Google Cloud

Analyzing unstructured data

Data Engineering on Google Cloud Platform

Notes:

25 slides + 1 lab: 1 hour

Remember?

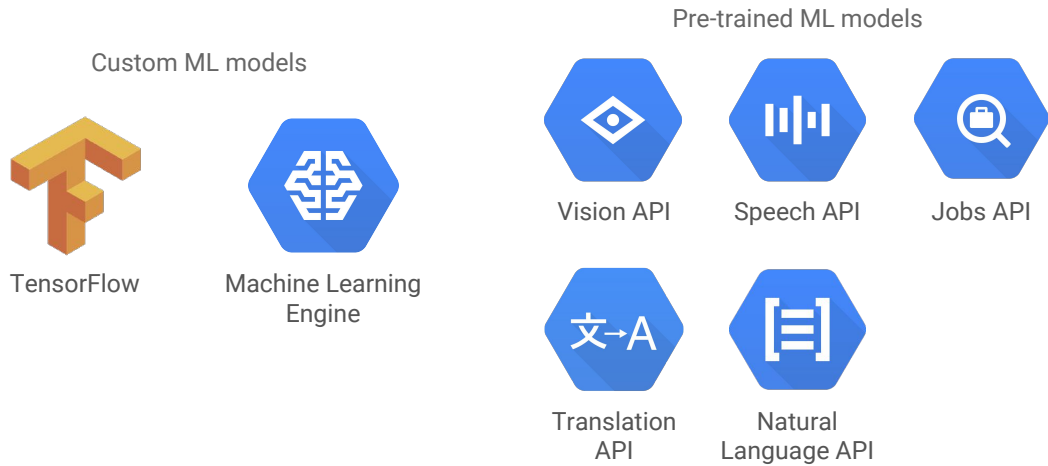| Human	Easy counting problems	Harder counting problems
Real-time insight into supply chain operations. Which partner is causing issues?	Did error rates decrease after the bug fix was applied?	Are programmers checking in low-quality code?
Drive product decisions. How do people really use feature X?	Which stores are experiencing long delays in payment processing?	Which stores are experiencing lacking of parking space?

Notes:

Compare these with the ones on the previous slide. Structured vs. unstructured.

Low-quality could be determined by bad sentiment in code reviews and often through programmer's own negative comments in the code. They are checking in the code because they need to move on to their next project …. But there are also tools out there that will look for code-smells. Those tools can be run at scale on Dataproc.

Agenda

Infuse your business with Machine Learning + Lab

Pretrained models are a fast way to magical experiences

Custom ML models

Pre-trained ML models

TensorFlow

Machine Learning Engine

Vision API

Speech API

Jobs API

Translation API

Natural Language API

Notes:

We'll look at TF & MLE next. But right now, let's talk about pre-trained ML models.

ML APIs are accessed through REST APIs -- no machine learning knowledge required

Invoke Vision API

The Vision API can work off an image in Cloud Storage or embedded directly into a POST message. I'll use C

请您爱护和保护卫生创建优美水环境

. That photograph is from http://www.publicdomainpictures.net/view-image.php?image=15842

Image (GCS/embedded)

JSON request

JSON response

```
# Running Vision API
import base64
IMAGE="gs://cloud-training-demos/vision/sign2.jpg"
vservice = build('vision', 'v1', developerKey=APIKEY)
request = vservice.images().annotate(body={
        'requests': [{
                'image': {
                    'source': {
                        'gcs_image_uri': IMAGE
                    }
                },
                'features': [{
                    'type': 'TEXT_DETECTION',
                    'maxResults': 3,
                }]
        }],
    })
responses = request.execute(num_retries=3)
print responses
```

{u'responses': [{u'textAnnotations': [{u'locale': u'zh', u'description': u'\u8bf7\u6
\u4f18\n\u7f8e\u6c34\u73af\u5883\n', u'boundingPoly': {u'vertices': [{u'y': 103, u'x
 {u'y': 654, u'x': 150}]}}, {u'description': u'\u8bf7', u'boundingPoly': {u'vertices
41, u'x': 322}, {u'y': 241, u'x': 178}]}}, {u'description': u'\u60a8', u'boundingPol
471}, {u'y': 241, u'x': 471}, {u'y': 241, u'x': 327}]}}, {u'description': u'\u7231\u

Notes:

This was a lab in the fundamentals course. They've done this already.

