101

Machine Learning

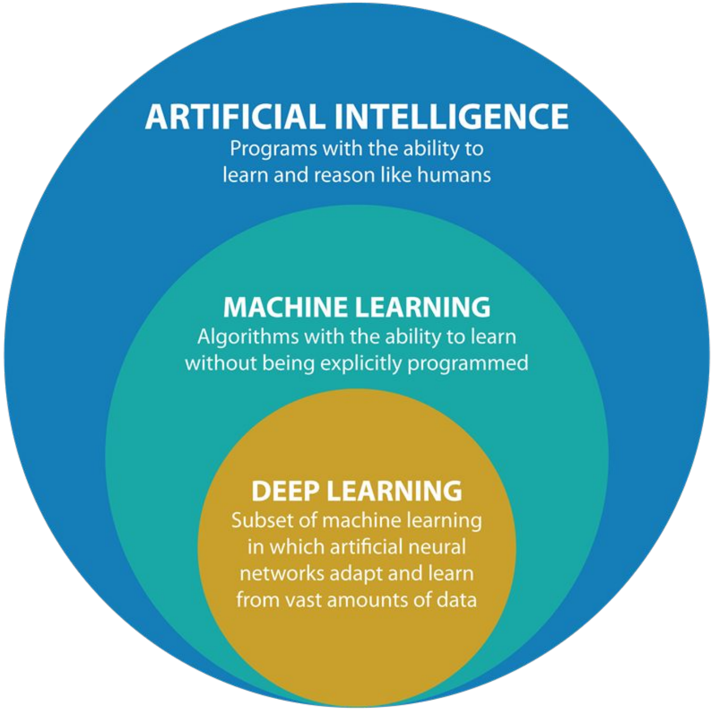
Meet your business challenges head on with
cloud computing services from Google

Agenda

- Machine Learning Overview
- Use-case of Machine Learning
- How to create Machine Learning
- Machine Learning and Data Scientist
- AI Services in Google Cloud Platform
- Use-case of AI Service in Google Cloud Platform
- Demonstration

Machine Learning Overview

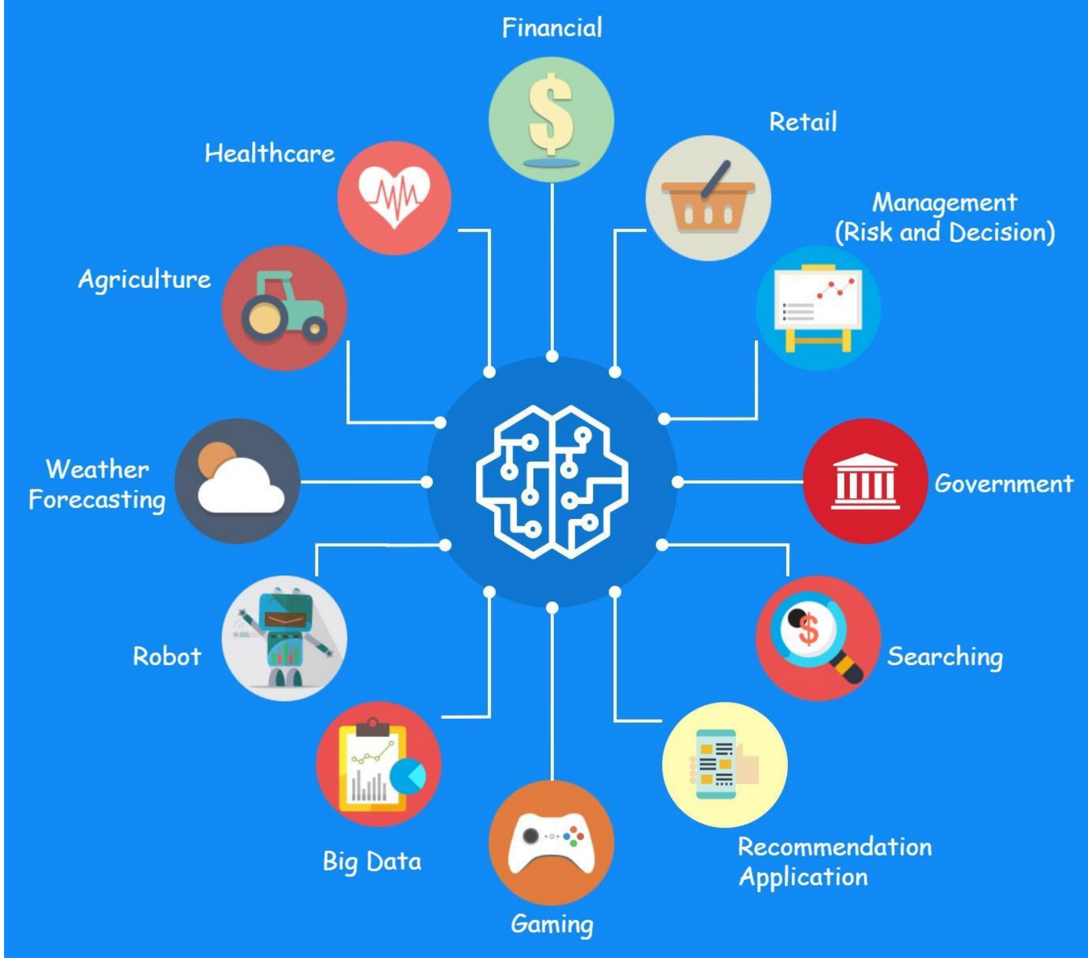
Machine Learning Overview



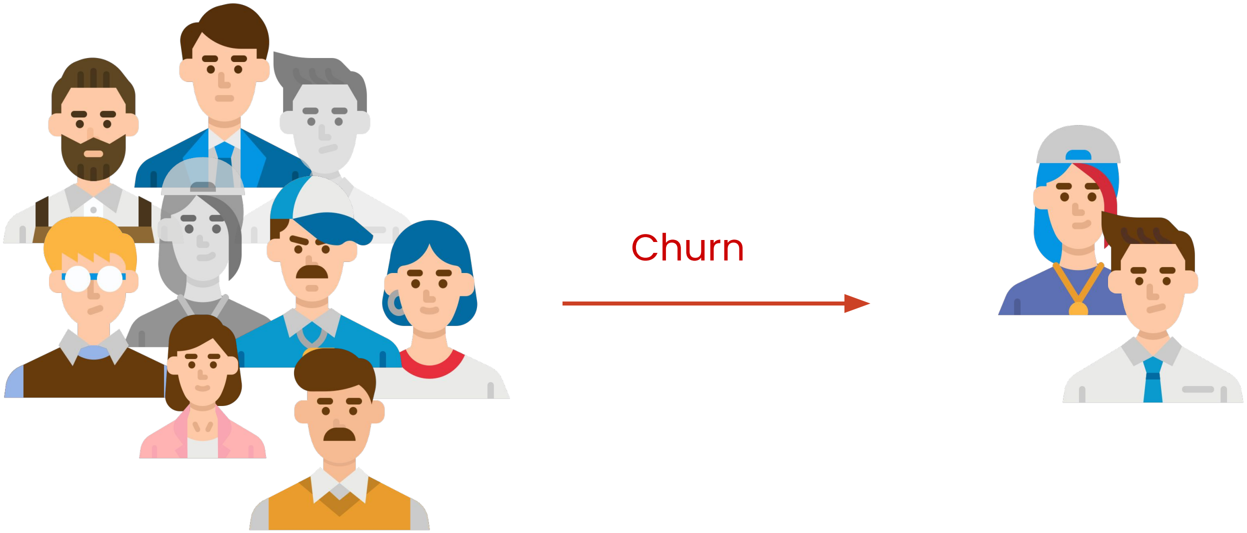
*" **Machine learning (ML)** is a subfield of artificial intelligence (AI). The goal of ML is to make computers learn from the data that you give them. "*

- Google Cloud -
(AI Platform Service)

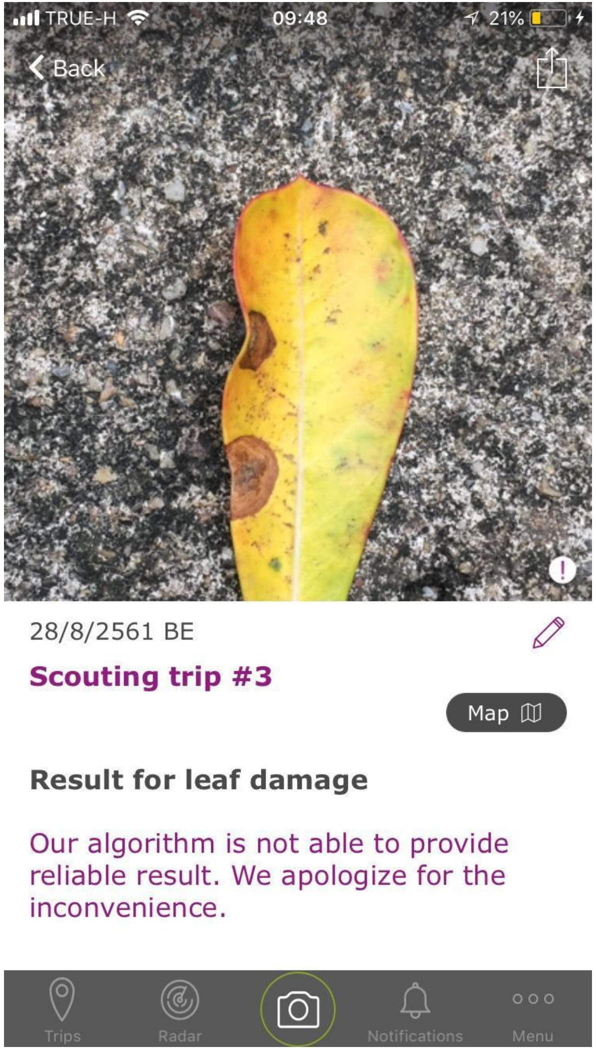
Industries applied
Machine Learning



Use case of
Machine Learning


Churn


9:41 AM
100%
Back
Selected disease result
Brown Rust
Puccinia Triticina
91%
Trips
Radar
Notifications
Menu


TRUE-H
09:48
21%
Back
28/8/2561 BE
Scouting trip #3
Map
Result for leaf damage
Our algorithm is not able to provide reliable result. We apologize for the inconvenience.
Trips
Radar
Notifications
Menu

