Data-based location recommendation for Houston restaurant owners

Capstone Project-The Battle of Neighborhoods

Location selection for restaurants is like gold mining

- The best option is not apparent
- Success location selection is a combination of:

    1. reasonable cost
    2. abundant customer traffic,
    3. nice neighborhood
    4. relatively low competition

- How confident are you to pick the best location?

Data-driven recommendation model for restaurant location is much needed

Comprehensive profile about a neighborhood can provide valuable insight on the recommendation:

- nearby venues (Foursquare API)
- demographics (age, gender, ethnics etc.)
- economics (income, housing etc.)
- family composition (couples, singles etc.)

Learn from similar cities using best models

- Factors impacting the success of restaurants share across big cities

- Learn from three metropolitans with similar size and variety (Los Angeles, Dallas and New York)

- Best models from Random Forest, Logistic Regression and Support Vector Machines

Data preparation

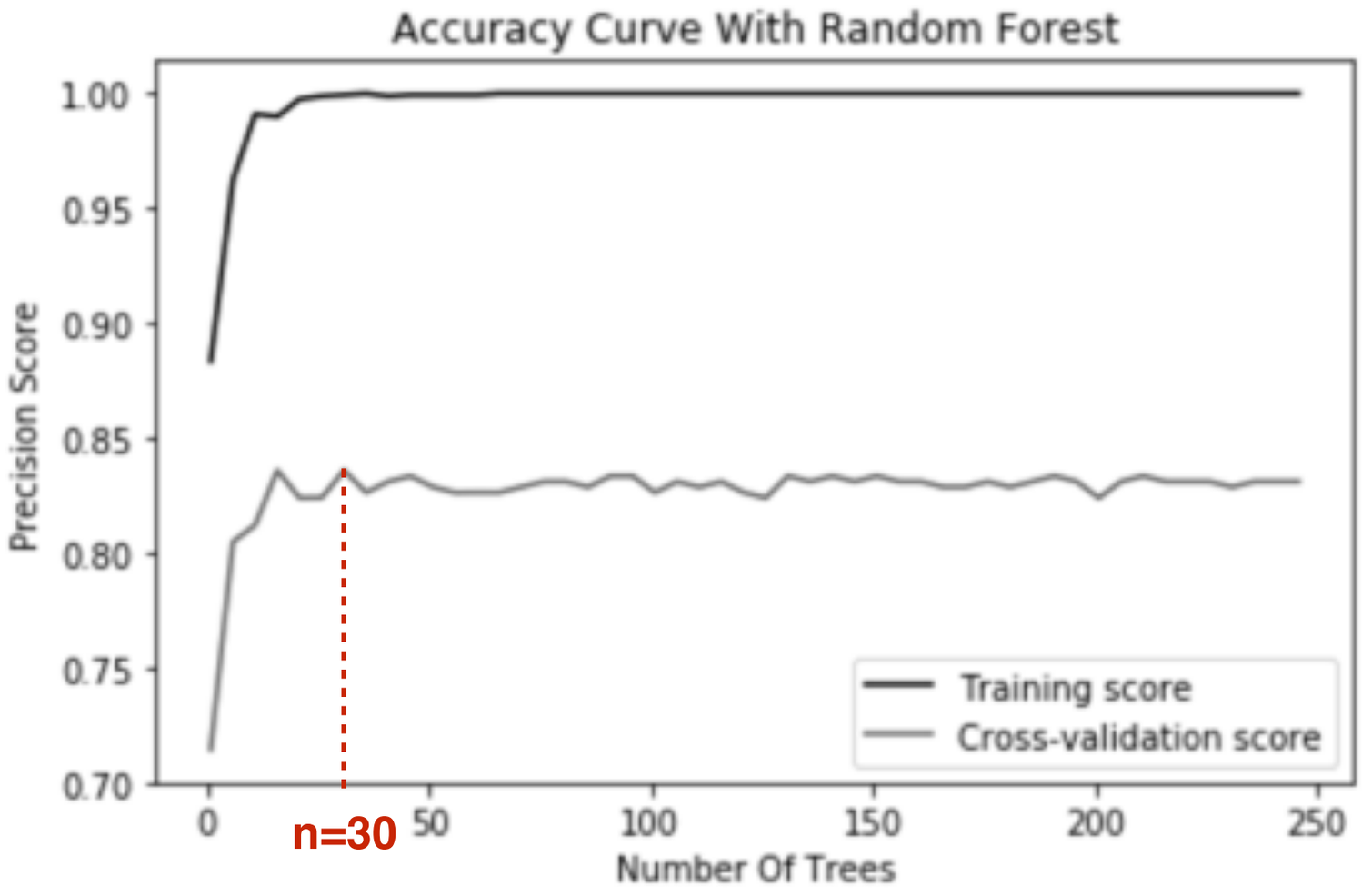
- Extract all neighborhood lists for Houston, Los Angeles, Dallas and New York from Wikipedia
- Geocode neighborhoods
- Annotate social-economic, demographic data
- Data cleaning (high missing rate, standardization, redundant information)

