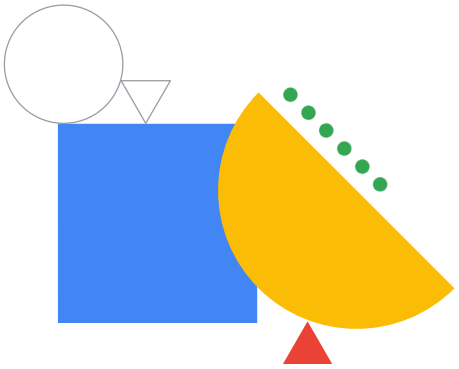
Google Cloud

Analytics and AI: Introduction



This is the final course in the Data Engineering on Google Cloud offering. While data has intrinsic value, that value can be augmented with Machine Learning. This course covers how to integrate Analytics and Machine Learning capabilities into your data pipelines.

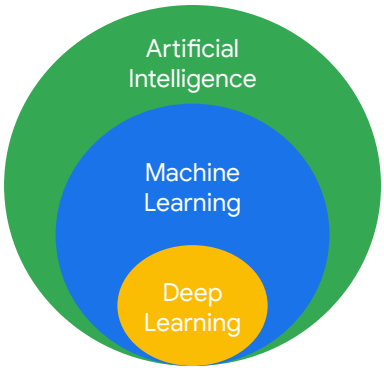
Agenda

What is AI?

From Ad-hoc Data Analysis to Data-Driven Decisions

Options for ML models on Google Cloud

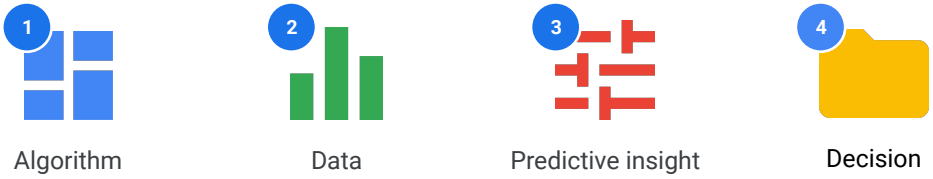
Machine Learning is a type of AI, and deep learning is a type of machine learning



Class of problems we can solve when **computers think/act like humans**

A very common question asked is: What's the difference between AI, Artificial Intelligence, Machine Learning and Deep Learning? Well, one way to think about it is AI is the discipline like something like physics. AI refers to machines that are capable of acting autonomously, machines that think. AI has to do with the theory and methods to build machines that can solve problems by thinking and acting like humans.

ML is a way to use standard algorithms to derive predictive insights from data and make repeated decisions



So you've heard a lot about Machine Learning or ML, let's start with a definition. What is ML? ML is a way to use standard algorithms to derive predictive insights from data and make repeated decisions.

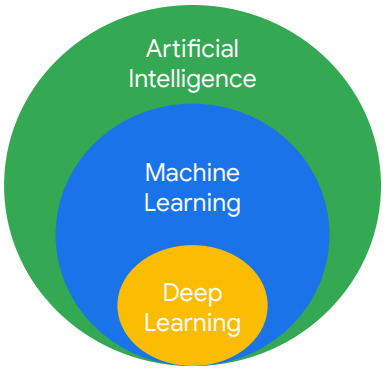
You do this using algorithms that are relatively general and applicable to a wide variety of datasets.

Think of a typical company and how they use their data today. Perhaps they have a dashboard that business analysts and decision-makers view on a daily basis, or report that's read on a monthly basis. This is an example of a backward-looking use of data, looking at historical data to create reports and dashboards. This is what people tend to mean when they talk about BI or Business Intelligence. A lot of data analytics is backward-looking, nothing wrong with that. But instead we're going to use ML or Machine Learning to generate forward-looking or predictive insights.

Of course, the point of looking at historical data might be to make those decisions. Perhaps business analysts examined the data and they suggest new policies or rules. They could suggest for example that it’s possible to raise the price of a product in a certain region. Now, that business analyst is making a predictive insight but is that scalable? Can the business analyst make such a decision for every single product in every single region? And can they dynamically adjust the price every second? Now, here's where the computers get involved. In order to make decisions around predictive insights repeatable, you need ML. You need a computer program to derive those insights for you. So ML is about making predictive insights from data, many of

them at a time. It's about scaling up BI and decision-making.

Why are Machine Learning and Deep Learning so exciting?



Class of problems we can solve when **computers think/act like humans**

Scalably solve those problems using **data** examples (**not custom code**)

Even when that data consists of **unstructured data** like images, speech, video, natural language text, etc.

