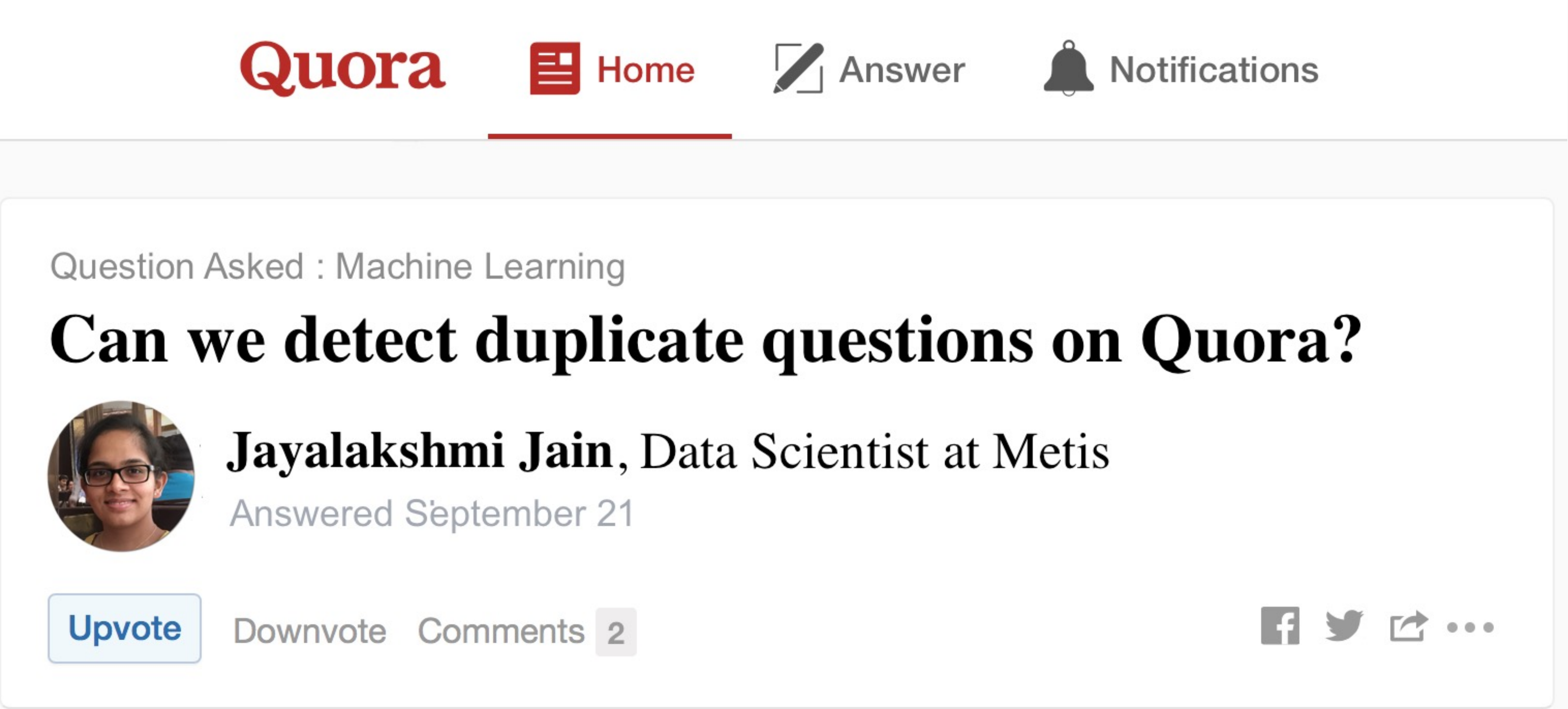
