

Introduction to Machine Learning

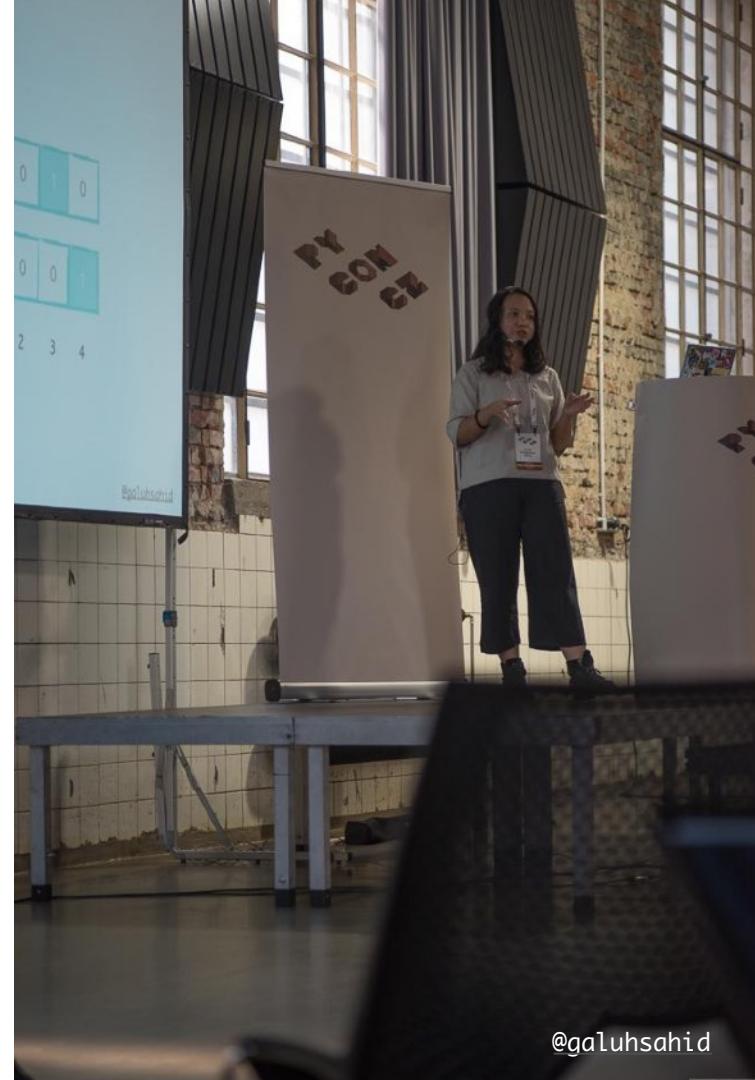
Nusantech Webinar - June 28, 2020

Galuh Sahid

@galuhsahid | galuh.me

Hi! I'm Galuh.

- Data Scientist at Gojek
- Google Developer Expert in Machine Learning
- Co-host podcast Kartini Teknologi (kartiniteknologi.id)



