

Transitioning from Academia to Data Science

Donald Lee-Brown — Insight Data Science



Insight Data Science



INSIGHT DATA SCIENCE FELLOWS PROGRAM

An intensive 7 week post-doctoral training fellowship
bridging the gap between academia & data science

APPLY NOW



READ THE WHITE PAPER



Want to be notified of future dates? [CLICK HERE](#)



Alex Smolyanskaya, Data Scientist at Stitch Fix,
Insight 2015 Fellow

insightdatascience.com

Insight Fellows are Data Scientists and Data Engineers at:

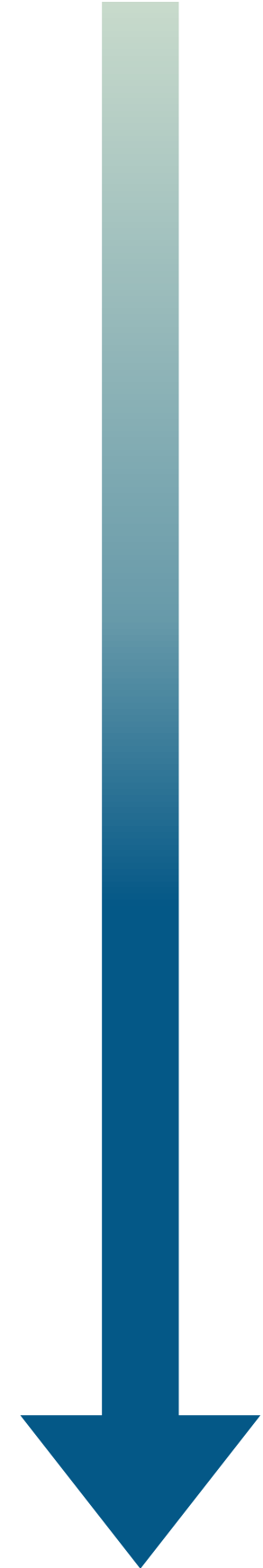


Seattle • Silicon Valley • New York • Boston
+ many others...

