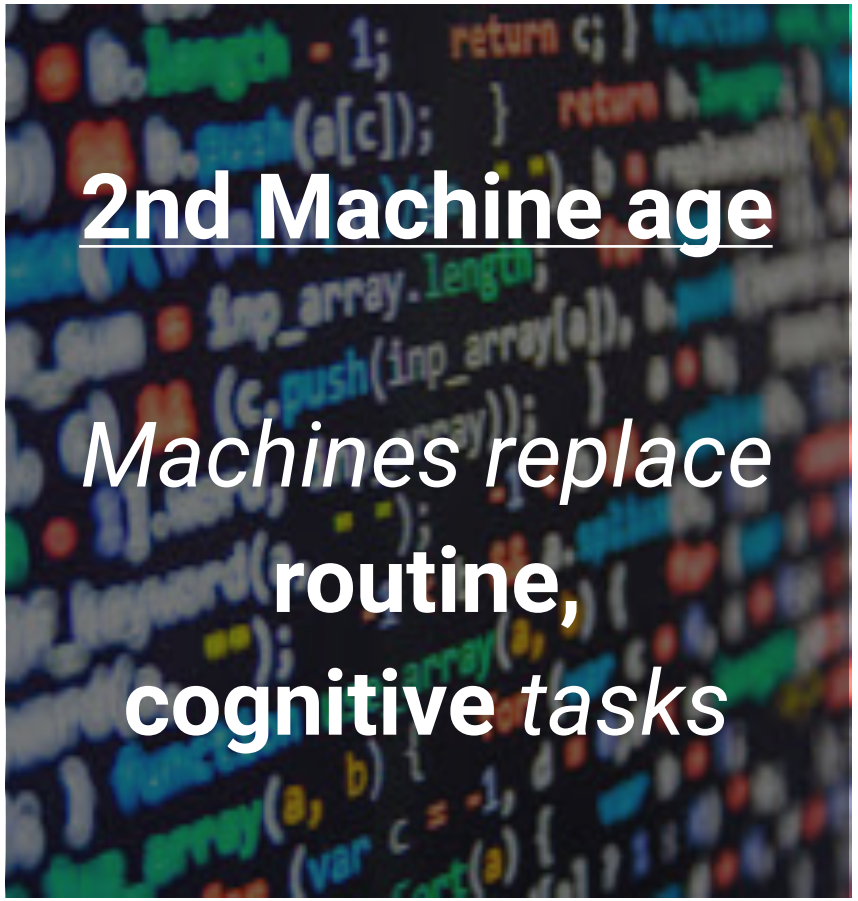
Machines learning your jobs

Taichi Hoshino

13 Dec 2016

Is your job at risk of being automated?

Machines and Humans



“47% of American jobs are at high risk of being automated”

- *Frey and Osborne (2013)* -

The process and libraries

1

***Scraped** 1.5 million job descriptions*

indeed

Se

amazon web services

mongoDB.

2

***NLP** to extract features*

python
Natural Language Analyses with NLTK

gensim

3

***Classification** of jobs into 500 categories*

scikit learn

4

*Interactive **front-end** development*

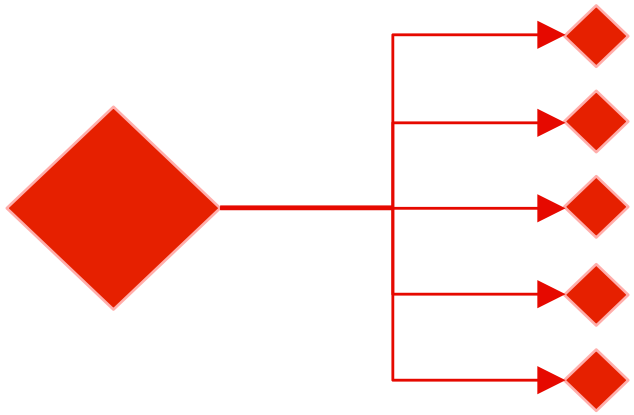
Bootstrap

The data science problem

How do you build a model to accurately classify job descriptions into one of **500** job categories?

The data science problem

How do you build a model to accurately classify job descriptions into one of **500** job categories?

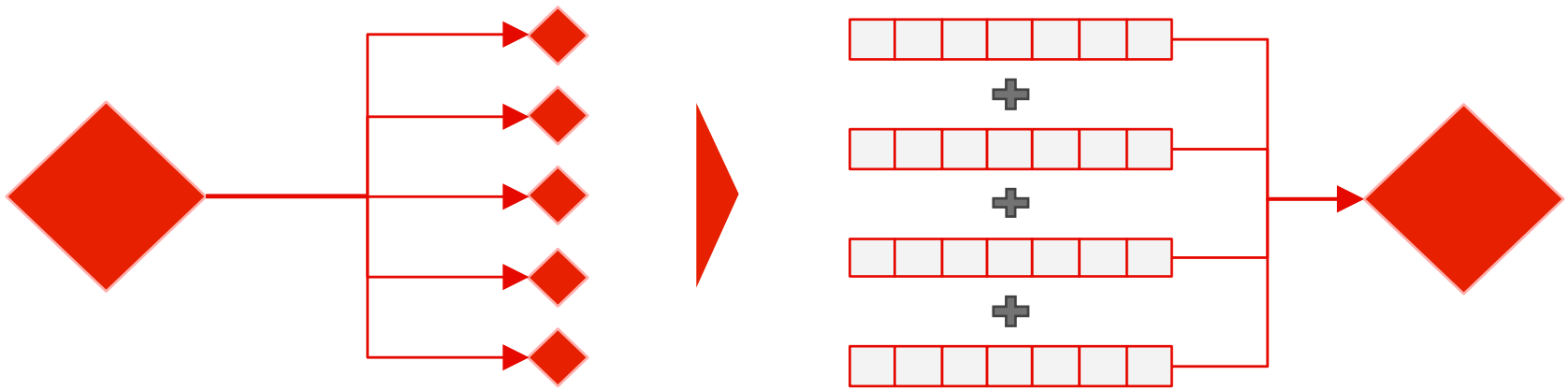


High recall model
Accuracy = 65%
Recall = 66%
Precision = 65%

Top candidates
Candidates = 5

The data science problem

How do you build a model to accurately classify job descriptions into one of **500** job categories?