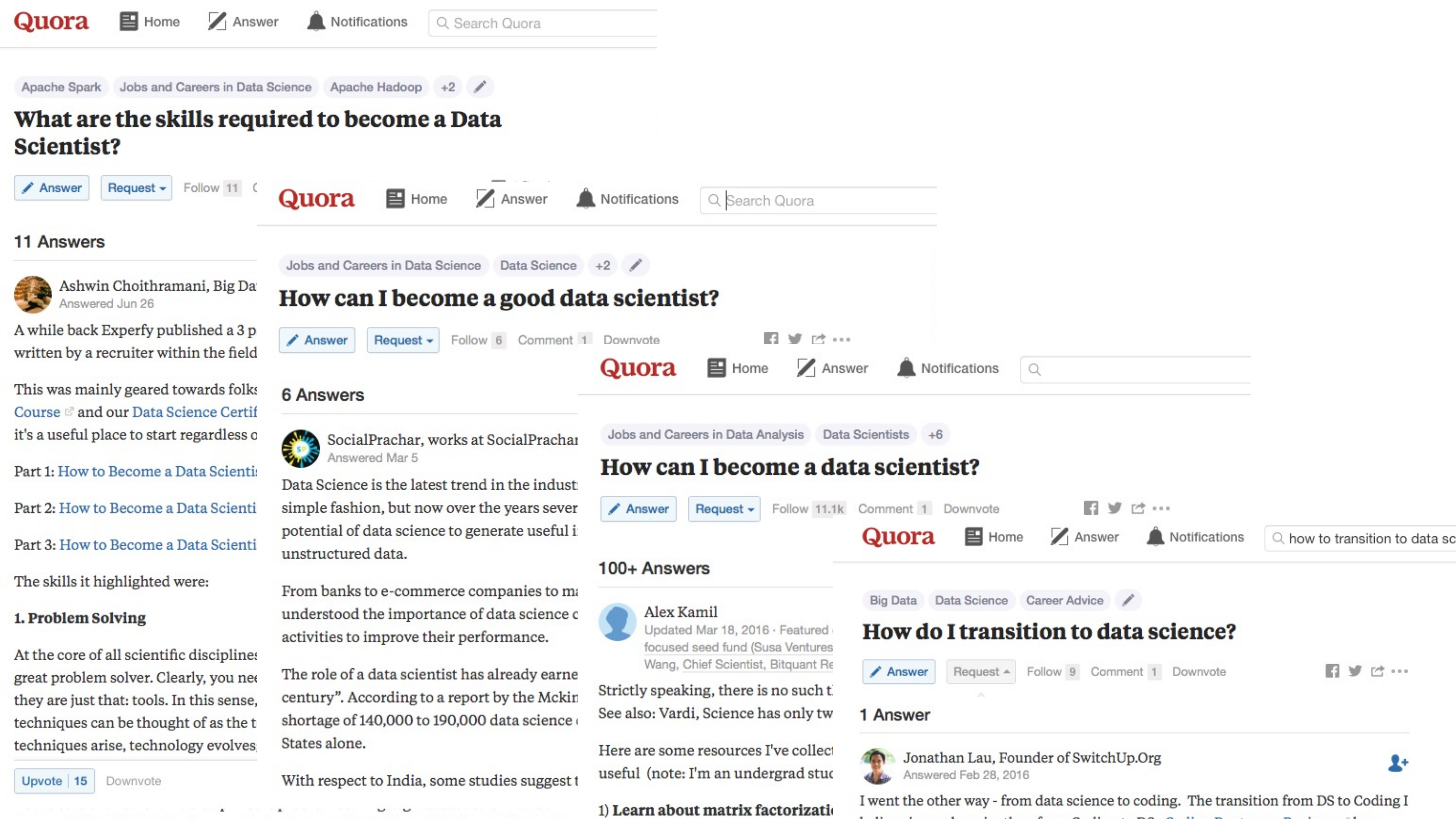