10^{170}

Possible positions

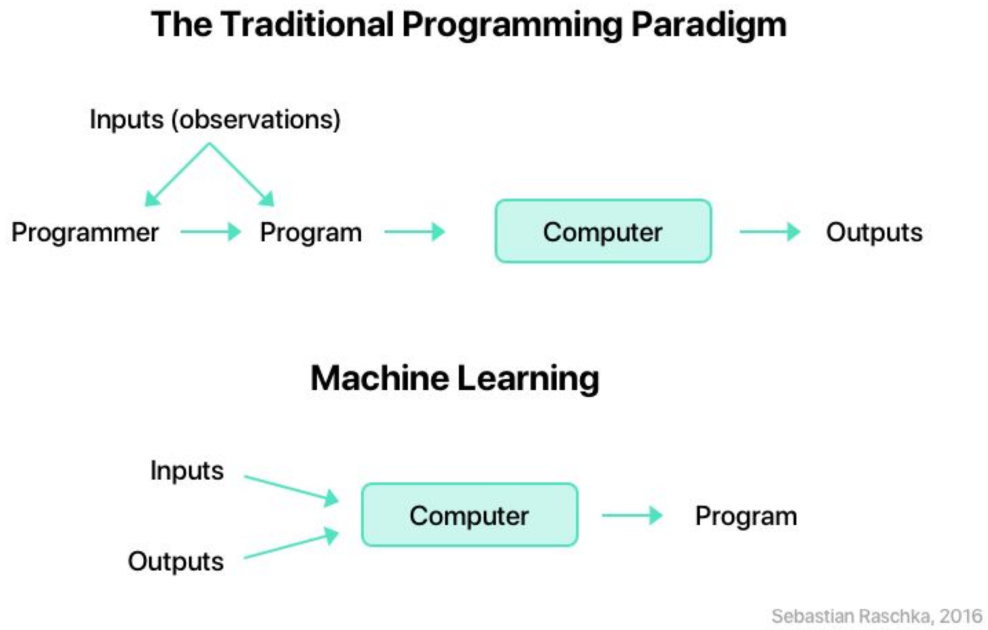
30M

Trained games

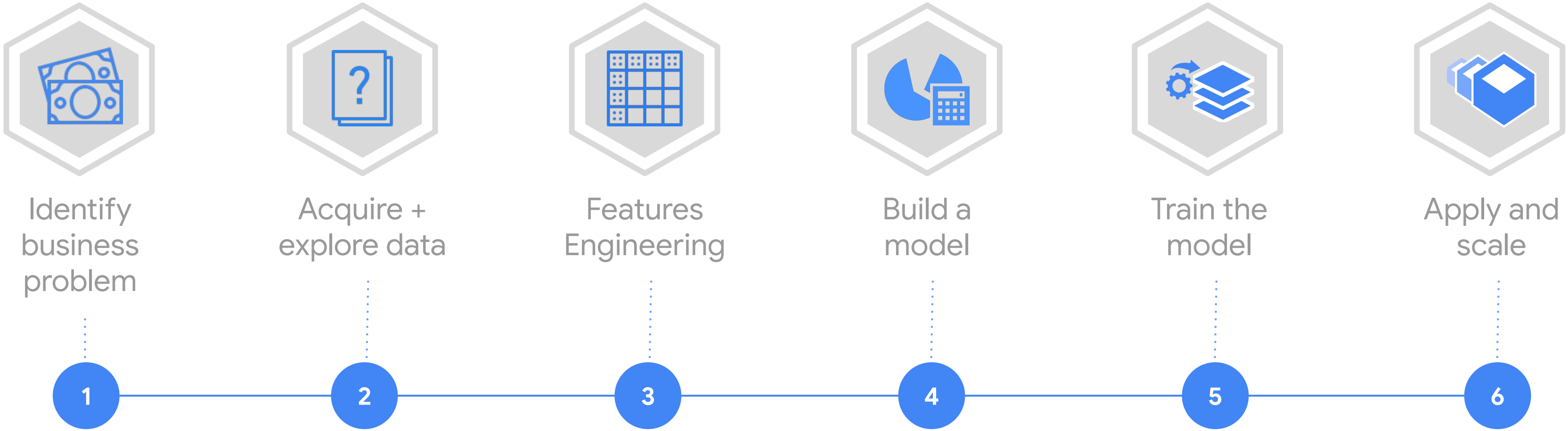

Lane Status
Direction: Left curve
Curvature 707.1 m
Off center: Right 0.5m
Detected viehicles
car

How to create Machine Learning

Machine Learning Approached



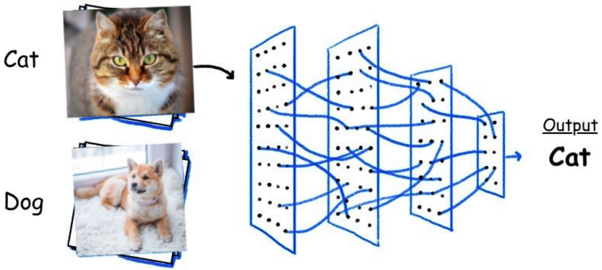
To build a Machine Learning Model



a. Identify business problem

Regression

Classification



b. Data Collection / Data Acquisition

Google

kaggle

IMAGENET

NYC OpenData

c. Feature Engineering

	A	B	C	D	E	F	G	H	I	J
3	Age	Party	Gender	Income		Age	Party 1	Party 2	Gender 1	Income
4	20	Rep	Male	45000		20	1	0	1	45000
5	25	Dem	Male	39000		25	0	1	1	39000
6	45	Ind	Male	56000		45	0	0	1	56000
7	35	Rep	Female	49000		35	1	0	0	49000
8	50	Dem	Female	41000		50	0	1	0	41000
9	55	Ind	Female	42000		55	0	0	0	42000
10	39	Rep	Male	58000		39	1	0	1	58000
11	48	Dem	Male	55000		48	0	1	1	55000
12	30	Ind	Male	46000		30	0	0	1	46000
13	27	Rep	Female	42000		27	1	0	0	42000
14	47	Dem	Female	37000		47	0	1	0	37000
15	21	Ind	Female	25000		21	0	0	0	25000
16	48	Rep	Male	75000		48	1	0	1	75000
17	24	Ind	Male	43000		24	0	0	1	43000
18	28	Ind	Female	40000		28	0	0	0	40000
19	40	Dem	Female	31000		40	0	1	0	31000

c. Feature Engineering (Cont.)

	Country	Age	Salary	Purchased
1	France	44	72000	No
2	Spain	27	48000	Yes
3	Germany	30	54000	No
4	Spain	38	61000	No
5	Germany	40		Yes
6	France	35	58000	Yes
7	Spain		52000	No
8	France	48	79000	Yes
9	Germany	50	83000	No
10	France	37	67000	Yes

