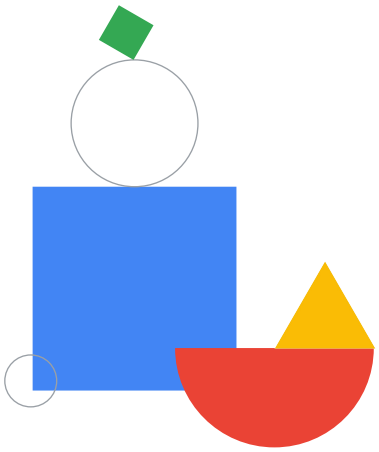
Google Cloud

Deriving Insights From Unstructured Data Using Machine Learning



Agenda

01 ML Drives Business Value

02 How does ML on unstructured data work?

03 Choosing the Right ML Approach

04 Pre-built AI Building Blocks
- Demo: Vision API
- Lab: Extract, Analyze, and Translate Text from Images with the Cloud ML APIs

05 Customizing Pre-built Models with AutoML
- Demo: AutoML Vision

06 Building a Custom Model
- Lab: Training with Pre-built ML Models using the Vision API and AutoML

ML Drives Business Value

Examples of real-world ML use cases

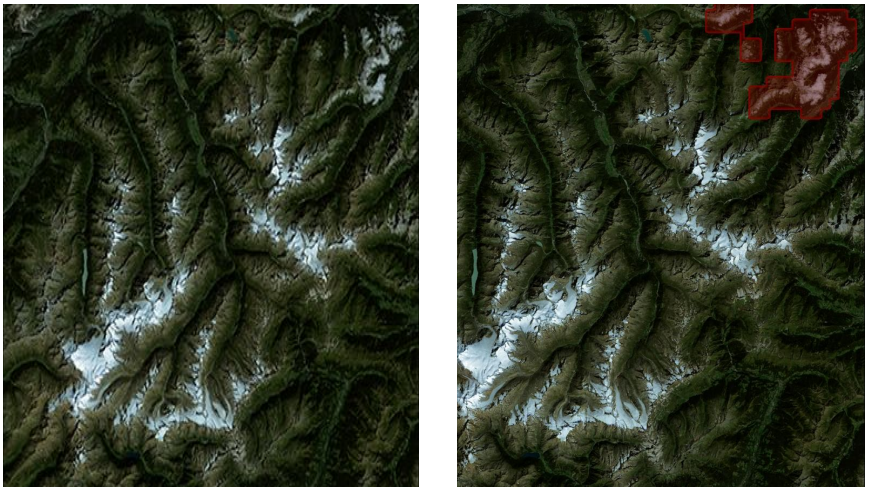
AUCNET	ocado	GIPHY	UNIQLO
Custom image model to price cars	Build off NLP API to route customer emails	Use Vision API as-is to find text in memes	Use Dialogflow to create a new shopping experience

Uniqlo (uni-clo) designed a shopping chatbot using A Dialogflow UI. Dialogflow is a google-owned company which specializes in building ML-based interfaces like intelligent chatbots.

Let's look at some other use cases for ML in business

[Resource]
Google NeXT Customer Success Stories
https://youtu.be/BwWg__HVfsM?t=4m41s

Clouds or snow-capped mountains?



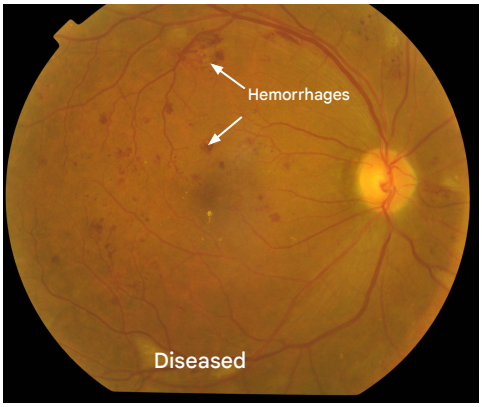
You might be Airbus, and use machine learning to differentiate between clouds and snow cover. If you're stumped like I am, the clouds are in the upper-right part of the right image highlighted in red.

https://cloud.google.com/blog/products/gcp/google-cloud-machine-learning-now-open-to-all-with-new-professional-services-and-education-programs

Empty or full?

You might be an economic forecast firm looking to track the global fleet of container ships via satellite imagery -- knowing the amount of cargo being carried might help improve your economic forecast days or months ahead of the official numbers.

Diagnosing diabetic retinopathy



Medical images are ripe for innovation. For example, You could diagnose medical conditions like Diabetic Retinopathy earlier when it's easier to treat and prevent blindness.

https://ai.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html

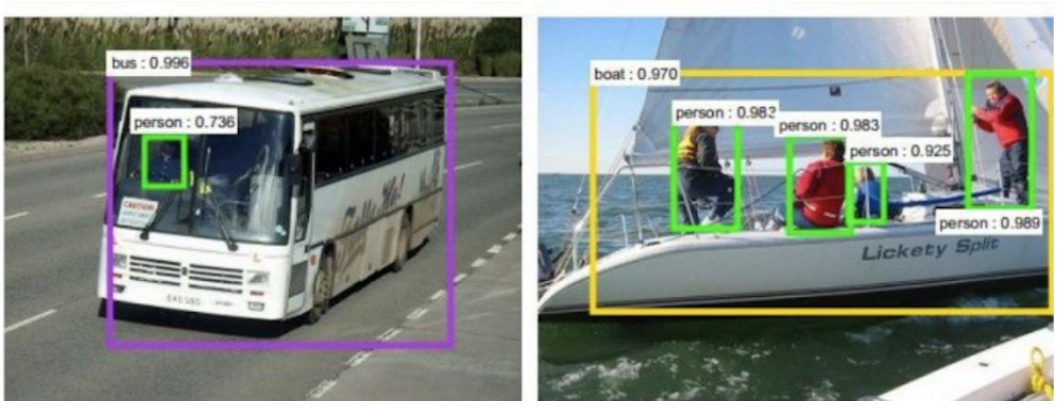
Recap: Image classification **automates tasks** that are easy (and not easy) for humans

The key takeaway is that ML can automate tasks that may save or assist a human team and, recently, the latest models are even out-performing humans in some domains.

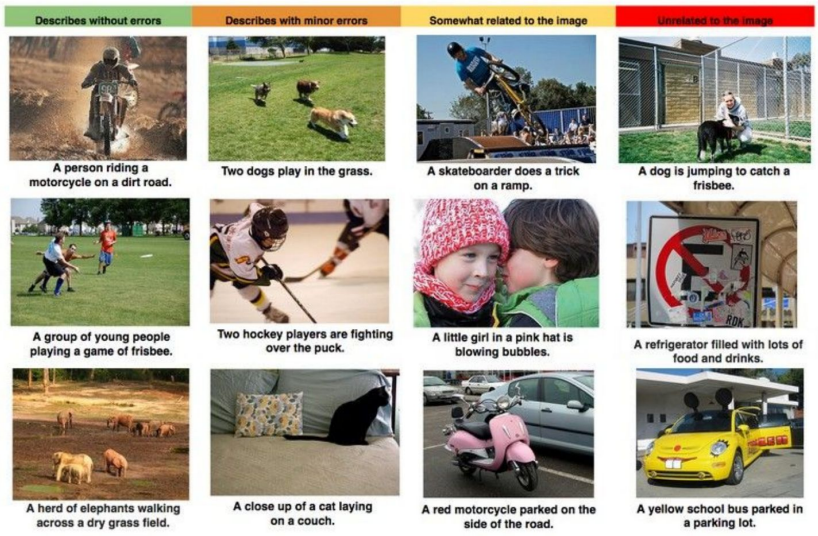
Machine learning can label images or video

Also, image classification as a field is more than just a binary classification tool. Later in this course you'll experiment with the pre-built Vision API model, which allows you to pass through a JSON request and get back a ranked list of associated labels for your image.