Notes:

https://pixabay.com/en/list-zettelbox-note-leaves-stack-1925395/ (cc0)
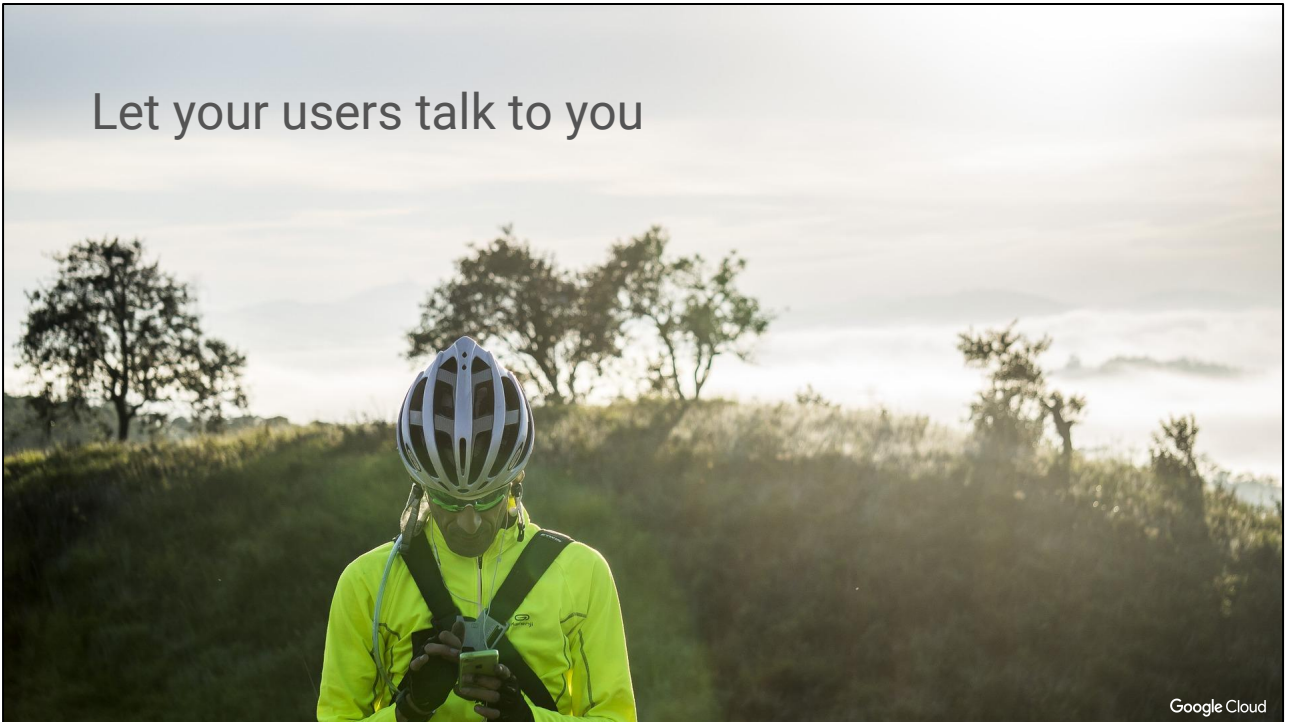
Notes:

idea : voice-enable your applications

Use Speech API

Just a REST call, so easy to incorporate

Voice-navigation?

Know your user/from their app/anticipate their need/carry on conversation = ASSISTANT story
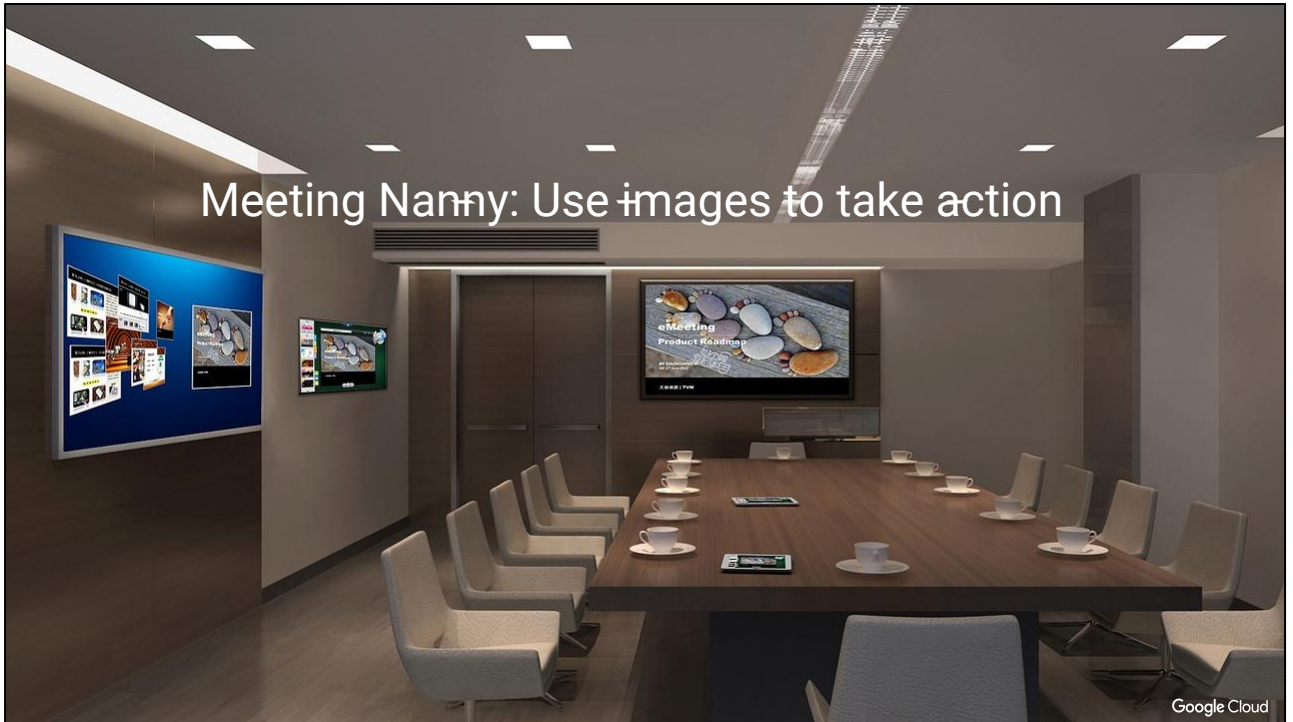
Go from speech to action with bots