Question Asked : Machine Learning

Can we detect duplicate questions on Quora?

Jayalakshmi Jain, Data Scientist at Metis

Answered September 21



Quora Home Answer Notifications Search Quora

Apache Spark | Jobs and Careers in Data Science | Apache Hadoop | +2

What are the skills required to become a Data Scientist?

Answer | Request | Follow 11 | (

11 Answers

Ashwin Choithramani, Big Da
Answered Jun 26

A while back Experfy published a 3 p
written by a recruiter within the field

This was mainly geared towards folks
Course and our Data Science Certif
it's a useful place to start regardless o

Part 1: How to Become a Data Scienti

Part 2: How to Become a Data Scienti

Part 3: How to Become a Data Scienti

The skills it highlighted were:

1. Problem Solving

At the core of all scientific disciplines
great problem solver. Clearly, you nee
they are just that: tools. In this sense,
techniques can be thought of as the t
techniques arise, technology evolves

Upvote 15 | Downvote

Quora Home Answer Notifications Search Quora

Jobs and Careers in Data Science | Data Science | +2

How can I become a good data scientist?

Answer | Request | Follow 6 | Comment 1 | Downvote

6 Answers

SocialPrachar, works at SocialPrachar
Answered Mar 5

Data Science is the latest trend in the indust
simple fashion, but now over the years sever
potential of data science to generate useful i
unstructured data.

From banks to e-commerce companies to ma
understood the importance of data science c
activities to improve their performance.

The role of a data scientist has already earne
century". According to a report by the Mckin
shortage of 140,000 to 190,000 data science
States alone.

With respect to India, some studies suggest t

Quora Home Answer Notifications

Jobs and Careers in Data Analysis | Data Scientists | +6

How can I become a data scientist?

Answer | Request | Follow 11.1k | Comment 1 | Downvote

100+ Answers

Alex Kamil
Updated Mar 18, 2016 · Featured
focused seed fund (Susa Ventures
Wang, Chief Scientist, Bitquant Re

Strictly speaking, there is no such t
See also: Vardi, Science has only tw

Here are some resources I've collect
useful (note: I'm an undergrad stud

1) **Learn about matrix factorizati**

Quora Home Answer Notifications how to transition to data sc

Big Data | Data Science | Career Advice

How do I transition to data science?

Answer | Request | Follow 9 | Comment 1 | Downvote

1 Answer

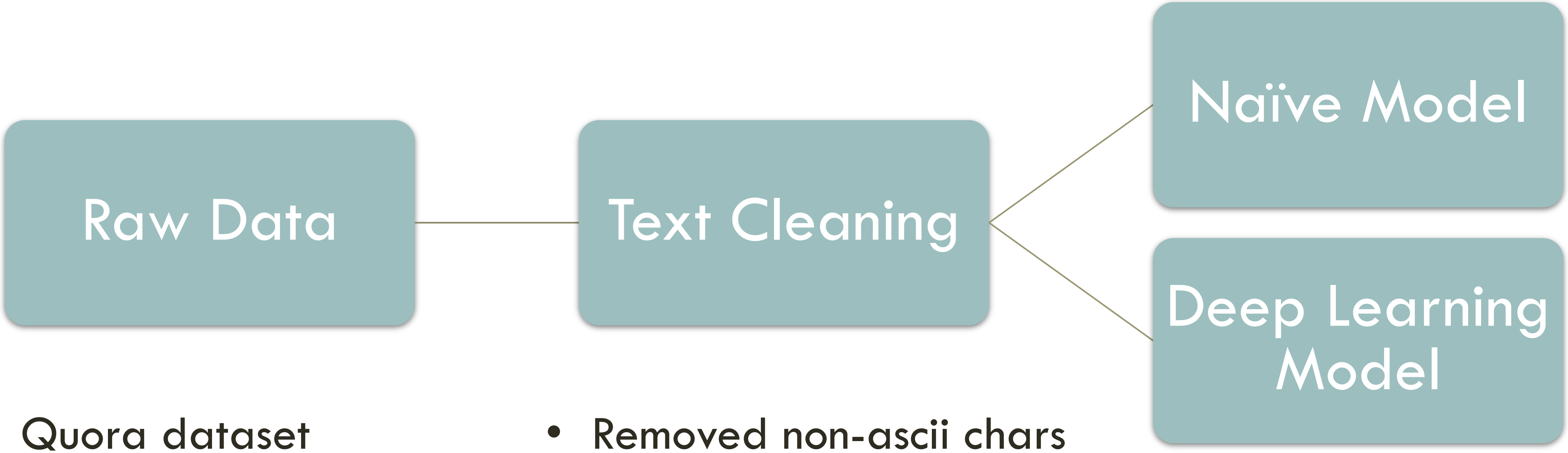
Jonathan Lau, Founder of SwitchUp.Org
Answered Feb 28, 2016