Machine learning can detect objects in images or video



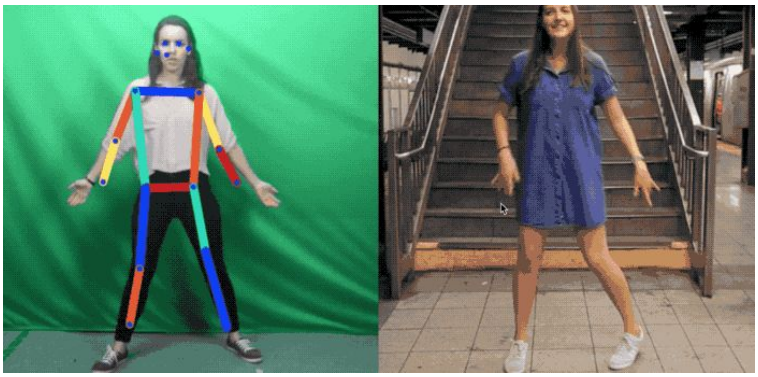
Also, if you have more than one subject matter in a photo, it can draw bounding boxes and classify pieces of an image as well.



Show and Tell: A Neural Image Caption Generator Vinyals et al 2015:
https://arxiv.org/abs/1411.4555

Modern image classification models can even generate captions describing what is going on in the image, like a map of dependencies like “two hockey players are fighting over a puck”. Here, it’s important to call out that even the best models can and will make mistakes in their predictions (like the road sign captioned as “a refrigerator filled with lots of food and drinks”.)

Having fun with ML: Pose detection



g.co/movemirror

As you saw, there were a lot of impressive uses for machine learning these days, like detecting objects in images, helping to detect diseases, and even enabling cars to drive themselves.

But AI can also be used in more playful ways too. Through a pose-estimation model, a Google AI Experiment called Move Mirror can match your real-time movements to hundreds of images of people doing similar poses from around the world. Feel free to try it out yourself and have some fun. Then, tune back into this course to learn how image classification models extract features like these from images.

https://www.blog.google/technology/ai/move-mirror-you-move-and-80000-images-move-you/
https://experiments.withgoogle.com/move-mirror
https://medium.com/tensorflow/move-mirror-an-ai-experiment-with-pose-estimation-in-the-browser-using-tensorflow-js-2f7b769f9b23

How Does ML on Unstructured Data Work?

Earlier in the course, you built custom models with BigQuery ML using columns of data as features and predicted the value of your labels. But what about unstructured data? How does ML learn the features?

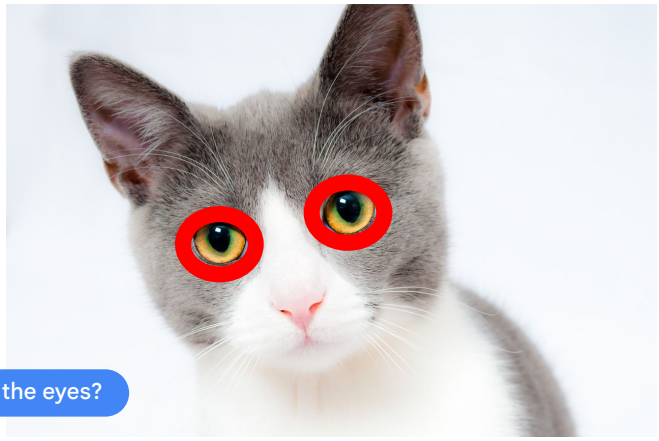
It’s best illustrated with a quick example.

What do you see here? It’s a cat -- but how you know?

We know this is a cat, but how would you teach a machine?

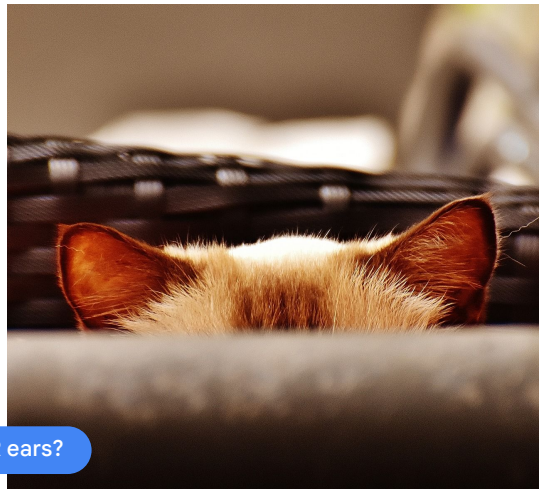
Your eyes have the benefit of many many years of evolution and intuition to allow you to perceive and interpret those pixels on the screen. How could we teach a machine to understand that this particular collection of pixels is a cat?

We know this is a cat, but how would you teach a machine?



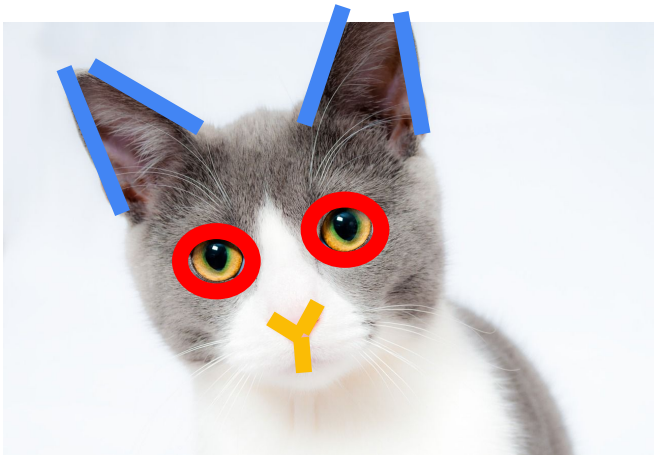
If you let yourself fall back into rules making, you might say look for cat-like eyes in the images.

What about this?



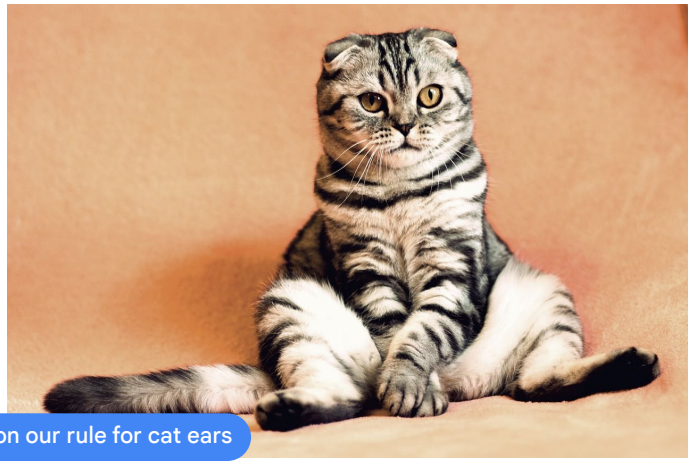
What about this image? Your brain still knows it’s likely a cat, but the machine now has no basis to go off of with our old rule of “look at the eyes”.

We've added more rules that we pre-define as "cat-like"



Okay, what if we added a bunch more hardcoded rules, like look for eyes, ears, and a nose?

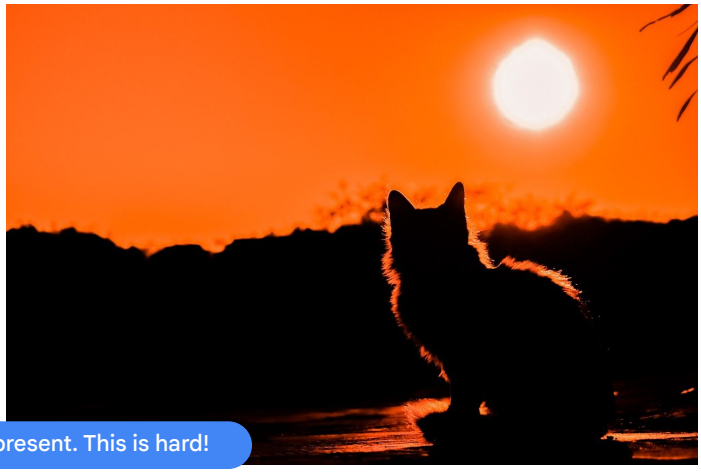
What about this?



Scores pretty low on our rule for cat ears

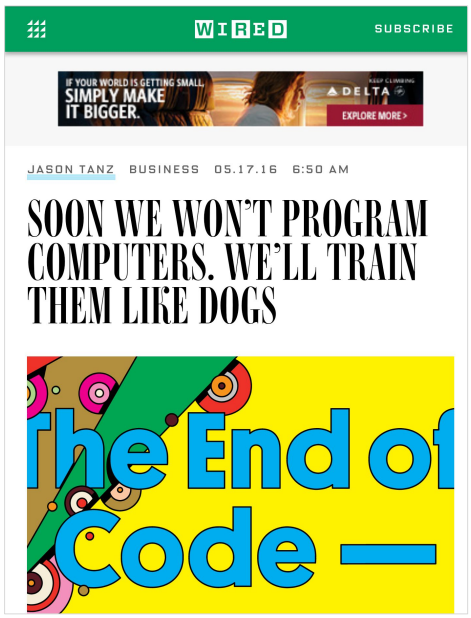
So, is this still a cat?

