# Chapter 2
# End to End Machine Learning

DuLi

2019 年 3 月 18 日

# 目录

# 1 Outline

Here are main steps you will go through:

1. Look at the big picture.
2. Get the data.
3. Discover and visualize the data to gain insights.
4. Prepare the data fpr machine learning algorithms.
5. Select a model and train it.
6. Fine-tune your model.
7. Present your solution.
8. Launch,monitor,and maintain your system.

# 2 Working with Real Data

Here are a afew places you can look to get data:

- Popular open data open repositories:
- UC Irvine Machine Learning Repositories.
- Kaggle datasets.
- Amazon's AWS datasets.
- Meta Portals(they list open data repositories)
- dataportals.org
- opendatamonitor.eu
- quandl.com
- Other pages listing many popular open data repositories

- Wikipedia's list of Machine Learning datasets.
- Quora.com question.
- Datasets subreddit.

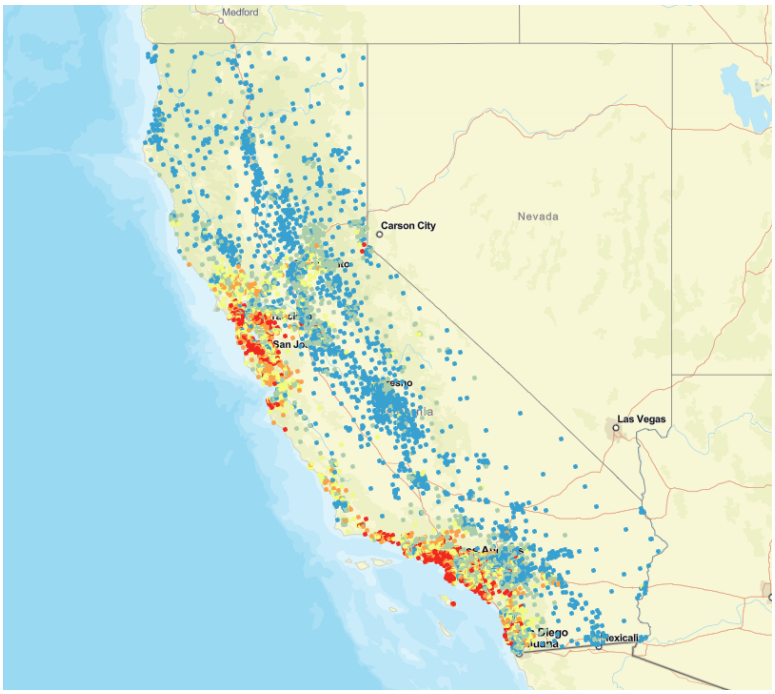


图 1: California Housing Prices Databases

# 3 Look at the Big Picture

## 3.1 Frame the Problem

The first question to ask you is what exactly is the bussiness objective;building a model is probably not the end goal.How do you expect use and benefit from this model?This is improtant because it will determine how you frame the problem,what algorithms you will select,what performance measure you will use to evaluate your model,and how much effort you should spend tweaking it. Your model output (a predicting of a district's median housing price) will be fed to another Machine Learning system along with

many other signals.This downstream system will determine whether it is worth investing in a given area or not.

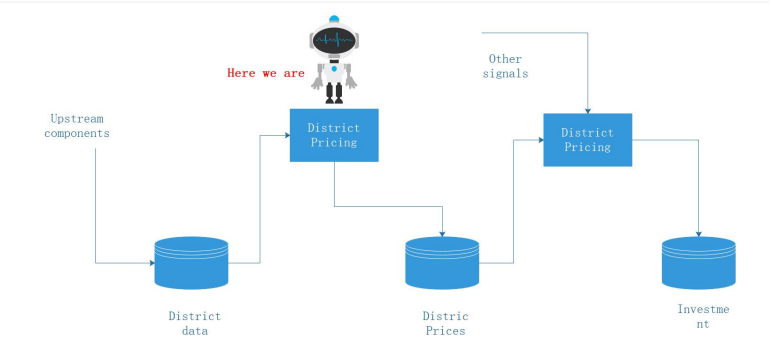


图 2: A machine learning pipeline for real estate investment