Machine Learning is a tool set like Newton's Laws of Mechanics. Just as you can use Newton's laws to figure out how long it'll take a ball to drop and when it falls off a cliff, you can use Machine Learning to scalably solve certain kinds of problems using data examples. But without the need for any custom code.

Deep Learning is a type of Machine Learning that works even when the data consists of unstructured data, like images, speech, video, natural language text and so on. One kind of deep learning is image classification. A machine can learn how to classify images into categories when it's shown lots of different examples. And the really cool thing about deep learning is that oftentimes in a really complex problems it can do better than at it. The basic difference between Machine Learning and other techniques in AI is that in Machine Learning machines learn. They don't start out intelligent, they become intelligent.

Keller Williams uses AutoML Vision to automatically recognize common elements of house furnishings and architecture

AutoML Vision

Google Cloud

Keller Williams, a U.S. real estate company, uses AutoML Vision to automatically recognize specific features of houses like built-in bookcases.

This helps agents get houses listed faster and buyers find houses that meet their needs.

Neil Dholakia, Chief Product Officer says "By training a custom model to recognize common elements of furnishings and architecture, customers can automatically search home listing photos for specific features like granite countertops like 'modern.'"

This application of machine learning quickly allows Keller Williams realtors to record a video walkthrough of a new home and use the object detection capabilities of AutoML Vision to find and tag key aspects of the home that customers would want to search on.

A big benefit for their organization is that they already had many existing images and videos of home walkthroughs already. They simply fed them into the pre-built AutoML Vision model and customized it. All without writing a line of code. You'll learn more about AutoML Vision and practice creating models with it later in this course.

https://cloud.google.com/blog/products/gcp/empowering-businesses-and-developers-do-more-ai

Kewpie uses ML to sort out the bad potatoes in baby food

kewpie

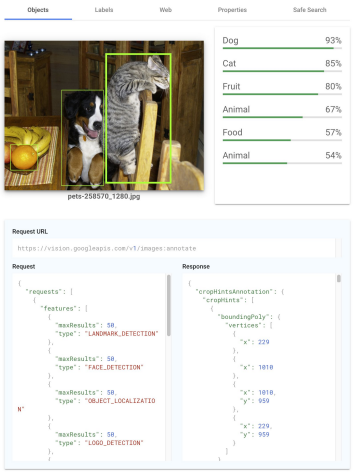
Original process required humans to identify low-quality ingredients, which was expensive and stressful.

Machine learning was used to replicate the quality control process.

Kewpie manufactures baby food. In this case, quality is not necessarily a matter of safety—because the food itself is safe—but discoloration can concern parents. So Kewpie turned to Google and our partner Brainpad to build a solution that leverages image recognition to detect low-quality potato cubes. The ML algorithm enabled them to free people from the tiring work of inspection and focus on other important work.

https://www.blog.google/products/google-cloud/how-ai-can-help-make-safer-baby-food-and-other-products/

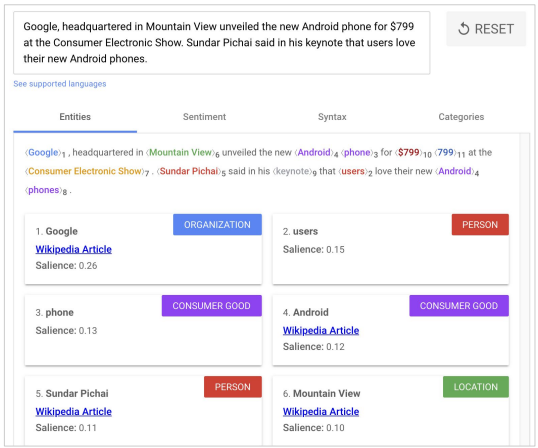
Play around with the power of AI yourself...



cloud.google.com/vision/

- Object detection
- Labeling and confidence
- Web lookup
- Pre-trained (call the API)

Try Google's natural language API



cloud.google.com/natural-language/

- Entity extraction
- Sentiment analysis
- Sentence structure
- Pre-trained (call the API)

Agenda

What is AI?

From Ad-hoc Data Analysis to Data-Driven Decisions

Options for ML models on Google Cloud

Imagine you're the owner of a bicycle rental business (in London). How do you stock enough bicycles?

