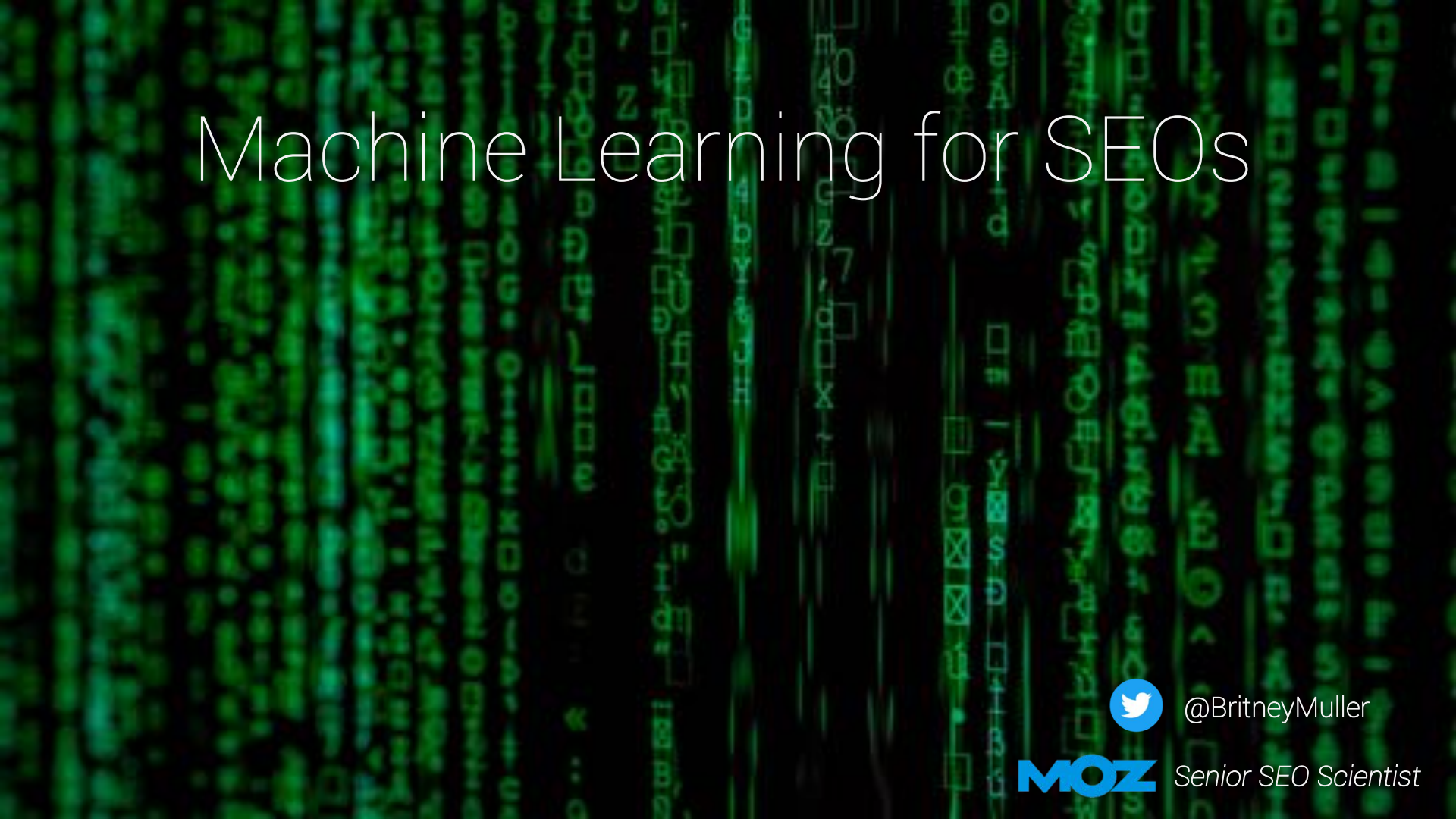
Machine Learning for SEOs



@BritneyMuller

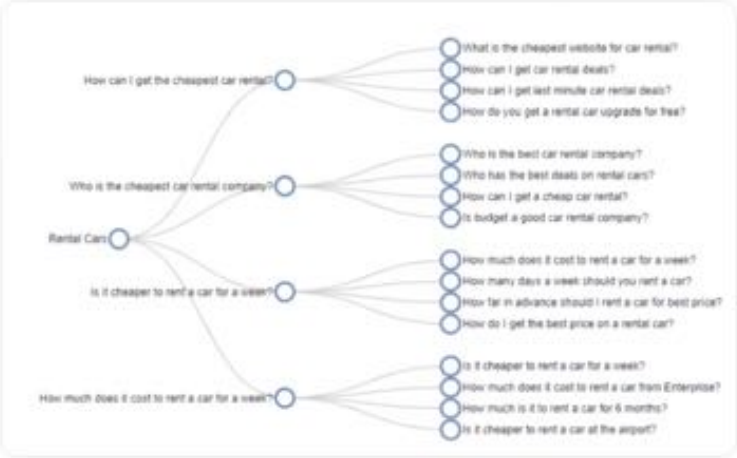
MOZ *Senior SEO Scientist*


You don't have to be a Data Scientist to think of the *next brilliant ML application*!

Matt Lacuesta
@MattLacuesta

Follow

One of the folks on our SEO team (9 months into their career) was inspired by @hamletbatista & @BritneyMuller to scrap G's paas and is learning python to pull some cool info. We're all excited about their progress and they just sent over some initial visuals coming from it.



3:57 PM - 16 Oct 2019 from Arvada, CO

What does your path from data to implementation look like?

To do better, we must think *differently*


MICROSOFT
MICROSOFT

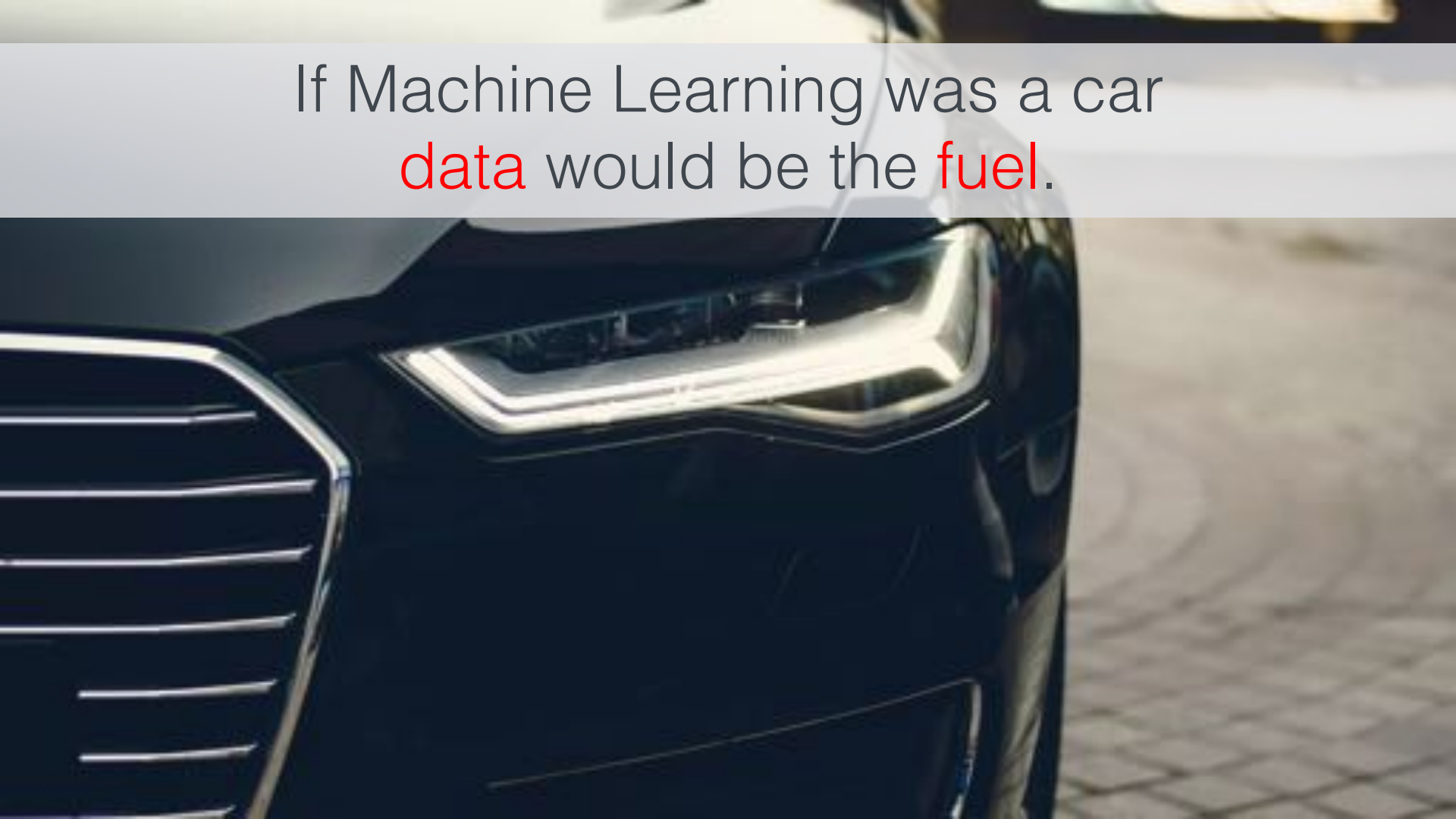
The Photonic Fence uses lasers to neutralize mosquitos


Machine learning
can be your laser beam!

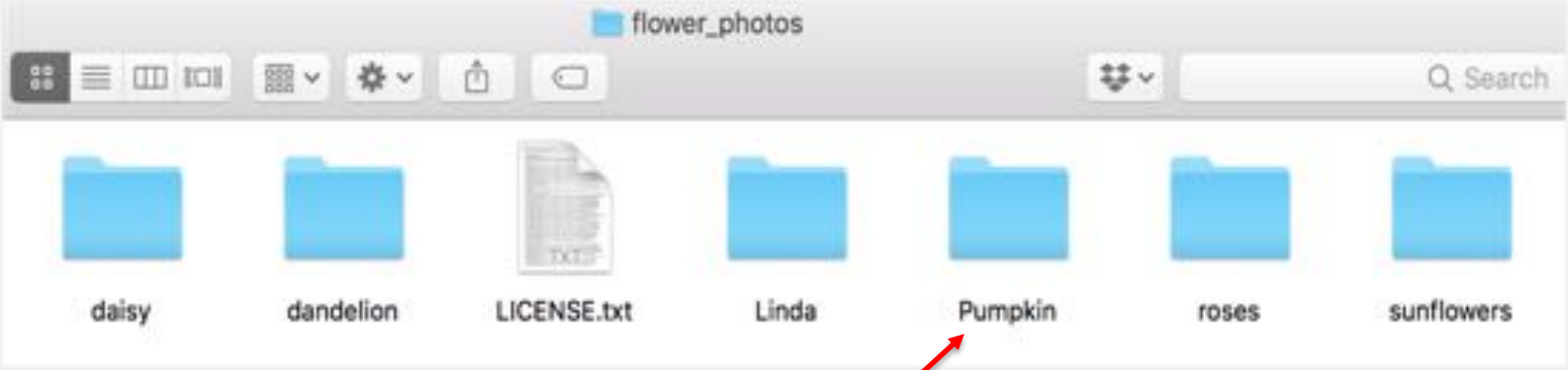
Machine Learning will free us up
to do more strategic work.

Machine Learning for SEOs

- How ML works
- BERT 101 
- SEO Applications
- Simple ML framework


If Machine Learning was a car
data would be the fuel.