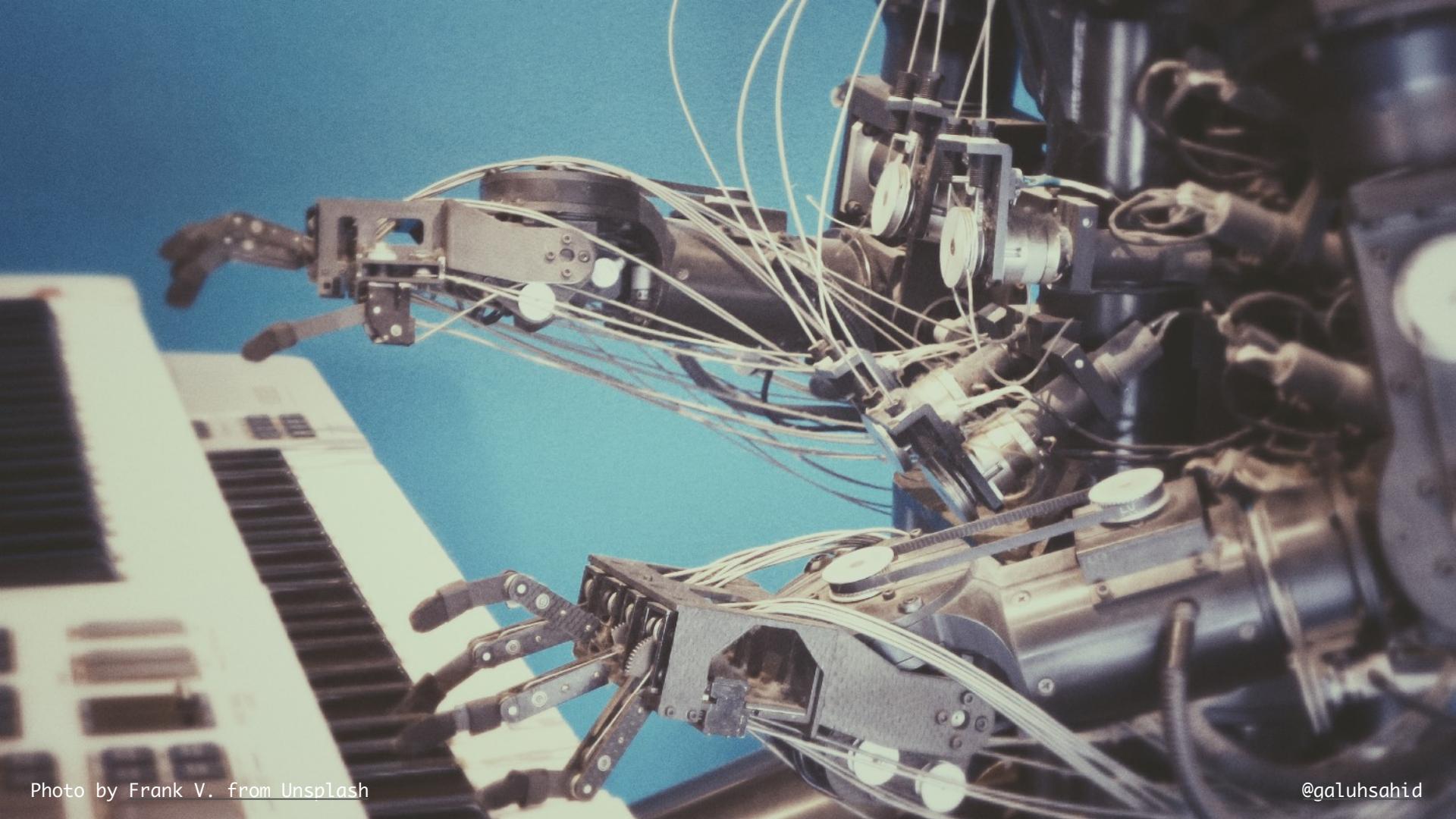
@galuhsahid

<https://bit.ly/intro-ml-nusantech>

Outline

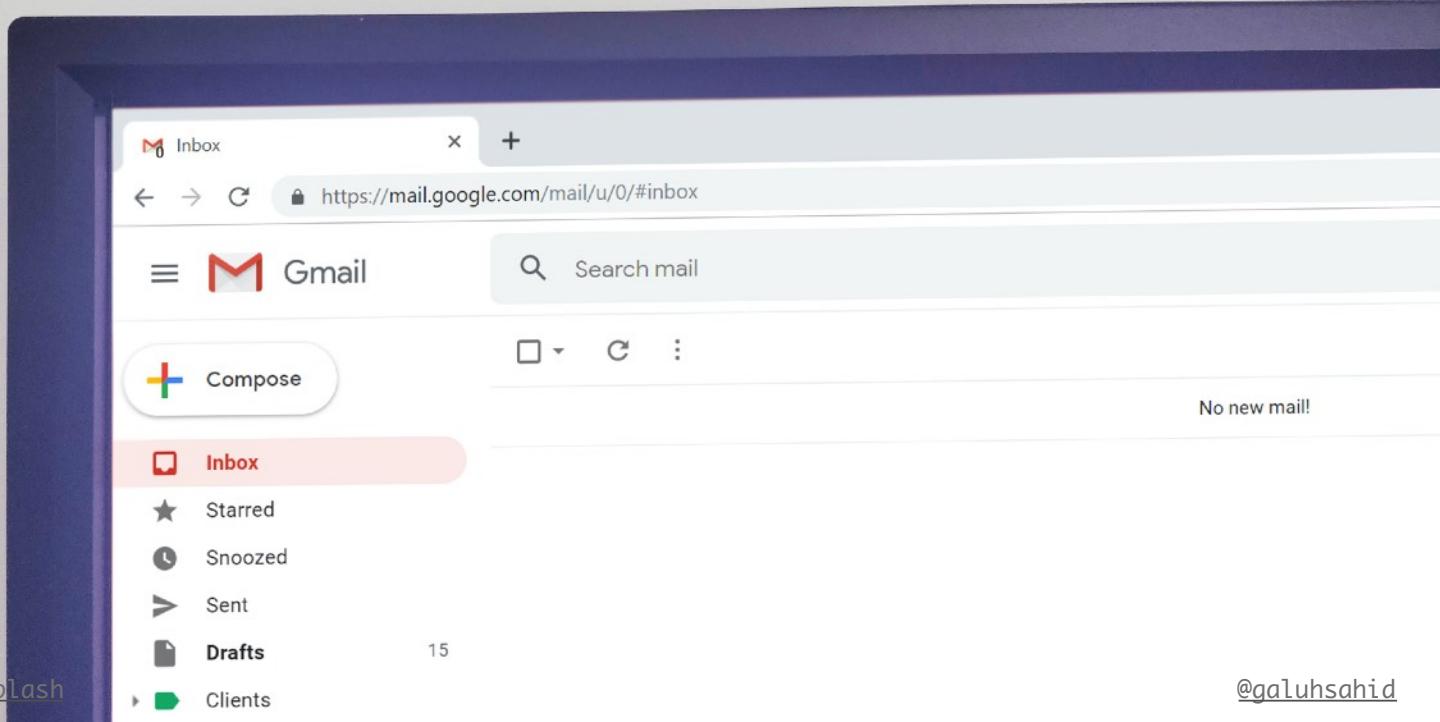
- Definition of machine learning & its difference with traditional programming
- Machine learning flow
 - Defining ML problem
 - Acquiring, getting to know, & preparing your data
 - Training your model
 - Making predictions
- Tools & resources
- Demo











**It's an exciting time to learn about
machine learning!**

But... what is machine learning?

A field of study that gives computers the ability to learn without being explicitly programmed.

Arthur Samuel (1959)

How is machine learning different from traditional programming?





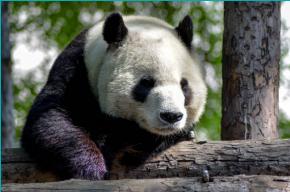
```
if pixel[5][7] is black and pixel [5][6]
is black and pixel [5][8] is black and ...:
    if pixel[6][7] is black and pixel[6][7]
    is black and ...:
        return "panda"
...
...
...
else:
    return "cat"
```



```
if pixel[5][7] is black and pixel [5][6]
is black and pixel [5][7] is black and ...:
    if pixel[6][7] is black and pixel[6][7]
    is black and ...:
        return "panda"
...
...
...
else:
    return "not cat"
```



