Detecting Housing Price Shifts Using Advanced Machine Learning Methods

Abha Jhaveri, Nicholas Colonna, Nicole Lange

STEVENS INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Introduction: Problem

In 2008, the housing market crashed and shocked the entire financial system. The housing bubble triggered prices to decline more than 40% and led to mortgage defaults that caused millions of foreclosures. By forecasting future shifts in the housing market with our model, we hope to detect a possible crash before it happens. This could help save investors, homeowners and prospective buyers from huge losses, like what happened in 2008.

Data Understanding

- 371 monthly observations from 2/1/87-12/1/17
- Dependent Variable: Direction of monthly returns for the Case Shiller Home Price Index
- 10 independent variables encompassing macroeconomic indicators, housing market trends, and consumer preferences
- Source: St. Louis Federal Reserve (FRED)

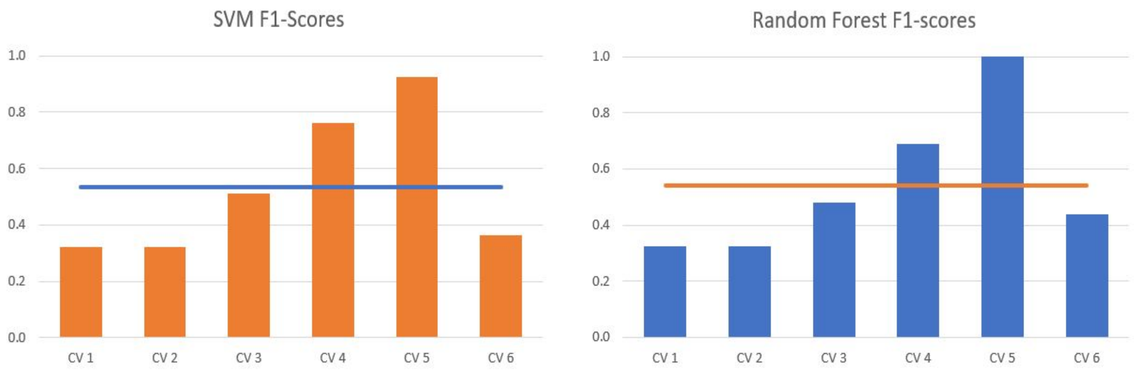
Data Preparation

- Classification of Case Shiller Index
  - Computed the median monthly return of the index
  - Classified as 1 (Up) if greater than the median, 0 (Down) otherwise
- Data Split
  - Training set: 2/1/87-12/1/13 (323 observations)
  - Testing set: 1/1/14 -12/1/17 (48 observations)
- Additional Data Splits
  - Pre-Crisis(2/1/87-7/1/06)
  - Post-Crisis(8/1/06-12/1/17)

Modeling

- Features: (6 filtered via backward stepwise selection)
  - Workers Over 55, Monthly Supply of Houses, Population Growth, 10 Year Treasury Yield, Median Household Income, CPI Inflation
- Support Vector Machine Classifier
  - Kernel: radial basis function (rbf)
  - Gamma: $7 \cdot 10^{-8}$
  - C: 0.1
- Random Forest Classifier
  - Criterion: entropy
  - Max Depth: 3
  - # of Estimators: 4