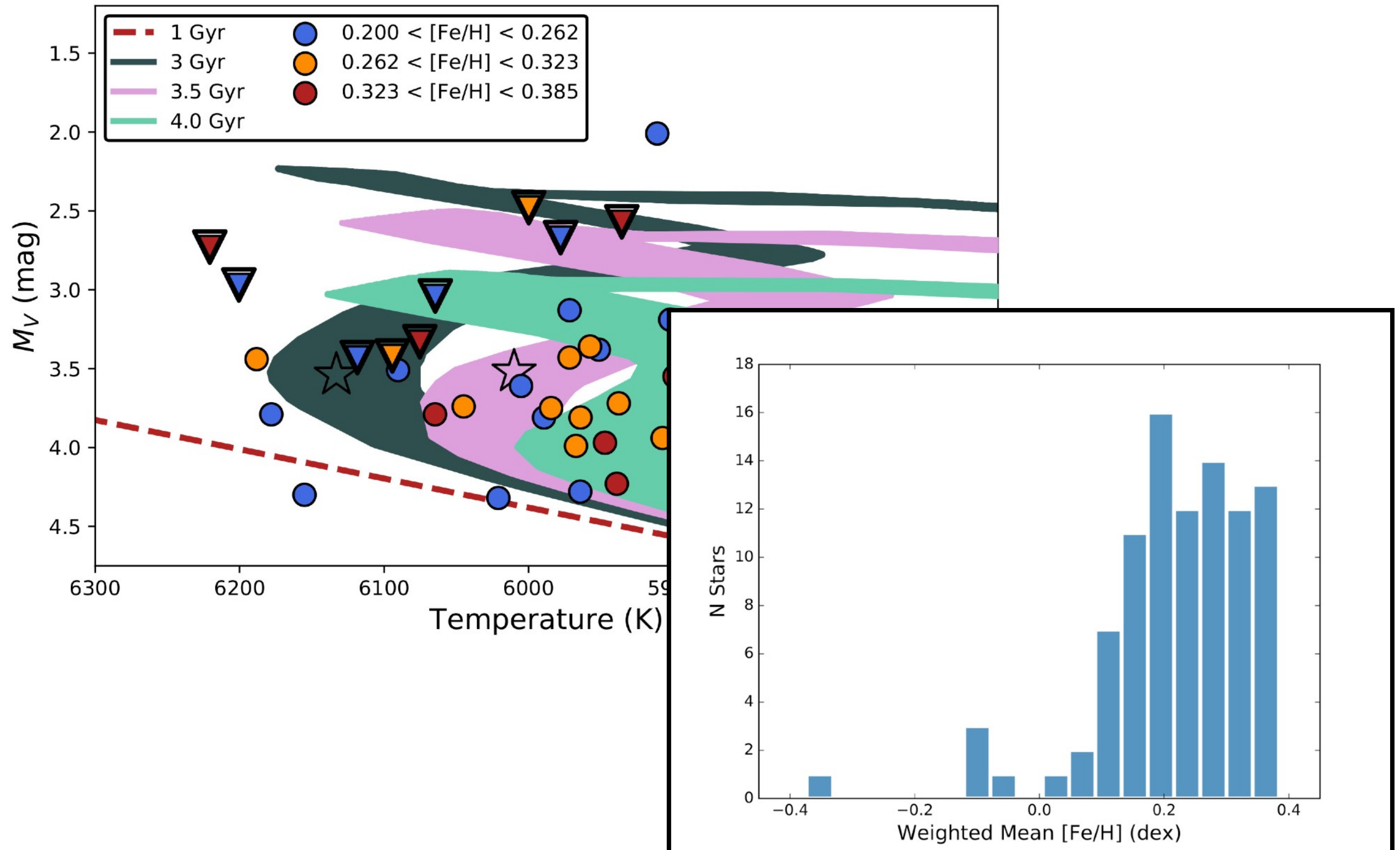
My Path

2013-
2018

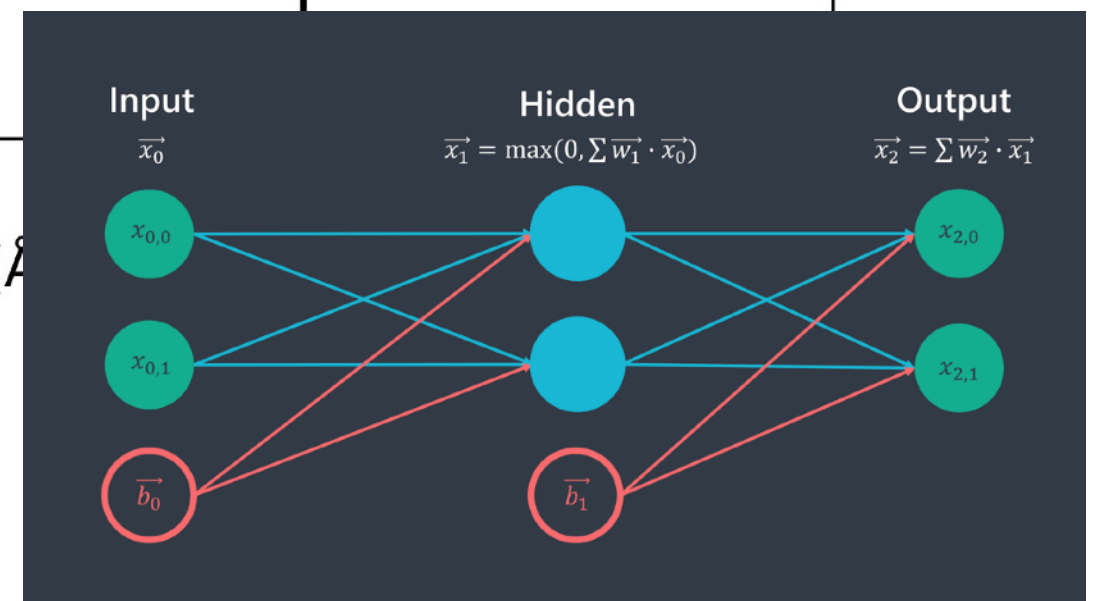