Or even this?



What about this?

Again, hardcoding rules completely fails us here and thats where deep learning comes into play.



WIRED's headline

"If you want to teach a neural network to recognize a cat, for instance, you don't tell it to look for whiskers, ears, fur, and eyes. You simply show it thousands and thousands of photos of cats, and eventually it works things out."

As this Wired article states, if you want to teach a neural network to recognize a cat you simply show it thousands of photos of cats and let it figure it out.

This is similar to how we teach children to recognize and classify new objects.

Google in 2012: Show the computer 10 million images, have it find cats

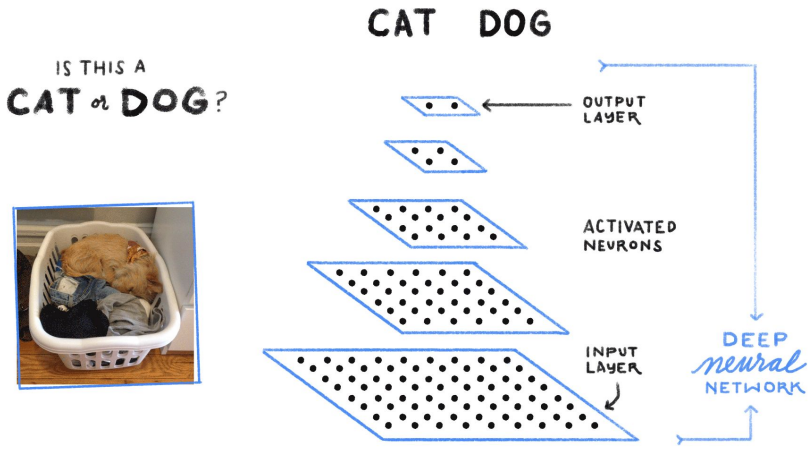
And in 2012 that's exactly what the Google Research team with Jeff Dean and Andrew Ng did. What you see here is what the deep learning neural network figured out what a cat is based on looking at over 10 million images and processing the model over 16,000 computers.

[resource]
2012 Research Paper 16,000 computers in a neural network to identify a cat face
https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/38115.pdf

Follow Google's AI research blog:
https://ai.googleblog.com/

https://googleblog.blogspot.com/2012/06/using-large-scale-brain-simulations-for.html

Modern AI applications use Deep Learning

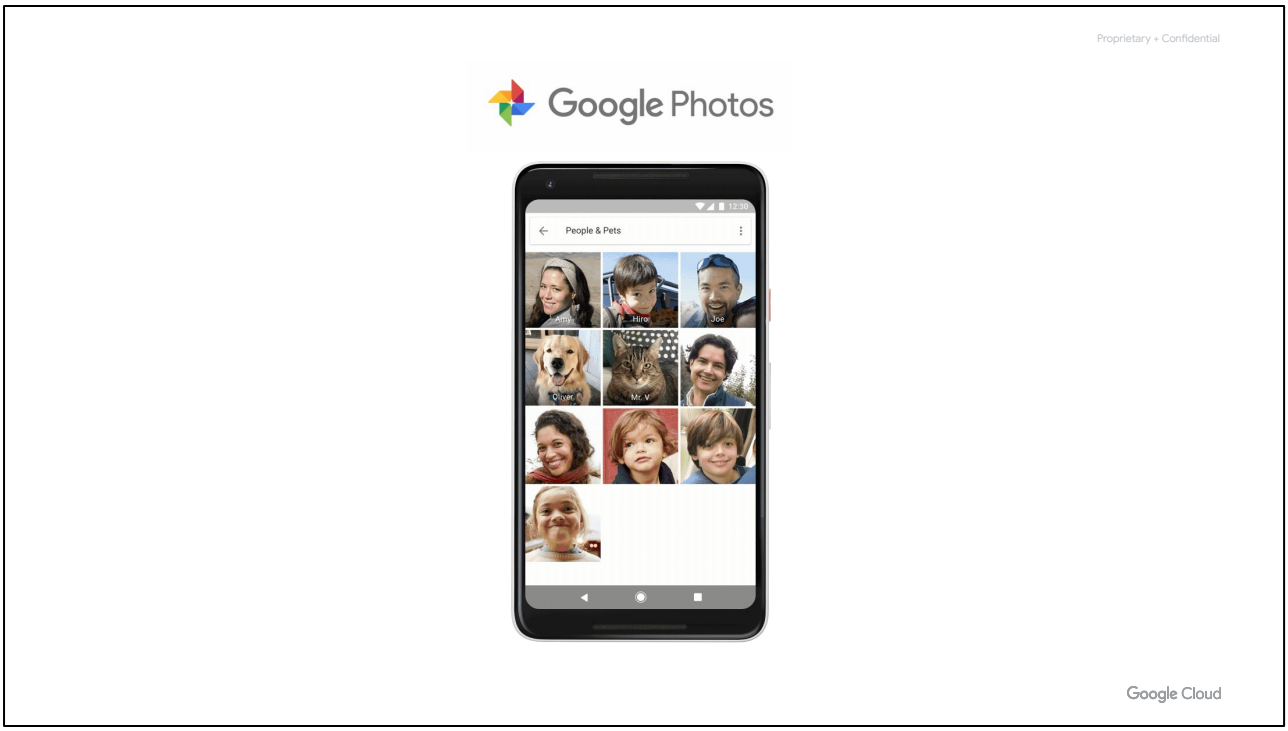


Now a familiar architecture for deep learning is the neural network which is a model inspired by our own human brain. Here it takes the input image and classifies it as a cat or dog.

Again, we're not telling the model to focus on looking for dog collars or cat whiskers, it builds its own recipe for determining the correct label to apply at the end.

And, as you can see from the image, modern ML models can scale and handle even tricky data points like this dog hiding in a laundry basket!

Google NN (gif source):
http://selmandesign.com/qa-on-machine-learning/



Machine Learning (specifically Deep Learning) is at the core of many Google Products like Google Photos, which can classify and group photos (like photos of your pets) together in an album.

Note that this is now multiple Machine Learning models in one:

- First it needs to identify whether the photo you just took is of a dog
- Second it needs to identify whether this dog is your pet based on comparing this photo to the history of photos you have with that specific dog

[resource]
Google Photos (gif source):
https://www.blog.google/products/photos/meow-its-even-easier-find-your-furry-friends-google-photos/

Google Photos Blog:
https://www.blog.google/products/photos/

Use **Deep Learning** when you can't explain the labeling rules

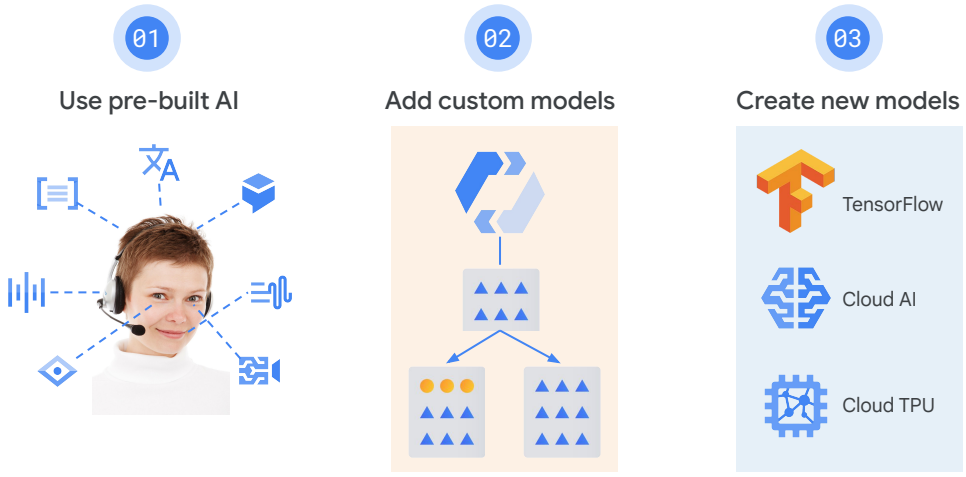
Deep Learning (remember that's a sub-discipline of Machine Learning) is incredibly useful when we as humans can't even map out our intuition about what makes a prediction correct or not. That's where we just show the computer ten thousand images of a cat and hope it figures it out. And with recent leaps in ML research at Google, computer vision has come a very long way.