I have no idea what I'm doing
gifbin.com


flower_photos
Search
daisy
dandelion
LICENSE.txt
Linda
Pumpkin
roses
sunflowers


```
Last login: Thu Jul  5 14:35:40 on ttys002
C1MRW3DQH3QK:~ britneymuller$ cd tensorflow-for-poets-2
C1MRW3DQH3QK:tensorflow-for-poets-2 britneymuller$ python
     --graph=tf_files/retrained_graph.pb      --image=tf_
fe.jpg
/anaconda2/lib/python2.7/site-packages/h5py/__init__.py:3
rsion of the second argument of issubdtype from `float` t
ecated. In future, it will be treated as `np.float64 == n
  from ._conv import register_converters as _register_con
2018-07-05 17:45:46.053777: I tensorflow/core/platform/cp
 Your CPU supports instructions that this TensorFlow bina
use: AVX2 FMA

Evaluation time (1-image): 0.202s

pumpkin (score=0.99691)
daisy (score=0.00172)
dandelion (score=0.00084)
roses (score=0.00027)
sunflowers (score=0.00010)
C1MRW3DQH3QK:tensorflow-for-poets-2 britneymuller$
```

IMAGENET

14,197,122 images, 21841 synsets indexed

Explore Download Challenges Publications Updates About

Not logged in. Login | Signup

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

Click here to learn more about ImageNet, Click here to join the ImageNet mailing list.

What do these images have in common? *Find out!*

Research updates on improving ImageNet data

prediction: sleeping bag

probability: 0.24296988546848297


BAK

bit.ly/rand-b

SEARCHLOVE

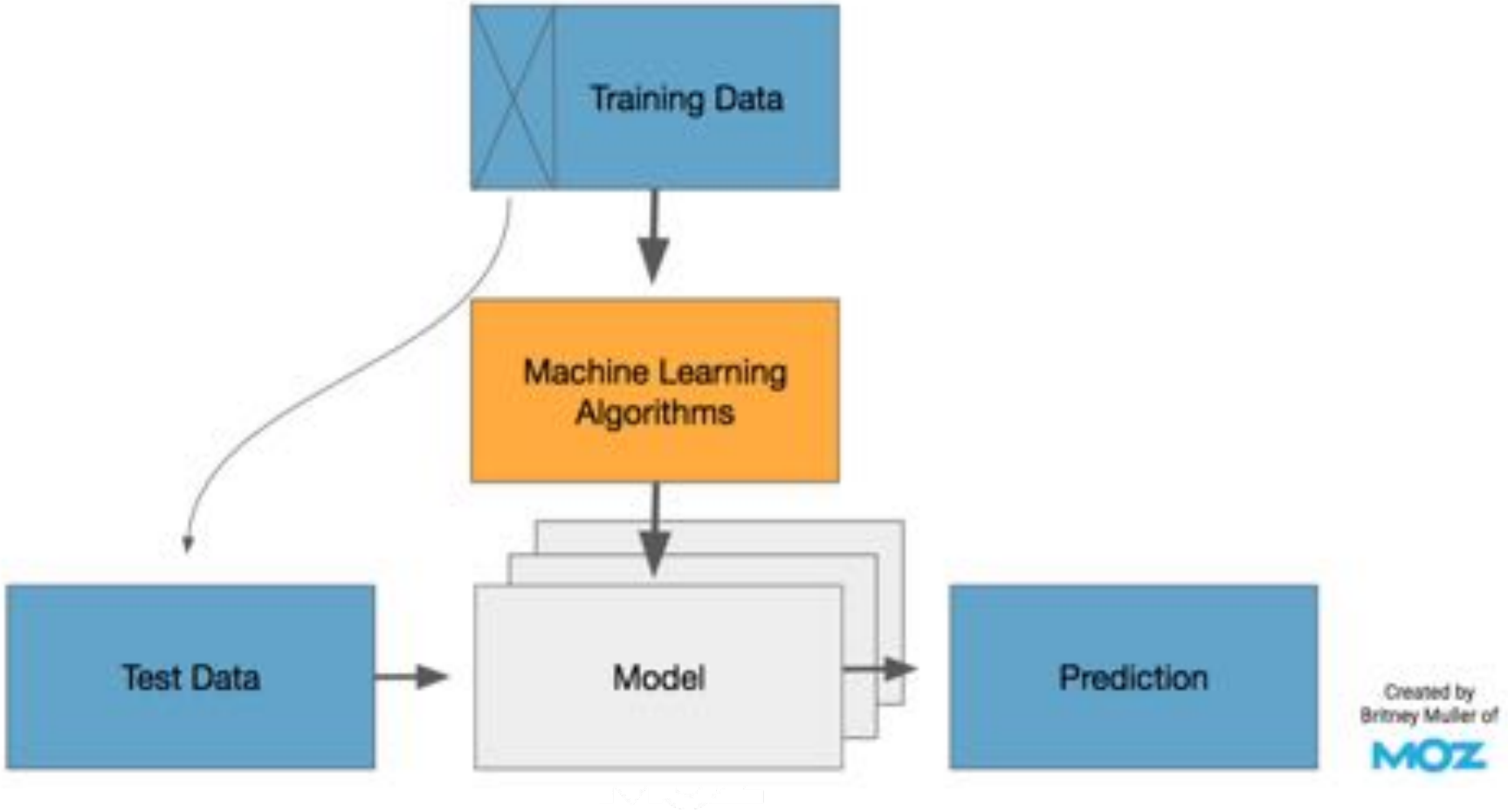
searchlove y
salves from 2016 vs. 2018:

google dobinn dow the wid and move a mountain
you you you you and me could calm a war down
you you you you and me could make it rain now
you you you and me could stop this love drought

and rub up and feel up on ya
give you some time to prove that i can trust ya again
i'm google's ownership of search has been remarkably stablet in funds,
but of much smaller amove you like i love you
step down, they don't love you like i love you
can't you see there's no other man above you

What is Machine Learning?

Machine Learning is a subset of AI that combines statistics & programming to give computers the ability to "learn" without explicitly being programmed.


Training Data
Machine Learning Algorithms
Test Data
Model
Prediction
Created by Britney Muller of MOZ

Select all squares with

street signs

VERIFY

ML doesn't solve well for soft/people skills:

For example, teachers, nurses, childcare

MOZ


Safe & effective ML **requires** diversity


GOOGLE REMOVES PRONOUNS FROM
GMAIL 'SMART COMPOSE' FEATURE
AFTER THEY ARE ACCIDENTALLY
SEXIST

Driverless cars ‘more likely to run over black people’

Gender and racial bias found in Amazon's facial recognition technology (again)

Research shows that Amazon's tech has a harder time identifying gender in darker-skinned and female faces

By James Vincent | Jan 25, 2019, 9:45am EST

SHARE

Machine Learning will free us up
to do more strategic work.

Machine Learning for SEOs

- ~~How ML works~~
- BERT 101 
- SEO Applications
- Simple ML framework


Wut?

Named entity recognition

Classification

Summarization

Machine translation

Sentiment Analysis

Question and answering

Sentence disambiguation

BERT combines and outperforms 10+ of the common NLP tools

BERT combines and outperforms 10+ of the common NLP tools

A pre-trained BERT model can be finetuned with just one additional output layer to create a SOTA model for wide range tasks such as question answering.

Sound familiar??

MOZ

Natural Questions --- Data

We like question answering as a testbed because

- Questions can be arbitrarily complex
  - require world knowledge
  - require reasoning about events
- Task is relatively easy to evaluate

This example requires us to know that disabling telephony implies that you cannot make a call.

Question: *Can you make and receive calls in airplane mode?*

*Airplane mode, aeroplane mode, flight mode, offline mode, or standalone mode is a setting available on many smartphones, portable computers, and other electronic devices that, when activated, **suspends radio-frequency signal transmission by the device, thereby disabling Bluetooth, telephony,** and Wi-Fi. GPS may or may not be disabled, because it does not involve transmitting radio waves.*

Answer: *No*

Google books

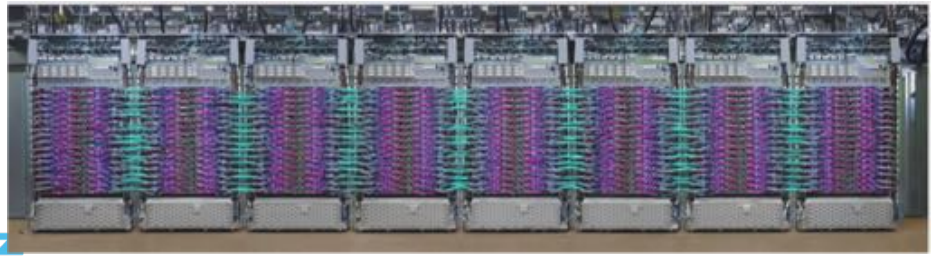
+

WIKIPEDIA

+

+

+



GLUE results (General Language Understanding Evaluation), gluebenchmark.com

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
$BERT_{BASE}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
$BERT_{LARGE}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

Table 1: GLUE Test results, scored by the GLUE evaluation server.

large improvements over state of the art (SOTA) on wide variety of language tasks

What BERT can't do

What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models

Allyson Ettinger
Department of Linguistics
University of Chicago
aettinger@uchicago.edu

Abstract

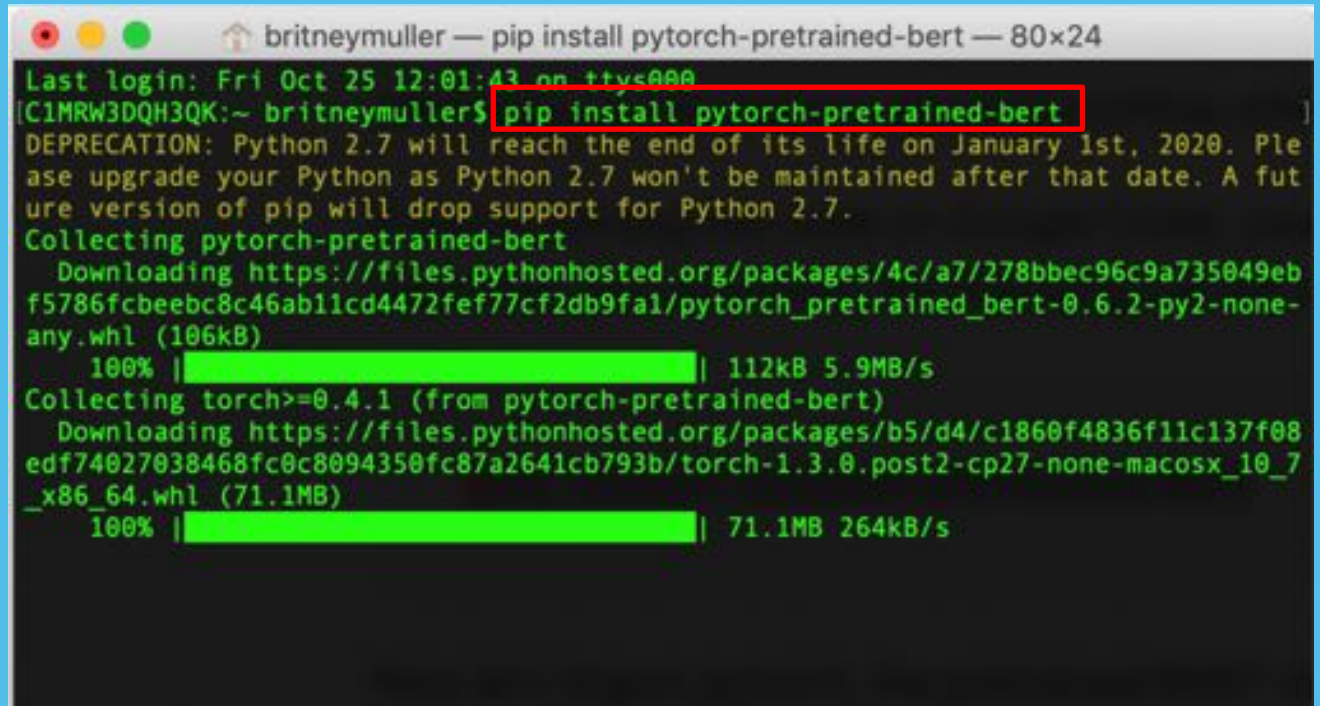
Pre-training by language modeling has become a popular and successful approach to NLP tasks, but we have yet to understand exactly what linguistic capacities these pre-training processes confer upon models. In this paper we introduce a suite of diagnostics drawn from human language experiments, which allow us to ask targeted questions about information used by language models for generating predictions in context. As a case study, we apply these diagnostics to the popular BERT model, finding that it can generally distinguish good from bad completions involving shared category or role reversal, albeit with less sensitivity than humans, and it robustly retrieves noun

controlled to ask targeted questions about linguistic capabilities, and they are designed to ask these questions by examining word predictions in context, which allows us to study LMs without any need for task-specific fine-tuning.