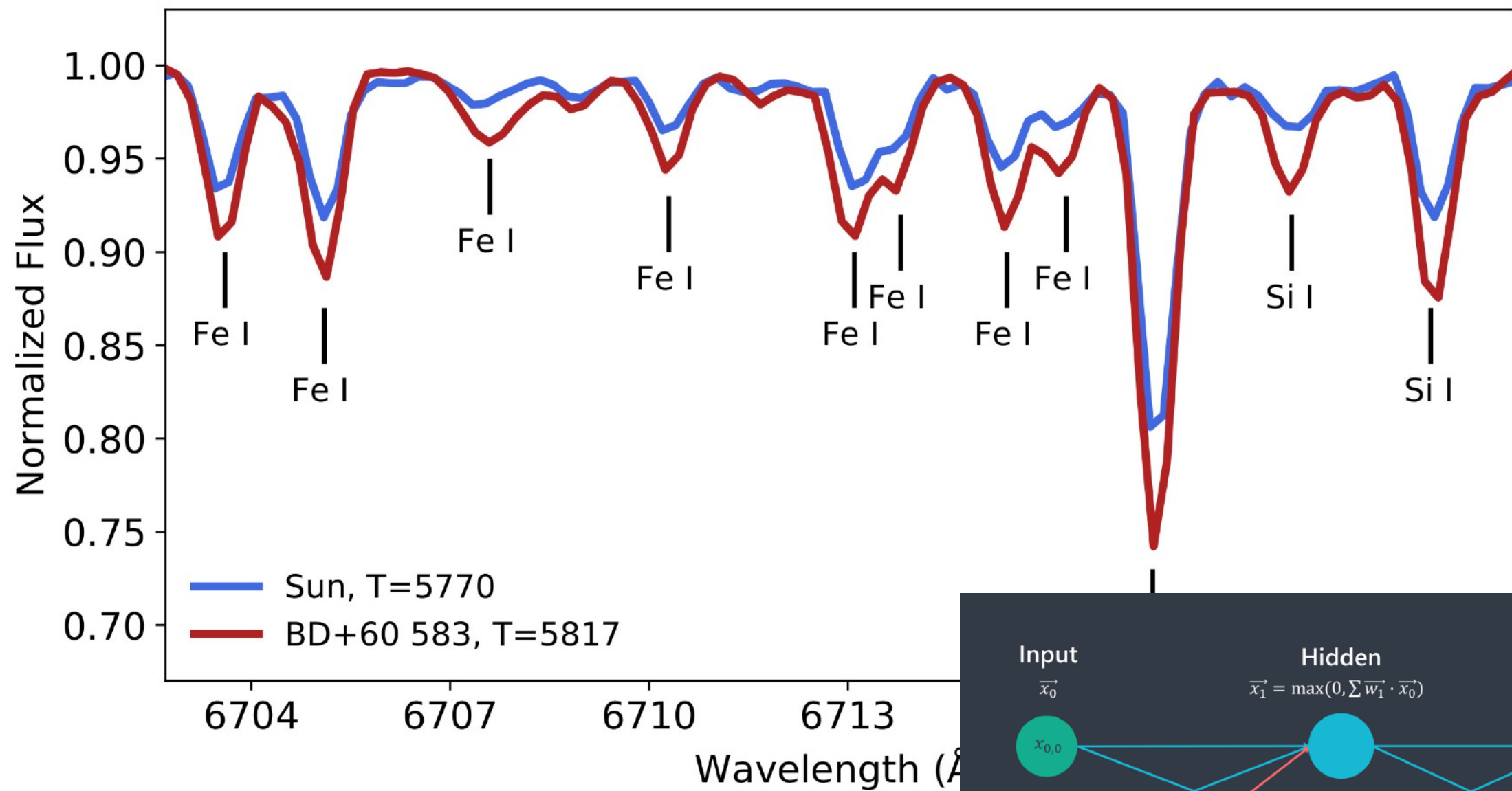
KU THE UNIVERSITY OF
KANSAS