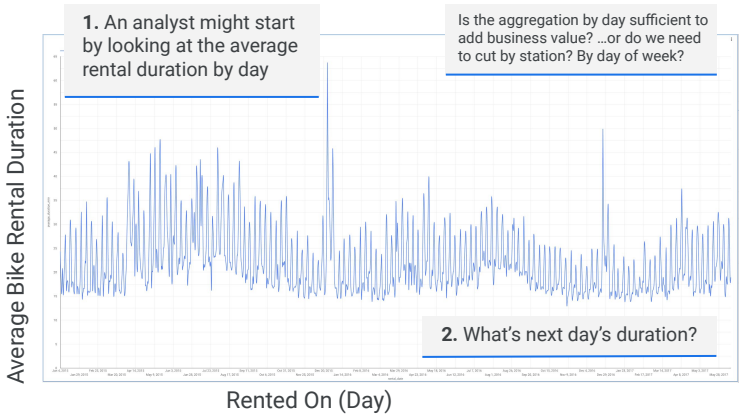
Commuter Bikes

If rental is likely to be for a **short duration**, we need to have commuter bikes in stock

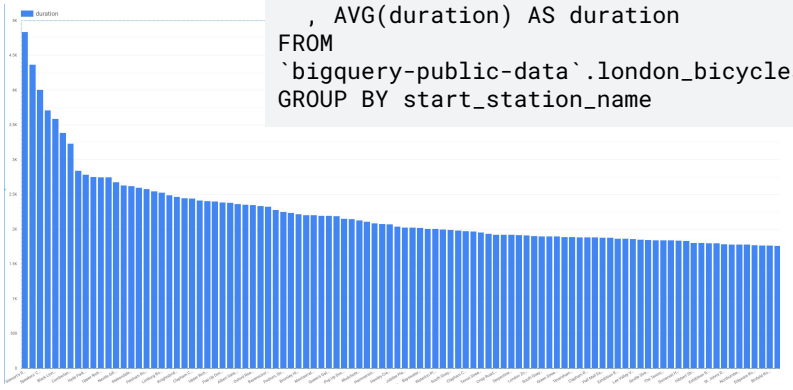
Road Bikes

If rental is likely to be for a **long duration**, we need to have road bikes in stock

You hire a data analyst to help get you insights on how to keep the right bicycles in stock

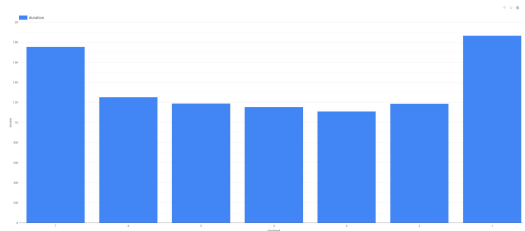


Does the duration of a rental vary by station?

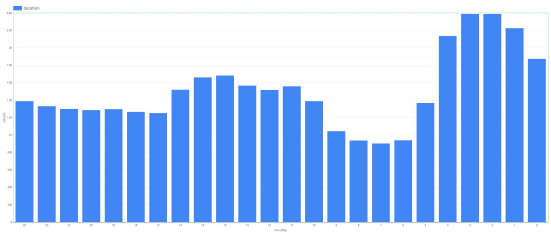



```
SELECT
  start_station_name
  , AVG(duration) AS duration
FROM
`bigquery-public-data`.london_bicycles.cycle_hire
GROUP BY start_station_name
```

How about the day of the week? Hour of day?



```
SELECT
  EXTRACT(dayofweek
  FROM
    start_date) AS dayofweek,
  AVG(duration) AS duration
FROM
 `bigquery-public-data`.london_bicycles.cycle_hire
GROUP BY
  dayofweek
```



```
SELECT
  EXTRACT(hour
  FROM
    start_date) AS hourofday,
  AVG(duration) AS duration
FROM
 `bigquery-public-data`.london_bicycles.cycle_hire
GROUP BY
  hourofday
```

This ad-hoc analysis is great but...

- A lot of manual, repetitive work involved for the data analyst
- Any decisions made will be based on hunches on how all these factors interact
- Wouldn't it be better if we could automate this analysis?

… what we need is an ML model to be able to make predictions

Goal: Augment our dashboards with predicted values e.g. prediction for the duration of a rental

As an example, Google augments Google Cloud cost dashboards (descriptive) with forecasted (predictive) usage costs.

https://cloudplatform.googleblog.com/2018/07/predict-your-future-costs-with-google-cloud-billing-cost-forecast.html

Use the ML model to anticipate what type of bike/how many to stock at your locations

- The ML model takes some of the drudgery out of ad-hoc analysis to help you make truer data-driven decisions.
- Can build a ML model in BigQuery or Vertex AI.