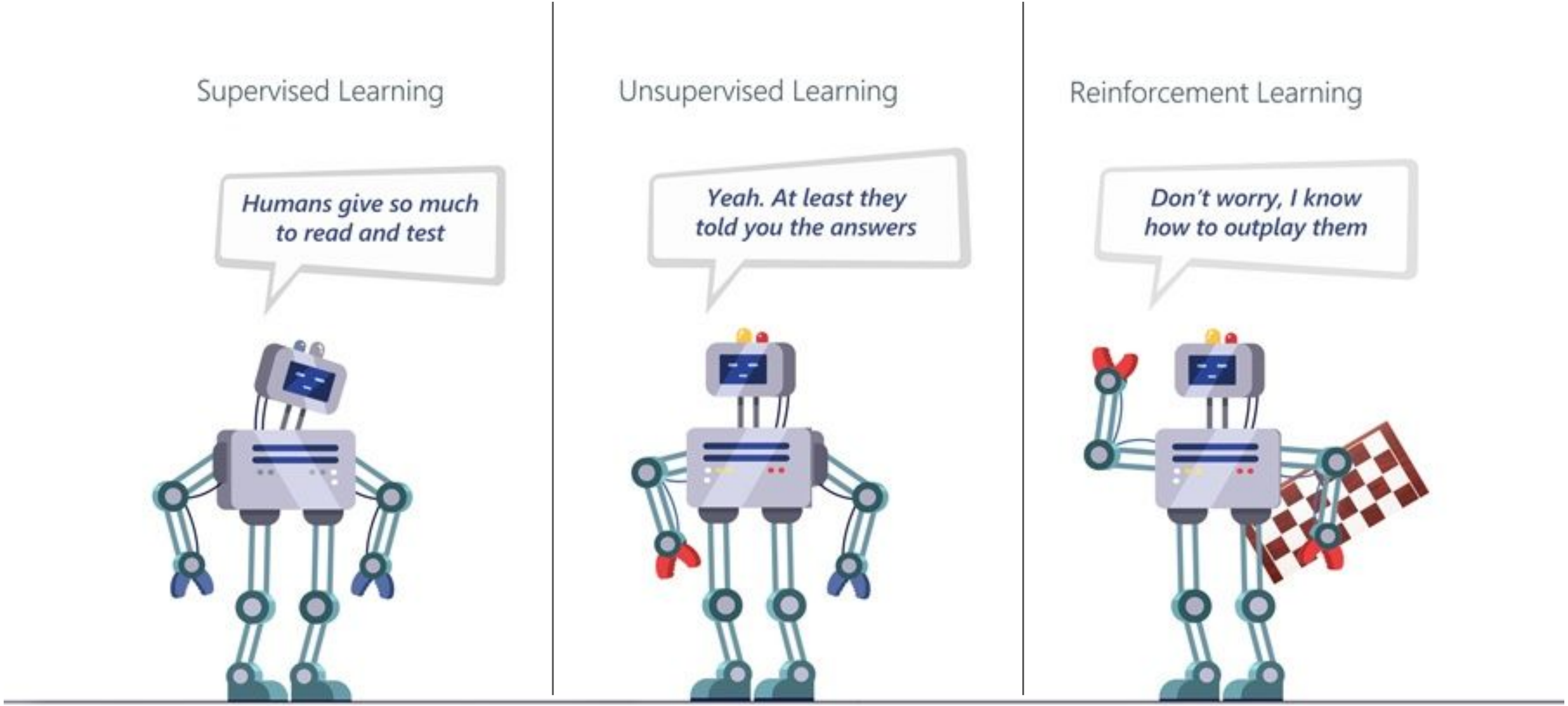
Standardisation

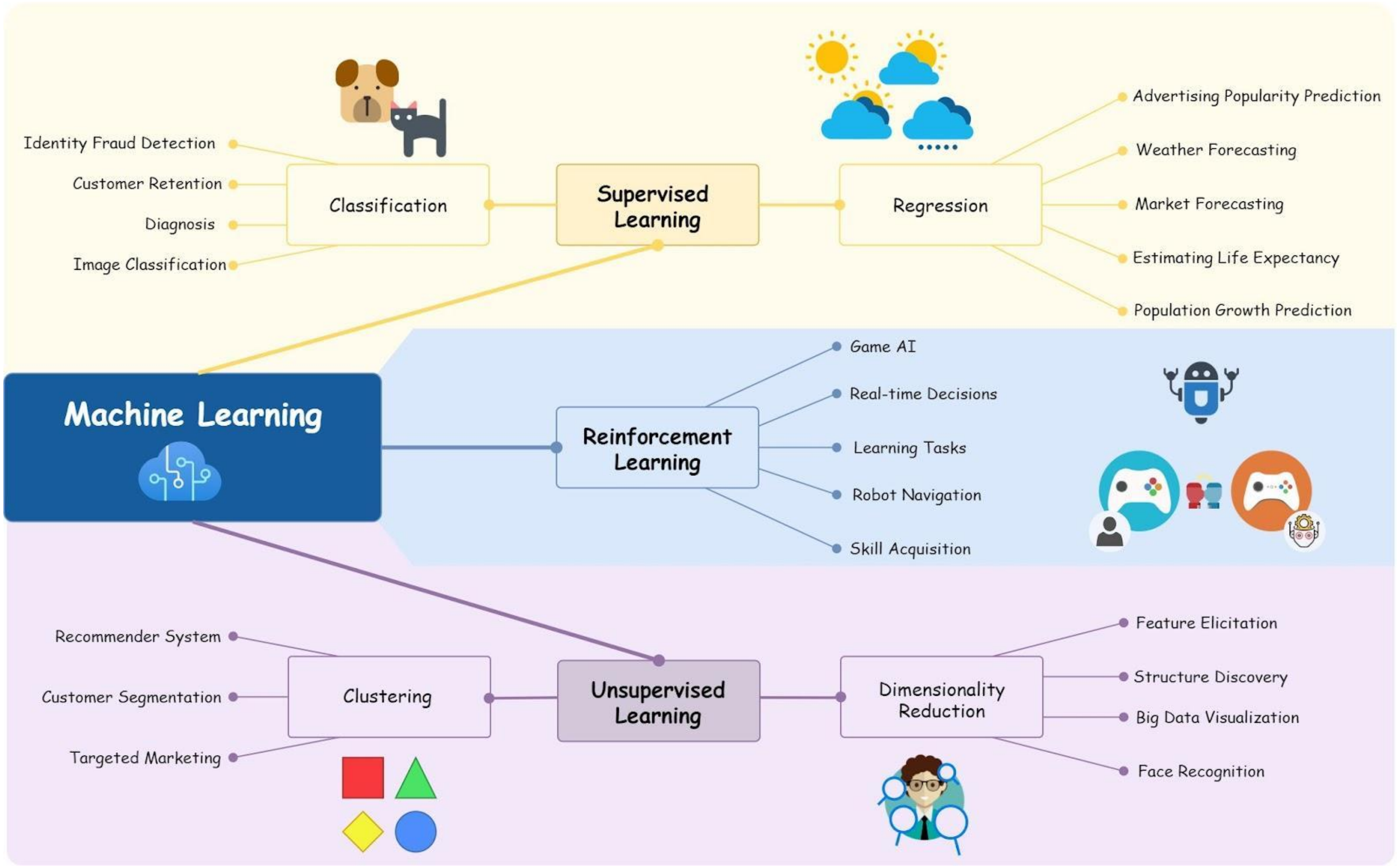
	Age	Salary
0	0.758874	7.494733e-01
1	-1.711504	-1.438178e+00
2	-1.275555	-8.912655e-01
3	-0.113024	-2.532004e-01
4	0.177609	6.632192e-16
5	-0.548973	-5.266569e-01
6	0.000000	-1.073570e+00
7	1.340140	1.387538e+00
8	1.630773	1.752147e+00
9	-0.258340	2.937125e-01

Max-Min Normalization

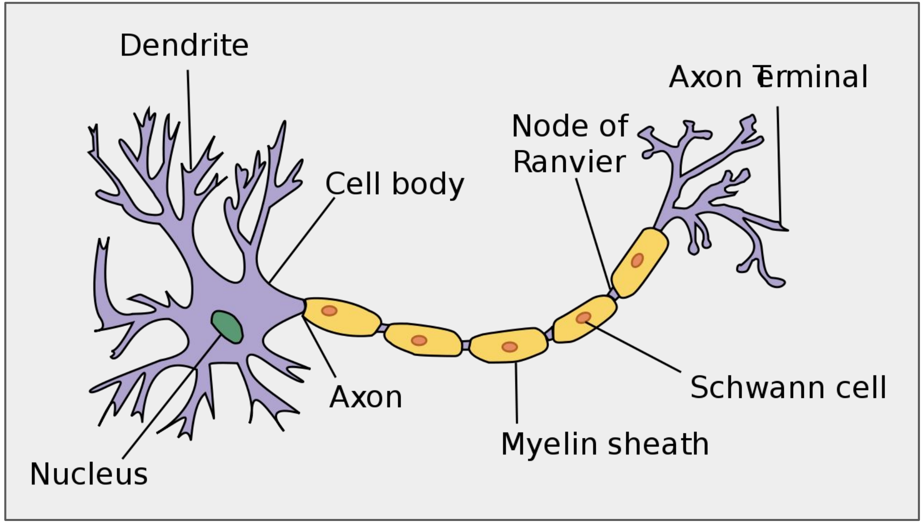
	Age	Salary
0	0.739130	0.685714
1	0.000000	0.000000
2	0.130435	0.171429
3	0.478261	0.371429
4	0.565217	0.450794
5	0.347826	0.285714
6	0.512077	0.114286
7	0.913043	0.885714
8	1.000000	1.000000
9	0.434783	0.542857

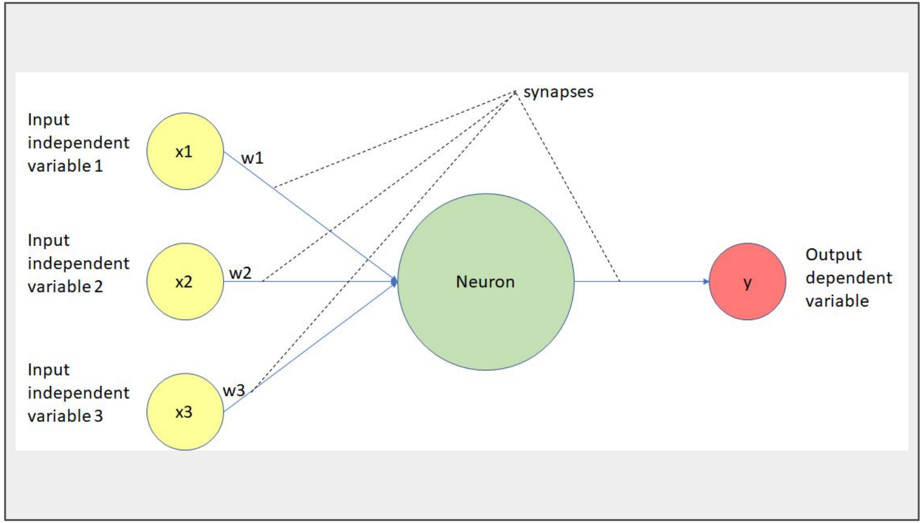
d. Build a Model : Machine Learning Type




Machine Learning
Supervised Learning
Classification
Identity Fraud Detection
Customer Retention
Diagnosis
Image Classification
Regression
Advertising Popularity Prediction
Weather Forecasting
Market Forecasting
Estimating Life Expectancy
Population Growth Prediction
Reinforcement Learning
Game AI
Real-time Decisions
Learning Tasks
Robot Navigation
Skill Acquisition
Unsupervised Learning
Clustering
Recommender System
Customer Segmentation
Targeted Marketing
Dimensionality Reduction
Feature Elicitation
Structure Discovery
Big Data Visualization
Face Recognition

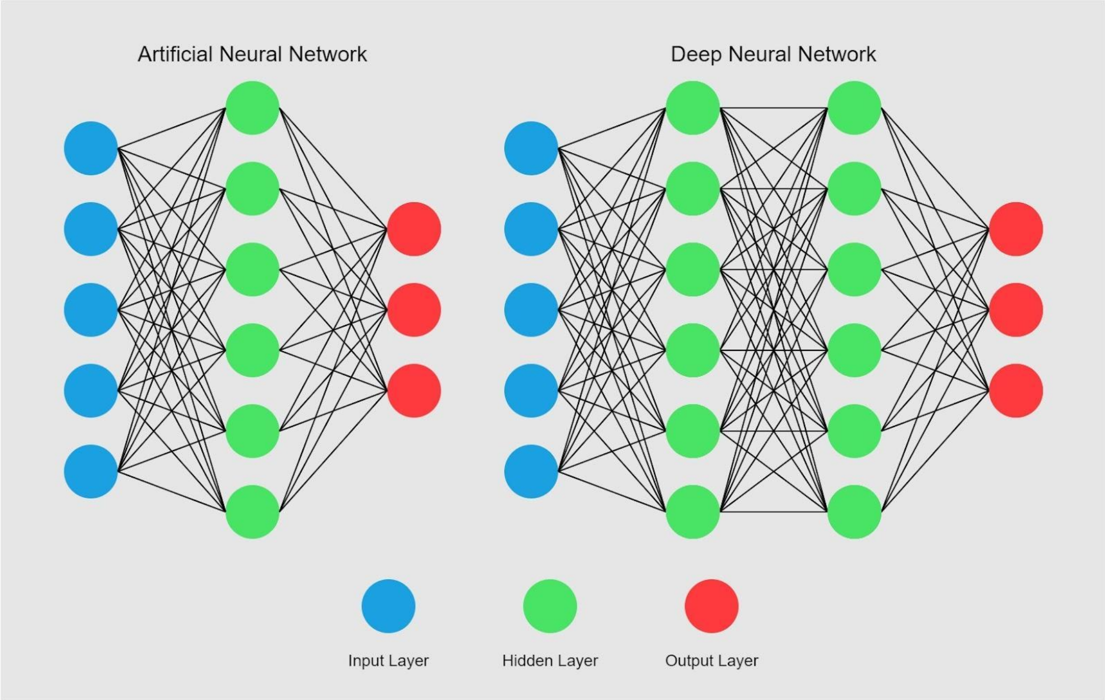
d. Build a Model : Neural Network Concept





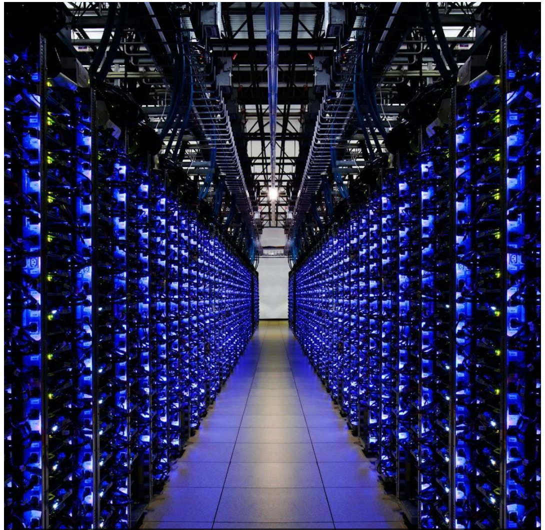
Neuron Network is the network formed by simulating the process of human brain. When human learns something, it will be sent to different parts of the brain to process and generate the result.

d. Build a Model : Neural Network Concept (cont.)