My Path



My Path



"Data Science"

https://www.google.com/search?q=angel+devil&espv=2&biw=1324&bih=694&tbm=isch&tbo=u&source=univ&sa=X&ved=0ahUKewja4NDgnbLLAhULwWMKHXYeBKEQs...

Clipart

Project

Icon

Word Cloud

Infographic

Table

Trial	Block #1	Block #2
#1	15 cm	10 cm
#2	40 cm	15 cm
#3	10 cm	35 cm
Average	21.2 cm	20.2 cm

Hacking Skills

Machine Learning

Math & Statistics Knowledge

Data Science

Traditional Research

Danger Zone!

Substantive Expertise

Byte

Business Data Network

Online Science

Service Learning System

Intelligence

Success Web Processing Analytics Manage

Storage Process Mining Net

BIG Data

Technology Cyberspace Media

Private

Model

Research Commerce Worldwide

Analysis Security Data Cleaning

Trend Teamwork Economic

big

hadoop

using

show

talk

cassandra

visualization

microwork

mapreduce

programming

apache

learn

platform

crowdsourcing

analytics

large

provides

research

recommendation

distributed

case

building beyond hidden

visualizations

find

source

machine

optimization

tutorial

design

applications

time

describe

problem

hive

mining

world

new web

real-time

development

experience

data

business

analysis

analytics

develop

learning

management

including new

models

problems

related

using

years

last

strong

working

team

plus

statistical

skills

technical

research

software

systems

support

projects

science

related

Pattern

recognised

Machine Learning

Statistical

Visualisation

DATA science

Data Mining

MODERN DATA SCIENTIST

Math & Statistics

Programming & Database

Communication & Visualization

Human Problem Solving & Creativity

Topical Analytical Techniques

Pattern Recognition

Statistical Analysis

Machine Learning

Big Data

Cloud Computing

Visualization

Communication

Human Problem Solving & Creativity

linked in (1).ipynb

linkedin.ipynb

Show All

My Path

coursera

Explore ▾

Q What do you want to learn?