Beyond these advantages, our diagnostics distinguish themselves from existing tests for LMs in two primary ways. First, these tests have been chosen s
sensitivit
patterns
Second,
guistic c
syntactic
we have
inference, semantic roles and event knowledge,

Context	Match	Mismatch
A robin is a ____	*bird*	*tree*
A robin is not a ____	*bird*	*tree*

You can play around with BERT today:



On-page SEO for NLP

July 17, 2018 by Justin

Traditional on-page SEO guidance is to target a primary phrase, its near-related terms, and its longtail variants by using them in the text and placing them in strategic locations on the page [i.e., title, headings, early in content, throughout content]. However, writing for Natural Language Processing, or NLP, requires some additional steps and considerations.

Managing on-page SEO for Google's NLP capabilities requires a basic understanding of the limitations of its parser and the intelligence behind the logic. In practical terms, this is technical SEO for content understanding. Writing for NLP requires clear, structured writing and an understanding of word relationships.

Brief Introduction to NLP

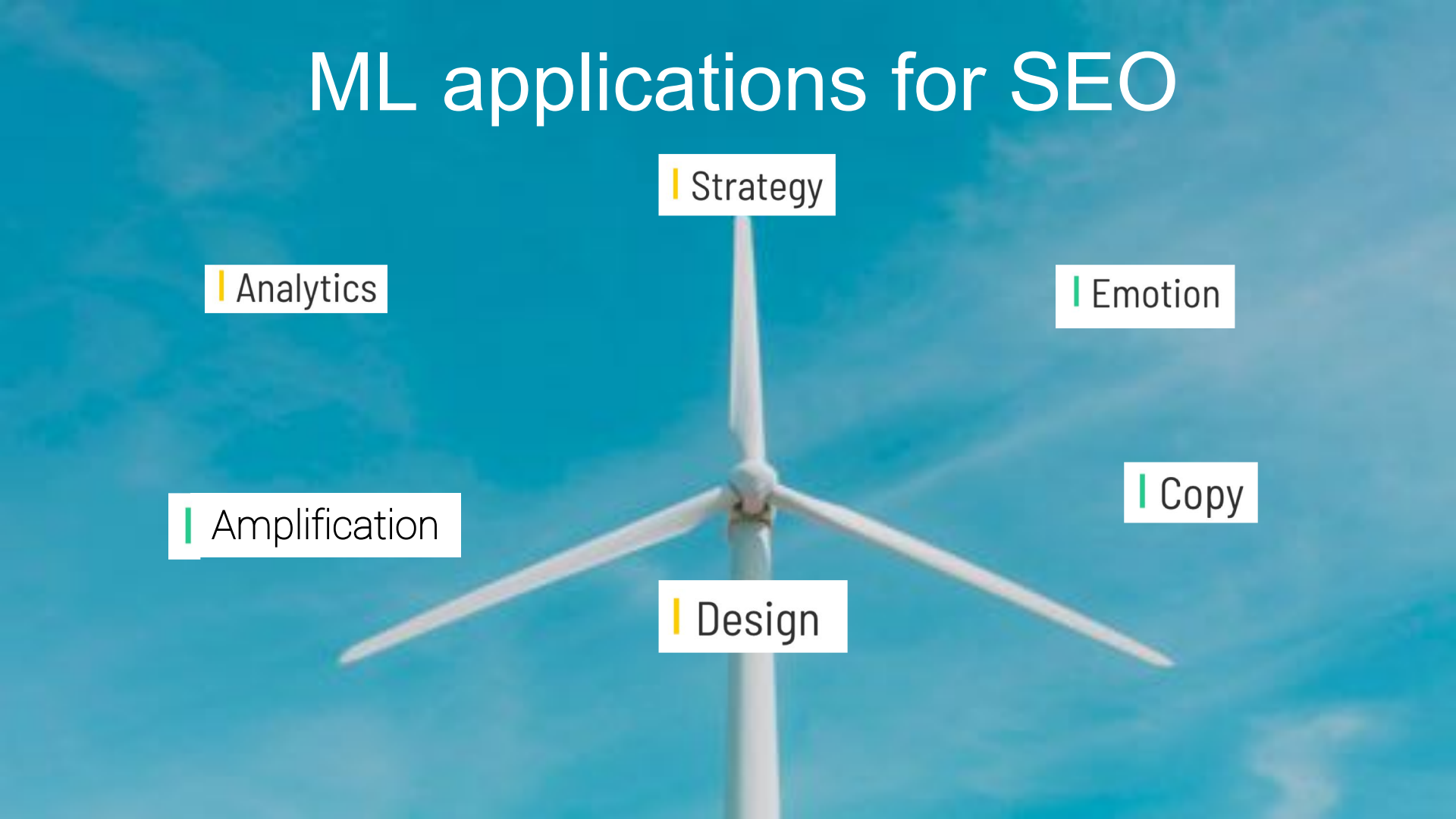
The are many aspects to Natural Language Processing, but we only need a basic understanding of its core components to do our job well as SEOs. In short, NLP is the process of parsing through text, establishing relationships between words, understanding the meaning of those words, and deriving a greater understanding of words. I'll briefly go through the major components and vocab you'll need.

Major Components of Natural Language Processing

Machine Learning for SEOs

- ~~How ML works~~
- ~~BERT 101~~ 
- SEO Applications
- Simple ML framework

ML applications for SEO



Uncover the ONE keyword/topic with the highest ROI potential

(rolling analysis)

Monthly Volume

Volume Distribution (Low - High)

Learn more about Volume Score

SERP Features

Learn more about SERP features

Difficulty

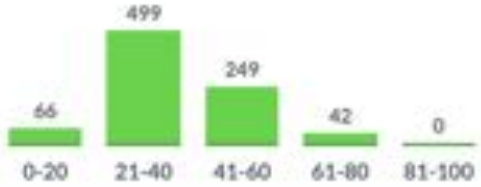
Learn more about Difficulty Score

Organic CTR

Learn more about Organic CTR Score

Priority

Learn more about Priority Score



Volume: any

SERP Feature: any

Difficulty: any

Organic CTR: any

My Score: any

Priority: any

SERP Age: any

Country: any

Reset filters

Power BI Basics for SEO & PPC Marketers

6 videos • 4,324 views • Last updated on Jul 9, 2019

Seer Interactive SUBSCRIBED

1 

Lesson 1: Creating a Relationship Between Paid & Organic | Power BI Basics for Digital Marketers
Seer Interactive

2 

Lesson 2: Using Data to Provide Insight to Clients & Teams | Power BI Basics for Digital Marketers
Seer Interactive

3 

Lesson 2.5: Getting the Data to Tell a Story & Work for YOU | Power BI Basics for Digital Marketers
Seer Interactive

4 

Lesson 3: Looking at your Competitors using Big Data | Power BI Basics for Digital Marketers
Seer Interactive

5 

Lesson 4: How LONG is the Long Tail? | Power BI Basics for Digital Marketers
Seer Interactive

6 

Calling all Data Geeks! Come Work at Seer!
Seer Interactive

FACETS

pair-code.github.io/facets

FACETS

pair-code.github.io/facets

Strategy

Navigate via rolling business & market data

DATA DRIVEN GROWTH WITH PYTHON

Predicting Sales

Forecasting the monthly sales with LSTM

Barış Karaman Follow
Jun 9 · 8 min read ★

This series of articles was designed to explain how to use Python in a simplistic way to fuel your company's growth by applying the predictive approach to all your actions. It will be a combination of programming, data

2. Machines are best equipped to make trading decisions in the short and medium term

"Machines have the ability to quickly analyze news feeds and tweets, process earnings statements, scrape websites, and trade on these instantaneously."

J.P.Morgan
Asset Management

MOZ

Source: news.efinancialcareers.com/ca-en/285249/machine-learning-and-big-data-j-p-morgan

SparkToro

My audience frequently talks about cro Search

YouTube Channels Followed & Subscribed-To

10% of this audience subscribe to these 4 channels

Google Analytics
Welcome to the official channel for Google Analytics, where you'll find videos and product tips for Analytics, Data Studio, Optimize, Surveys, and Tag...

Moz
Moz: Marketing Analytics Software for SEO, Links, Social, and Brand - A Vibrant Online Marketing Community - Resources for Learning Inbound Marketing

HubSpot
HubSpot is a leading growth platform with thousands of customers around the world. Comprised of Marketing Hub, Sales Hub, Service Hub, and a powerful ...

Search Engine Land
Search Engine Land is a must read hub for news and information about search engine marketing, optimization and how search engines such as Google, Yaho...

See All YouTube Channels They Follow

SparkToro

My audience frequently talks about cro Search

Podcasts Followed, Listened, & Subscribed-To

13% of this audience follows these 4 podcasts

The Search Engine Journal Show
Search Engine Nerds is a bi-weekly show put on by Search Engine Journal. Search Engine Journal (SEJ) covers the marketing industry, focusing on SEO, P...

SEO 101 on WebmasterRadio.fm
SEO 101 is Search Engine Optimization from the very beginning. SEO 101 will teach you SEO from Square one. Hosts Ross Dunn of StepForth Web Marketing ...

Behind the Numbers: eMarketer Podcast
"Behind the Numbers" is a freewheeling, daily conversation about digital media and marketing, and how digital is transforming business—and even life. ...

Marketing Over Coffee Marketing Podcast
Marketing over coffee is a weekly discussion of what's new in marketing with John Wall and Christopher PennMarketing over coffee is a weekly discussio...