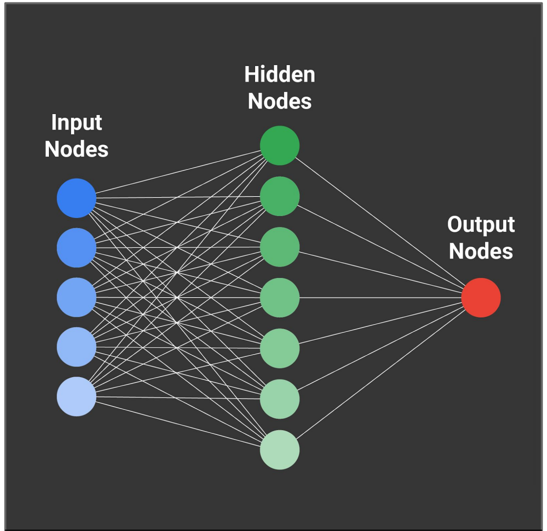


Neural Network is one of Algorithm of Machine Learning which has complex structure

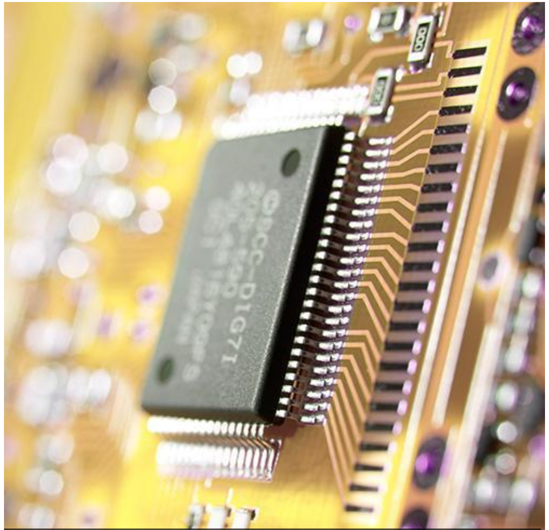
e. Train the Model : Key to Successful ML



Large Datasets

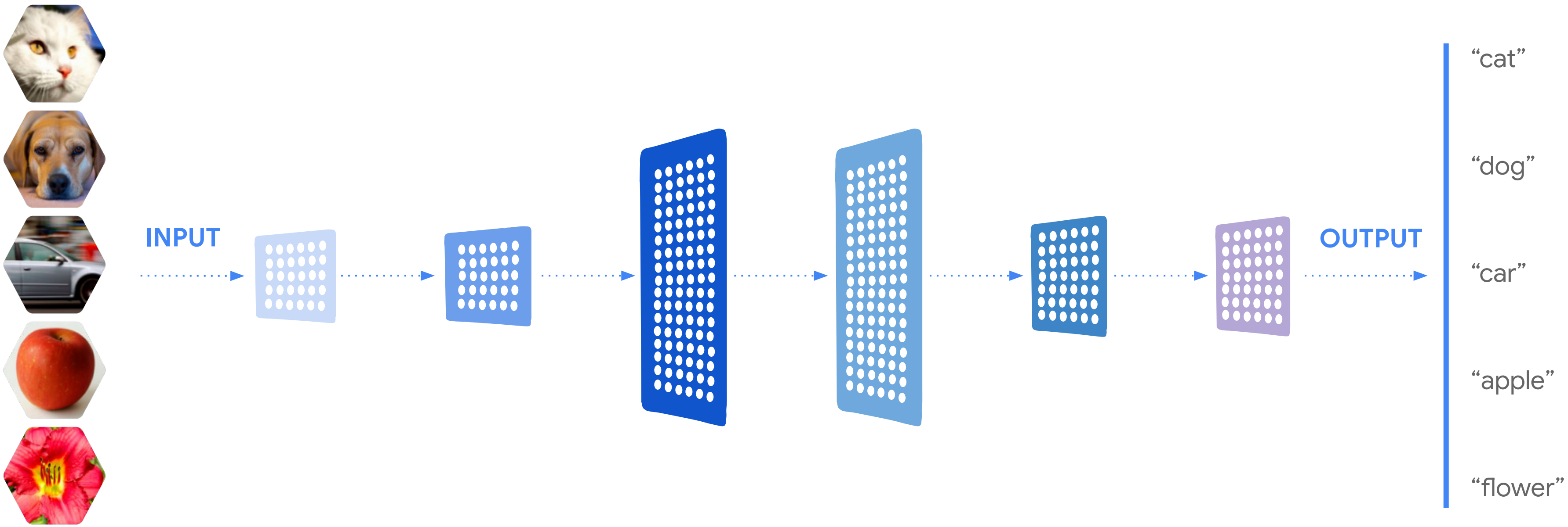


Good ML Models



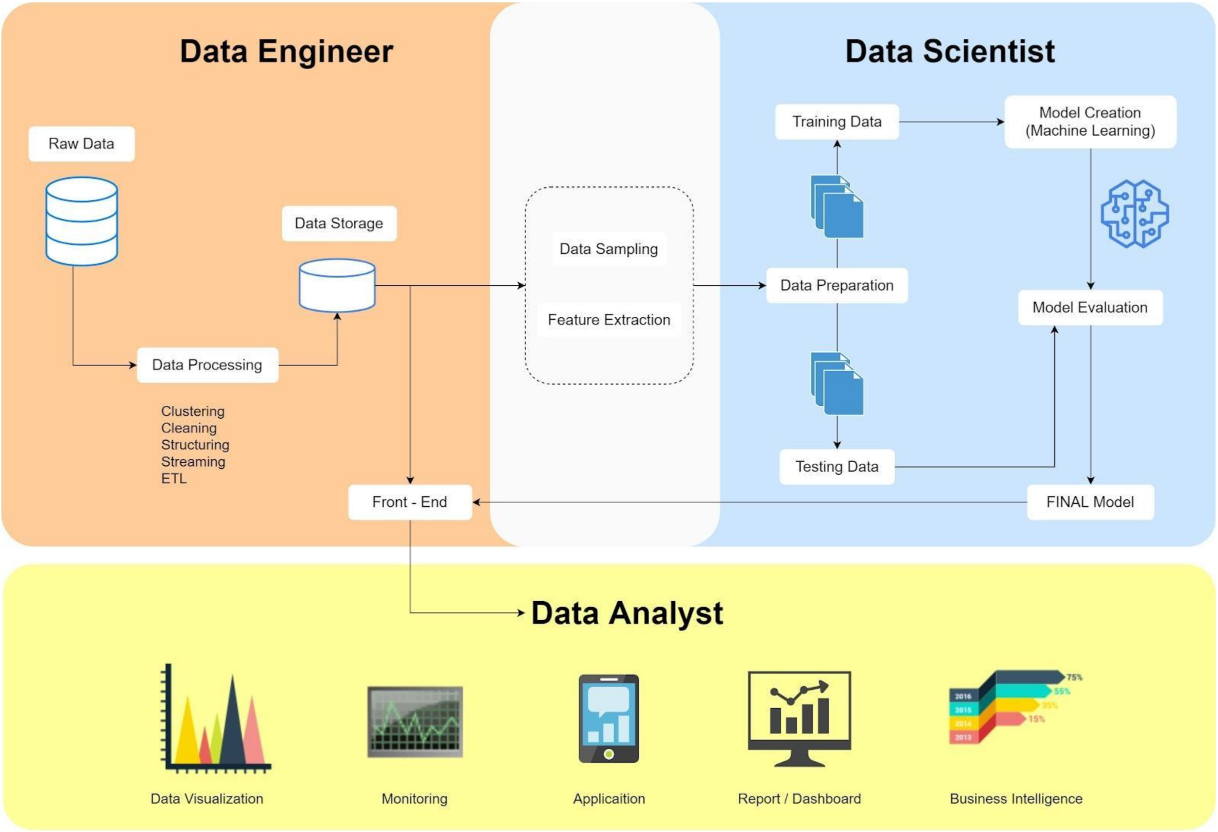
Lots of Computation

f. Apply and scale

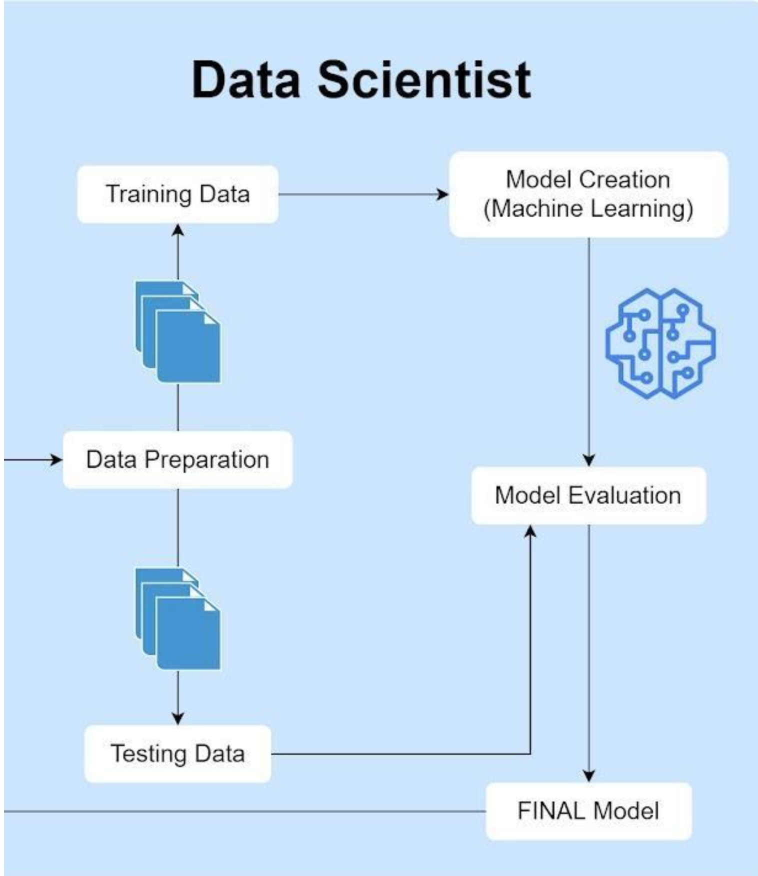


Machine Learning and Data Scientist

Machine Learning and Data Scientist



Machine Learning and Data Scientist



AI Services in Google Cloud Platform

AI Service in Google Cloud

“**End-to-end platform** for data science and machine learning ”