Data



Answers

Panda

Panda

Cat

Cat

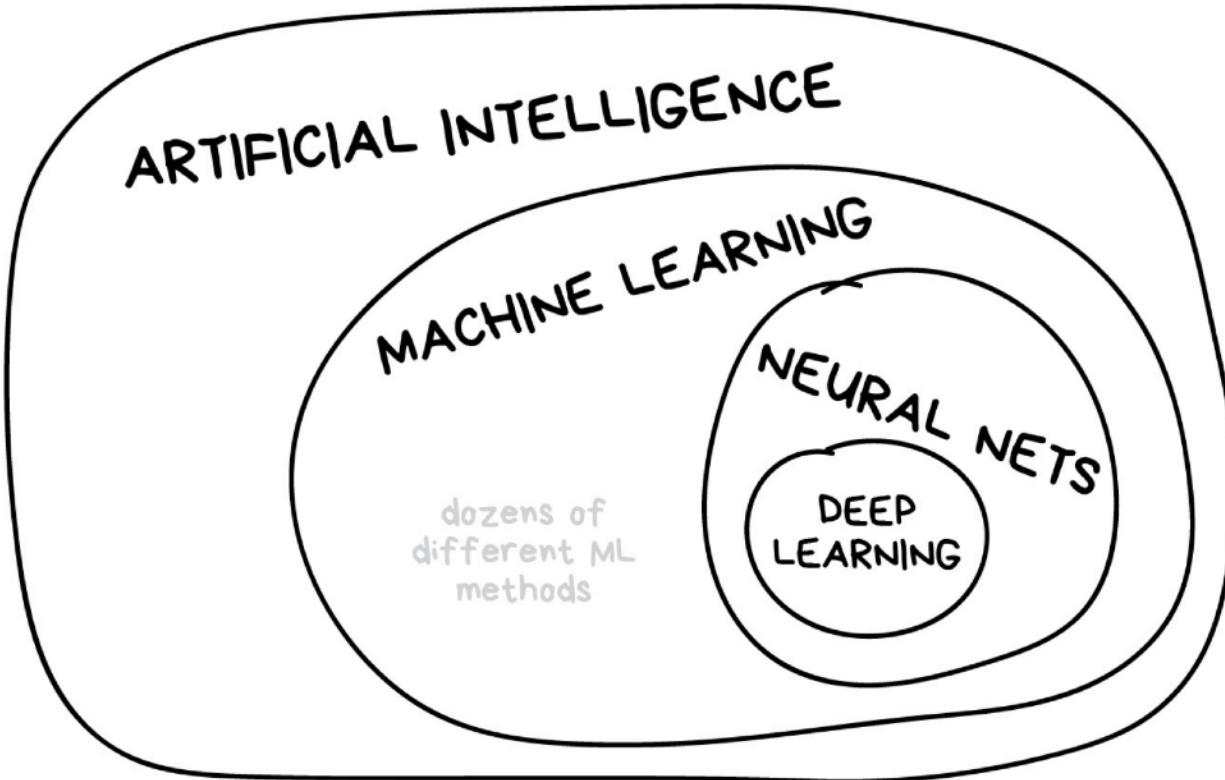


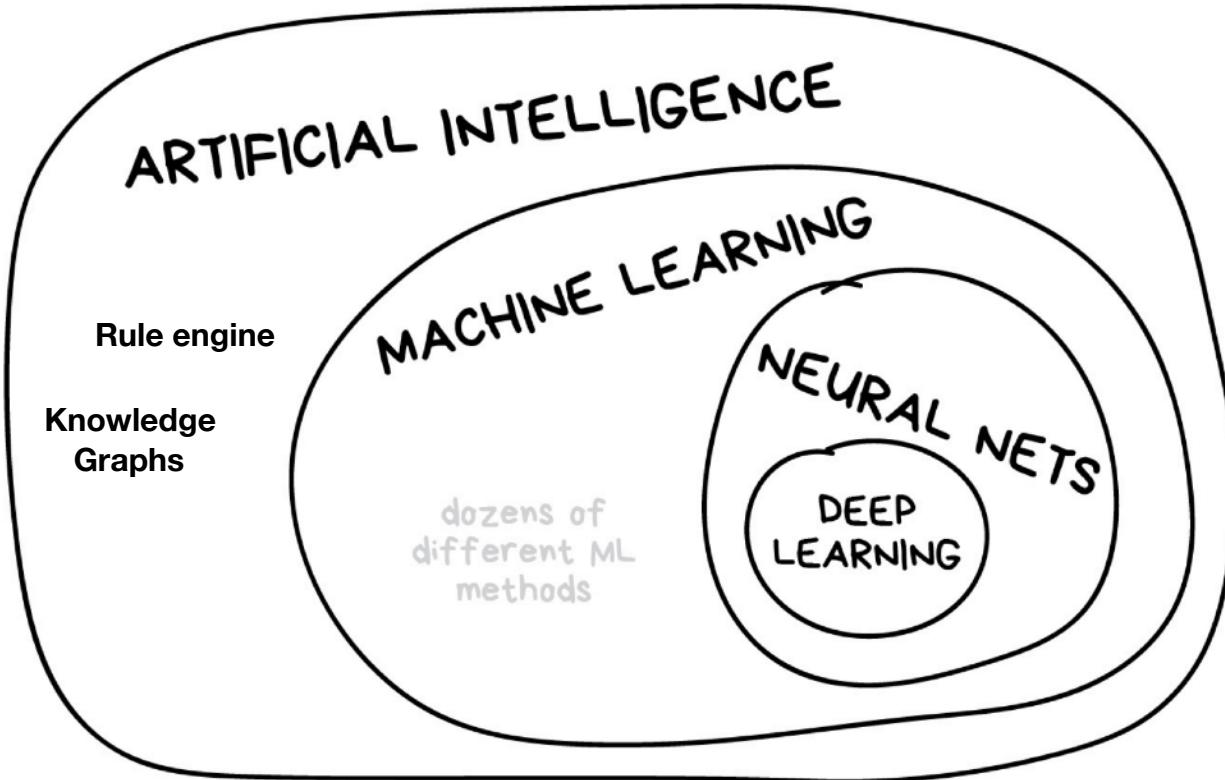


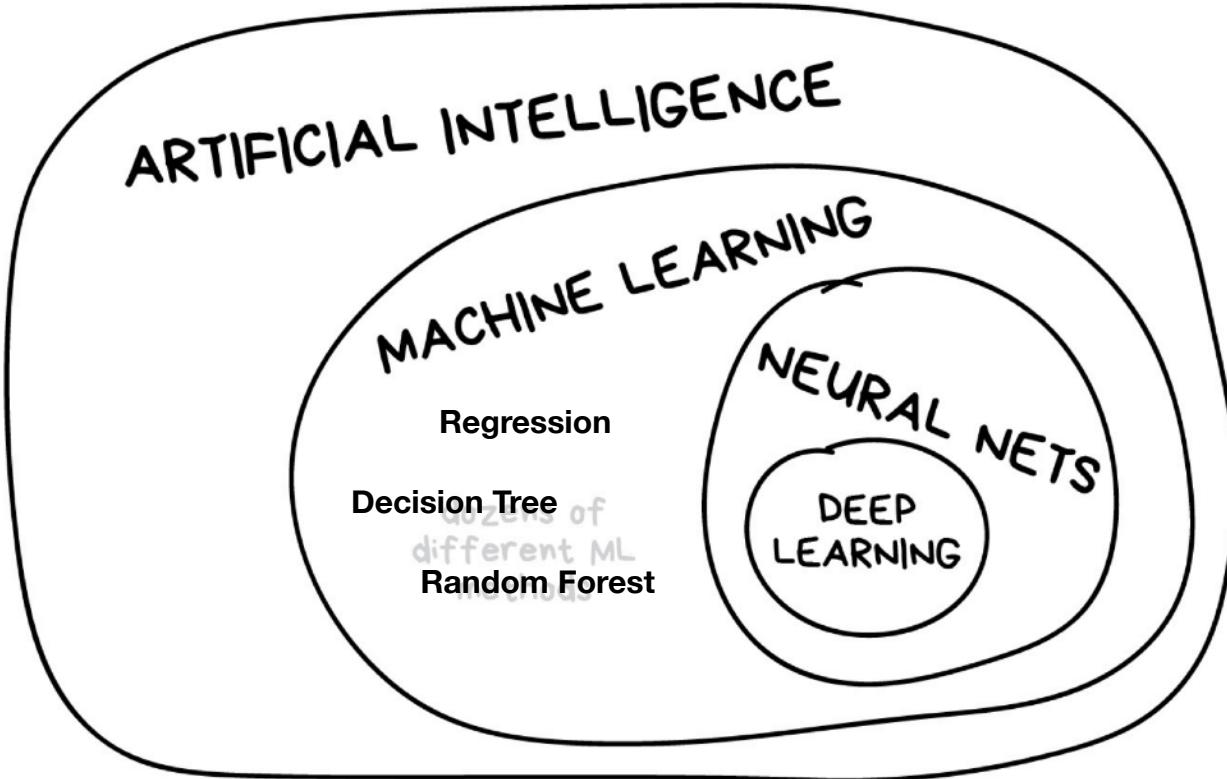
Cat

?

**Machine learning? Artificial
intelligence? Neural networks? Deep
learning?**







Machine learning flow



Step #1

Define your machine learning problem

Type #1

Binary classification

Classifies input into one of two categories

Binary Classification

Spam or not spam

DEAR GOOD FRIEND, ⏪ Spam ✖



Mr. Alassane Bello alassanebelloa@gmail.com via yahoo.com
to ▾

Mon, Jun 8, 11:40 PM (13 days ago)



This message seems dangerous

Similar messages were used to steal people's personal information. Avoid clicking links, downloading attachments, or replying with personal information.