See All 54 Podcasts They Follow

SparkToro

My audience | frequently talks about | cro | Search

Audience Insights

21.2% are concentrated in these **2** geographies

- New York
- Greater Boston Area

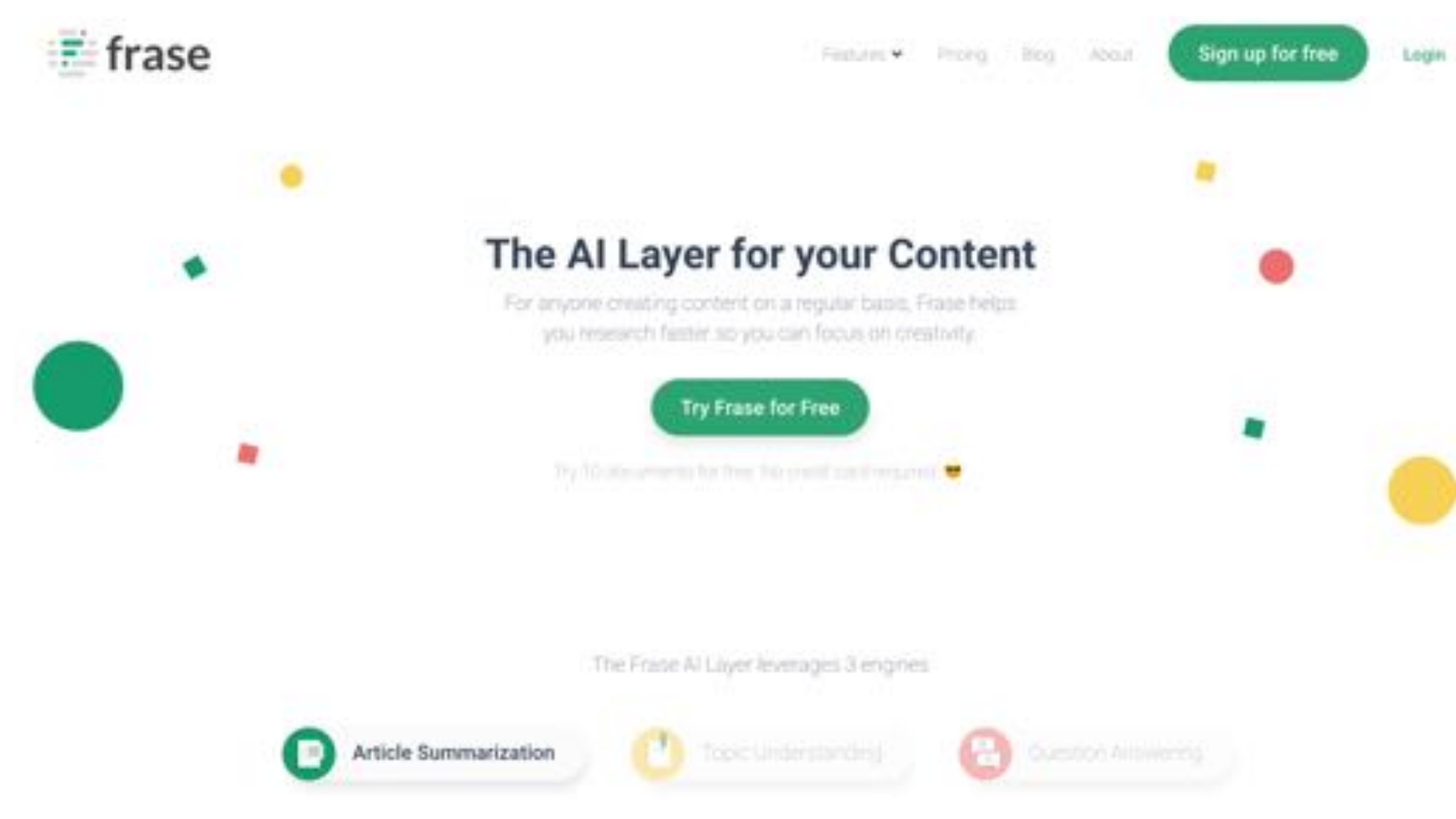
See All 35 Geographies

Frequently used hashtags in shares and content

Show More

Process

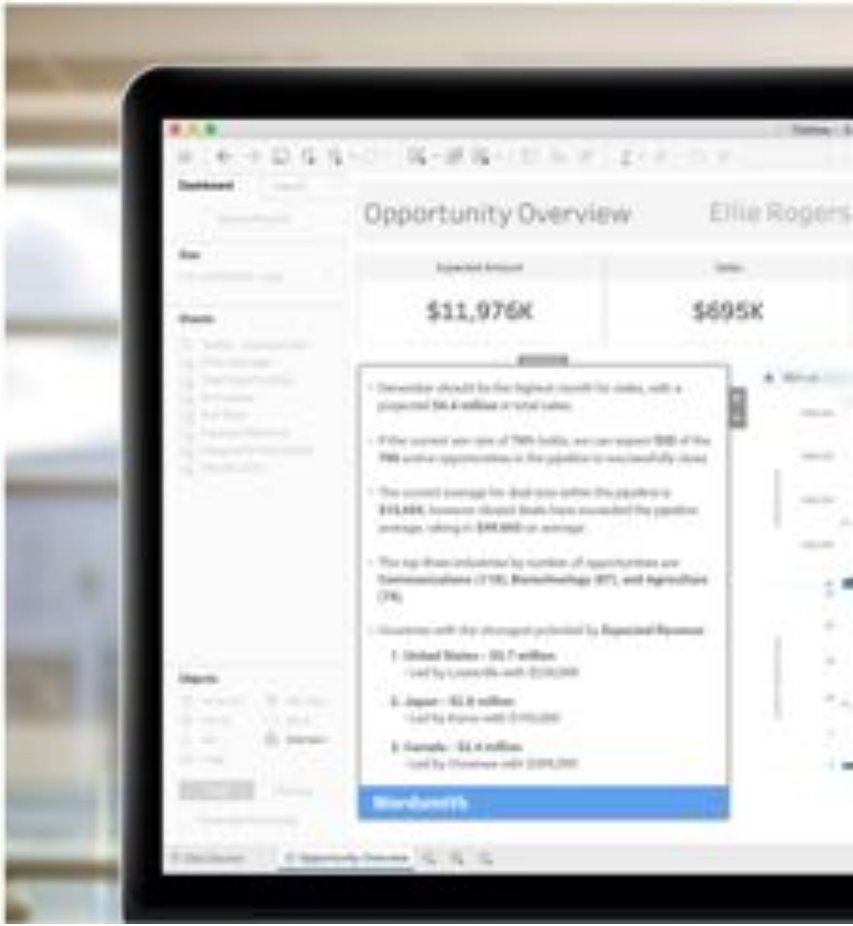
Add deep insights to your current processes


frase
Features
Pricing
Blog
About
Sign up for free
Login
The AI Layer for your Content
For anyone creating content on a regular basis, Frase helps you research faster so you can focus on creativity.
Try Frase for Free
Try 10 documents for free. No credit card required.
The Frase AI Layer leverages 3 engines
Article Summarization
Topic Understanding
Question Answering

Extend the Power of Tableau with Wordsmith

Automatically generate easy-to-understand, written analysis using natural language generation within your Tableau dashboards.

REQUEST DEMO



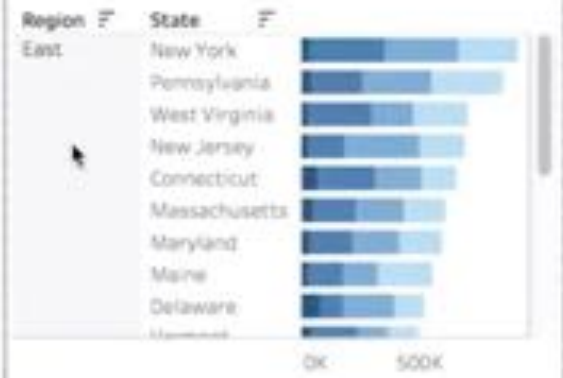
Credit Risk Overview

Wordsmith for Tableau

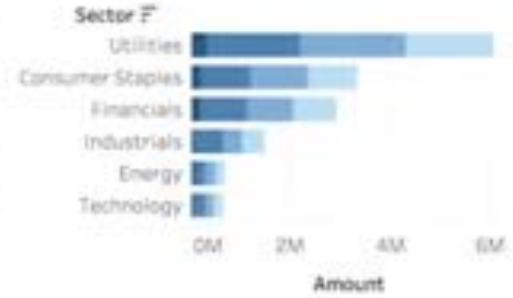
- New York had the most total outstanding loans ($913,187), while Delaware had the most loans that have been outstanding for over 90 days ($82,801).
- The average loan was overdue by 48 days, with Maine owning the lowest average (45.4 days) and Florida claiming the highest (51.7 days).
- Total loans that were overdue from 30 – 60 days had the largest percentage of the total amount of loans ($5.0 million, 32.9%).

Outstanding Loans

States Overdue



Loans by Sector



Automatic 301 Redirects

searchwilderness.com/mozcon-2019

library imports

```
In [2]: import pandas as pd
        import csv
        import time
        import requests
        import re
        from goose3 import Goose
        import spacy
        nlp = spacy.load('en_core_web_lg')
```

User inputs

```
In [3]: domain = "dakessianlaw.com"
        sf_crawl = "C:/Users/paul.Shapiro/Desktop/live-website-source.csv"
        csv_output = "C:/Users/paul.Shapiro/Desktop/301-redirect-matching.csv"
```

Extract historic URLs from Wayback Machine

Auto301Redirects.ipynb hosted with ❤ by GitHub view raw

Paul Shapiro

MOZ

Emotion

Rich customer understanding.

Faces | Objects | Labels | Logos | Web | Text | Properties | Safe Search

Seattle-SEO-Beers.jpeg

Face 1

Joy	Very Likely
Sorrow	Very Unlikely
Anger	Very Unlikely
Surprise	Very Unlikely
Exposed	Very Unlikely
Blurred	Very Unlikely
Headwear	Very Unlikely