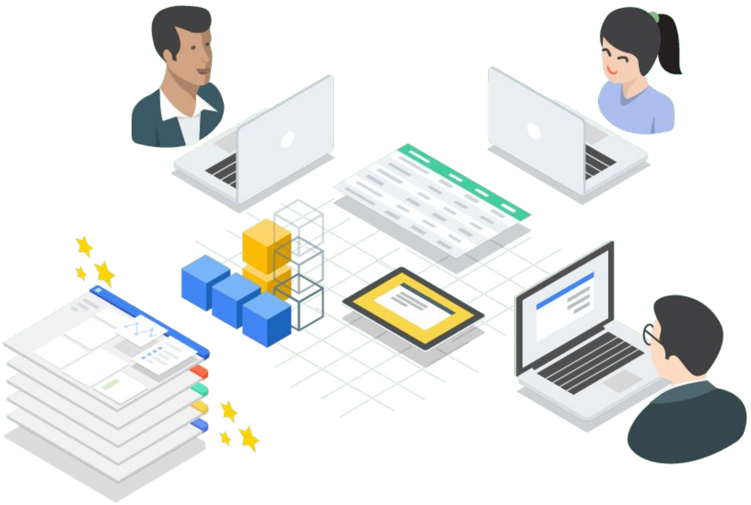
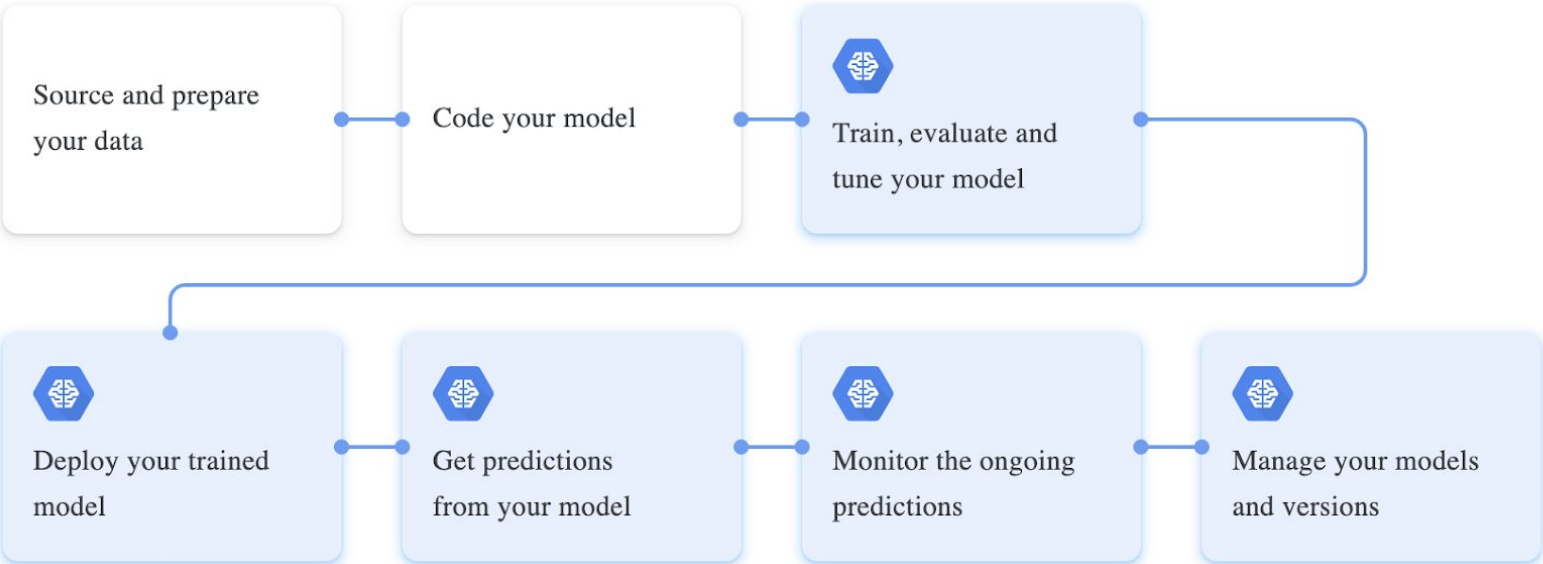