Browse > Data Science > Machine Learning

Machine Learning

Enroll

Starts Jan 07

Financial aid available

Offered By

Stanford

About Syllabus Reviews Instr

Overview

Repositories 17

Stars 2

Follow

Pinned repositories

ANNA

Automatically parameterize stellar spectra using a convolutional neural network

Python ★ 2 🍴 3

nesar/RockTheBoar

Convolutional Neural Network pipeline for pixel-wise



Donald Lee-Brown
dleebrown

My Path



2013-
2018

KU THE UNIVERSITY OF
KANSAS

June
2018



INSIGHT



Common Ground

The job of a scientist:

Common Ground

The job of a scientist:

- Start with a question

Common Ground

The job of a scientist:

- Start with a question
- Collect and clean data

Common Ground

The job of a scientist:

- Start with a question
- Collect and clean data
- Use programming and statistics knowledge to discriminate between signal and noise

Common Ground

The job of a scientist:

- Start with a question
- Collect and clean data
- Use programming and statistics knowledge to discriminate between signal and noise
- Answer the question and connect to other work

Common Ground

The job of a scientist:

- Start with a question
- Collect and clean data
- Use programming and statistics knowledge to discriminate between signal and noise
- Answer the question and connect to other work
- Convey results to scientific community

Common Ground

The job of a data scientist:

- Start with a question
- Collect and clean data
- Use programming and statistics knowledge to discriminate between signal and noise
- Answer the question and **distill actionable insights**
- Convey results to **the team/company/investors**

Who are Data Scientists?

Data science is a very broad term

Analytics

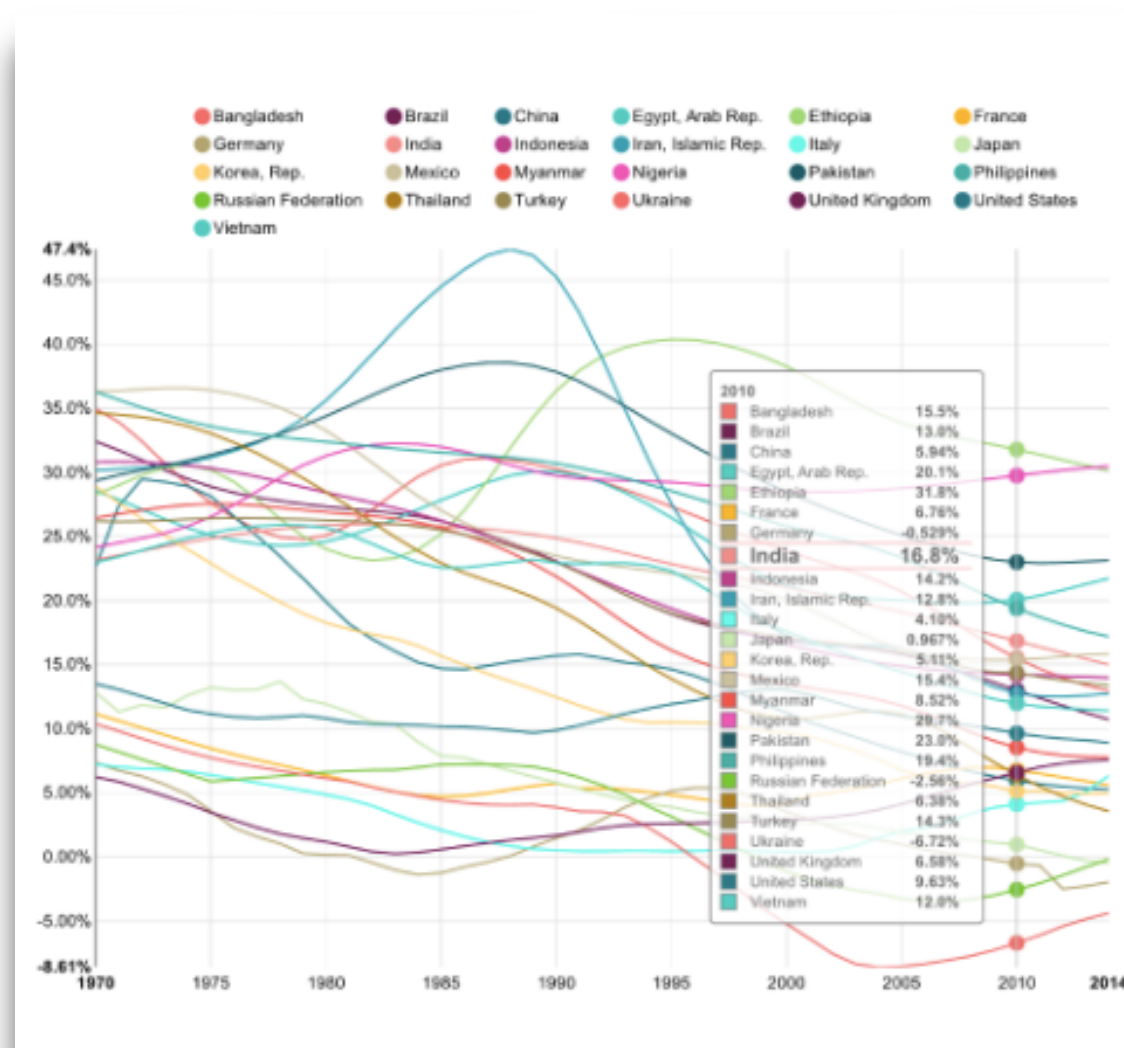
Data
Products

Machine
Learning

Engineering

Who are Data Scientists?