Roll: 2° Tilt: -14° Pan: 13°

Confidence 95%

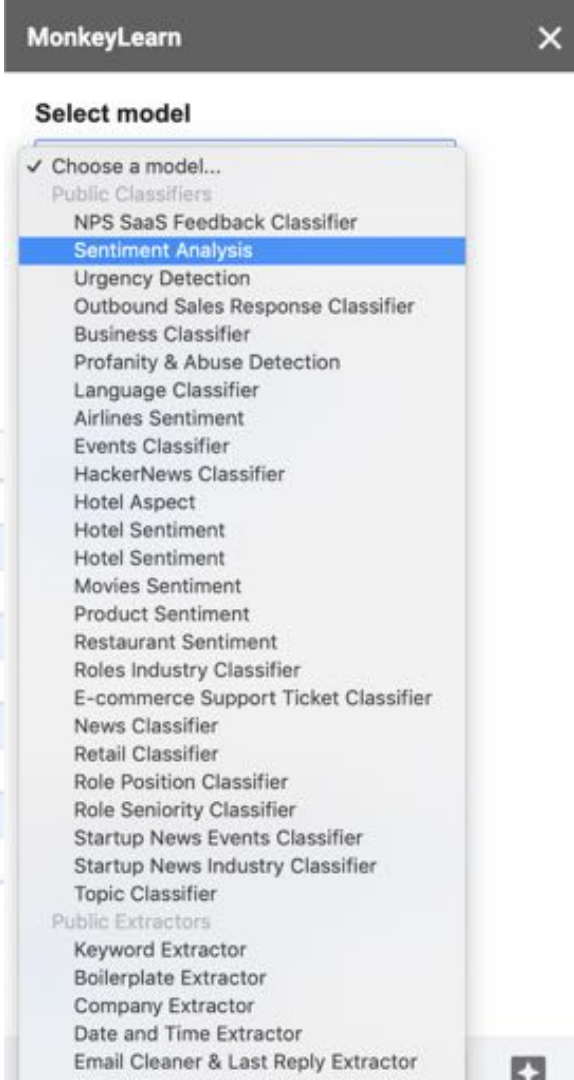
Request

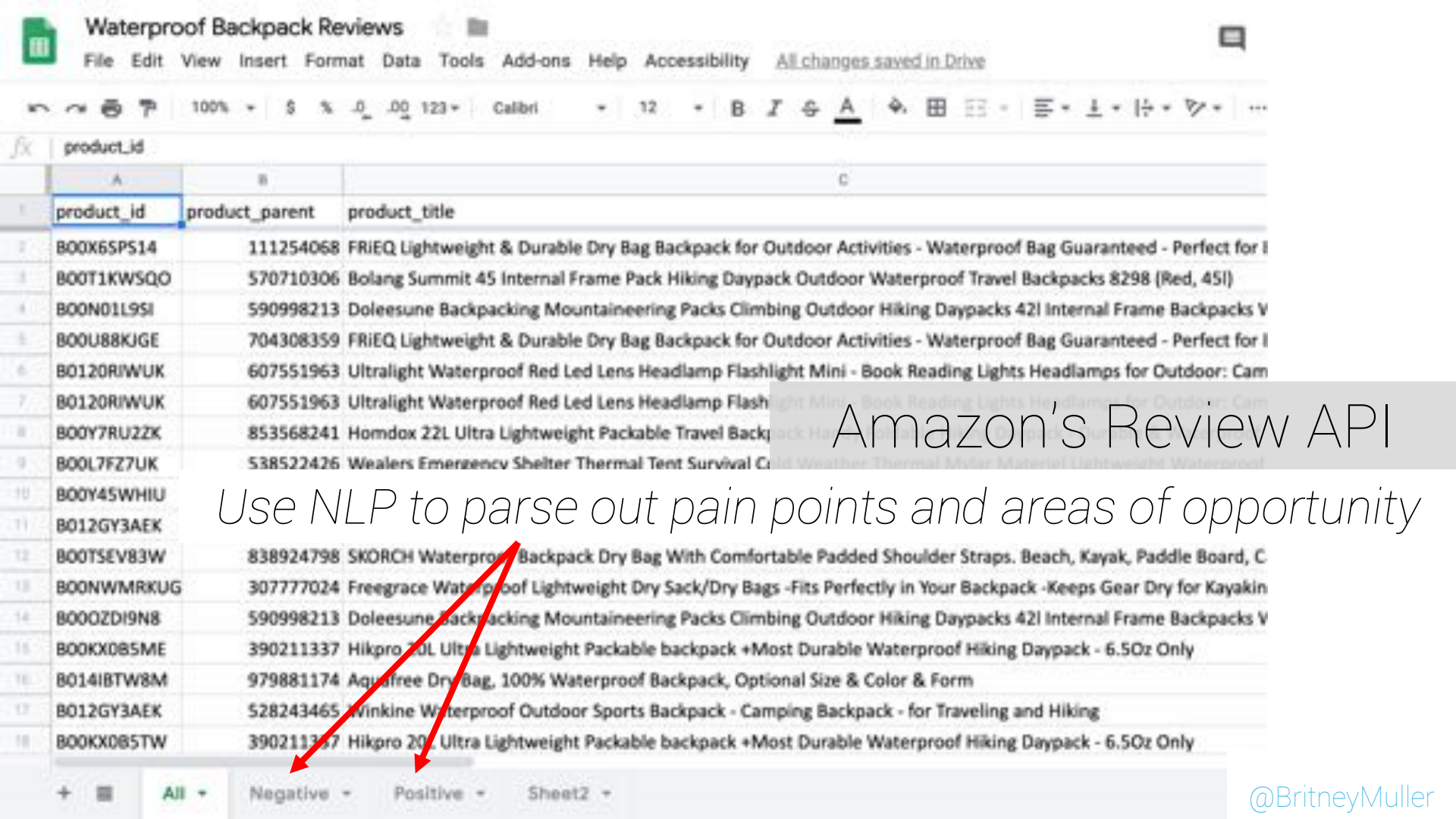
```
{
  "requests": [
    {
      "features": [
        {
          "maxResults": 50,
          "type": "LANDMARK_DETECTION"
        },
        {
          "maxResults": 50,
          "type": "FACE_DETECTION"
        },
        {
          "maxResults": 50,
          "type": "OBJECT_LOCALIZATION"
        },
        {
          "maxResults": 50,
          "type": "LOGO_DETECTION"
        },
        {
          "maxResults": 50
```

Response

```
        },
        "type": "LEFT_EYE_BOTTOM_BOUNDARY"
      },
      {
        "position": {
          "x": 651.6909,
          "y": 435.73477,
          "z": 1.7427447
        },
        "type": "LEFT_EYE_LEFT_CORNER"
      },
      {
        "position": {
          "x": 658.44446,
          "y": 436.48486,
          "z": -0.66775113
        },
        "type": "LEFT_EYE_PUPIL"
      },
      {
        "position": {
```

B	C	D	
@JonathanAufray	Jonathan Aufray	11 of the Most Effective #SEO Tools You Need to Know About: - S	Positive
@honmahimawari	本間ひまわり	おはござ～(・ワ・)ノ 昨日は色々動けて楽しかった!! またにじ	Neutral
@sspencer	Stephan Spencer	One of the advantages of #DigitalMarketing is having detailed #	Positive
@Red_Web_Design	Red Website Design	10 Wonderful Blog Posts for a First-Class Marketing Strategy	Neutral
@Startup_Nerd	The Startup Nerd	11 of the Most Effective #SEO Tools You Need to Know About: - S	Positive
@Startup_Nerd	The Startup Nerd	11 of the Most Effective #SEO Tools You Need to Know About: - S	Positive
@Startup_Nerd	The Startup Nerd	11 of the Most Effective #SEO Tools You Need to Know About: - S	Positive
@rustybrick	Barry Schwartz	ICYMI: Check out my interview with @dr_pete of @Moz about M	Positive
@rustybrick	Barry Schwartz	Check out my interview with @dr_pete of @Moz about Mozcast,	Positive





Amazon's Review API

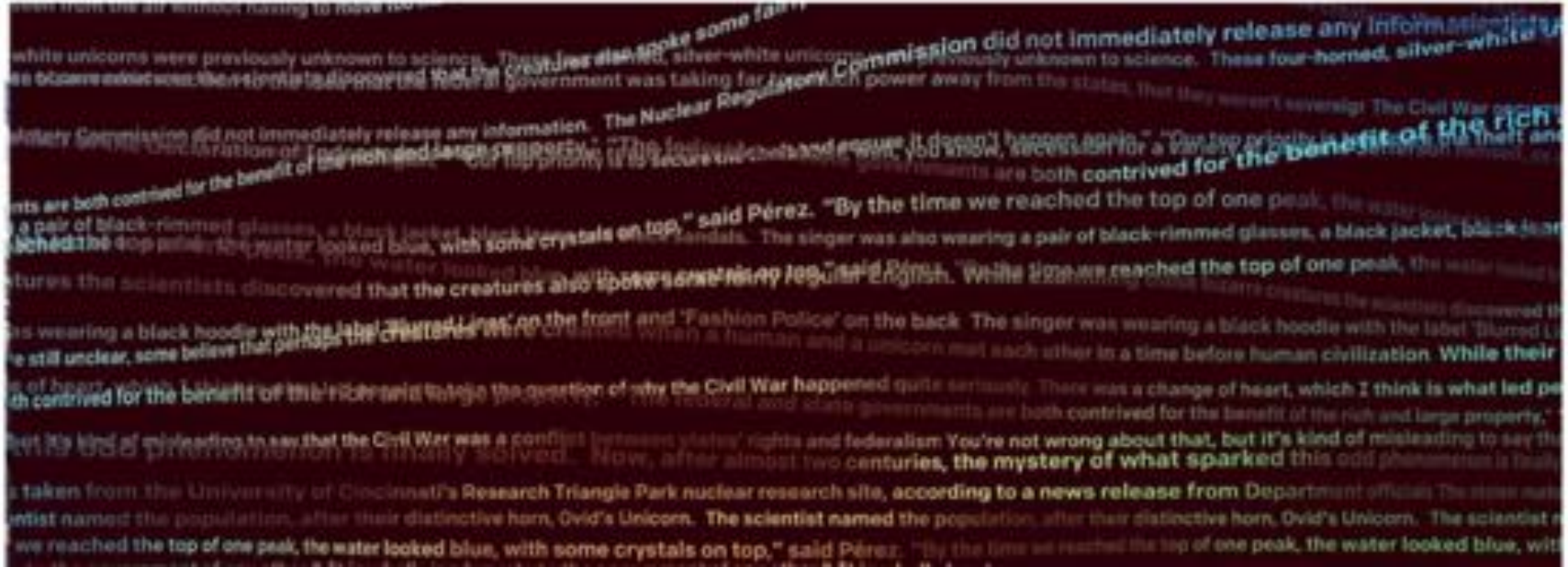
Use NLP to parse out pain points and areas of opportunity

@BritneyMuller


Copy
Automate meaningful content

OpenAI built a text generator so good, it's considered too dangerous to release

Zack Whittaker @zackwhittaker / 4 months ago



Talk to Transformer

See how a modern neural network completes your text. Type a custom snippet or try one of the examples. Learn more below.

Follow @AdamDanielKing for updates and other demos like this one.

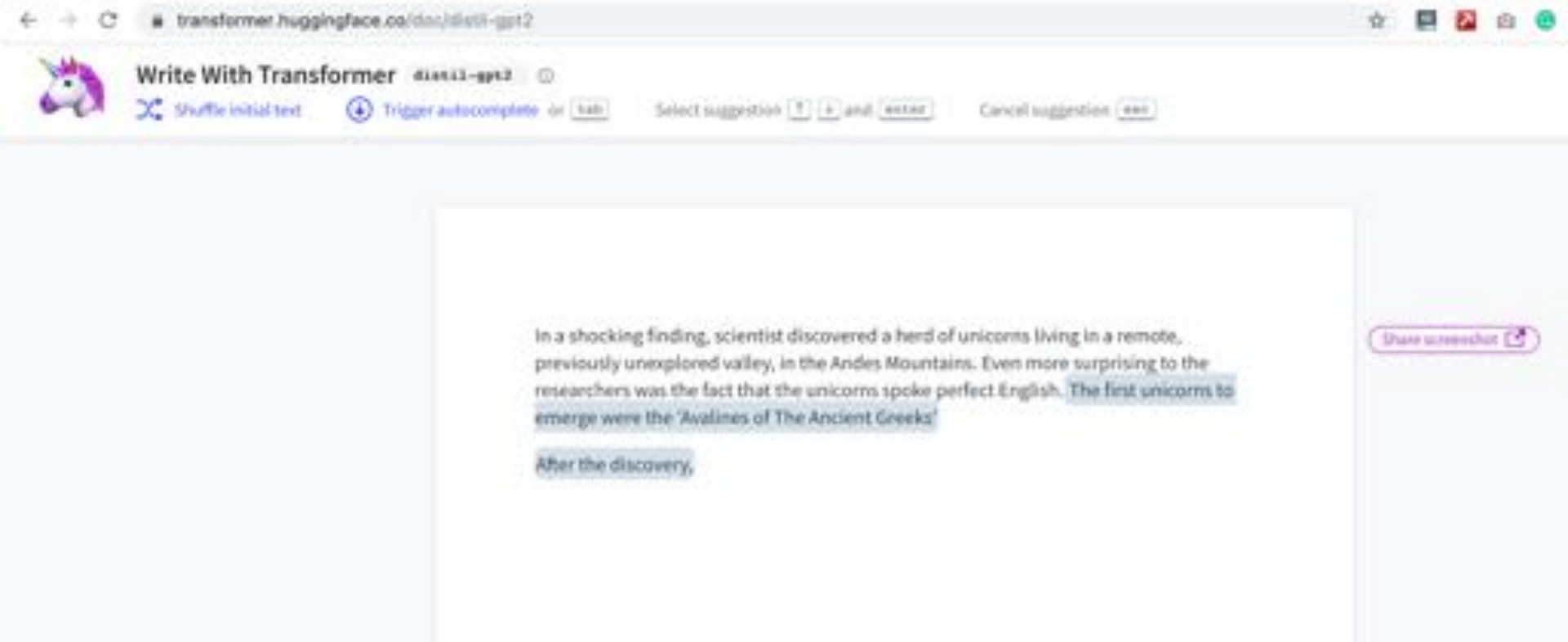
Custom prompt

Every year, 400+ marketers and top-rated speakers spend two days sharing insights into content strategy, promotion, conversion and analytics.