Choosing the Right ML Approach

We've covered machine learning throughout this course but that was just one approach -- custom model building. In our first challenge you may recall you inherited a Spark ML job that your data science team created to predict housing rentals. The second challenge had you create a forecasting model with BigQuery ML from scratch -- although you saved time by coding and running it in just SQL.

It's now time to zoom back out on approaches to machine learning and even look at some ways we can apply it without having to write code at all.

Artificial Intelligence application strategy



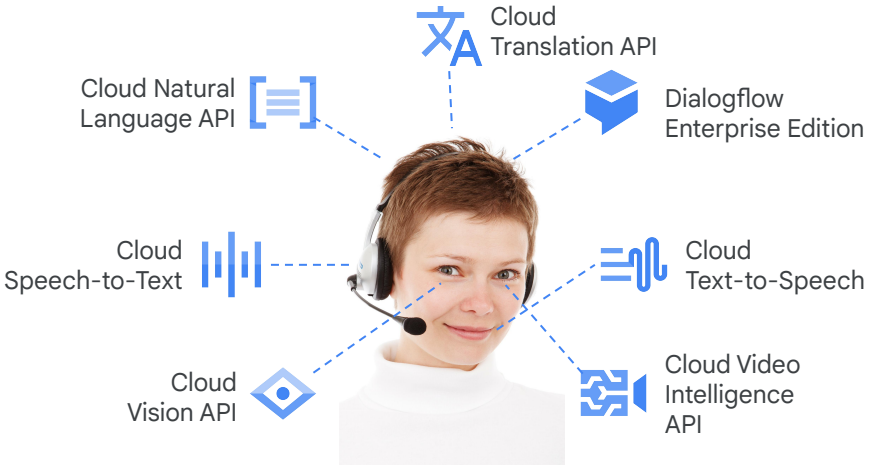
There are three approaches to AI that you should consider. You have already seen and built custom models with BQML and we have a separate set of courses on TensorFlow for even deeper model building. A good rule of thumb is to consider custom model building only when you have a lot of data like 100,000+ to millions of examples.

But what if you don't? Consider using pre-built AI which are models like the Video Intelligence and Cloud Vision APIs that you saw before. In addition, if you're looking to build a chatbot start with DialogFlow, which is a full-fledged application with ML built-in.

But what about if the building blocks don't work well for the specificity you need on your data? That's when you consider AutoML as a good candidate. It can even work with just a little bit of data like 10-100 images per label.

This lesson covers each of these approaches in detail.

Do you need a custom model?



Pre-built models are offered as services. In many cases these building blocks can be used to create the application you want without the expense or complexity of creating your own models.

Cloud Speech-to-Text converts audio to text for data processing. The Cloud Natural Language API recognizes parts of speech called entities and sentiment. Cloud Translation converts text in one language to another. Dialogflow Enterprise Edition is used to build chatbots to conduct conversations. Cloud Text-to-Speech converts text into high quality voice audio. The Cloud Vision API is for working with and recognizing content in still images. And the Cloud Video Intelligence API is for recognizing motion and action in video.

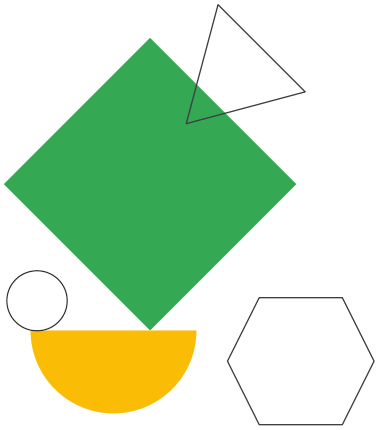
Pre-built AI Building Blocks

Let’s examine these pre-built models with a demo of the Cloud Vision API

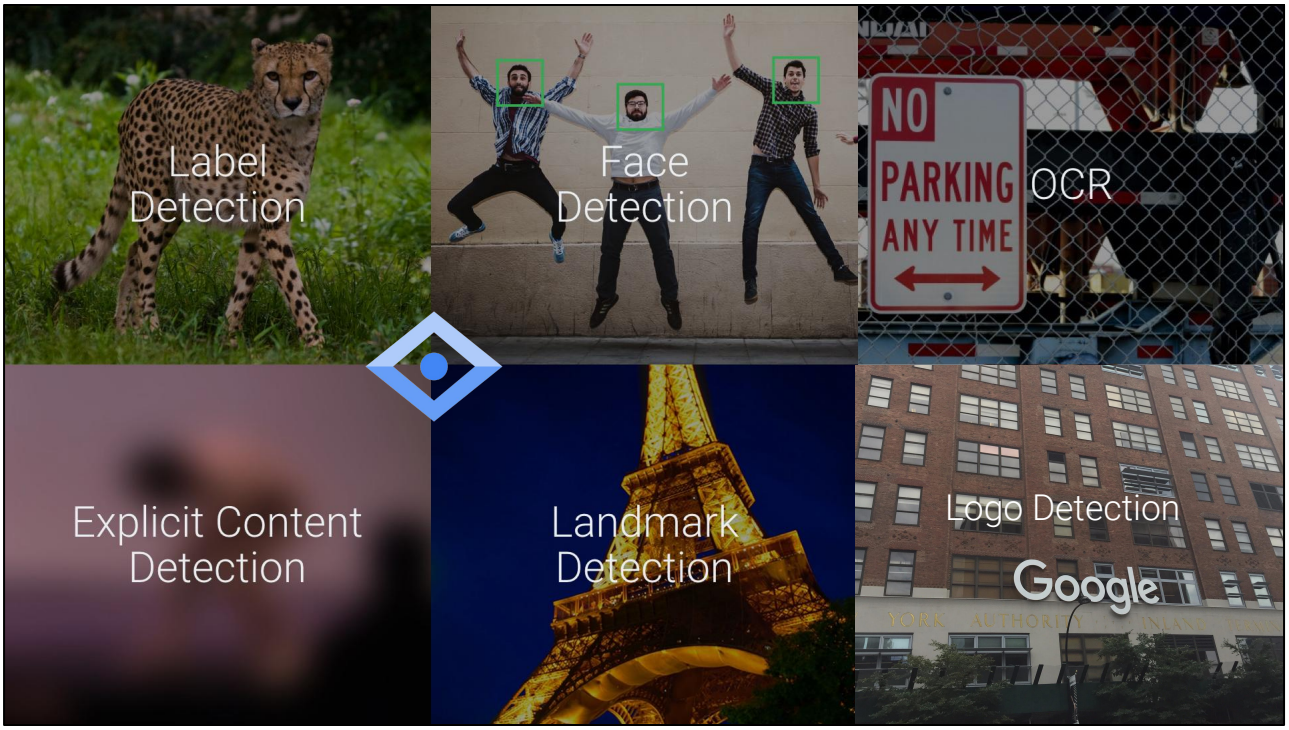
Demo

Vision API

Out-of-the-box, pre-built ML models



Refer to
https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/data-to-insights/demos/vision-api.md



Logo Detection: We can detect popular product logos within an image.

Google Cloud Cloud OnBoard

Face detection

```
"faceAnnotations" : [
  {
    "headwearLikelihood" : "VERY_UNLIKELY",
    "surpriseLikelihood" : "VERY_UNLIKELY",
    rollAngle" : -4.6490049,
    "angerLikelihood" : "VERY_UNLIKELY",
    "landmarks" : [
      {
        "type" : "LEFT_EYE",
        "position" : {
          "x" : 691.97974,
          "y" : 373.11096,
          "z" : 0.000037421443
        }
      },
  ...
    ],
    "boundingPoly" : {
      "vertices" : [
        {
          "x" : 743,
          "y" : 449
        },
        ...
```

```
"detectionConfidence" : 0.93568963,
"joyLikelihood" : "VERY_LIKELY",
"panAngle" : 4.150538,
"sorrowLikelihood" : "VERY_UNLIKELY",
"tiltAngle" : -19.377356,
"underExposedLikelihood" : "VERY_UNLIKELY",
"blurredLikelihood" : "VERY_UNLIKELY"
```

Here's an example of what the JSON response looks like for face detection - it's a picture a Googler on our team took with 2 other teammates on a trip to Jordan.