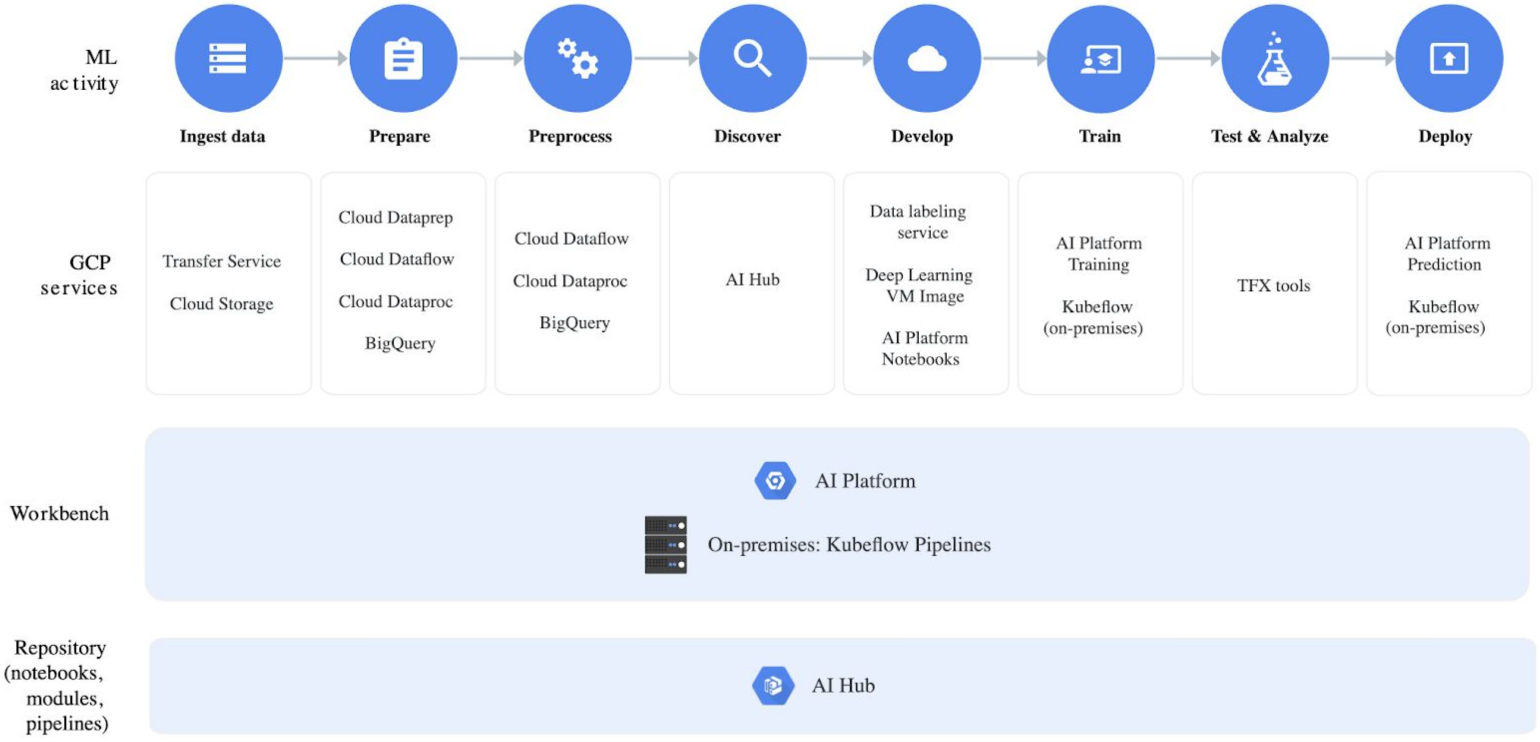
Image from Google

End-to-End in AI Service

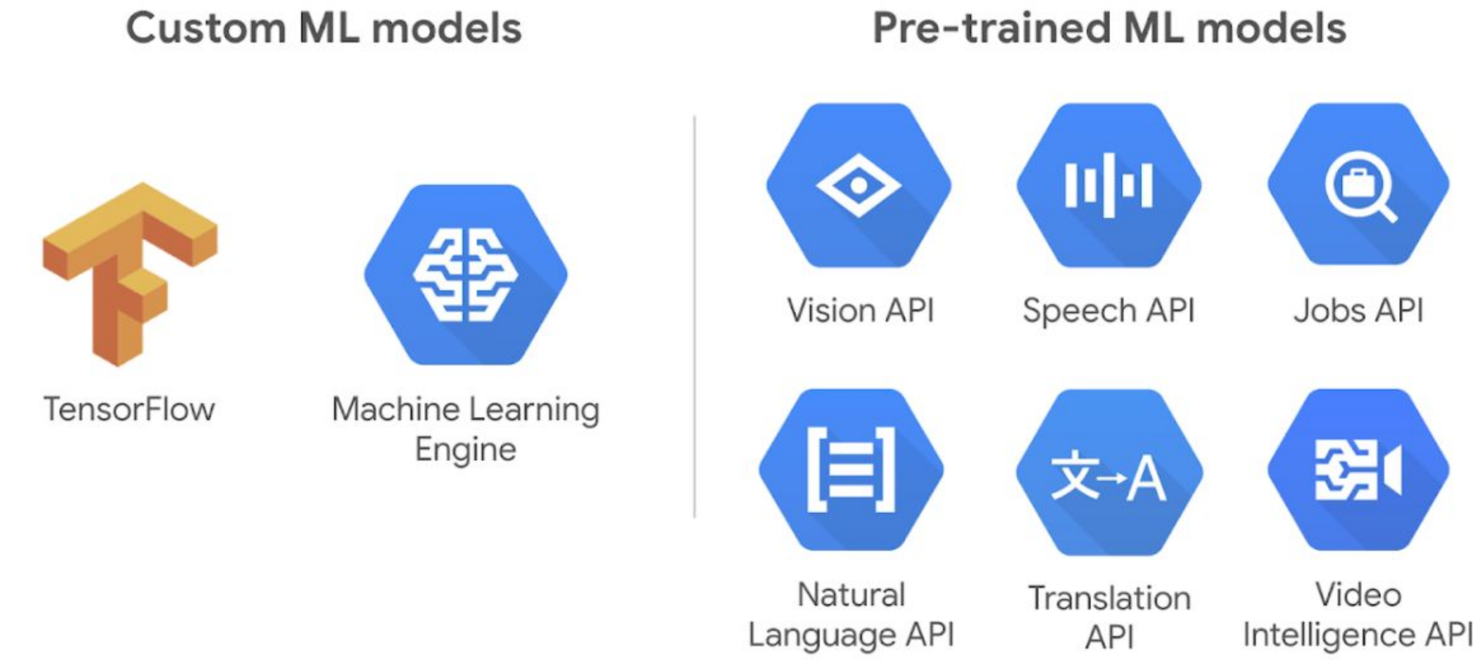


ML workflow

Machine Learning Development



Custom ML vs Pre-trained ML



Service for Custom ML

AI Platform

AI Platform Data Labeling Service

Cloud AutoML

AutoML Translation

AutoML Vision

Cloud TPU

Recommendations AI

Advanced Solutions Lab

Cloud Text-to-Speech

Dialog Flow Enterprise Edition

AI Hub

AutoML Video Intelligence

AutoML Natural Language

AutoML Tables

Google Cloud Platform Icons

AI and Machine Learning

AI Platform

Cloud Vision API

Cloud Speech-to-Text

Cloud Video Intelligence API

Cloud AutoML

Cloud TPU

Cloud Natural Language API

Cloud Translation API

Cloud Jobs API

Advanced Solutions Lab

Cloud Text-to-Speech

Dialog Flow Enterprise Edition

AI Hub

AutoML Video Intelligence

AutoML Natural Language

AutoML Tables

AutoML Translation

AutoML Vision

Recommendations AI

Cloud Inference API

AI Platform Data Labeling Service

AutoML Video Intelligence

AutoML Natural Language

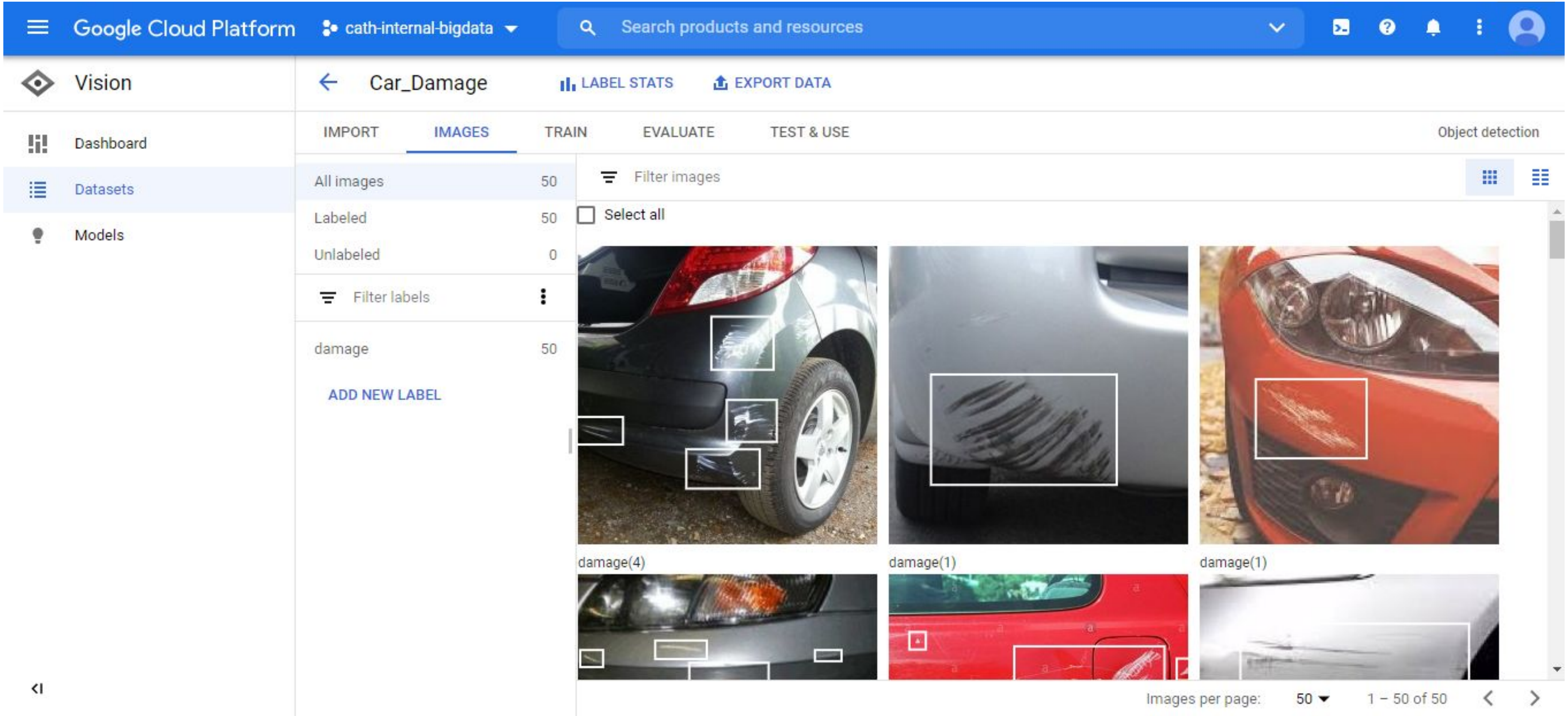
AutoML Tables

AutoML Translation

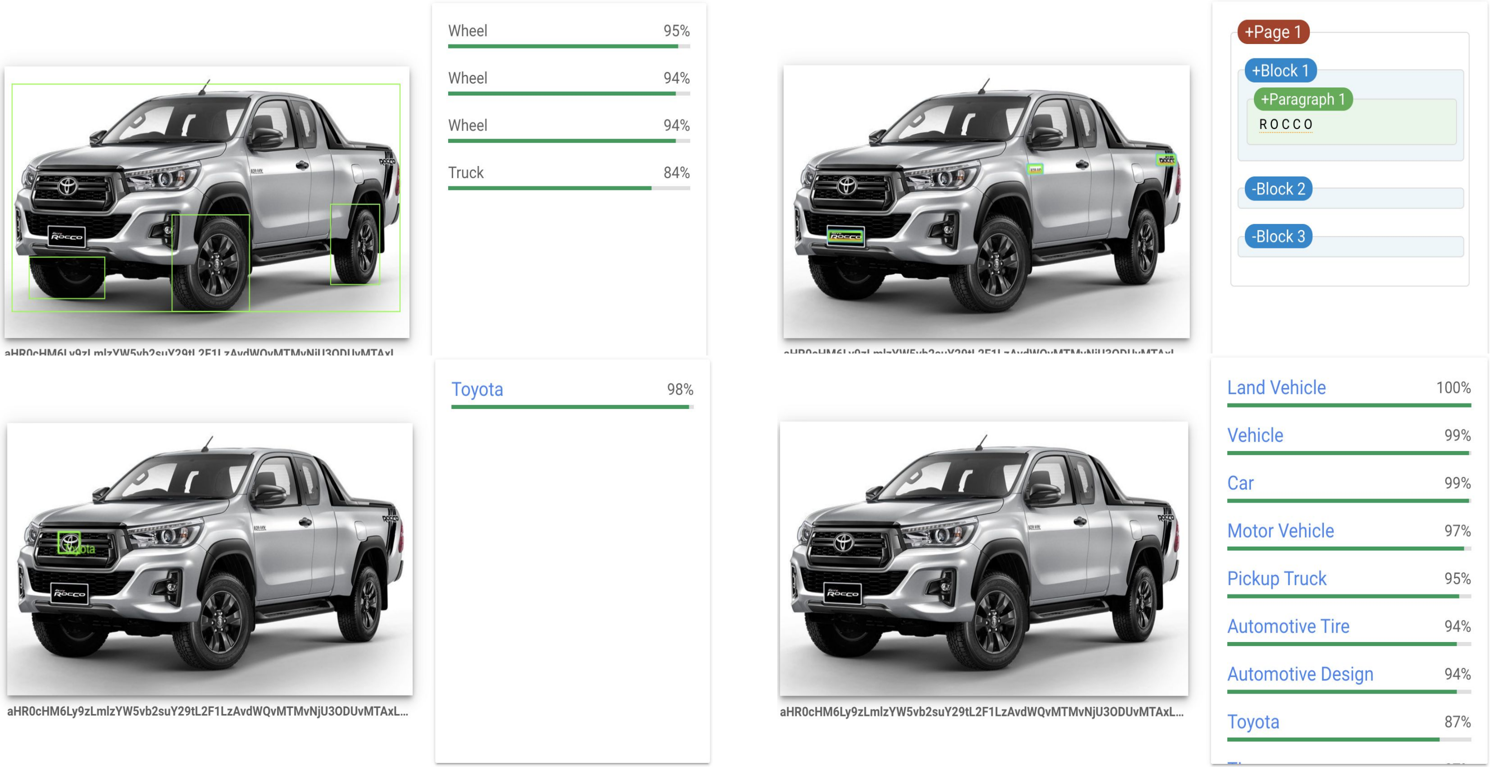
AutoML Vision

Use case of AI Services in Google Cloud Platform

Example: AutoML Vision

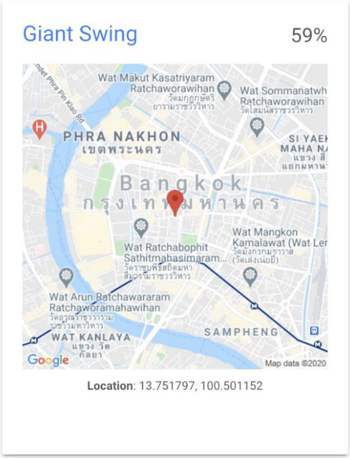


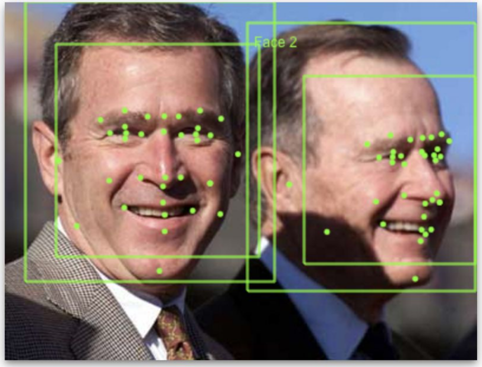
Example: Vision API



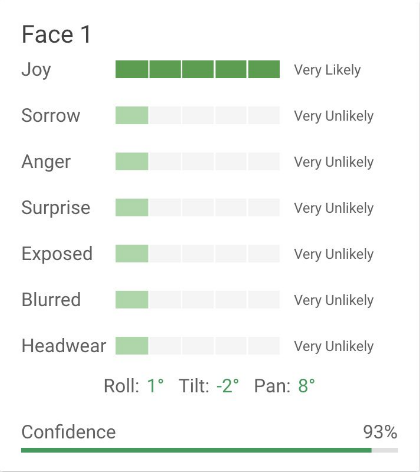
Example: [Vision API]