I went the other way - from data science to coding. The transition from DS to Coding I

OBJECTIVE

- To detect duplicate questions on Quora using Deep Learning techniques

OVERVIEW



Raw Data

Text Cleaning

Naïve Model

Deep Learning Model

- Quora dataset
- 400K+ question pairs
- Every pair is labeled as duplicate or not by Quora moderators

- Removed non-ascii chars (e.g. $\ddot{a}$, $\grave{a}$, $\hat{a}$), stop words (e.g the, is, or)
- Lemmatization {danced, dancing, dances} → dance

NAÏVE MODEL

Step 1 : Vectorizes the sentences

Step 2 : Measure the similarity

ISSUES WITH NAÏVE MODEL

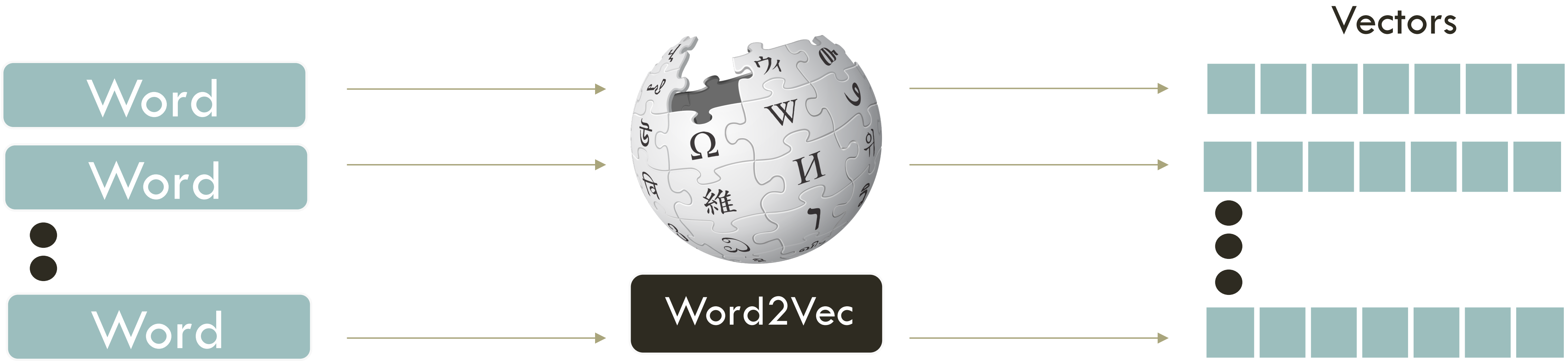
- Falters when synonyms or associated concepts are used
- Doesn't consider the ordering of the words