It returns an object for each face found in an image. You can see as part of the FaceAnnotations some pretty cool attributes like headwearLikelihood and joyLikelihood.

Google Cloud Cloud OnBoard

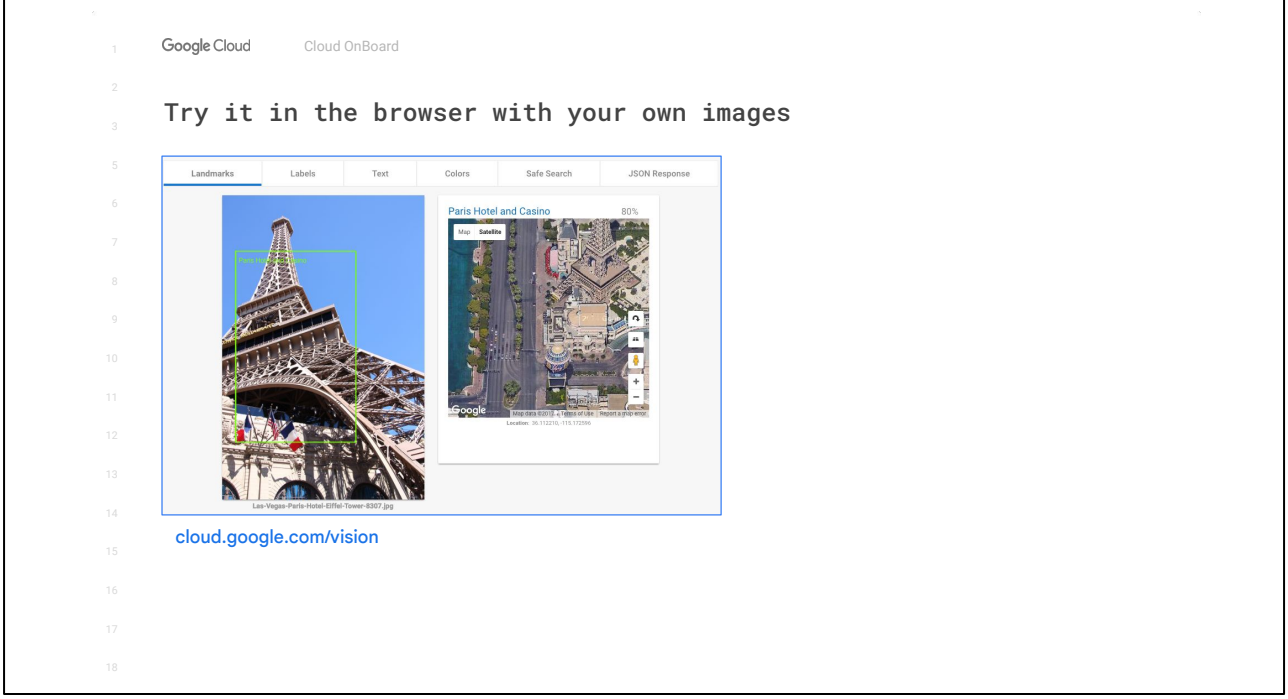
Web annotations

```
{
  "entityId": "/m/0gff2yr",
  "score": 5.92256,
  "description": "ArtScience Museum"
}
```

```
{
  "entityId": "/m/016ms7",
  "score": 1.44038,
  "description": "Ford Anglia"
}
```

```
{
  "entityId": "/m/0h898pd",
  "score": 7.4162,
  "description": "Harry Potter (Literary Series)"
}
```

The Vision API can also provide annotations on the image by looking up where and in what context this image or visually similar ones have appeared on the web. Take the image of this car which may not be immediately recognizable to some of you. But to others (and to Cloud Vision through web annotations) it's identified as the flying car from the Harry Potter series.



Take a minute and try Cloud Vision yourself directly in your browser. Navigate to cloud.google.com/vision and scroll down to try the API with your own image to upload.

https://cloud.google.com/vision

Google Cloud Cloud OnBoard

The Translation API supports 100+ languages

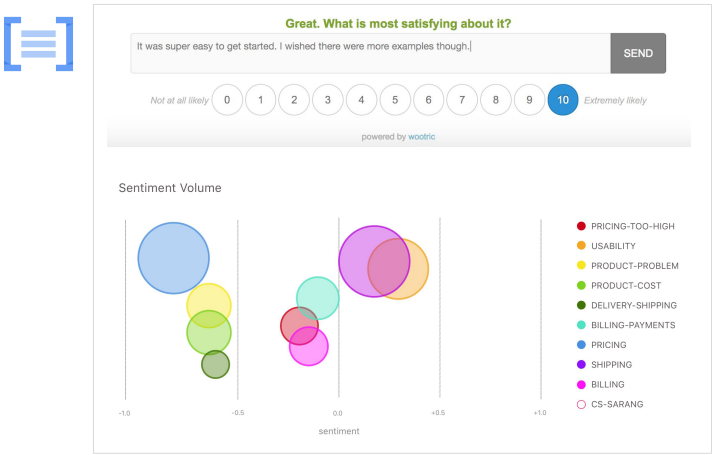
https://cloud.google.com/translate/

Another pre-built ML solution is the Translate API which underlies the product shown above. Simply place your camera over a sign, and it gets auto-translated for you. This is a combination of Vision API (to do optical character recognition) and Translate API (to do actual translation). Vision API supports 90+ languages, can detect the language of source text, and is highly scalable.

You can try this one too on the web.

Google Cloud Cloud OnBoard

Wootric uses the Cloud Natural Language API (entity and sentiment) to make sense of qualitative customer feedback

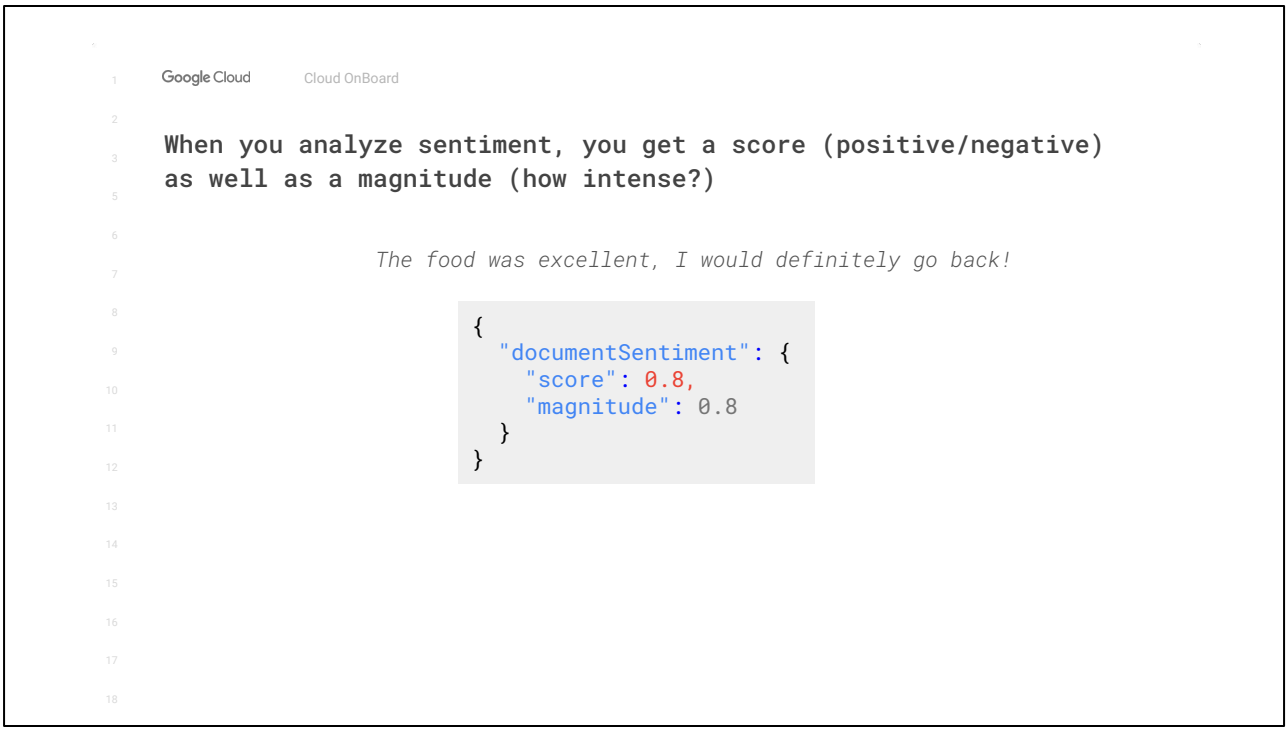


WOOTRIC
WIN CUSTOMERS FOR LIFE

Here's an example of a Google Cloud customer who uses the Cloud Natural Language API.