- Analytics, experimentation, visualization, and data stories

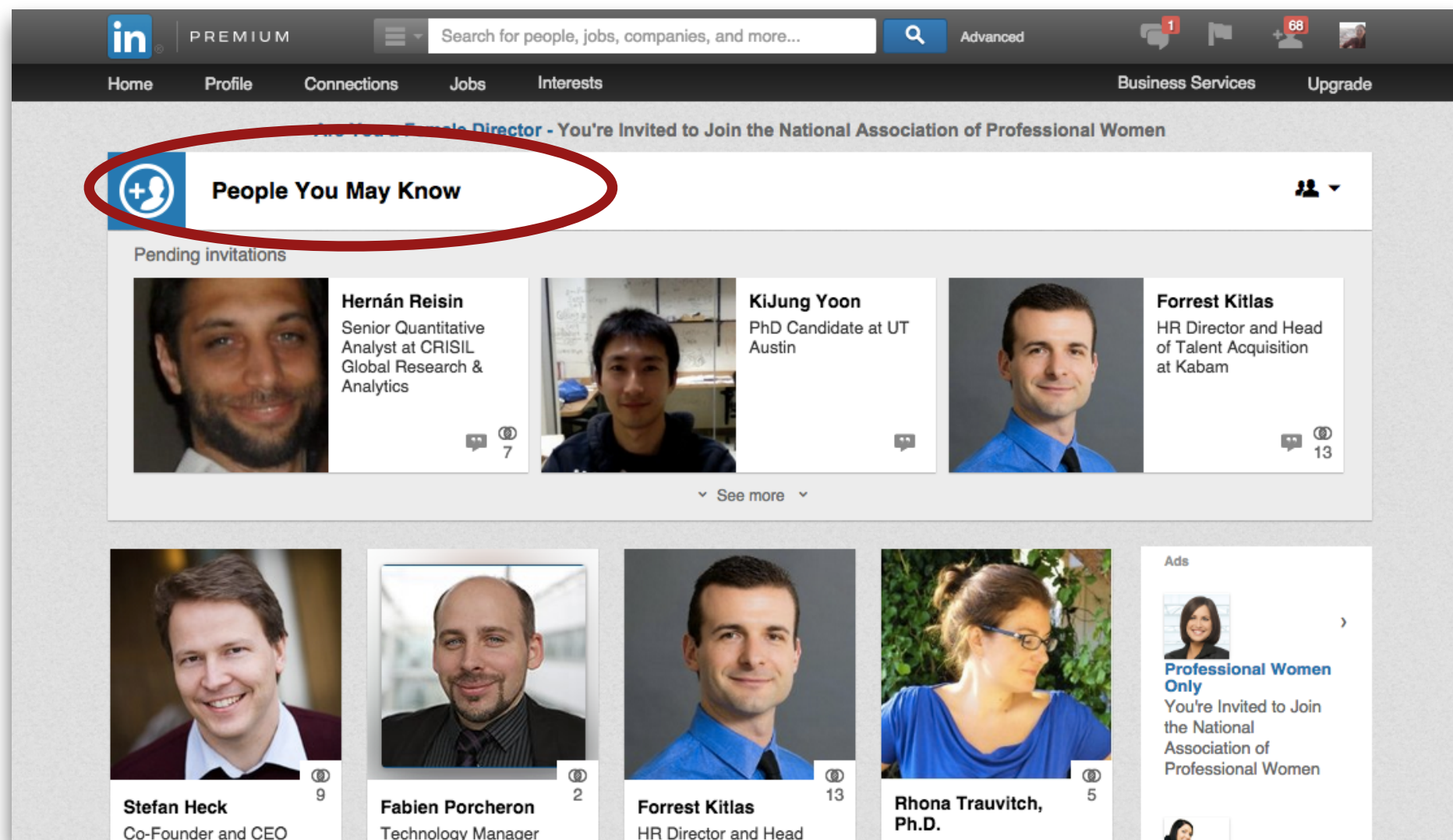


source: <https://github.com/airbnb/caravel>



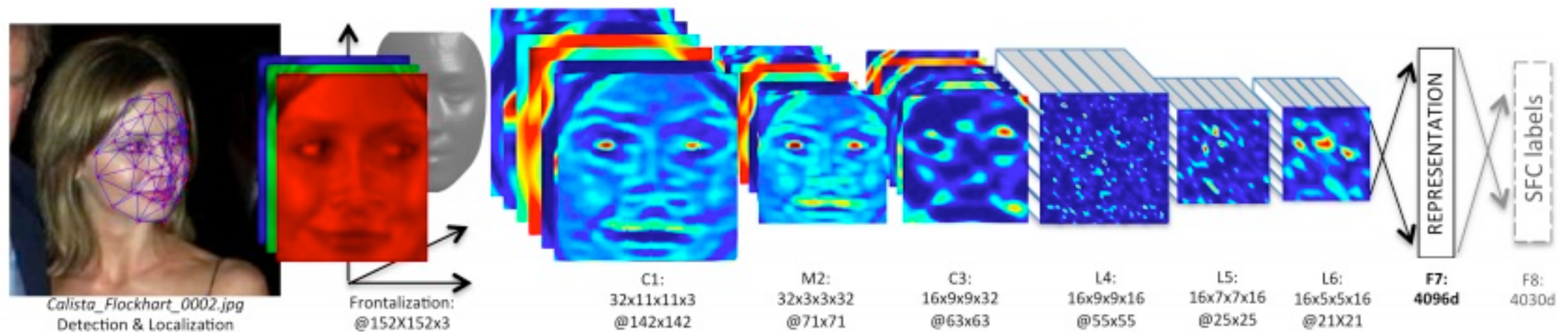
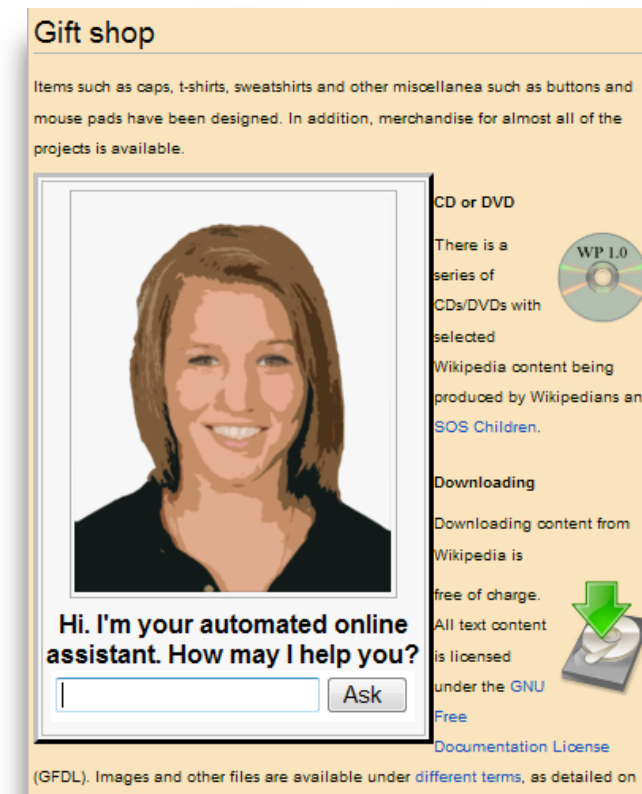
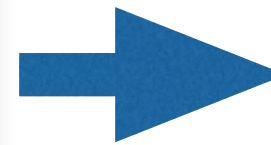
Who are Data Scientists?

- Data Products and Dashboards



Who are Data Scientists?

- Machine Learning



Who are Data Scientists?

**Great data scientists are
fundamentally great scientists!**



**Smart
individuals
with the right
skills**

The Gap



**High impact
teams**

The Gap

Challenges academics face when making the transition

The Gap

Challenges academics face when making the transition



The Gap

Challenges academics face when making the transition



Results published
O(once a year)



**Results needed
O(weekly), MVP**

The Gap

Challenges academics face when making the transition



Results published
O(once a year)



**Results needed
O(weekly), MVP**

Analysis for
scientific community



**Results need to
impact bottom line**



**Smart
individuals
with the right
skills**

Bridging The Gap



Tools



Evidence



Team



Time



**High impact
teams**

Thinking about DS?

Great! But how to bridge the gap?

Thinking about DS?

Great! But how to bridge the gap?

- Learn industry tools and techniques

Thinking about DS?

Great! But how to bridge the gap?

- Learn industry tools and techniques
- Build evidence with projects

Thinking about DS?

Great! But how to bridge the gap?

- Learn industry tools and techniques
- Build evidence with projects
- Figure out where you want to go!

Thinking about DS?

Learn industry tools and techniques

- Programming [python]
- Machine Learning [sklearn]
- Statistics [scipy]
- Databases [SQL]
- Front end tools [dash]
- Distributed Computing [spark]

kaggle

coursera

 stackoverflow

Thinking about DS?

Build evidence with projects

- Highlight relevant skills
- Communicate actionable insights
- Show creativity as a scientist
- Write blog posts
- Put code on Github
- Contribute to open source libraries

GitHub

Medium



Thinking about DS?

Understand the landscape of DS

- Big company vs. startup
- Engineering vs. prototyping
- Analytics vs. ML
- Company culture
- Diversity and inclusion

The Meetup logo, featuring the word "Meetup" in a red, cursive script font.The LinkedIn logo, consisting of the word "Linked" in black and "in" in white inside a blue square.The Glassdoor logo, featuring a green square icon with a white outline and the word "glassdoor" in black, followed by a trademark symbol.

Thinking about DS?

Great! But how to bridge the gap?

- Learn industry tools and techniques
- Build evidence with projects
- Figure out where you want to go!