https://dialogflow.com/

https://cloud.google.com/dialogflow-enterprise/

Notes:

https://cloud.google.com/dialogflow-enterprise/

Notes:

Image from https://pixabay.com/en/interior-design-tv-multi-screen-828545/
(cc0)

From:
https://g3doc.corp.google.com/java/com/google/corp/bizapps/rews/spaces
aver/g3doc/gvc.md?cl=head

We do occupancy detection via motion detection (by the VC camera) and by
call ID matching. Every 30 seconds, the VC unit sends a Pubsub notification
whether motion was detected or not. It also sends a Pubsub notification when
a call started or ended together with whether the call ID matched the meeting
ID.
If motion is detected between 6 and 8 minutes after the meeting start time, the
room counts as occupied. Otherwise, it's empty.

Notes:

- World's largest online only grocery supermarket
- Goal for best customer service
- Customers call or email their contact center (Social media, landline, email, SMS)
- Types: General feedback, refunds, redeliver, payment issues
- No forms or self categorization…all emails in a central mailbox
- Traditionally, each email gets addressed and routed. Can't scale, longer delays, poor experience
- Sifting through email is a repetitive task
- Ocado Technology w/ 1000+ developers, engineers, data scientists
- Used Natural Language processing: combines computer science, artificial intelligence, and computational linguistics
- Parse through the body of emails, tags and routes to help contact center reps determine the priority and context

Wootric collects both numeric and qualitative feedback

EASY TO COMPUTE NET PROMOTER SCORE

FREE FORM TEXT -- NOT AS EASY TO HANDLE

https://cloud.google.com/blog/big-data/2017/03/analyzing-customer-feedback-using-machine-learning

They use NLP API and custom ML models to classify sentiment, topic, and support personnel

Sentiment Volume

WOOTRIC

● PRICING-TOO-HIGH
● USABILITY
● PRODUCT-PROBLEM
● PRODUCT-COST
● DELIVERY-SHIPPING
● BILLING-PAYMENTS
● PRICING
● SHIPPING
● BILLING
○ CS-SARANG

-1.0   -0.5   0.0   +0.5   +1.0

sentiment
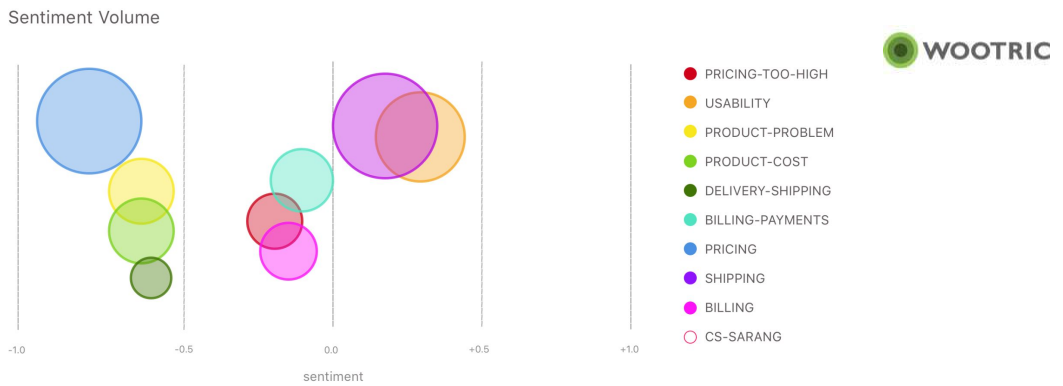
Notes:

Support personnel is not shown, but the idea is that if the feedback mentions "jessica", they know who is being talked about.

Lab 6: Add Machine Learning (ML)

- Simplify machine learning tasks using the NLP API

- Incorporate several machine learning services from the Natural Language API

- Use sentiment analysis and entity analysis, to produce meaningful results from unstructured data.

cloud.google.com