Wootric is a ML-driven customer feedback platform that helps businesses improve their customer service. They collect millions of free text customer survey responses each week. The use the Natural Language API to automate the text processing and sentiment analysis. In this visual you see the volume of the feedback on the vertical axis and the sentiment on the horizontal axis. Lastly, the coloring of the circles indicates which bucket of feedback that response was automatically classified into (like Usability feedback or Pricing feedback). This allows Wootric and similar organizations to intelligently route and prioritize customer feedback in real time.

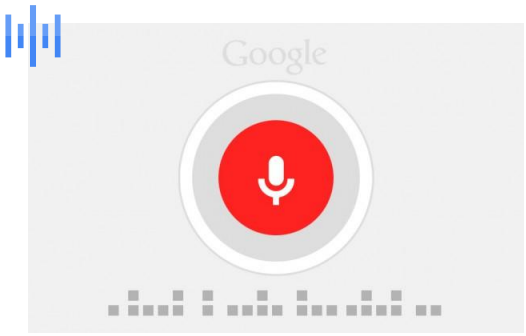
https://cloud.google.com/blog/big-data/2017/03/analyzing-customer-feedback-using-machine-learning



You can try out the pre-built model and get sentiment score for your own free text using the link provided. Note that you'll get a sentiment score (how positive or negative the text was) as well as a magnitude which indicates how intense of a feeling.

Google Cloud Cloud OnBoard

The Cloud Speech API can be used to transcribe audio to text



http://cloud.google.com/speech

Often multiple pre-built models are used together in a ML system. For example, say if you didn't have free text comments for user reviews but rather had audio from your customer's interactions with your call centers. If you wanted to get the sentiment of the customer's conversation you could first transcribe audio to text with the Cloud Speech API and then use the Natural Language API for sentiment as you saw before.

Google Cloud Cloud OnBoard

Like the Vision API, the Video Intelligence API can identify labels in a video, along with a timestamp

```
{
  "description": "Bird's-eye view",
  "language_code": "en-us",
  "locations": {
    "segment": {
      "start_time_offset": 71905212,
      "end_time_offset": 73740392
    },
    "confidence": 0.96653205
  }
}
```

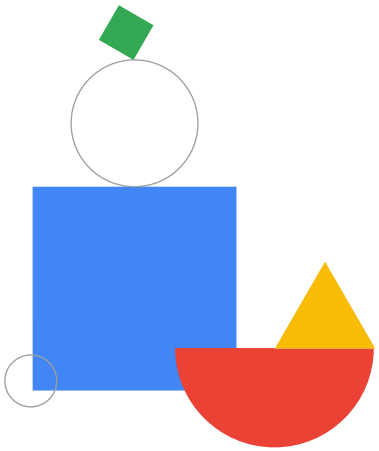
https://cloud.google.com/video-intelligence/

One last pre-built model you can leverage is the Video Intelligence API which is similar to the Cloud Vision API except for video instead of images. Here the API can identify labels within a video and when they occured as well as a confidence level. Here you see at 1:14 in the video the model is 96% confidence that this frame is a “bird’s eye view”

One real customer use case for this API are film companies who are looking to target and recommend movies to an audience based on similar movie-trailer watching behavior. These companies run all their movie trailers through the Video Intelligence API to label key features of the trailer (like “rugged” or “outer space” or “wild west”) and then can programmatically recommend similar movie trailers based on common themes.

Lab Intro

Extract, Analyze, and Translate Text from Images with the Cloud ML APIs



Lab objectives

01 Creating a Vision API request and calling the API with curl

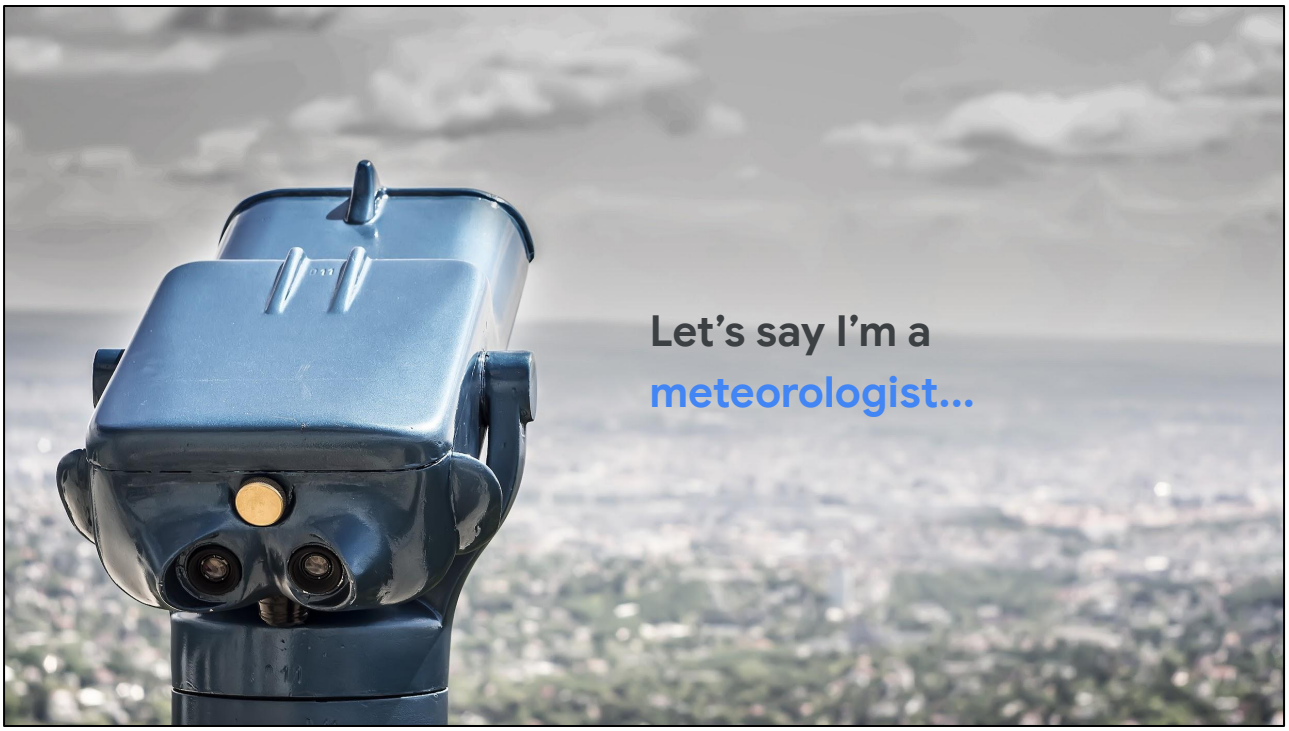
02 Using the text detection (OCR) method of the Vision API