Completion

Every year, 400+ marketers and top-rated speakers spend two days sharing insights into content strategy, promotion, conversion and analytics. And every year, this panel, and other key events, are sold out. The next conference is on July 31st with the latest info about registration and venue HERE .
You can also find the speakers, trainers, and other resources they will be using on this blog .
A few key things I'm going to cover in this post:

1. What is a good content strategy?
2. What are our goals for content?
3. What are the tools we need to start?


transformer.huggingface.co/doc/distil-gpt2
Write With Transformer
distil-gpt2
Shuffle initial text
Trigger autocomplete
or
tab
Select suggestion
↑
↓
and
enter
Cancel suggestion
esc
In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English. The first unicorns to emerge were the 'Avalines of The Ancient Greeks'
After the discovery,
Share screenshot

Automate Meta Descriptions

```
import csv
import os
from sumy.parsers.html import HtmlParser
from sumy.parsers.plaintext import PlaintextParser
from sumy.nlp.tokenizers import Tokenizer
from sumy.summarizers.lsa import LsaSummarizer as Lsa
from sumy.summarizers.luhn import LuhnSummarizer as Luhn
from sumy.summarizers.text_rank import TextRankSummarizer as TxtRank
from sumy.summarizers.lex_rank import LexRankSummarizer as LexRank
from sumy.summarizers.sum_basic import SumBasicSummarizer as SumBasic
from sumy.summarizers.kl import KLSummarizer as KL
from sumy.summarizers.edmundson import EdmundsonSummarizer as Edmundson
from sumy.nlp.stemmers import Stemmer
from sumy.utils import get_stop_words

LANGUAGE = "english"
SENTENCES_COUNT = 1
```

searchwilderness.com/mozcon-2019

Paul Shapiro

Here is the Python code I shared


```
[269] for url in low_ctr_pages:
        if not keywords_in_title[url]: #only URLs without the keywords in the title
          query = queries_titles['query'][url]
          title = queries_titles['title'][url]

          score, best_match = find_best_match2(query, title)
          print('Keyword: {query}'.format(query=query))

          #print(score, best_match)

          new_title = replace_query_in_title(title[0], query, best_match)

          print('Old Title: {title}\nNew Title: {new_title}\n'.format(title=title[0], new_title=new_title))

C→ Keyword: name badge ribbons
   Old Title: Badge Ribbons, 100 Pre-Printed + Custom Ribbon Titles -
   New Title: Name Badge Ribbons, 100 Pre-Printed + Custom Ribbon Titles -

   Keyword: name tag holders
   Old Title: Badge Holders, Name Tags & Name Badges from $0 -
   New Title: Badge Holders, Name Tag Holders & Name Badges from $0 -
```


Design
Automate visual content and image understanding

Video Generation

Story
Media
Music
Style
Format

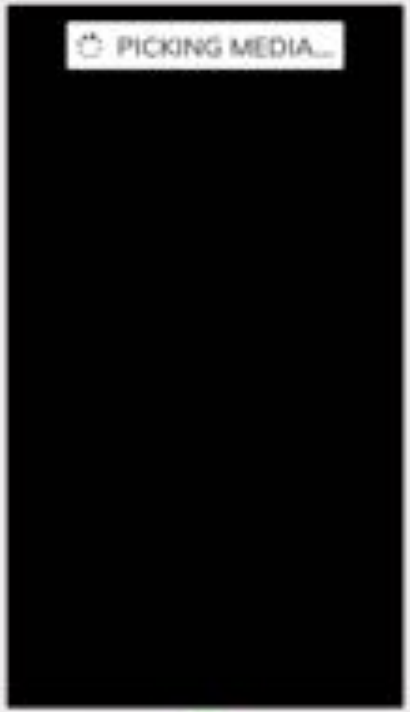
of SEO, from finding the terms and phrases (keywords) that can generate qualified traffic to your website, to making your site friendly to search engines, to building links and marketing the unique value of your site.

The world of search engine optimization is complex and ever-changing, but you can easily understand the basics, and even a small amount of knowledge can make a big difference. Free SEO education is also widely available on the web, including in guides like this! (Woohoo!)

Combine this information with some practice and you are well on your way to becoming a savvy SEO.

Here's what you'll find in the guide

01:17



Automate Image Understanding

Vision API

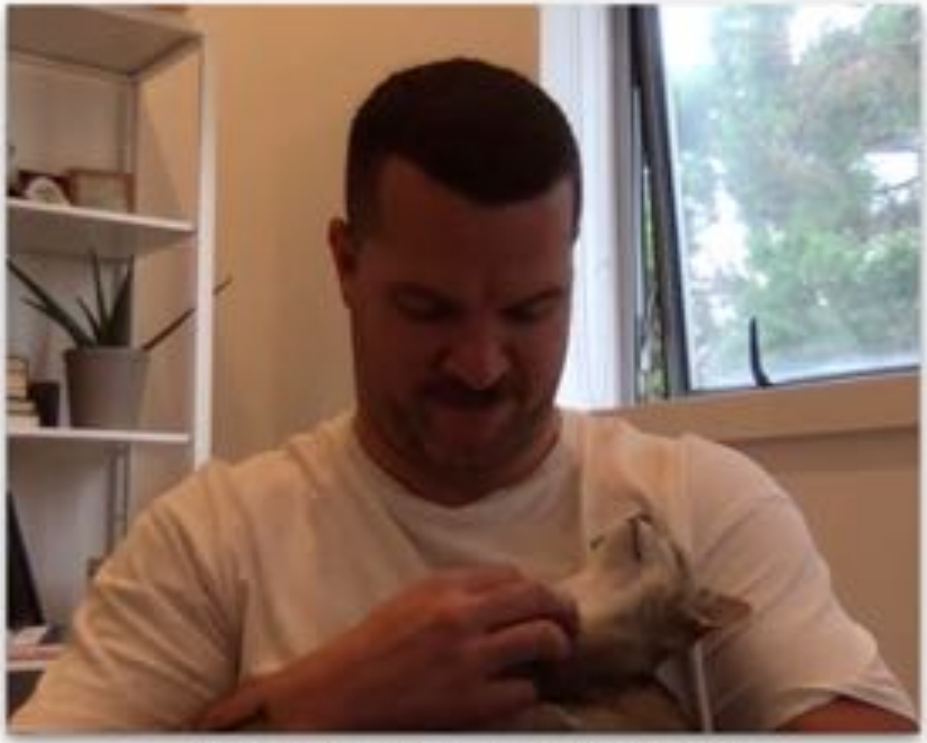
Google Cloud's Vision API offers powerful pre-trained machine learning models through REST and RPC APIs. Assign labels to images and quickly classify them into millions of predefined categories. Detect objects and faces, read printed and handwritten text, and build valuable metadata into your image catalog.

Objects | **Labels** | Web | Properties | Safe Search

Screen Shot 2019-09-19 at 12.01.06 PM.png

Nose	84%
Muscle	82%
Forehead	78%
Facial Hair	77%
Arm	76%
Jaw	70%
Finger	61%
Eating	59%

Objects | Labels | Web | Properties | **Safe Search**



Screen Shot 2019-09-19 at 12.01.06 PM.png

Adult	Unlikely
Spoof	Unlikely
Medical	Very Unlikely
Violence	Very Unlikely
Racy	Likely

Likelihood values are Unknown, Very Unlikely, Unlikely, Possible, Likely, and Very Likely

Faces | Objects | Labels | Logos | **Web** | Text | Properties | Safe Search

Seattle-SEO-Beers.jpeg

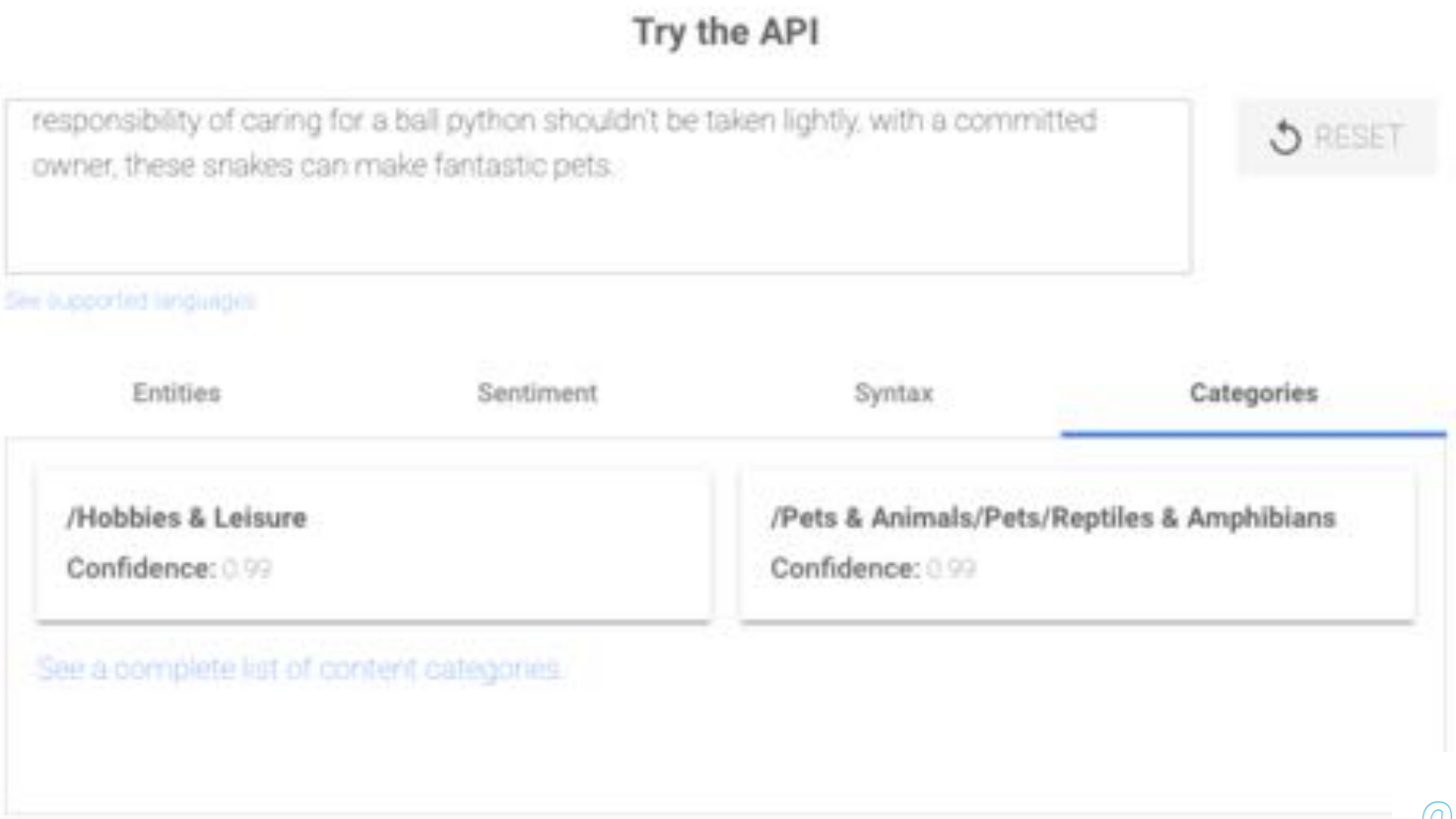
Web Entities

Rand Fishkin	5.178
Search Engine Optimization	0.6277
Moz	0.5355
Seattle	0.35805
Lost and Founder: A Painfu...	0.34755
Marketing	0.3153
SEO Beers	0.3009
Web search engine	0.1932
Scientist	0.1911