DEEP LEARNING MODEL

Step 1: Use word embeddings from pre-trained model

Step 2: Feed word embeddings to LSTM network

- Word embeddings map words to their underlying meaning



Vectors

Word

Word

Word

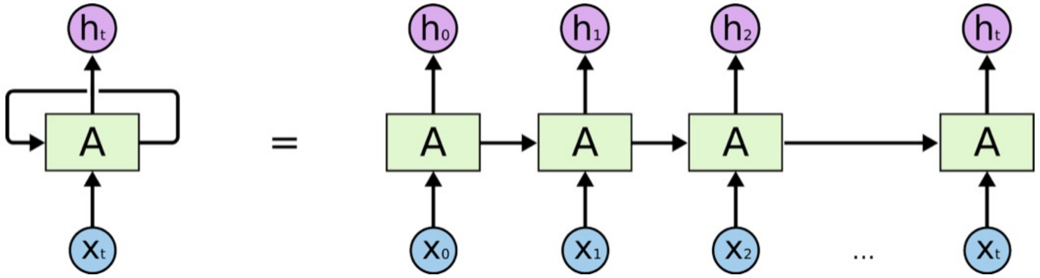
Word2Vec

DEEP LEARNING MODEL

Step 1: Use word embeddings from pre-trained model

Step 2: Feed word embeddings to LSTM network

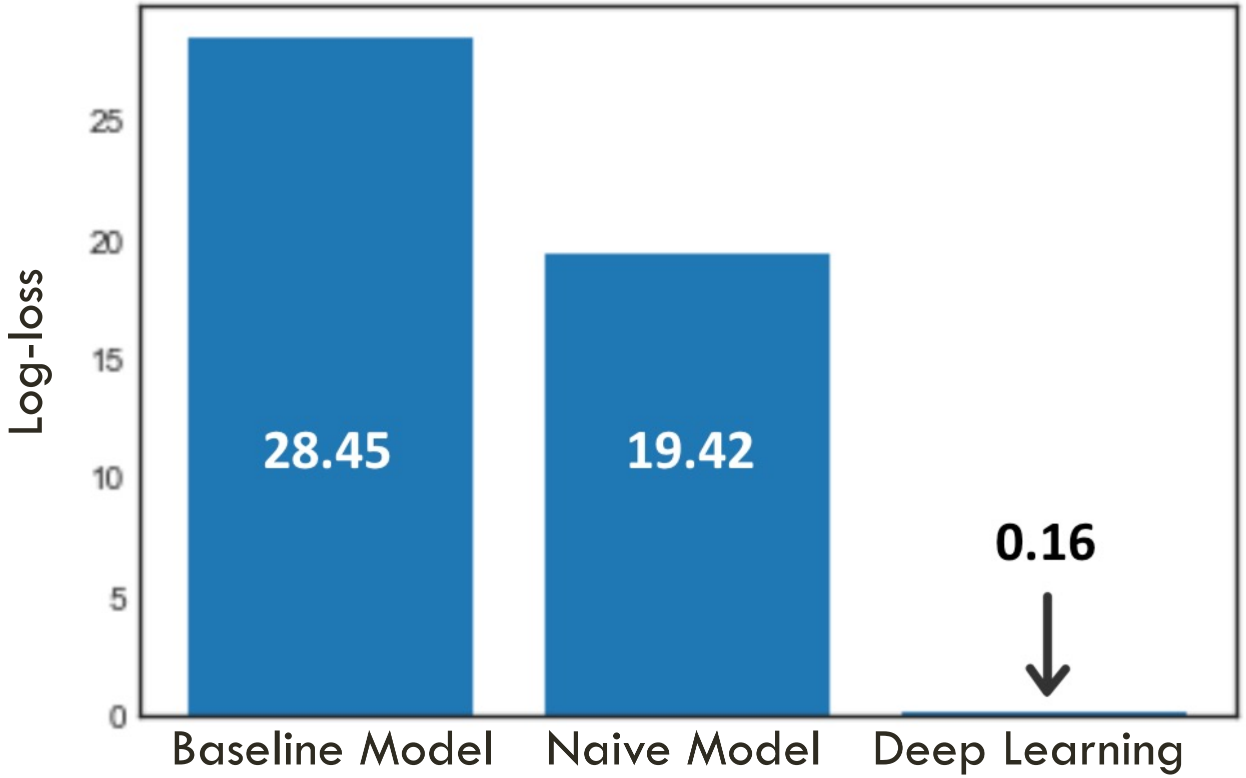
- Long-short term memory
- Special cells in Recurrent Neural Networks that remembers values over arbitrary intervals



An unrolled recurrent neural network.

MODEL EVALUATION

- Kaggle competition evaluated on log-loss
- Model is in the top 15% of all submissions



APPLICATIONS

Similar framework of Word Embeddings and LSTMs could be used for

- Sentiment Analysis
- Smarter plagiarism tools

THANK YOU

✉ jlakshmi235@gmail.com

Jlakshmi235

in jayalakshmi-jain