For data from cities other than Houston:

- Assign target label
- Data split: 80% training, 20% test

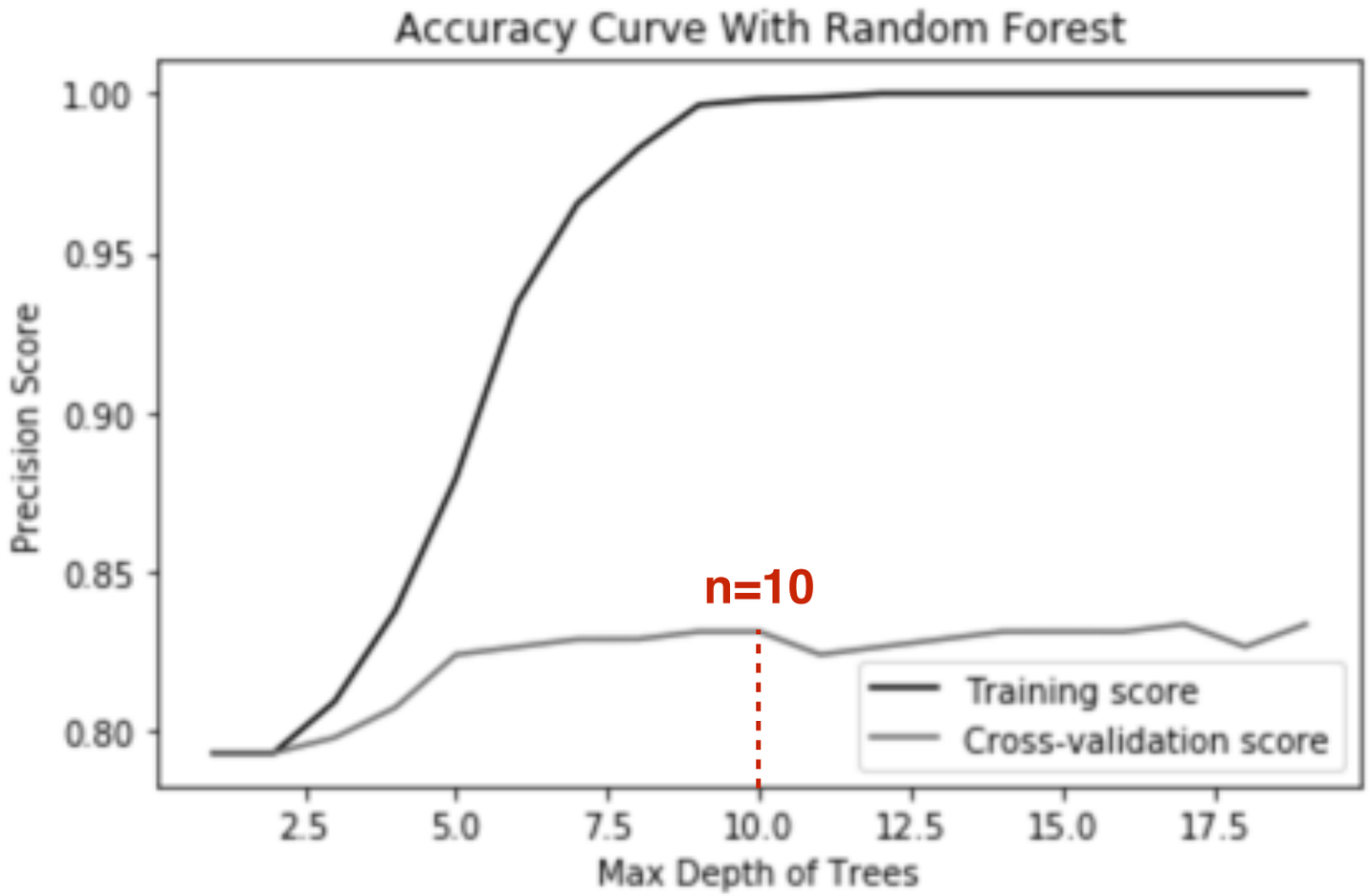
Model training and hyperparameter tuning —Random Forest

