

Predicting Error Values for Zillow's Sale Price Zestimate Through Machine Learning

Elizabeth Vincent

October 4, 2017

Abstract

This is my abstract, in which I tell you about my project.

1 Introduction

Something here about the competition, what the Zestimate is, what the logerror is, why using the logerror instead of RMSE

2 Methods

1. Data gathering
2. Data cleaning
3. Missingness
4. Imputation
 - City and zip based on unique city-zip or zip-city combinations, and based on long/lat coordinates
5. EDA
 - Correlation of missingness between vars
 - Correlation of logerror with missingness
 - Correlation of logerror with numeric vars
6. Analysis
 - Predict the mean
 - Linear model

- Random forest

3 Results

RMSE and R^2 for each method. Compare methods, which gave lowest RMSE and highest R^2

4 Discussion

Interpretation of what factors influence the logerror, which method was best.