unnamed (5).jpg

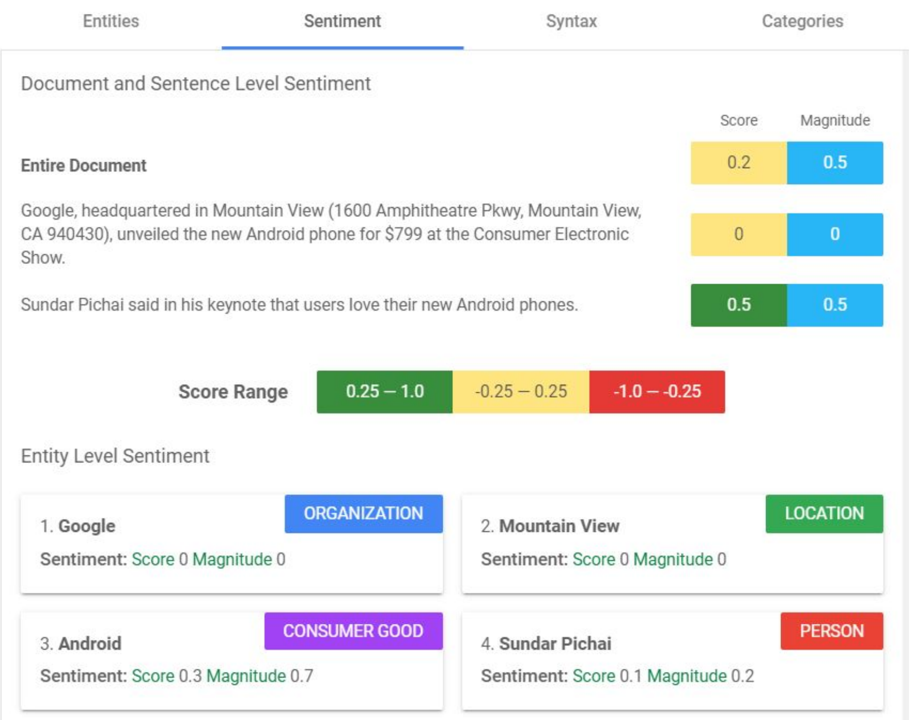


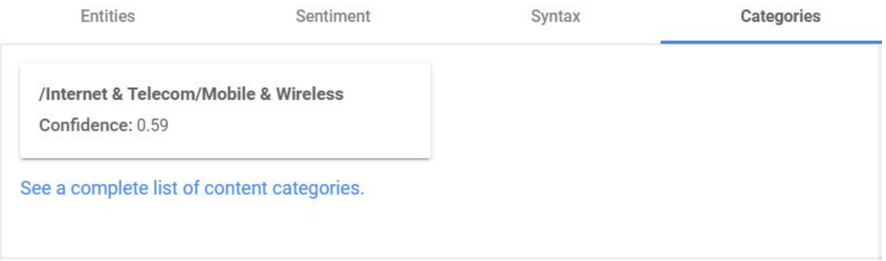


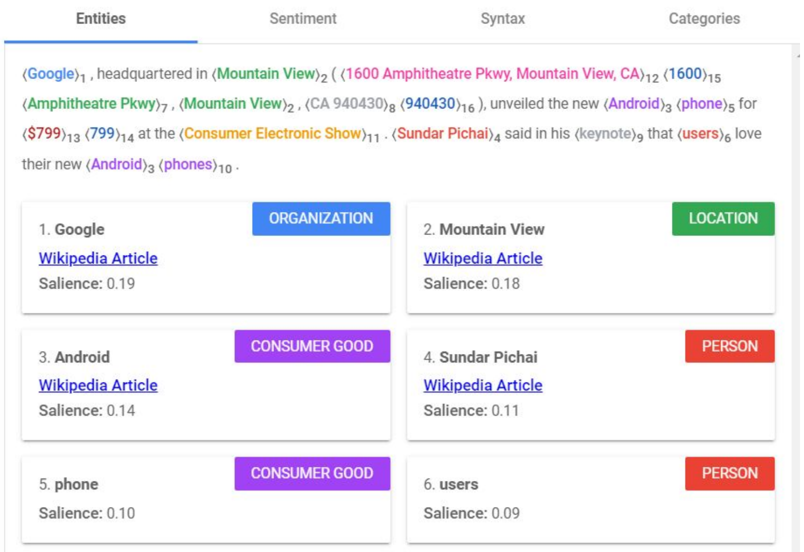
549000015140602.jpeg



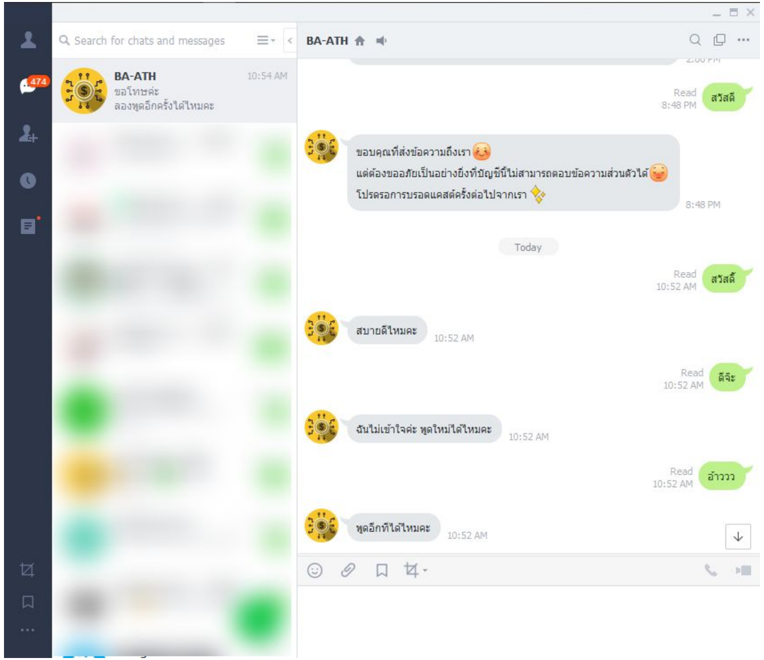
Example: Natural Language







Example: Dialogflow



Break 14:45

Identify Damaged Car Parts with Vertex AutoML Vision

1 hour 30 minutes   5 Credits

GSP972

Google Cloud Self-Paced Labs

BigQuery ML

What is BigQuery ML?

BigQuery + **Machine Learning**

1. Easy to build and run the model
2. No Python or Java are required
3. High speed of model development

BigQuery ML is a way to build custom models

Build a Custom Model

Cloud TPUs

Compute Engine

Cloud Dataproc

Kubernetes Engine

Cloud AI Platform

BigQuery ML

Build Custom Model (codeless)

AutoML

Call a Pretrained Model

Cloud Translation API

Cloud Vision API

Cloud Speech API

Cloud Video Intelligence API

Data Loss Prevention API

Cloud Speech Synthesis API

Cloud Natural Language API

Dialogflow

Working with BigQuery ML

1. Dataset
2. Create/train
3. Evaluate
4. Predict/classify

```
FROM
  ML.EVALUATE(MODEL
`bqml_tutorial.sample_model`,
     TABLE eval_table)
```

```
CREATE MODEL `bqml_tutorial.sample_model`
OPTIONS(model_type='logistic_reg') AS
SELECT
```

```
FROM
    ML.PREDICT(MODEL
`bqml_tutorial.sample_model`,
      table game_to_predict) )
AS predict
```

Where was this article published?

1. Techcrunch
2. GitHub
3. NY Times

Unlikely Partnership in House Gives Lawmakers Hope for Border Deal

Representatives Nita M. Lowey and Kay Granger are the first women to lead the House Appropriations Committee. Their bond gives lawmakers optimism for the work to come.

By EMILY COCHRANE

Fitbit's newest fitness tracker is just for employees and health insurance members

Fitbit has a new fitness tracker, but it's one that you can't buy in stores. The company quietly uncorked the Inspire on Friday, releasing its first product that is available only to co...

1 hour ago **Jon Russell**

Downloading the Android Studio Project Folder

FTC Engineering edited this page on Sep 19, 2017 · 1 revision

Downloading the Android Studio Project Folder

SQL query to extract data

**no clusters, no indexes, ad hoc query!*

```
SELECT
  url, title
FROM
  `bigquery-public-data.hacker_news.stories`
WHERE
  LENGTH(title) > 10
  AND LENGTH(url) > 0
LIMIT 10
```

url	title
http://www.bbc.co.uk/news/business-27732743	Vodafone reveals direct government wiretaps
https://www.kickstarter.com/projects/appdocu/a...	Doc – App: The Human Story
http://www.starwebworld.com/android-jelly-bean...	Android Jelly Bean: Streaming Audio Through th...
http://www.myplanetdigital.com/digital_strateg...	Why Canadian Tech Entrepreneurs Need to Man/Wo...
http://startupislandconference.com/index.html	StartupConference June 13. - 16. 2013, HVAR Cr...
http://kopimism.org/	Kopimism Hactivism Meetup Tomorrow (Sunday) in...
http://unearthedgadget.com/xbox-live-gold-2/14...	Xbox Live Gold Membership Is It Really Worth -...
https://evertale.com	Evertale changes the way people remember
http://www.racketboy.com/retro/commodore-amiga...	Commodore Amiga: A Beginner's Guide
http://www.extremetech.com/extreme/156393-cold...	Cold fusion reactor "independently verified"

Use regex to get source + train on words of title

txtclass_words

LINK SHARING

```sql
WITH extracted AS (
SELECT source, REGEXP_REPLACE(LOWER(REGEXP_REPLACE(title, '[^a-zA-Z0-9 $.-]', ' ')), "  "
FROM
  (SELECT
    ARRAY_REVERSE(SPLIT(REGEXP_EXTRACT(url, '.*://(.[^/]+)/'), '.'))[OFFSET(1)] AS source
    title
  FROM
    `bigquery-public-data.hacker_news.stories`
  WHERE
    REGEXP_CONTAINS(REGEXP_EXTRACT(url, '.*://(.[^/]+)/'), '.com$')
    AND LENGTH(title) > 10
  )
)
, ds AS (
SELECT ARRAY_CONCAT(SPLIT(title, " "), ['NULL', 'NULL', 'NULL', 'NULL', 'NULL']) AS words
extracted
WHERE (source = 'github' OR source = 'nytimes' OR source = 'techcrunch')
)
SELECT
source,
words[OFFSET(0)] AS word1,
words[OFFSET(1)] AS word2,
words[OFFSET(2)] AS word3,
words[OFFSET(3)] AS word4,
words[OFFSET(4)] AS word5
FROM ds
```

Run | Save query | Save view | More

This query will process 204.

