Housing Price Prediction

Alice Cheng
Yuanqi Sun
Vanessa Zoghbiharb
Shivam Sonar

Motivation

Everyone plans on having their own house one day. But what kind of house do you want, or what features would you like in it? Also, how much will said features add to the cost?

Using data that is available to the general public, we built a web application to estimate house prices based on specific features.

Data Set - Ames Housing

Obtained from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010

Total variables :   82

Categorical variable: 23 nominal and 23 ordinal

Data set size: 2930 observations

Models tested:

1) Linear Regression   3) Lasso Regression

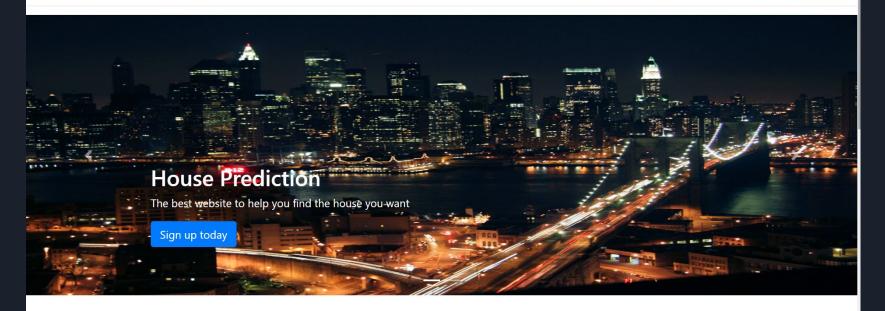2) Decision Tree   4) Random Forest

Website Demonstration

http://datascienceteam.pythonanywhere.com/

Data Gathering & Preprocessing

Integrity Checks

- Properties that had blanks in one or more feature(s) were removed

Feature Engineering

- Categorical features of properties were encoded into dummy variables

Standardization

- The dummy variables of the categorical features are not large

Model Selection

- K-fold cross validation
  - 20 % Testing
  - 80% Training
- Cross Validation
  - Random Forest
- Select K Best
  - Linear Regression select top 11 of 48 features

Model Selection

Model	MSE	R^2
Linear Regression	591,475,207	0.92
Lasso Regression	**586,711,534**	**0.92**
Random Forest Regression	792,726,275	0.89
Decision Tree Regression	1,929,023,367	0.62

Use of MySQL

Database for storing account details that can be imported onto the website.

```
mysql> use datascienceteam$try
Database changed
mysql> SELECT * FROM try
    -> ;
+-----------------+-------------+----------------+-------------+--------------+----------------+-------------+
| Overall_Qual    | Utilities   | Neighborhood   | Bldg_Type   | House_Type   | Misc_Feature   | SalePrice   |
+-----------------+-------------+----------------+-------------+--------------+----------------+-------------+
| 'Overall_Qual'  | 'Utilities' | 'Neighborhood' | 'Bldg_Type' | 'House_Style'| 'Misc_Feature' | 'SalePrice' |
| '6'             | 'AllPub'    | 'NAmes'        | '1Fam'      | '1Story'     | 'NA'           | '215000'    |
| '5'             | 'AllPub'    | 'NAmes'        | '1Fam'      | '1Story'     | 'NA'           | '105000'    |
| '6'             | 'AllPub'    | 'NAmes'        | '1Fam'      | '1Story'     | 'Gar2'         | '172000'    |
| '7'             | 'AllPub'    | 'NAmes'        | '1Fam'      | '1Story'     | 'NA'           | '244000'    |
| '5'             | 'AllPub'    | 'Gilbert'      | '1Fam'      | '2Story'     | 'NA'           | '189900'    |
| '6'             | 'AllPub'    | 'Gilbert'      | '1Fam'      | '2Story'     | 'NA'           | '195500'    |
| '8'             | 'AllPub'    | 'StoneBr'      | 'TwnhsE'    | '1Story'     | 'NA'           | '213500'    |
| '8'             | 'AllPub'    | 'StoneBr'      | 'TwnhsE'    | '1Story'     | 'NA'           | '191500'    |
| '8'             | 'AllPub'    | 'StoneBr'      | 'TwnhsE'    | '1Story'     | 'NA'           | '236500'    |
| '7'             | 'AllPub'    | 'Gilbert'      | '1Fam'      | '2Story'     | 'NA'           | '189000'    |
| '6'             | 'AllPub'    | 'Gilbert'      | '1Fam'      | '2Story'     | 'NA'           | '175900'    |
```

Use of MySQL

Database for storing account details that can be imported onto the website.

```
mysql> SELECT * FROM users;
+-----+----------+--------------------------------------------------------------+------------------+----------+-----
-------+------+
| uid | username | password                                                     | email            | settings | tra
cking | rank |
+-----+----------+--------------------------------------------------------------+------------------+----------+-----
-------+------+
|   1 | adasdqwd | $5$rounds=535000$86HXkc8PYLs1ZG3l$/or9R2eQadU93wOsRKf2aARz8DRcat/ksHeTRRPtoK. | test@gmail.com   | NULL     | /ac
count/ | NULL |
|   2 | test     | $5$rounds=535000$YFTYdfp.sdpHhCCs$eyZmXs0WkzEJgvRJki5/CCO4E8gkNzzW6sKhVfW1TG3 | test@gmail.com   | NULL     | /ac
count/ | NULL |
|   3 | test1    | $5$rounds=535000$KUFTlkOjPmhV5rbw$FivDel3gTRIoVVjPvm325c.d9RXl7//RBcEWCr2/l78 | test@gmail.com   | NULL     | /ac
count/ | NULL |
|   4 | test4    | $5$rounds=535000$wyaCqNacnsTCSiqG$Tj3GGvPnbTFa3kGlD0HsFeQdlNhAqoN40khl13SXJ77 | test@gmail.com   | NULL     | /ac
count/ | NULL |
|   5 | test11   | $5$rounds=535000$k9NCdioI8PkNtzk6$B.SAq5CEt7mQcl63UOEVAIHk4WqTWVa8km4bR2TlWi2 | test1@gmail.com  | NULL     | /ac
count/ | NULL |
|   6 | test5    | $5$rounds=535000$WR5kmWf3UZvGrXr4$/45aQiImKrABPGby482M5qMmfiQ1sovWgqdcGLPtEe0 | test@gmail.com   | NULL     | /ac
count/ | NULL |
|   7 | test6    | $5$rounds=535000$96YTz0D8Zc3qLm6f$t5zHb8YyXMfNJTxWD.vxaG9ql.c0.qnLXV.kslqLeA3 | test@gmail.com   | NULL     | /ac
count/ | NULL |
+-----+----------+--------------------------------------------------------------+------------------+----------+-----
-------+------+
```

Thank You!