03 Using the Translation API to translate text from your image

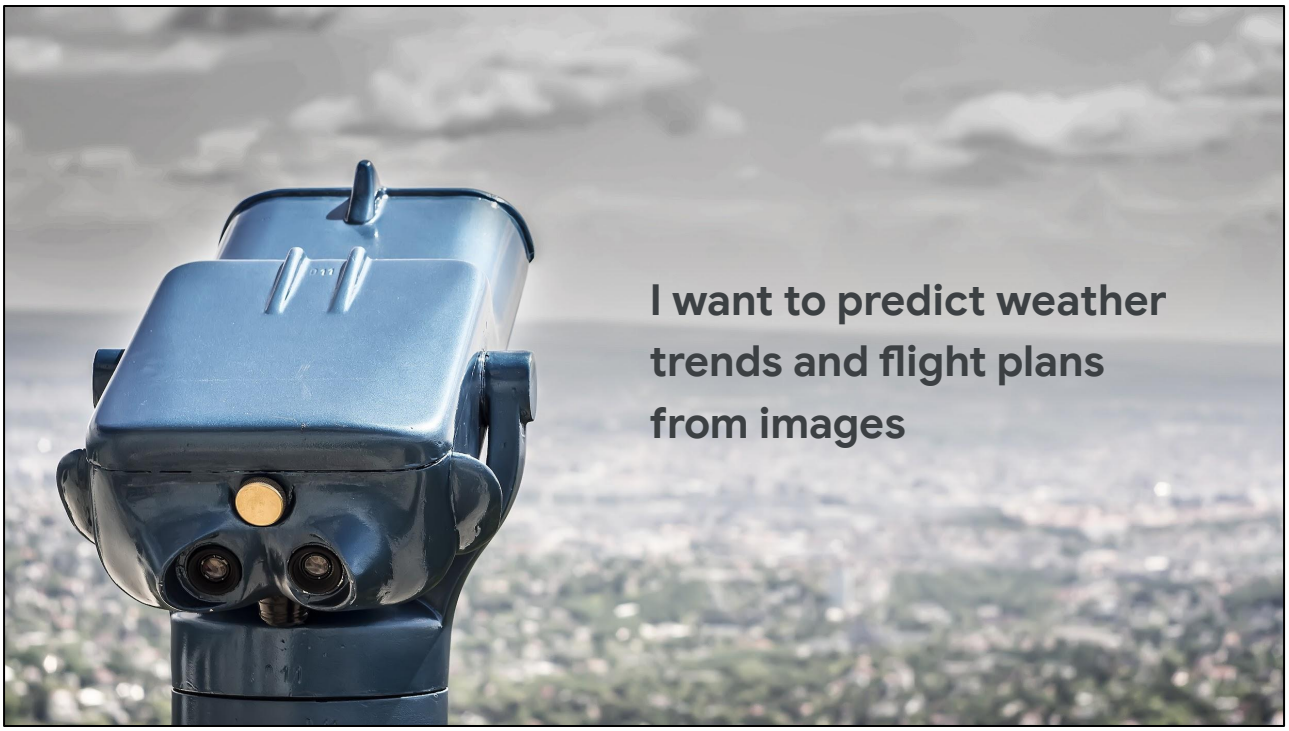
04 Using the Natural Language API to analyze the text

Customizing Pre-built Models with AutoML

We'll continue our tour of pre-built models with high performing custom ML models with one surprising twist -- you can build them with no code. Welcome to AutoML.



Let’s say I’m a meteorologist...



I want to predict weather trends and flight plans from images

There are 10+ different types of clouds

There are more than 10 different types of clouds

There are 10+ different types of clouds

And each type

There are 10+ different types of clouds

Could represent different patterns of weather

There are 10+ different types of clouds

That I may want to know ahead of time before it strikes

There are 10+ different types of clouds

If I’m a pilot or farmer or travel agency or even just someone who wants to know if it’s going to rain today or not.

They all indicate different weather patterns

If we want to predict on weather, this means we need to identify not just that there's a cloud, but also what *type* of cloud.

Yes, it's a cloud, but what type?

Cumulonimbus

versus

Cirrus

For example, a cirrus cloud is usually associated with fair weather, whereas a Cumulonimbus cloud usually foreshadows rain.

So let's revisit our Vision API and see how well it does on this new problem.

Let's try the Vision API

Here we show the image to the Vision API

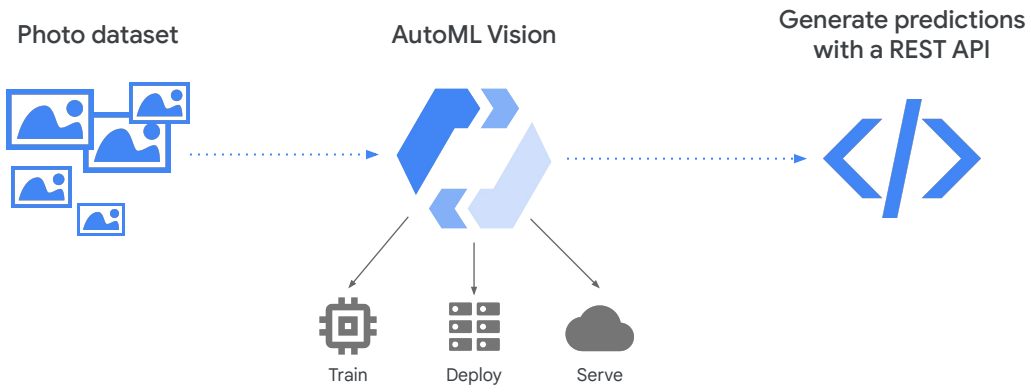
Let's try the Vision API

Sky	99%
Cloud	98%
Daytime	97%
Blue	97%
Atmosphere	93%

After I uploaded the image to the Vision API here are the results for the Labels it inferred.

The pre-trained model likely was never taught to recognize cloud types to this granularity. We need something a little more custom that we can train ourselves.

AutoML to the rescue



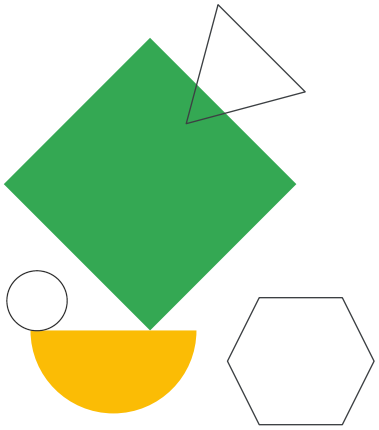
If you need to extend the capabilities of the AI building blocks, consider AutoML.

Here we will use our own labelled dataset of cloud images and train a custom model using AutoML Vision.

Demo

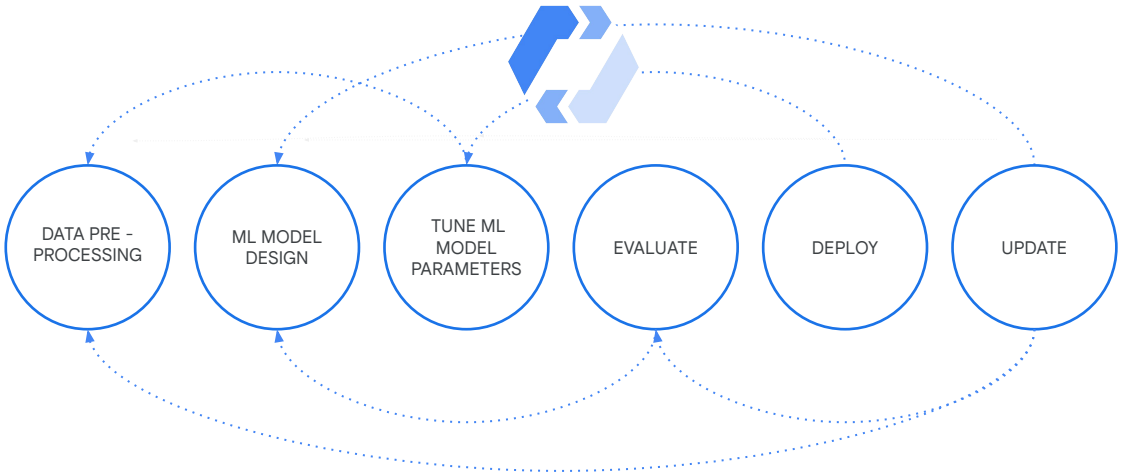
AutoML Vision

Codeless ML for unstructured datasets



Refer to https://github.com/GoogleCloudPlatform/training-data-analyst/blob/master/courses/data-to-insights/demos/auto-ml-vision.md

Codeless model building with AutoML



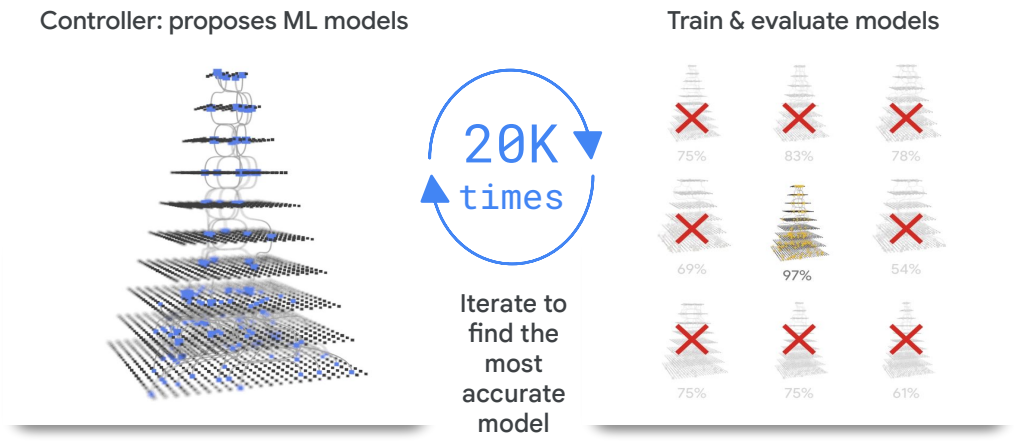
The big takeaway here is not what we did, but what we *didn't* do.