Smart
individuals
with the right
skills



INSIGHT



High impact
teams

Program Timeline

7 Weeks

4-8 Weeks

Week 1

Weeks 2-4

Weeks 5-7

Weeks 8+



Intro
Week

Data
Project

Company
Visits

Job
Interviews



7 Weeks

\$0

Full time

need based
scholarship

Thank you!

INSIGHT

Apply and find out more on the web:

(applications due in March for summer session)

insightdatascience.com

Email me:

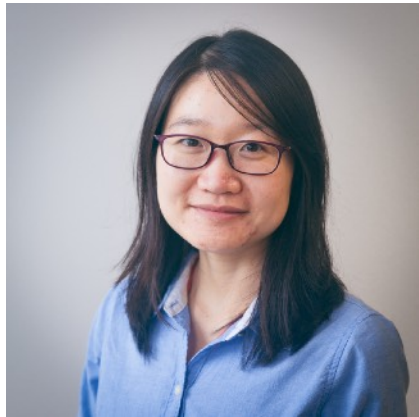
donald@insightdatascience.com

Hang out with us Tuesday at 7pm!

605 1st Ave, Suite 200

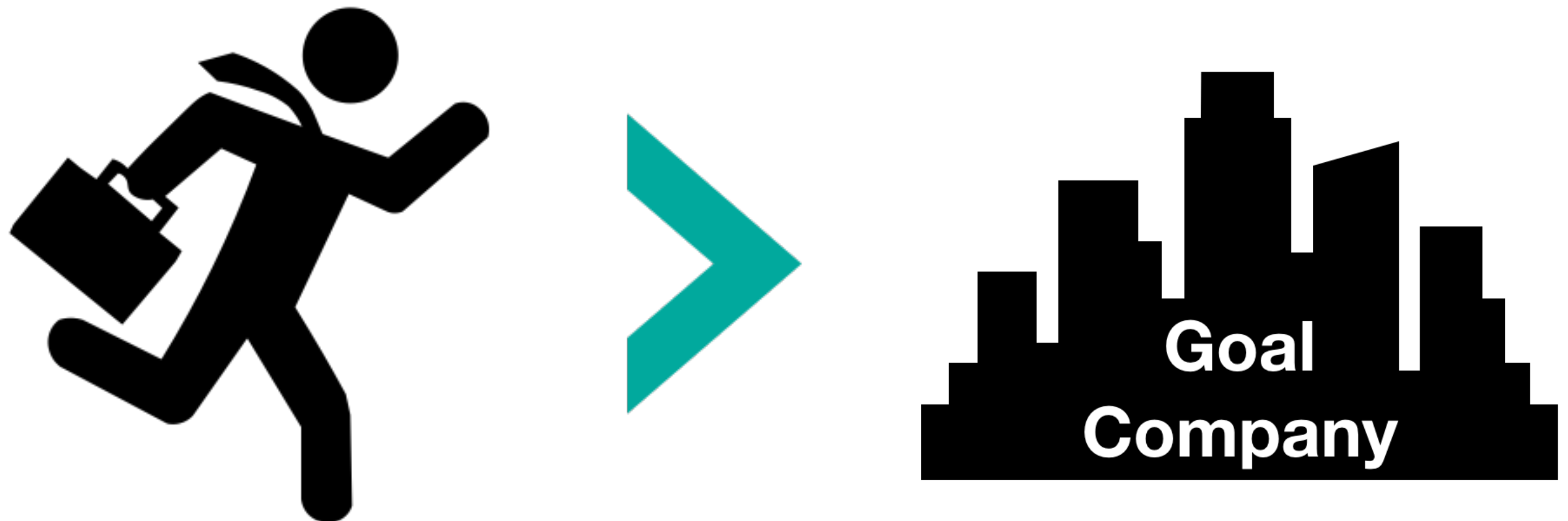
Light food, meet data scientists, panel on data ethics

Projects



Yanxia Li (Insight Fellow Fall 2018)

"Company Snapshot"



Glassdoor reviews serve as **road signs** pointing toward our goal company!