Looks safe



ATTENTION PLEASE,

I AM A BANKER AND MANAGER OF OUR BANK. I WANT TO TRANSFER SUM OF \$20.6 MILLION US INTO YOUR ACCOUNT BASED ON COMMISSION. IF YOU CAN HELP ME TO RECEIVE THIS MONEY BASE ON COMMISSION, THEN KINDLY RESPOND IMMEDIATELY, SO THAT I WILL SEND DETAILS TO YOU AND LET YOU KNOW YOUR COMMISSION OK

THIS IS A RISK FREE TRANSACTION, SO DO NOT BE AFRAID OF HAVING PROBLEM. BECAUSE WE ARE NOT GOING TO HAVE ANY PROBLEM. PLEASE, KINDLY ANSWER IMMEDIATELY THROUGH THIS MY PRIVATE E-MAIL ADDRESS: alassanebello@outlook.com: SO THAT I WILL SEND FURTHER DETAILS TO YOU OK.

YOURS FAITHFULLY,
MR. ALASSANE BELLOW.

Reply

Forward

@galuhsahid

Type #2

Multi-class classification

Classifies input into one of more than two categories

Multi-class Classification

Language prediction

The screenshot shows the homepage of lemonde.fr. At the top, there's a navigation bar with icons for back, forward, and search, followed by the URL 'lemonde.fr'. To the right of the URL is a language selection dropdown set to 'French' with an English option and a Google Translate link. Below the header, the 'Le Monde' logo is prominently displayed. On the left, there's a thumbnail of the newspaper and a link to 'Consulter le journal'. The main content area features several news cards with titles and timestamps. The first card, under 'ACTUALITÉS', is titled 'Chronique | « La fable du colibri, selon laquelle une somme de petites actions individuelles pourrait résoudre la crise écologique, porte une belle idée mais elle est aussi dangereuse »' and is timestamped at 11:02. The second card, under 'ÉCONOMIE', is titled 'Record pour la vente d'une guitare de Kurt Cobain' and is timestamped at 10:49. The third card, under 'OPINIONS', is titled 'Mort de Pascal Clément, ex-ministre de la justice' and is timestamped at 10:34. The fourth card, under 'CULTURE', is titled 'Climat : dernier jour pour la convention citoyenne' and is timestamped at 09:35. The fifth card, under 'M LE MAG', is titled 'Un an après la mort de Steve, une marche organisée' and is timestamped at 09:04. A 'Voir plus' link is visible on the far right. At the bottom, there are two large images: one showing a crowd of people at a political rally and another showing a person wearing a mask.

lemonde.fr

French English

Google Translate

Le Monde

Consulter le journal

ACTUALITÉS

ÉCONOMIE

VIDÉOS

OPINIONS

CULTURE

M LE MAG

SERVICES

11:02 Alerte

Chronique | « La fable du colibri, selon laquelle une somme de petites actions individuelles pourrait résoudre la crise écologique, porte une belle idée mais elle est aussi dangereuse »

10:49 Record pour la vente d'une guitare de Kurt Cobain

10:34 Mort de Pascal Clément, ex-ministre de la justice

09:35 Climat : dernier jour pour la convention citoyenne

09:04 Un an après la mort de Steve, une marche organisée

Voir plus

La convention citoyenne pour le climat doit rendre ses propositions cet après-midi