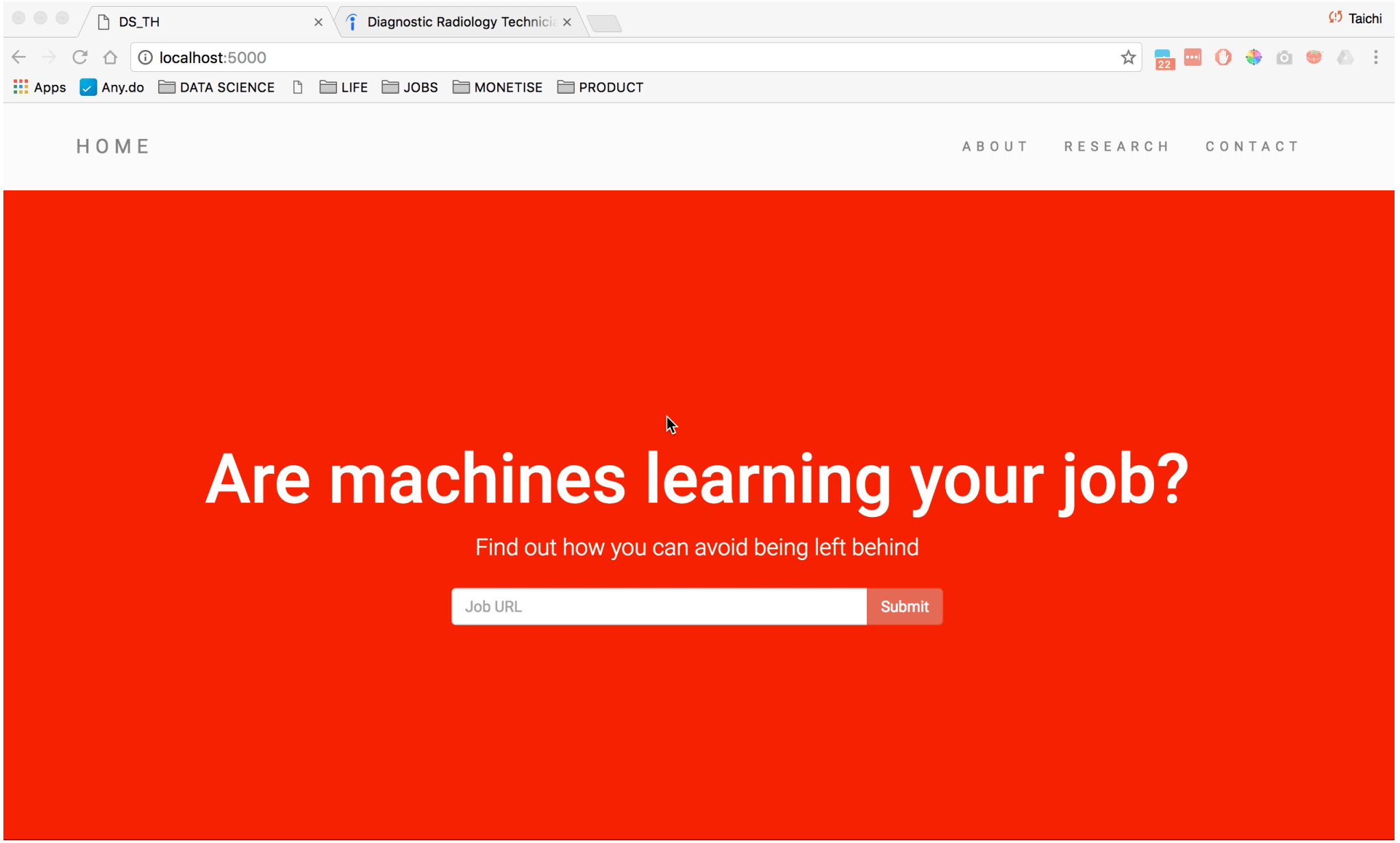


High recall model
Accuracy = 65%
Recall = 66%
Precision = 65%

Top candidates
Candidates = 5

Feature eng.
Candidate vectors
Doc vectors
Diff vectors
Similarity vectors

Precision model
Accuracy = 83%
Recall = 84%
Precision = 86%


DS_TH
Diagnostic Radiology Technici
Taichi
localhost:5000
Apps
Any.do
DATA SCIENCE
LIFE
JOBS
MONETISE
PRODUCT
HOME
ABOUT
RESEARCH
CONTACT
Are machines learning your job?
Find out how you can avoid being left behind
Job URL
Submit

What next?

- **Get more granular** – **binary standardizations** allows accurate, high multi-class text classifications
- **Understand use case** – explore the **pivotal moments** for workers to utilize a tool like this
- **Convert** – into a **plug-in** to sit on top of career and vocational websites

Taichi Hoshino

taichi27@gmail.com

github.com/th27

www.linkedin.com/in/taichi-hoshino