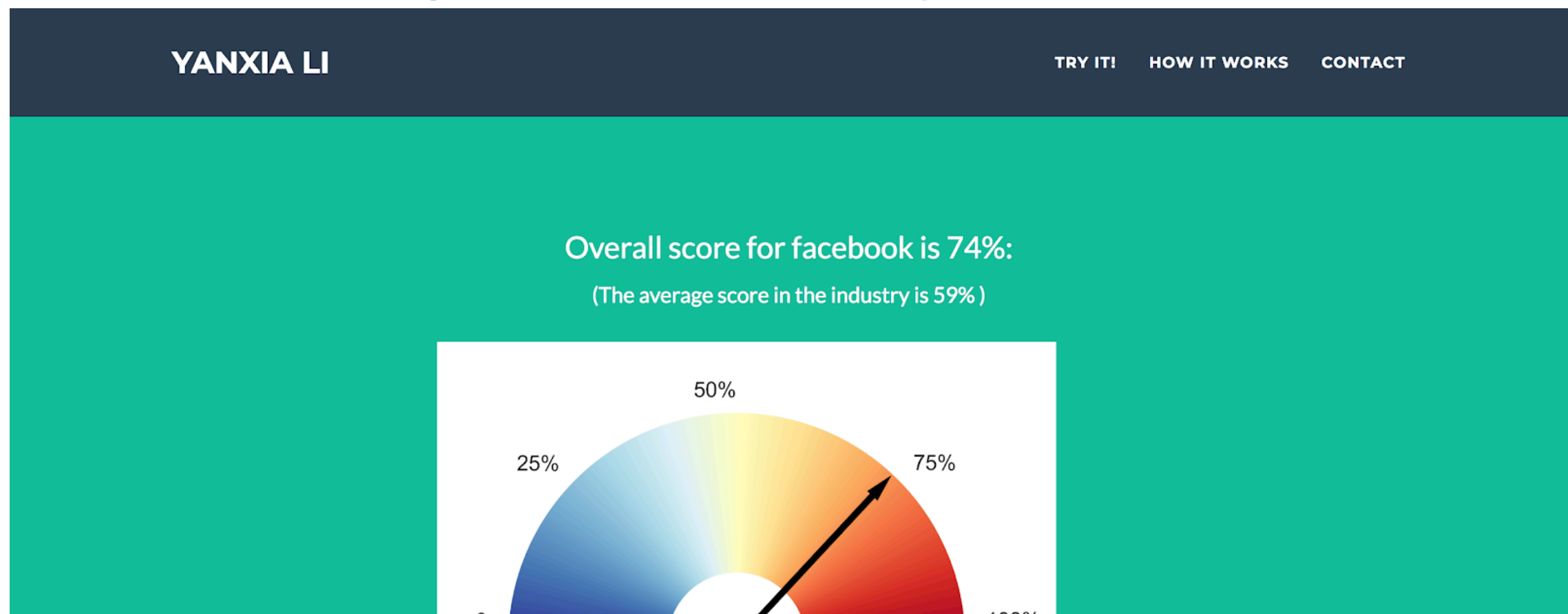
Projects

CompanySnapshot as a solution

Thinknum Glassdoor Data

- 2008 - 2018
- Unstructured
- Total 2.6 million reviews, 35 features, 4931 IPO companies

Topic Modeling + Sentiment Analysis -> Intuitive Web App



Projects



Nitya Mandyam (Insight Fellow Fall 2018)

“YouTopia”

Menu Search

Bloomberg

Sign In

Technology

Ban Sought in U.S. for YouTube Ads ‘Masquerading’ as Kids’ Shows

By [Todd Shields](#)

October 21, 2016, 12:01 AM EDT

YouTube pulls ads on 2 million inappropriate children's videos

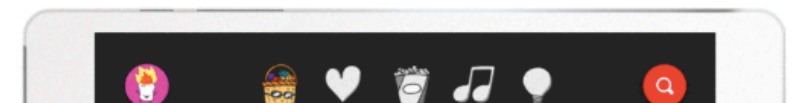
It's also investigating autocomplete search after suggestions for child exploitation content

HOME > DIGITAL > NEWS

SEPTEMBER 13, 2018 10:41AM PT


YouTube Kids App Now Lets Parents Handpick All Videos and Channels

By [TODD SPANGLER](#)



Projects

Word Embeddings + Supervised Learning -> Chrome Extension





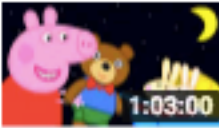

The video player shows a scene from a Peppa Pig episode. George Pig is standing on a dark, rocky surface, looking distressed. A large, pink, tentacle-like creature is visible in the foreground. The video player interface includes a progress bar at 5:35 / 11:02, a 'Settings' button, and a 'Transcript' overlay on the right side.

Transcript

- 00:03 evil t-rex is breathing fire George pig
- 00:10 and blue t-rex have come to put out the
- 00:13 fire with their special powers
- 00:16 it's not working they need backup here
- 00:23 comes a firetruck to help
- 00:31 Oh deal if George got too close to the
- 00:35 fire George needs some medicine and

English (auto-generated)

Up next AUTOPLAY

-  Peppa Pig English Episodes | Peppa Pig...
Peppa Pig - Official Ch
17M views
-  Peppa Pig English Episodes in 4K |...
Peppa Pig - Official Ch
24M views
-  Peppa Pig English Full Episodes Compilation...
Peppa Pig English
4.6M views
-  PEPPA PIG gets a new toy House in this Kids...

peppa pig english episode #Peppa Pig Crying Compilation #Peppa Pig Doctor, spiderman, Truck #pepp

4,140 views

peppa pig english episodes
Published on Oct 21, 2016

peppa pig english episode #Peppa Pig Crying Compilation #Peppa Pig Doctor, spiderman, Truck #pepp .

SHOW MORE

SUBSCRIBE 33