As we talked about at the beginning, typically when building a custom model, you have several steps which are complex and time intensive.

With AutoML, we don't have to do any of that!

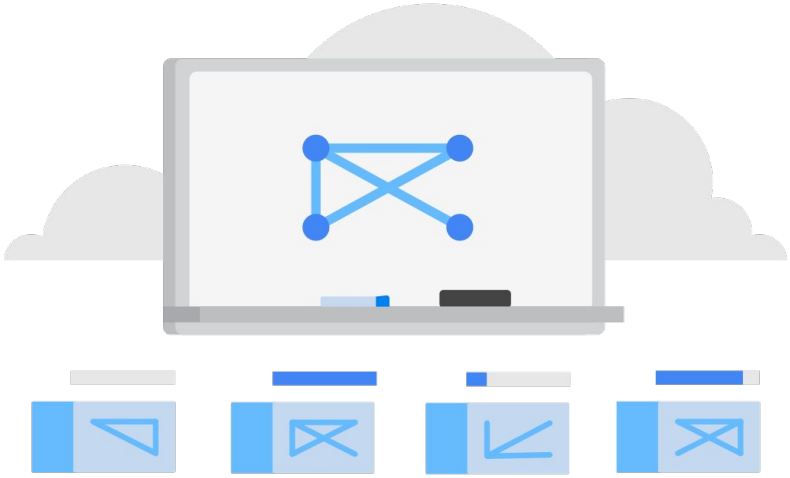
However, as great as AutoML is, it can't solve every ML problem which is why you learned how to build a CNN by hand earlier in the course.

AutoML is built with Neural Architecture Search



Behind-the-scenes, AutoML is powered by the latest ML research. While your model trains, the AutoML platform actually trains and evaluates multiple models and compares them against eachother. This NASNet approach or Neural Architecture Search produces an ensemble of ML models and chooses the best one.

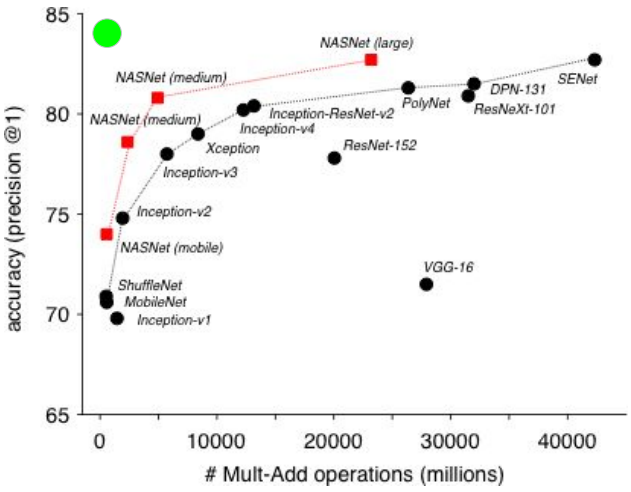
ML models creating ML models



Like Go and self driving cars… deep learning is now doing… deep learning as you saw with the neural architecture search. But how well does a codeless model you build with AutoML compare with some of the other image classification models you may have heard of?

Image: https://cloud.google.com/automl/

Inception network for image classification



This graph is a refresh of the best models for ImageNet published in 2017.

The X axis is model size.
The Y axis is accuracy.

The best possible model position would be the green dot. What you want is a small model (left on the X) with great accuracy (top of the Y).

Our world has been trending towards big, heavy, models that aren't exactly like a brain. That's the **black** line.
AutoML, powered by NASNet is shown in the red line.The AutoML networks are smaller and more efficient.

I'll provide a link to the Google AI blog so you can track the latest developments.

https://ai.googleblog.com/2017/11/automl-for-large-scale-image.html

AutoML Vision vs Vision API

Attribute	AutoML Vision	Vision API
Objective	Enabling **developers with no ML expertise** to build state of the art ML models for images	Enabling **ML practitioners** to harness power of Google's ML for images
Primary use case	Classification	Face detection, OCR, Object detection etc.
Data requirements	Images with labelled data	Just Images (may or may not required labelled data)
Output format	Labels with probability	As per the problem
Custom requirements	Can't be customized	Can be used for any custom made solutions
Efforts	Low for solution designing	High for end to end model development
Status	In public beta program	Publically available

Here's a final recap on when you should use the Vision API vs AutoML Vision.

Recall that with AutoML vision you are primarily doing classification on image data that you are providing to the service. We did this in our demo where we needed the model to learn a label, the type of cloud, that the Vision API didn't know. AutoML requires no coding experience while the Vision API requires you to be familiar with invoking APIs with a programming language of your choice.

Building a Custom Model

Earlier in the course you saw a few ways to create a custom ML model. We'll briefly review them here and point you to additional resources where you can practice building more yourself.

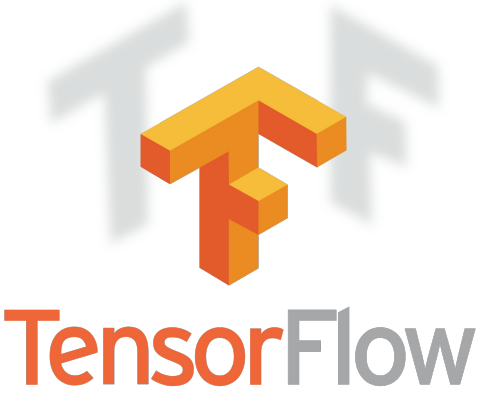
Train and run ML in the familiar BigQuery UI

One of the easiest ways to create a custom model on a structured dataset from scratch is to try out BigQuery ML. Earlier in the course you practiced:

- Creating a machine learning dataset and identifying features and labels
- Choosing the right model type for your dataset and what you are trying to predict or infer
- Providing any custom model options
- Training the model and evaluating its performance
- Inspecting what the model learned about the weight of each feature
- And predicting on unknown future data

I'll provide additional resources and links for you to practice and learn more about BigQuery ML.

Create custom ML models with TensorFlow



Lastly we mentioned that ML engineers often create their own models using open source libraries like TensorFlow running on Google Cloud. The value of these models can be huge if you build and train them correctly or minimal if they are not done well. If you're looking for experience building TensorFlow models on Google Cloud, checkout our ML on Google Cloud specialization in our additional resources section of the course.

Lab: Classify images with ML two-ways using pre-built models

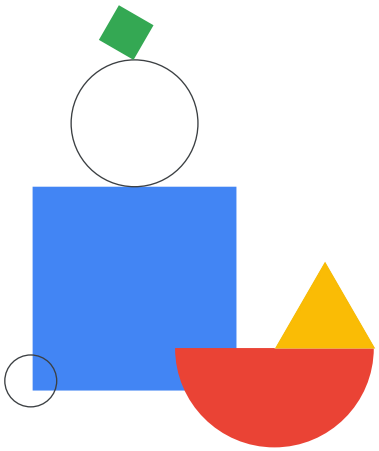
How can I leverage pre-trained models for image classification? What about my own image datasets?

ML APIs

AutoML

Lab Intro

Training with Pre-built ML Models using the Vision API and AutoML



Lab objectives

01 Setup API key for ML Vision API

02 Invoke the pretrained ML Vision API to classify images

03 Review label predictions from Vision API

04 Train and evaluate custom AutoML Vision image classification model

05 Predict with AutoML on new image