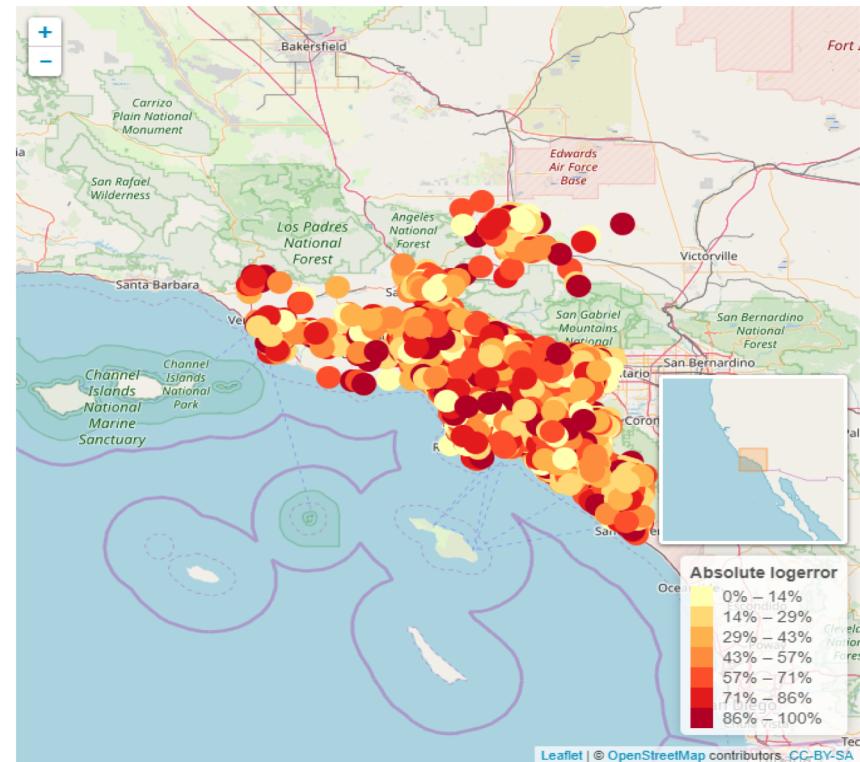
- Zillow and Trulia: only 50% estimated prices to be within the 5% range of their actual sold prices
- Traditional Economic Model
 - Hedonic Price Model for spatial correlation in transaction prices (Basu 1998)
 - Grey relational analysis and weighted synthesis for house price prediction (Tan 2016)
 - Geographically weighted regression with a non-Euclidean distance metric (Lu 2014)
- Machine Learning Method
 - Classification and regression by random forest (Liaw 2002)
 - Neural network ensemble and cross validation (Clark 2018)
 - Deep learning in neural networks (Schmidhuber 2015)
 - Very deep convolutional networks (Simonyan 2014)
- Only physical features!

Existing Solutions

- Data Preprocessing
 - Data cleaning, normalization and transformation(Kotsiantis 2006)
 - Dealing with textual attributes, such as features of a neighborhood quality (Ahmed 2016)
 - Support vector machine(SVM), least squares SVM and partial least squares (PLS) on Boston housing data (Mu 2014)

Project Features

- A housing price prediction model based on multiple factors of the property
- An interactive visualization product