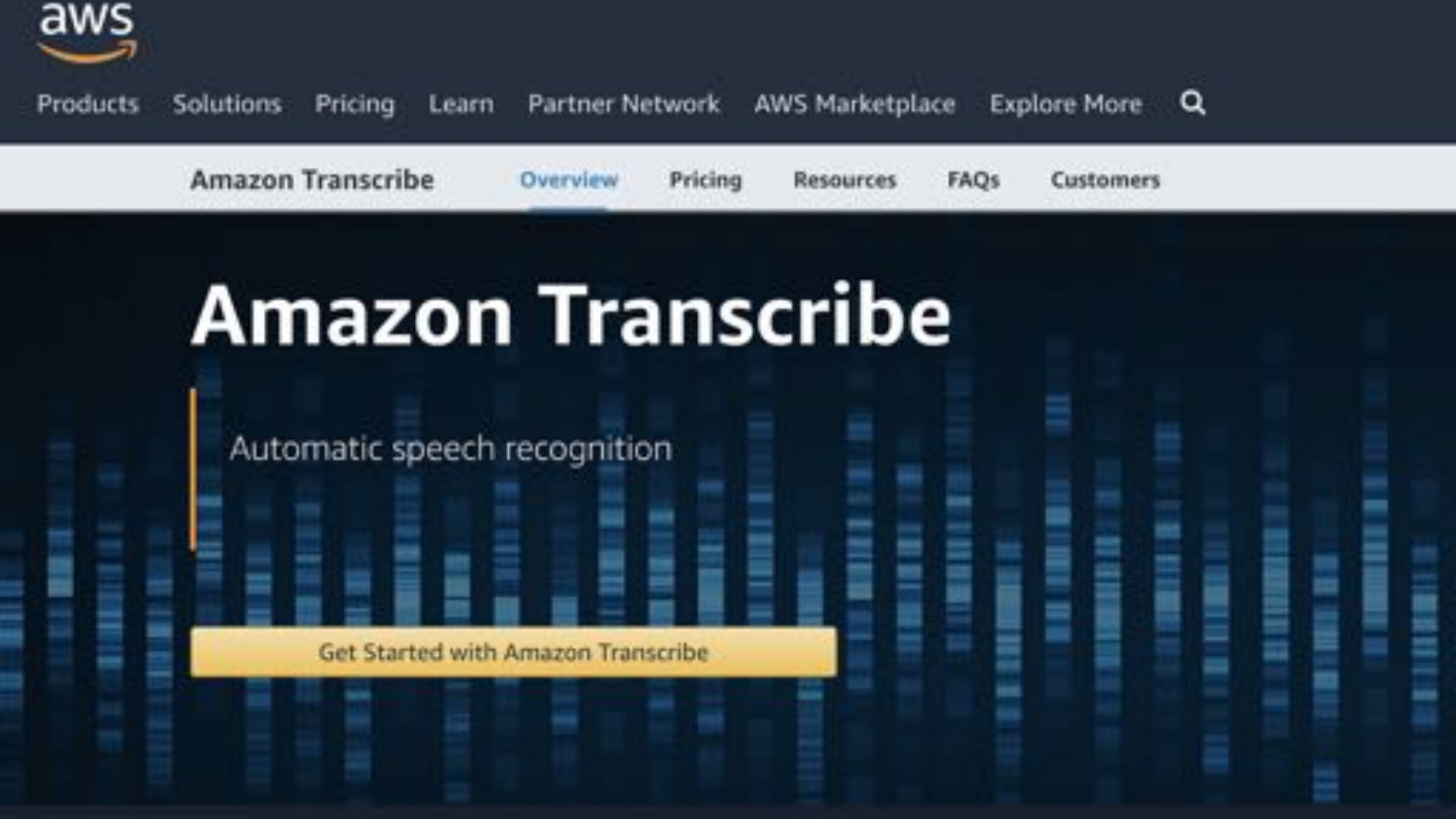
Pages with Matched Images

Machine Learning is becoming more accessible &
will free us up to work on higher level strategy.

Natural Language API demo




Automate Transcriptions

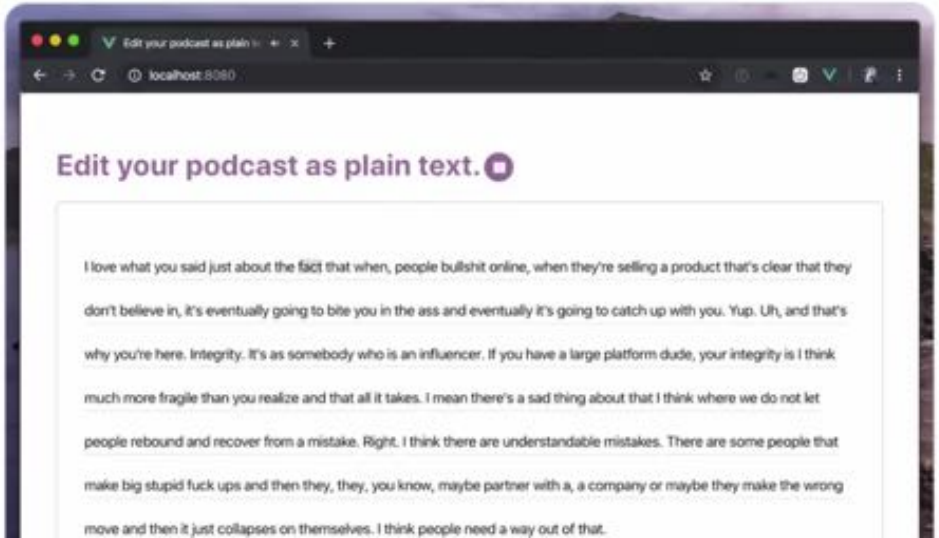

aws
Products
Solutions
Pricing
Learn
Partner Network
AWS Marketplace
Explore More
Amazon Transcribe
Overview
Pricing
Resources
FAQs
Customers
Amazon Transcribe
Automatic speech recognition
Get Started with Amazon Transcribe

Armagan Amcalar
@dashersw

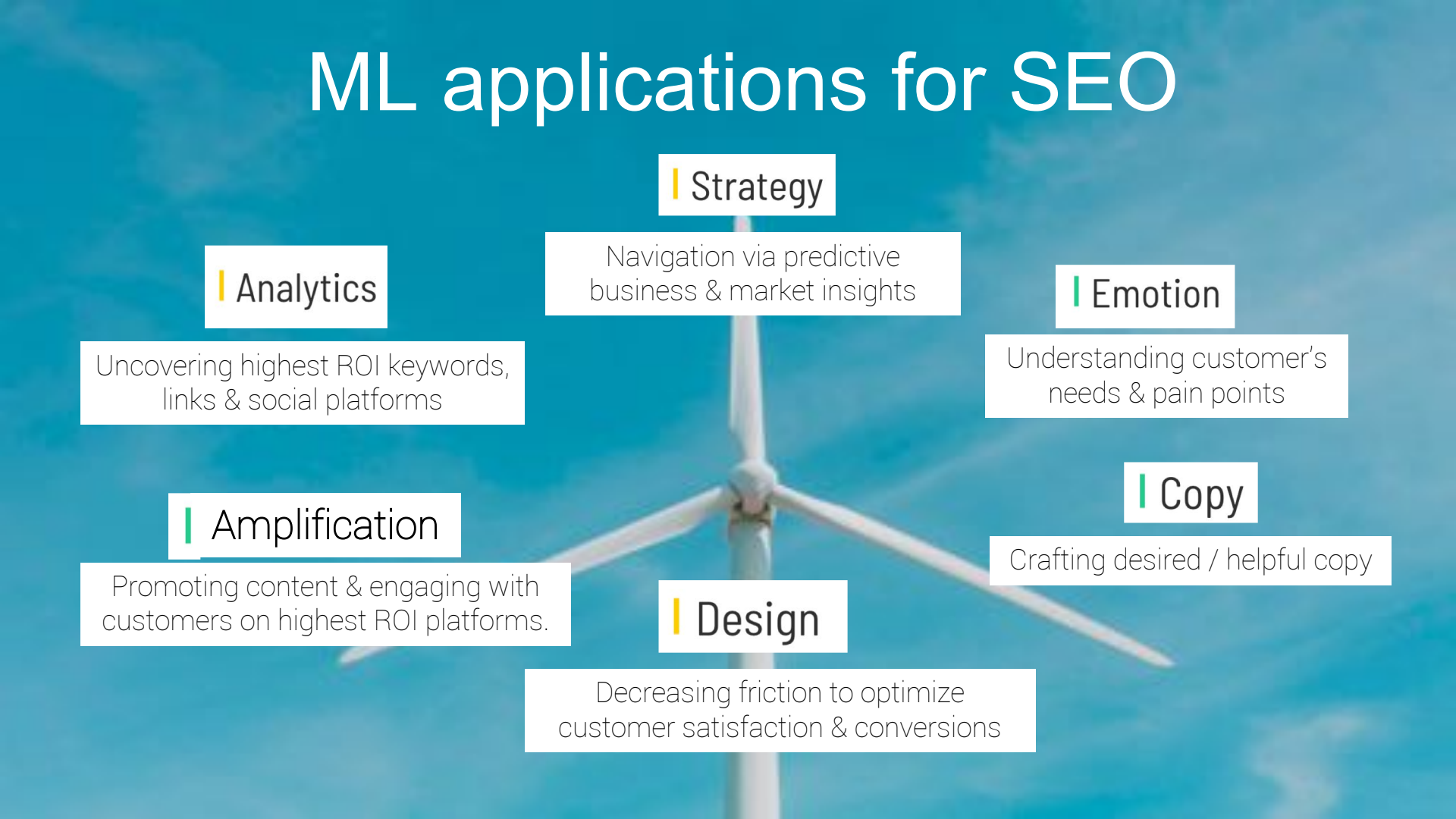
🥳 I'm super excited to share with you my newest prototype: a real-time plain text editor 📝 for podcasts 🎙 !

✂️ Cut out the parts you don't want, or
🧩 rearrange topics without any hassle.

🌊 Waveforms are so 20th century! 🤸



ML applications for SEO



Strategy

Navigation via predictive business & market insights

Analytics

Uncovering highest ROI keywords, links & social platforms

Emotion

Understanding customer's needs & pain points

Copy

Crafting desired / helpful copy

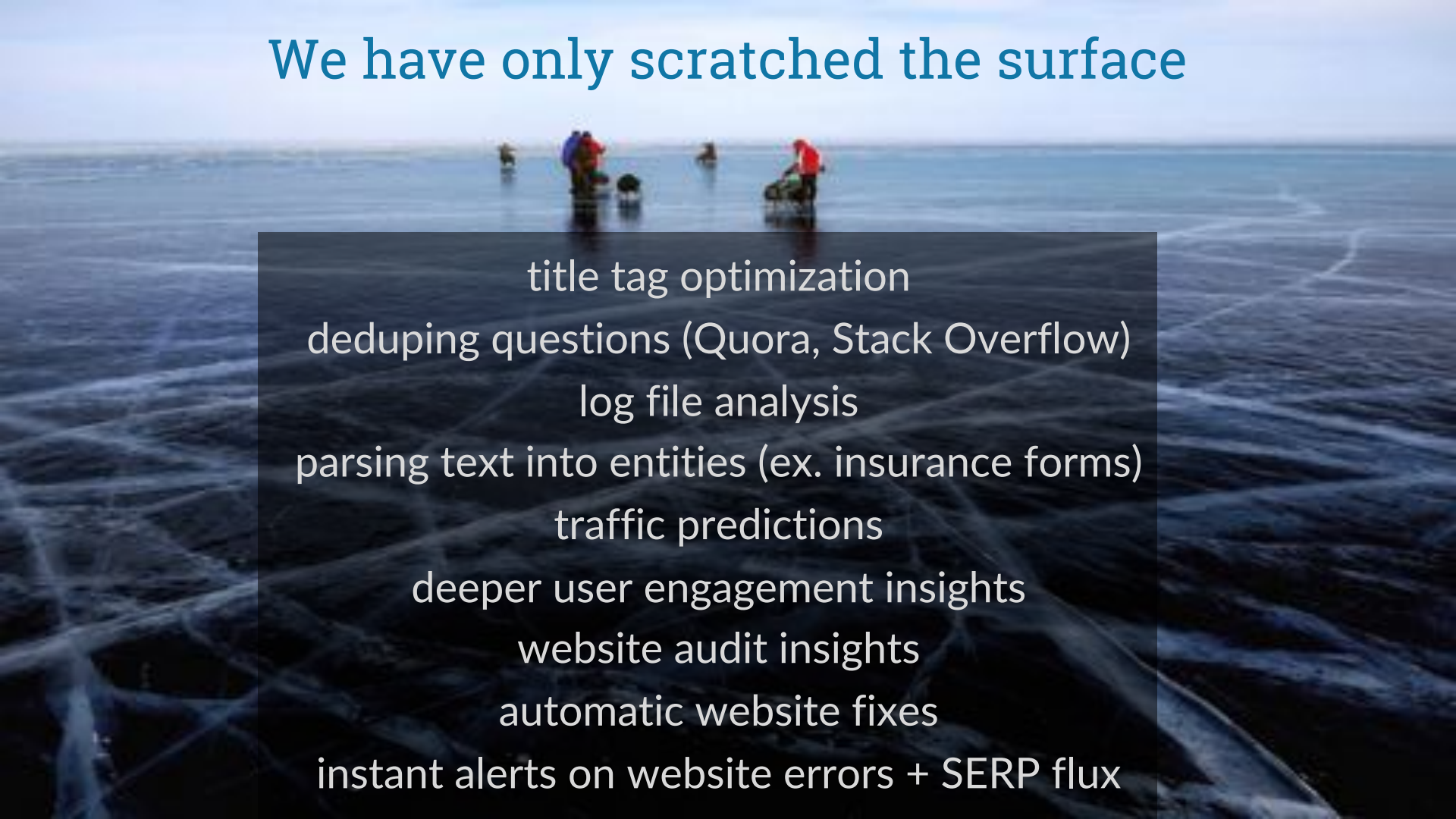
Amplification

Promoting content & engaging with customers on highest ROI platforms.