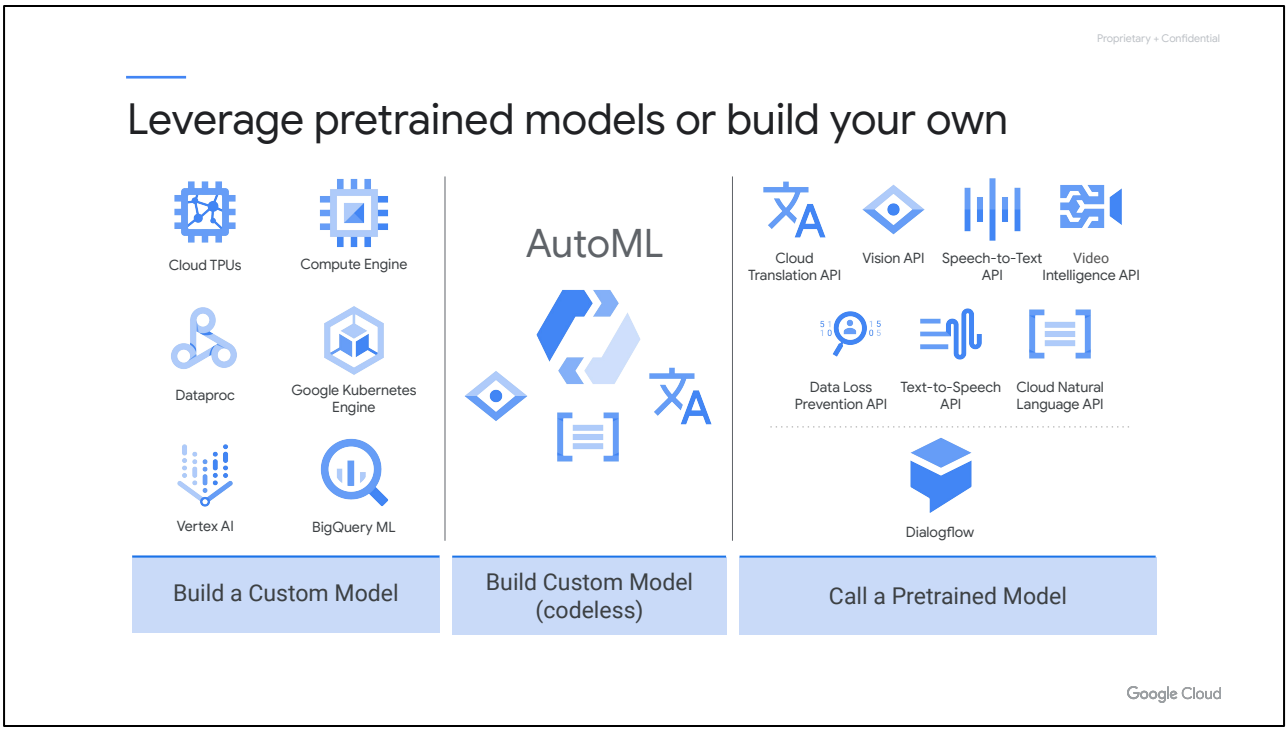
```
CREATE OR REPLACE MODEL
  bike_model.model_bucketized TRANSFORM(* EXCEPT(start_date),
  IF
    (EXTRACT(dayofweek
      FROM
        start_date) BETWEEN 2 AND 6,
      'weekday',
      'weekend') AS dayofweek,
    ML.BUCKETIZE(EXTRACT(HOUR
      FROM
        start_date),
      [5, 10, 17]) AS hourofday )
OPTIONS
  (input_label_cols=['duration'],
    model_type='linear_reg') AS
SELECT
  duration,
  start_station_name,
  start_date
FROM
  `bigquery-public-data`.london_bicycles.cycle_hire
```

Agenda

What is AI?

From Ad-hoc Data Analysis to Data-Driven Decisions

Options for ML models on Google Cloud



Different options exist when it comes to leveraging machine learning.

Advanced users who want more control over the building and training of their ML models will use tools that offer the levels of flexibility that they're looking for. This can involve developing custom models through an ML library, like TensorFlow, that's supported on Vertex AI. This option works well for data scientists with the skills and the need to create a custom TensorFlow model.

But increasingly, you don't have to do that. Google makes the power of machine learning available to you even if you have a limited knowledge machine learning. You can use AutoML to build on Google's machine learning capabilities to create your own custom machine learning models that are tailored to your specific business needs. And then, integrate those models into applications and websites. All without running a line of TensorFlow code.

Alternatively, Google has a range of pre-trained, meaning you don't need to bring your own data, machine learning models that are ready for immediate use within applications in ways that the respective APIs are designed to support. Such pre-trained models are excellent ways to replace user input with machine learning.

Module summary

- AI's impact on industry is huge.
- Predictive modeling takes data-driven decision making to a new level.
- The typical data science workflow.