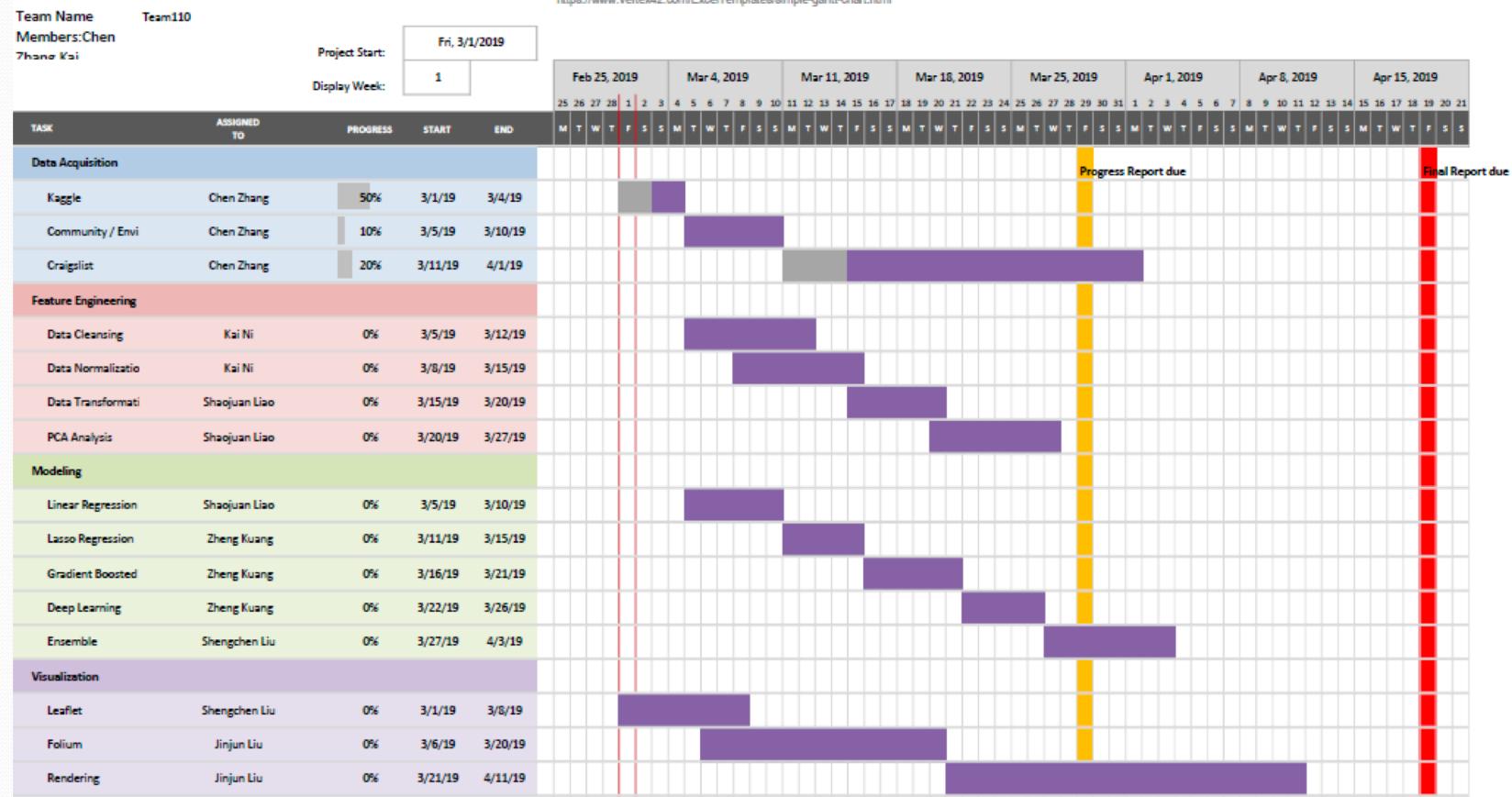


Project Plan

- Phase I
 - Use Zillow's data for Los Angeles area in 2016-17 to develop a prototype model
 - Scrape additional data
 - Midterm goal
- Phase II
 - Enhance model using additional data.
 - Present result using interactive visualization map
 - Draft report

Plan of Activities

House Price Prediction in Los Angeles



Evaluation of Success

- Data splitted by 80/20
- Mean Absolute Error (MAE) and log error

$$\logerror = \log(Estimate) - \log(SalePrice) \quad (1)$$

Mean Absolute Error (MAE) is defined as :

$$MAE = \sum_{i=1}^n |y_i - x_i| / n \quad (2)$$

Risks and Costs

- Potential Risks
 - Data parsing, cleaning and quality problem
 - Model may not work as expected, overfitting
- Expense and Cost
 - Data: free to obtain online
 - Cost: use of AWS

Selected References

- Ahmed**, E. and Moustafa, M., 2016, House price estimation from visual and textual features, arXiv preprint, arXiv:1609.08399
- Liaw**, A. and Wiener, M., 2002, Classification and regression by randomForest, R news, 2(3), 18-22
- Simonyan**, K. and Zisserman, A., 2014, Very deep convolutional networks for large-scale image recognition, arXiv preprint, arXiv:1409.1556
- Kotsiantis**, S. B., Kanellopoulos, D., and Pintelas, P. E., 2006, INTERNATIONAL JOURNAL OF COMPUTER SCIENCE VOLUME 1 NUMBER 1, 2006 ISSN 1306-4428
- Mu**, J., Wu, F., and Zhang, A., Housing Value Forecasting Based on Machine Learning Methods;
- Clark**, Stephen D., and Nik Lomax. "A mass-market appraisal of the English housing rental market using a diverse range of modelling techniques." Journal of Big Data 5.1 (2018): 43.
- Schmidhuber**, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.
- Basu**, S. & Thibodeau, T.G. The Journal of Real Estate Finance and Economics (1998) 17: 61.
- Lu**, Binbin, Martin Charlton, Paul Harris, and A. Stewart Fotheringham. "Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data." International Journal of Geographical Information Science 28, no. 4 (2014): 660-681.
- Tan**, W. and Chou, T., 2016, Combine Grey Relational Analysis and Weighted Synthesis for Housing Price Prediction, BESSH/Full Paper Proceeding, 2016, 117(2), 18-27