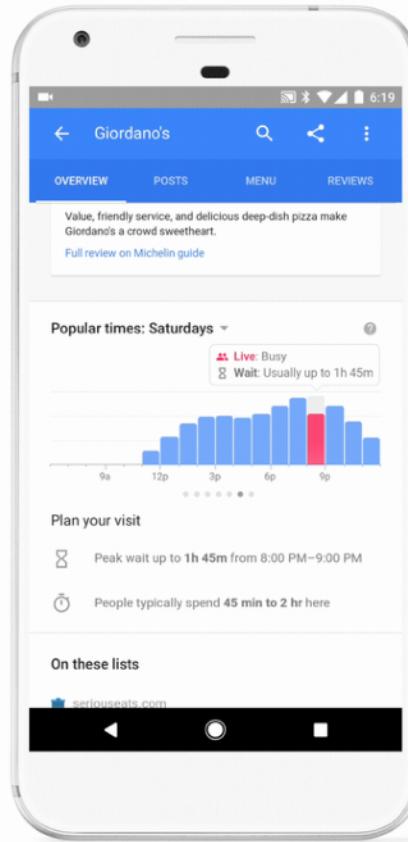
Type #3

Regression

Predicts a value on a continuous scale

Regression

Visit duration & wait time estimates



Type #4

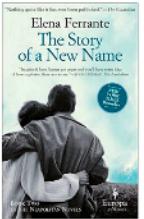
Catalog organization

Produces a set of result to present to users

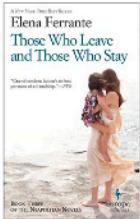
Recommender System

What other items do customers buy after viewing this item?

Page 1 of 3



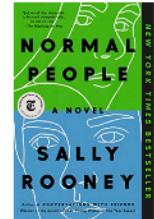
[The Story of a New Name: Neapolitan Novels, Book Two](#)
by Elena Ferrante
 1,612 Paperback
74 offers from \$3.74



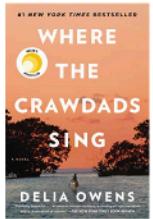
[Those Who Leave and Those Who Stay: Neapolitan Novels, Book Three](#)
by Elena Ferrante
 1,423 Paperback
81 offers from \$5.93



[The Story of the Lost Child: Neapolitan Novels, Book Four](#)
by Elena Ferrante
 1,604 Paperback
56 offers from \$5.77



[Normal People: A Novel](#)
by Sally Rooney
 3,194 Paperback
50 offers from \$5.50



[Where the Crawdads Sing](#)
by Delia Owens
 50,410 Paperback
#1 Best Seller in Mothers & Children Fiction
Hardcover
140 offers from \$5.12



[Just Mercy: A Story of Justice and Redemption](#)
by Bryan Stevenson
 5,873 Paperback
#1 Best Seller in Criminology
Hardcover
60 offers from \$3.99



[The Vanishing Half: A Novel](#)
by Brit Bennett
 100 Paperback
\$18.02



Type #5

Generative model

Focuses on generating data rather than classifying or organizing it

Generative Model

Generating a new face



Generating a new text

talktotransformer.com

Cooking instructions

Making an omelette is simple!

1.

GENERATE ANOTHER

Completion

Making an omelette is simple!

1. Cook the rice

Use extra-virgin olive oil, freshly ground black pepper and powdered onion salt.

2. Heat your butter in a small non-stick frying pan.

3. Once the pan is hot and the oil is melted, add the onions and butter and sauté for 5 minutes.

Life advice

List of the best advice I ever heard:

1.

GENERATE ANOTHER

Completion

List of the best advice I ever heard:

1. Work to your strengths

In my old job, I had a tendency to take on the perspective that my approach was the best thing for my role. The only reason I was able to adapt to change was because I was incredibly confident about my expertise. I could see past the times when things didn't work and moved on.

(This isn't how we work today. What are the problems? How are we going to make things better?)

When I decided to take the time to focus on business school, I realized that the problem wasn't technical and I could take it personally. When I work to my strengths, I'm going to change the approach and effectively capitalize on the answer

Step #2

**Acquire, get to know, & prepare your
data**

Step #2

Acquire, get to know, & prepare your data

You need to know:

- What are the types of data that you can use
- Where to get them
- How to get to know your data
- How to prepare your data



Kementerian PPN - Bappenas

There is no description for this organization

Dataset Tematik

Produksi Pangan Strategis (Juta Ton)_Padi

Informasi Metadata

Elemen	Isi
Last Updated	February 1, 2019, 2:12 PM (UTC+07:00)
Created	February 1, 2019, 2:12 PM (UTC+07:00)

Nama Berkas	Ekstensi	View
Data Series Produksi Pangan Strategis (Juta Ton)_Padi	CSV	0

Data Type #1

Tabular

series	kode_wilayah	wilayah	tahun	nilai
Produksi Pangan	1	Indonesia	2014	70.6
Produksi Pangan	1	Indonesia	2015	75.4
Produksi Pangan	1	Indonesia	2016	79.1
Produksi Pangan	1	Indonesia	2017	77

Data Type #2

Text

Titania (satelit)

Artikel ini bukan mengenai [Titian](#) (satelit).