Design

Decreasing friction to optimize customer satisfaction & conversions

We have only scratched the surface



title tag optimization

deduping questions (Quora, Stack Overflow)

log file analysis

parsing text into entities (ex. insurance forms)

traffic predictions

deeper user engagement insights

website audit insights

automatic website fixes

instant alerts on website errors + SERP flux

Machine Learning for SEOs

- ~~How ML works~~
- ~~The Google Effect~~ ~~also…~~ 
- ~~SEO Applications~~
- Simple ML framework

Simple ML Framework

- What would you like to solve for?
- Do you have labeled data to help train a model?
- If not, can you start to collect data to help solve for your problem?
- Consider what data you currently have access to.

bit.ly/ml-framework

ML Problem Framing Worksheet

(This worksheet was transcribed from the original provided by Kshitij Gautam.)

Exercise 1: Start Clearly and Simply

Write what you'd like the machine learned model to do.

We want the machine learned model to...

Example: We want the machine learned model to predict how popular a video just uploaded now will become in the future.

Tips: At this point, the statement can be qualitative, but make sure this captures your real goal, not an indirect goal.

Exercise 2: Your Ideal Outcome

Your ML model is intended to produce some desirable outcome. What is this outcome, independent of the model itself. Note that this outcome may be quite different from how you assess the model and its quality.

ALGORITHMIA | AI LAYER

Product Pricing Resources Sign In Try It For Free Get Your Demo

Operationalize Your Machine Learning Life Cycle

The AI Layer abstracts your models from infrastructure,
so you can do more of what matters as you scale.

Get Your Demo Try It For Free

kaggle Search kaggle Competitions Datasets Kernels Discussion Learn

Competitions

Documentation InClass

General InClass Sort by Grouped

All Categories Search competitions

18 Active Competitions

Spooky Author Identification
Share code and discuss insights to identify horror authors from their writings
Playground · 7 months ago · literature, linguistics, multiclass classification

$25,000
1,244 teams

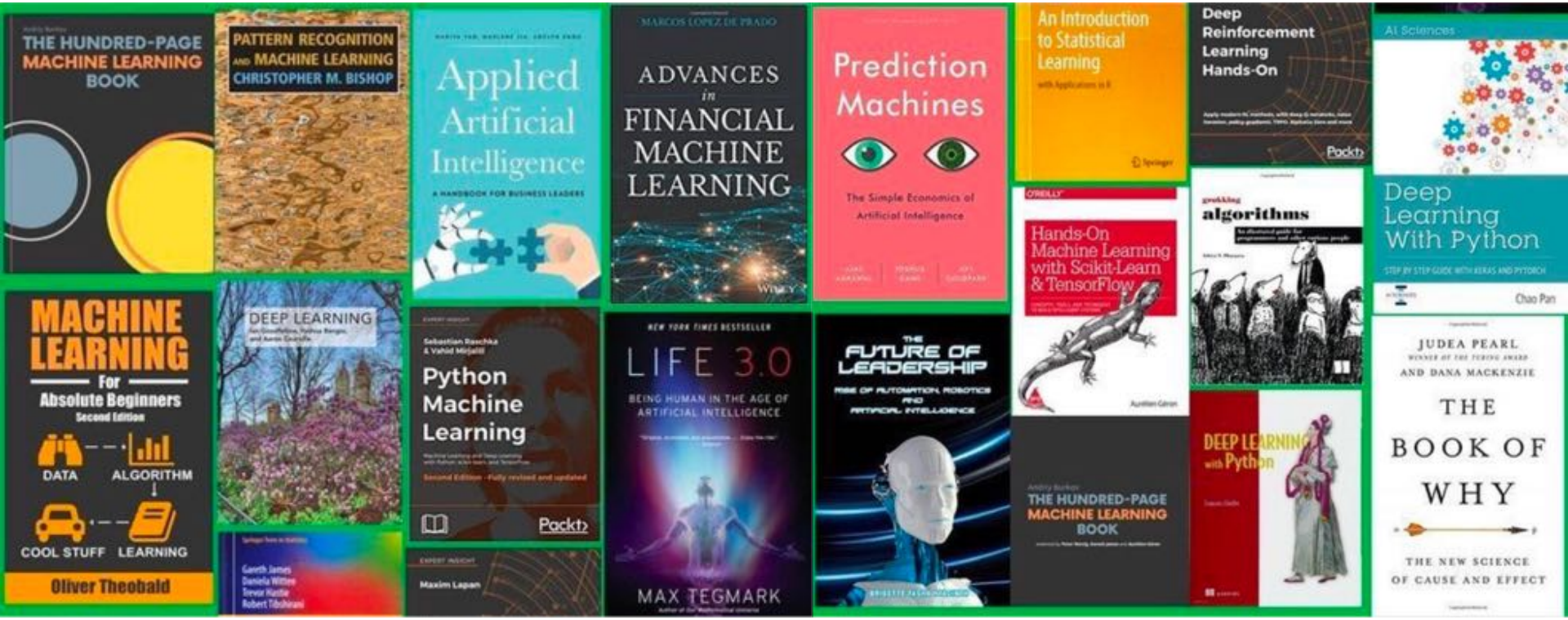
Passenger Screening Algorithm Challenge
Improve the accuracy of the Department of Homeland Security's threat recognition algorithms
Featured · 7 months ago · terrorism, image data, object detection

$1,500,000
518 teams

Getting Started

- Search ‘Harvard CS109’ in GitHub
- Learn Python in 10 Mins
- Google CodeLabs – Break things!!!
- MNist --The “Hello World!” of Machine Learning
- Colab Notebooks OR Jupyter Notebooks
- Learn With Google AI
- Image-net.org
- Kaggle
- MonkeyLearn

Top ML Books



MOZ

Free ML Books: bit.ly/free-ml-books

(With a free Data Science Central account)

- Statistics: New Foundations, Toolbox, and Machine Learning Recipes
- Classification and Regression in a Weekend
- Online Encyclopedia of Statistical Science
- Azure Machine Learning in a Weekend
- Enterprise AI - An Application Perspective
- Applied Stochastic Processes

Machine Learning Toolkit for SEO (MLTS)

Repository for ideas resources, APIs, data, models, and data munging for testing ML theories useful for SEO

Repositories 1 | Packages | People 5 | Teams | Projects

Find a repository... | Type: All | Language: All | New

MLTS

Machine Learning Toolkit for SEO

Jupyter Notebook | Apache-2.0 | 15 | 54 | 0 | 1 | Updated 2 minutes ago

Top languages

Jupyter Notebook

People 5 >

Python For Data Science *Cheat Sheet*

Python Basics

Learn More Python for Data Science Interactively at www.datacamp.com

Variables and Data Types

Variable Assignment

```
>>> x=5
>>> x
5
```

Calculations With Variables

```
>>> x+2
7
>>> x-2
3
>>> x*2
10
>>> x**2
25
>>> x%2
1
>>> x/float(2)
2.5
```

Sum of two variables
Subtraction of two variables
Multiplication of two variables
Exponentiation of a variable
Remainder of a variable
Division of a variable

Types and Type Conversion

str()	'5', '5.45', 'True'	Variables to strings
int()	5, 3, 1	Variables to integers
float()	5.0, 1.0	Variables to floats
bool()	True, True, True	Variables to booleans

Asking For Help

```
>>> help(str)
```

Lists

Also see NumPy Arrays

```
>>> a = 'is'
>>> b = 'nice'
>>> my_list = ['my', 'list', a, b]
>>> my_list2 = [[4,5,6,7], [3,4,5,6]]
```

Selecting List Elements

Index starts at 0

Subset

```
>>> my_list[1]
>>> my_list[-3]
```

Select item at index 1
Select 3rd last item

Slice

```
>>> my_list[1:3]
>>> my_list[1:]
>>> my_list[:3]
>>> my_list[:]
```

Select items at index 1 and 2
Select items after index 0
Select items before index 3
Copy my_list

Subset Lists of Lists

```
>>> my_list2[1][0]
>>> my_list2[1][:2]
```

my_list[list][itemOfList]

List Operations

```
>>> my_list + my_list
['my', 'list', 'is', 'nice', 'my', 'list', 'is', 'nice']
>>> my_list * 2
['my', 'list', 'is', 'nice', 'my', 'list', 'is', 'nice']
>>> my_list2 > 4
True
```

List Methods

```
>>> my_list.index(a)
>>> my_list.count(a)
>>> my_list.append('!')
>>> my_list.remove('!')
>>> del(my_list[0:1])
>>> my_list.reverse()
>>> my_list.extend('!')
>>> my_list.pop(-1)
>>> my_list.insert(0,'!')
>>> my_list.sort()
```

Get the index of an item
Count an item
Append an item at a time
Remove an item
Remove an item
Reverse the list
Append an item
Remove an item
Insert an item
Sort the list

Libraries

Import libraries

```
>>> import numpy
>>> import numpy as np
```

Selective import

```
>>> from math import pi
```

pandas — Data analysis
scikit-learn — Machine learning
NumPy — Scientific computing
matplotlib — 2D plotting

Install Python

ANACONDA
Leading open data science platform powered by Python

spyder
Free IDE that is included with Anaconda

jupyter
Create and share documents with live code, visualizations, text, ...

Numpy Arrays

Also see Lists

```
>>> my_list = [1, 2, 3, 4]
>>> my_array = np.array(my_list)
>>> my_2darray = np.array([[1,2,3],[4,5,6]])
```

Selecting Numpy Array Elements

Index starts at 0

Subset

```
>>> my_array[1]
2
```

Select item at index 1

Slice

```
>>> my_array[0:2]
array([1, 2])
```

Select items at index 0 and 1

Subset 2D Numpy arrays

```
>>> my_2darray[:,0]
array([1, 4])
```

my_2darray[rows, columns]

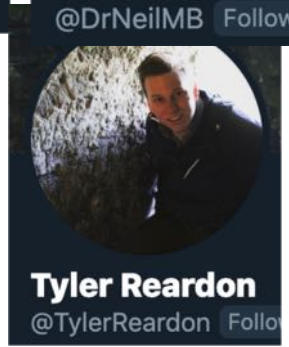
Numpy Array Operations

```
>>> my_array > 3
array([False, False, False,  True], dtype=bool)
>>> my_array * 2
array([2, 4, 6, 8])
>>> my_array + np.array([5, 6, 7, 8])
```

Advanced Resources

- Yearning Learning (free book preview by Andre Ng)
- Neural Networks & Deep Learning
- Correlation vs Causation (by Dr. Pete!)
- Exploring Word2Vec
- The Zipf Mystery
- BigML
- Targeting Broad Queries in Search
- Project Mosaic Books
- Algorithmia
- LDA
- Learn Python
- Massive Open Online Courses
- Coursera Machine Learning
- RAY by Professors at UC Berkeley
- How to eliminate bias in data driven marketing
- TensorFlow Dev Summit 2018 [videos]
- NLP Sentiment Analysis
- Talk 2 Books
- The Shallowness of Google Translate
- TF-IDF
- LSI

People to follow



MOZ

ML for SEOs Takeaways:

1. ML can shorten the path between data ←--→ insights
2. An ML model is only as good as its training data
3. Consider the data you have & what you could do with it
4. YOU can create an ML model today
5. Diversity is paramount in ML moving forward.
6. ML will help us level up as an industry

MOZ

The Data Science Team at Moz is innovating in this space to make your journey from data to insights more efficient

think differently

MOZ

think differently

What will you solve for?

MOZ

Thank you!

@BritneyMuller

britney@moz.com

prediction: cowboy hat, ten-gallon hat

probability: 0.9708651900291443