Query results | SAVE RESULTS | EXPLORE IN DATA STUDIO

37293	nytimes	the	socratic	shrink	NULL	NULL
37294	nytimes	still	stuck	in	a	climate
37295	nytimes	as	unlimited	data	plans	are
37296	nytimes	disney	s	neuroscience	advertising	lab
37297	nytimes	hold	that	thought	the	google

Create model

Query to extract training data

```
CREATE OR REPLACE MODEL advdata.txtclass
OPTIONS(model_type='logistic_reg',
input_label_cols=['source'])
AS


WITH extracted AS (
...
)
, ds AS (
SELECT ARRAY_CONCAT(SPLIT(title, " "), ['NULL', 'NULL',
'NULL', 'NULL', 'NULL']) AS words, source FROM extracted
WHERE (source = 'github' OR source = 'nytimes' OR source
= 'techcrunch')
)


SELECT
source,
words[OFFSET(0)] AS word1,
words[OFFSET(1)] AS word2,
words[OFFSET(2)] AS word3,
words[OFFSET(3)] AS word4,
words[OFFSET(4)] AS word5
FROM ds
```

Evaluate model

```
SELECT * FROM ML.EVALUATE(MODEL advdata.txtclass)
```

precision	recall	accuracy	f1_score	log_loss	roc_auc
0.783	0.783	0.79	0.783	0.858	0.918

(BQML splits the training data and reports evaluation statistics on the held-out set)

Actual labels / Predicted labels	github	nytimes	techcrunch	% samples
github	88.8%	5.29%	5.9%	37.83%
nytimes	6.34%	70.92%	22.74%	31.26%
techcrunch	5.54%	19.35%	75.11%	30.9%

Predict using trained model

```
SELECT * FROM ML.PREDICT(MODEL advdata.txtclass,(
     SELECT 'government' AS word1, 'shutdown' AS word2, 'leaves'
AS word3, 'workers' AS word4, 'reeling' AS word5
     UNION ALL SELECT 'unlikely', 'partnership', 'in', 'house',
'gives'
     UNION ALL SELECT 'fitbit', 's', 'fitness', 'tracker', 'is'
     UNION ALL SELECT 'downloading', 'the', 'android', 'studio',
'project'
))
```

"Batch prediction"

Row	predicted_source	word1	word2	word3	word4	word5
1	nytimes	government	shutdown	leaves	workers	reeling
2	nytimes	unlikely	partnership	in	house	gives
3	techcrunch	fitbit	s	fitness	tracker	is
4	techcrunch	downloading	the	android	studio	project

Supported models in BigQuery ML

- Linear regression
- Binary logistic regression
- Multiclass logistic regression
- Matrix Factorization
- K-means clustering
- Time series
- Boosted Tree
- Deep Neural Network (DNN)
- AutoML Tables
- TensorFlow model importing

1) Classification model

Linear Classifier (Logistic regression)

```
#standardsql
CREATE OR REPLACE MODEL flights.ontime
OPTIONS
  (model_type='logistic_reg', input_label_cols=['on_time']) AS
SELECT
  IF(arr_delay < 15, 1, 0) AS on_time,
  carrier,
  origin,
  dest,
  dep_delay,
  taxi_out,
  distance
FROM
  `cloud-training-demos.flights.tzcorr`
WHERE
  arr_delay IS NOT NULL
```

1) Classification model

DNN Classifier (alpha)

```
#standardsql
CREATE OR REPLACE MODEL flights.ontime
OPTIONS
  (model_type='dnn_classifier', hidden_units = [47,29,18],
  input_label_cols=['on_time']) AS
SELECT
  IF(arr_delay < 15, 1, 0) AS on_time,
  carrier,
  origin,
  dest,
  dep_delay,
  taxi_out,
  distance
FROM
  `cloud-training-demos.flights.tzcorr`
WHERE
  arr_delay IS NOT NULL
```

1) Classification model

xgboost Classifier (alpha)

```sql
#standardsql
CREATE OR REPLACE MODEL flights.ontime
OPTIONS
  (model_type='boosted_tree_classifier', input_label_cols=['on_time']) AS
SELECT
  IF(arr_delay < 15, 1, 0) AS on_time,
  carrier,
  origin,
  dest,
  dep_delay,
  taxi_out,
  distance
FROM
  `cloud-training-demos.flights.tzcorr`
WHERE
  arr_delay IS NOT NULL
```

2) Regression model

Linear Regression

```
CREATE OR REPLACE MODEL
  taxi.taxifare_dnn OPTIONS (model_type='linear_reg',
    labels=['fare_amount']) AS
SELECT
    fare_amount,
    hourofday, dayofweek,
    pickuplon, pickuplat, dropofflon, dropofflat,
    passenger_count
  FROM
    `taxi.taxi3m`
```

Google Cloud

2) Regression model

DNN Regression (alpha)

```
CREATE OR REPLACE MODEL
  taxi.taxifare_dnn OPTIONS (model_type='dnn_regressor',
    hidden_units=[144,89,55],
    labels=['fare_amount']) AS
SELECT
    fare_amount,
    hourofday, dayofweek,
    pickuplon, pickuplat, dropofflon, dropofflat,
    passenger_count
  FROM
    `taxi.taxi3m`
```

Google Cloud

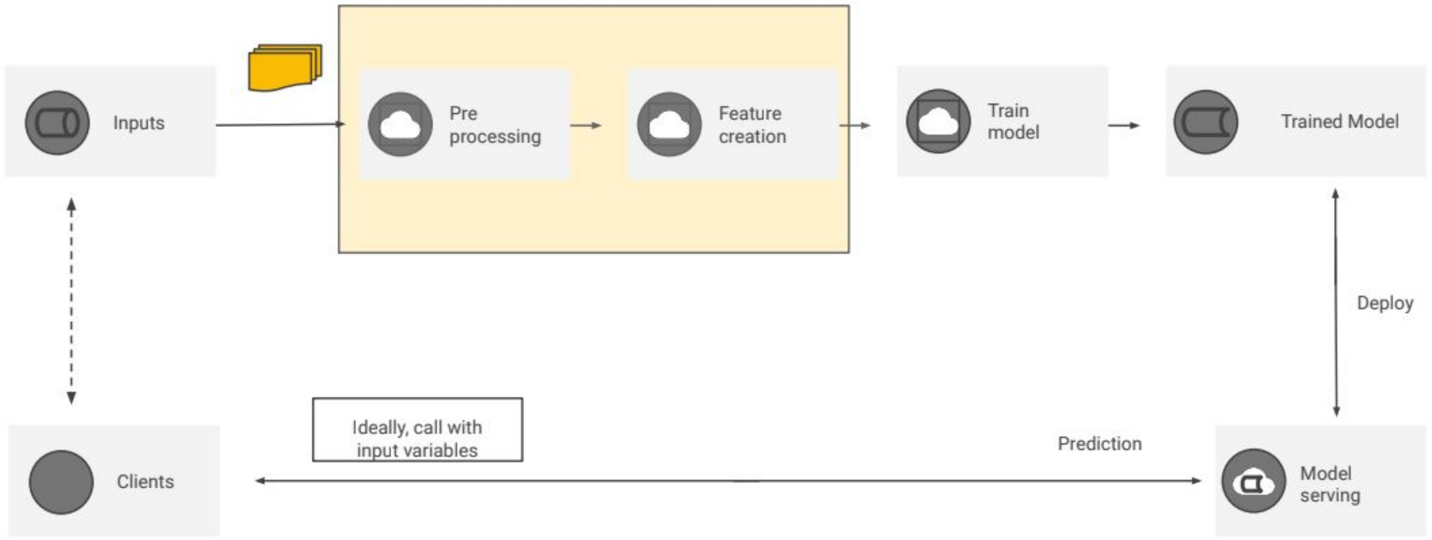
2) Regression model

xgboost Regression (alpha)

```
CREATE OR REPLACE MODEL
  taxi.taxifare_xgboost
  OPTIONS (model_type='boosted_tree_regressor',
    labels=['fare_amount']) AS
SELECT
    fare_amount,
    hourofday, dayofweek,
    pickuplon, pickuplat, dropofflon, dropofflat,
    passenger_count
  FROM
    `taxi.taxi3m`
```

Google Cloud

Use the transform clause



TRANSFORM ensures transformations are automatically applied during ML.PREDICT

```
CREATE OR REPLACE MODEL ch09edu.bicycle_model
OPTIONS(input_label_cols=['duration'],
        model_type='linear_reg')
AS

SELECT
  duration
  , start_station_name
  , CAST(EXTRACT(dayofweek from start_date) AS STRING)
        as dayofweek
  , CAST(EXTRACT(hour from start_date) AS STRING)
        as hourofday
FROM
  `bigquery-public-data.london_bicycles.cycle_hire`
```

```
SELECT * FROM ML.PREDICT(MODEL ch09edu.bicycle_model,(
  350 AS duration
  , 'Kings Cross' AS start_station_name
  , '3' as dayofweek
  , '18' as hourofday
))
```

```
CREATE OR REPLACE MODEL ch09edu.bicycle_model
OPTIONS(input_label_cols=['duration'],
        model_type='linear_reg')
TRANSFORM(
  SELECT * EXCEPT(start_date)
  , CAST(EXTRACT(dayofweek from start_date) AS STRING)
        as dayofweek
  , CAST(EXTRACT(hour from start_date) AS STRING)
        as hourofday
)
AS
SELECT
  duration, start_station_name, start_date
FROM
  `bigquery-public-data.london_bicycles.cycle_hire`
```

```
SELECT * FROM ML.PREDICT(MODEL ch09edu.bicycle_model,(
  350 AS duration
  , 'Kings Cross' AS start_station_name
  , CURRENT_TIMESTAMP() as start_date
))
```

BigQuery ML - Free Tier

Resource	Monthly free usage limits	Details
Storage	The first 10 GB per month is free.	BigQuery ML models and training data stored in BigQuery are included in the BigQuery storage free tier.
Queries (analysis)	The first 1 TB of query data processed per month is free.	Queries that use BigQuery ML prediction, inspection, and evaluation functions are included in the BigQuery analysis free tier. BigQuery ML queries that contain `CREATE MODEL` statements are not. BigQuery flat-rate pricing is also available for high-volume customers that prefer a stable, monthly cost.
BigQuery ML `CREATE MODEL` queries	The first 10 GB of data processed by queries that contain `CREATE MODEL` statements per month is free.	BigQuery ML `CREATE MODEL` queries are independent of the BigQuery analysis free tier, and only apply to BigQuery ML built-in models (models that are trained within BigQuery).

Google Cloud

QWIKLABS

LAB1

Predict Visitor Purchases with a Classification Model in BQML

https://www.qwiklabs.com/focuses/1794?locale=en&parent=catalog