Support Vector Classifier vs. Random Forest Classifier

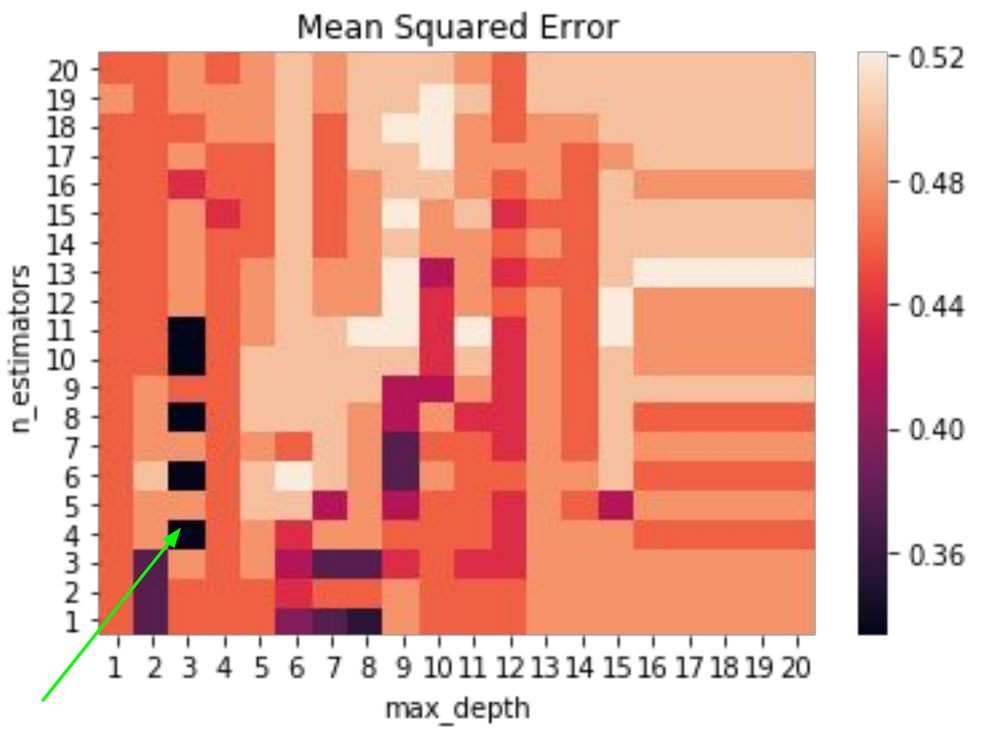


Results & Evaluation

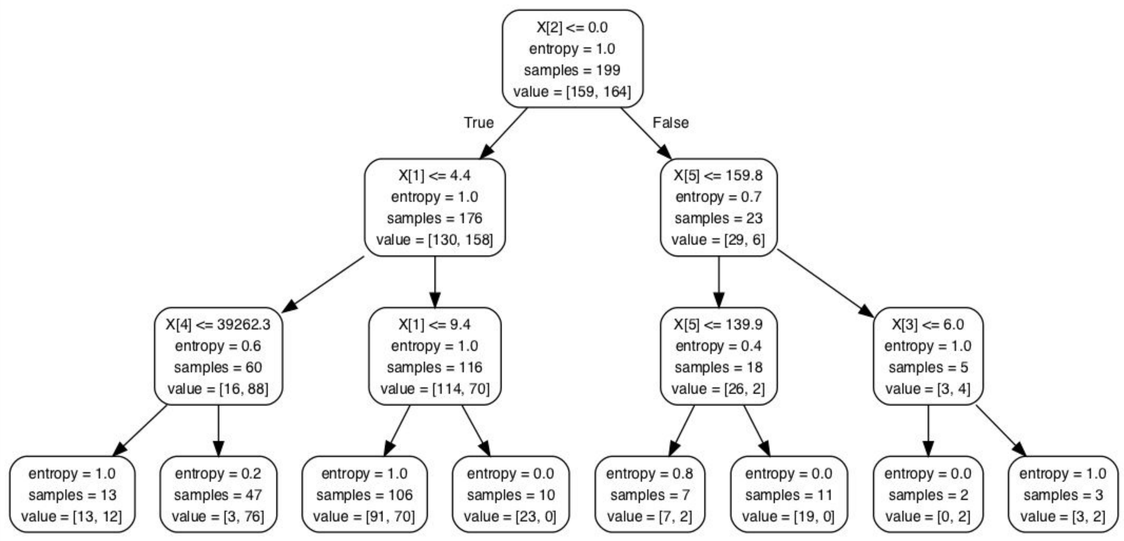
- Model performance on test set: 4yrs of economic data

	precision	recall	f1-score	support
0	0.6471	0.8462	0.7333	26
1	0.7143	0.4545	0.5556	22
avg / total	0.6779	0.6667	0.6519	48

- Selected the number of estimators and maximum depth that minimized MSE



- 1 of 4 Decision Trees Used by Random Forest Model



Conclusion

- Our model can predict the monthly directional movement of the housing market more than 50% of the time
- Random Forest is a strong algorithm for classification due to its majority vote methodology
- Reduces both variance and bias
- The applications of Random Forest are broad and not limited to a binary-state prediction

Deployment

- Confusion matrix on test set

	Predicted Down	Predicted Up
Actual Down	22	4
Actual Up	12	10