n_estimators: number of trees



Model training and hyperparameter tuning —Random Forest

maximum depth



Find best models based on accuracy on test data

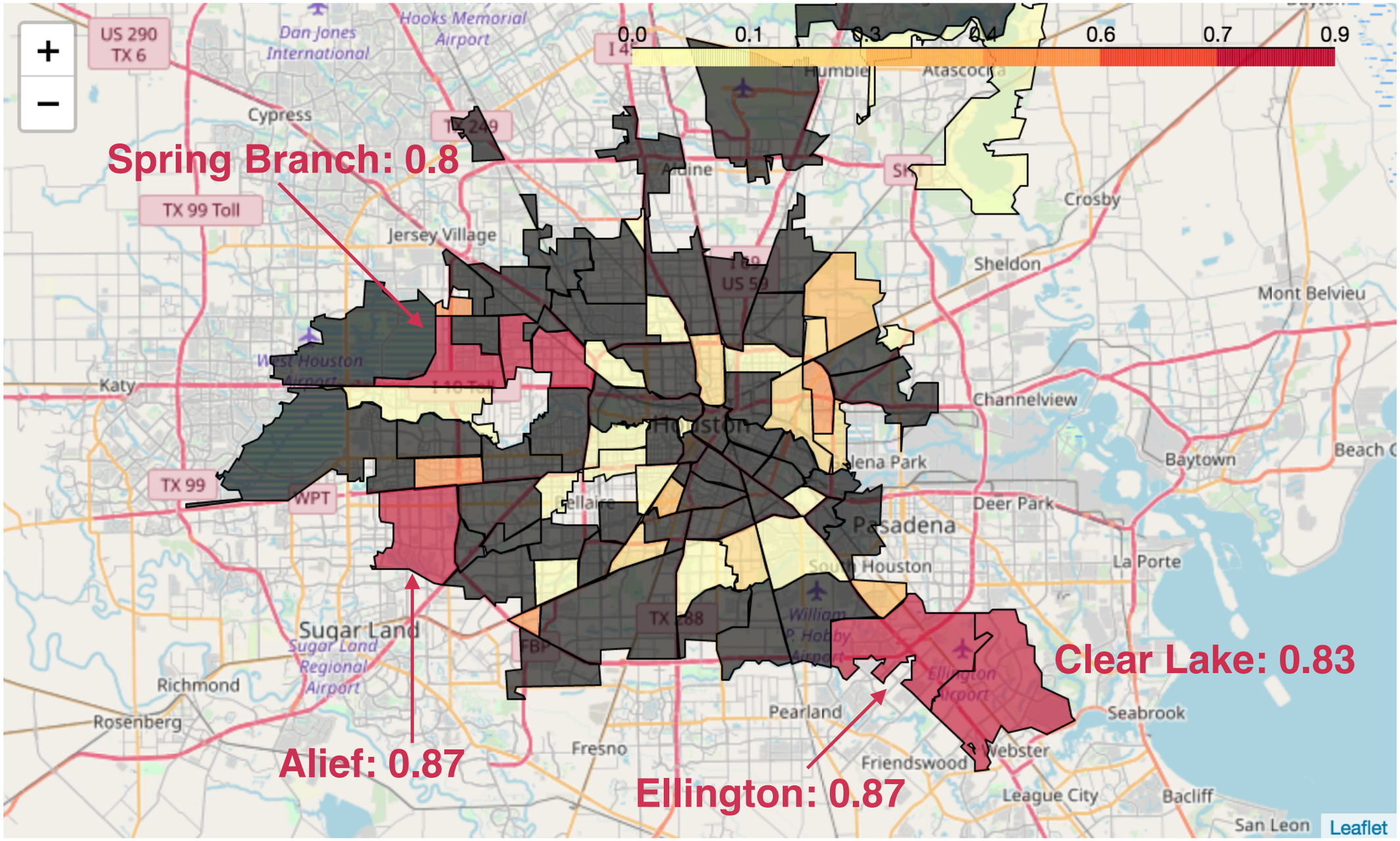
Best models found by grid search for optimal combination of hyperparameters, the accuracies are:

Random Forest: 0.858 ✅

Logistic Regression: 0.830

Support Vector Machines: 0.802

Apply best model to predict neighborhoods in Houston



Understand the underlying predictors in the model

	Top features	Top neighbors	Bottom neighbors
*	Mexican Restaurant	0.440476	0.0000
*	Vietnamese Restaurant	0.007937	0.0000
	Chinese Restaurant	0.000000	0.0000
	Italian Restaurant	0.000000	0.0000
	Asian Restaurant	0.000000	0.0000
	Seafood Restaurant	0.000000	0.0000
	American Restaurant	0.000000	0.0000
*	ACS Demographics/Race and ethnicity/Not Hispanic or Latino: Asian alone/Percentage	0.153000	0.1104
*	ACS Housing/Value of owner-occupied housing units/$300,000 to $399,999/Percentage	0.051500	0.0466
*	ACS Demographics/Population by age range/Female: 30 to 34 years/Percentage	0.076333	0.0718

Top neighborhoods have more:

- Mexican/Vietnamese Restaurant
- Asian residents
- Housing priced $300k-$400k
- Female aged 30-34

Conclusions and future direction

- Data-drive machine learning model can uncover some implicit features that do not come apparent
- Potential important features not included: crime data, competition, taste etc.
- Further refine to specific blocks within a neighborhood