Artikel ini berisi uraian tentang satelit planet Uranus. Untuk karakter dalam sandiwara karya William Shakespeare yang berjudul *Impian di Tengah Musim*, lihat [Titania](#).

Titania adalah satelit alami terbesar planet [Uranus](#) sekaligus satelit alami terbesar kedelapan di [Tata Surya](#) dengan diameter 1.578 km.

Titania ditemukan oleh [William Herschel](#) pada tahun 1787 dan dinamai menurut nama [seorang ratu peri](#) dalam sebuah sandiwara komedi karya [William Shakespeare](#), yakni *Impian di Tengah Musim*. Orbitnya terletak di dalam [magnetosfer](#) Uranus.

Titania terdiri atas kendungan es dan bebatuan yang kurang lebih sama banyak, dan kemungkinan terdirerensiasi menjadi [inti](#) yang berbatu dan [mantel](#) yang terdiri dari es. Di [batas antara inti dan mantel](#) mungkin terdapat sebuah lapisan air. Permukaan Titania, yang relatif gelap dan sedikit berwarna merah, tampaknya terbentuk akibat tubukan dan proses [endogenik](#). Permukaan Titania juga penuh akan [kawah](#) tubukan yang diameternya dapat mencapai 326 km, tetapi kawan di permukaan Titania tidak sebanyak kawah yang terdapat di permukaan satelit terluar Uranus, [Oberon](#). Permukaan Titania pernah mengalami proses pelapisan kembali secara endogenik yang menghancurkan permukaannya yang tua dan penuh akan kawah. Di permukaan Titania juga terdapat rangkaian [ngarai](#) dan [gawir](#) besar, yang merupakan dampak dari pengembangan bagian dalamnya selama tahap akhir evolusinya. Sama seperti sebagian besar satelit Uranus lainnya, Titania kemungkinan terbentuk dari [piringen alkresi](#) yang mengelilingi planet tersebut setelah pembentukannya.

Spektroskopi inframerah pada tahun 2001 dan 2005 menunjukkan keberadaan [es](#) air dan [karbon dioksida](#) beku di permukaan Titania, yang mengindikasikan keberadaan [atmosfer](#) karbon dioksida. Pengukuran selama [okulasi](#) Titania terhadap suatu [bintang](#) menunjukkan batas maksimal tekanan atmosfer sebesar 10–20 [nbar](#). Uranus dan satelit-satelitnya hanya pernah dipelajari dari dekat oleh wahana [Voyager 2](#) pada Januari 1986. Wahana tersebut mengabadikan beberapa citra Titania, yang memungkinkan pemetaan 40% permukaan satelit tersebut.

Daftar isi [sembunyikan]

- 1 Sejarah
- 2 Orbit
- 3 Komposisi dan struktur dalam
- 4 Ketamponakan permukaan

Eddy Omonrups @eddy_omonrup · Sep 10
kebakaran di lahan gambut bat-bat kab. banjar semakin parah, jarak pandang hanya kurang lebih 10 s/d 20m, gmna pihak terkait sudah hampir 1 minggu tpp ada titik api baru @jckonw @osinmasinsipost @benarkab
@4tbsKaise #kebakaran2019 #kebakaranhutan

1:16 | 89 views

Romadholi @Romadholi · Sep 14
Bahan kami menemukan beberapa titik api yang nampanknya baru saja dibakar.
#kebakaranhutan
#kebakaranHutanMakinMengigil

1:16 | 89 views

Titania



Citra belahan selatan Titania yang diabadikan oleh Voyager 2 pada 24 Januari 1986.

Penemuan

Ditemukan oleh [William Herschel](#)

Tanggal penemuan 11 Januari 1787^[1]

Penamaan

Nama alternatif [Uranus II](#)

Karakteristik orbit

Sumbu semi-mayor 435.910 km^[2]

kabut asap [kebakaran hutan](#) dan lahan ([karhutla](#)) yang wilayah kota dan kabupaten dinyatakan ada dalam bahaya.

akan oleh Pusat Pengendalian Pembangunan [Kra](#) Kementerian Kehutanan dan Lingkungan Hidup [Jung](#)an Hidup dan Kehutanan Provinsi Riau.

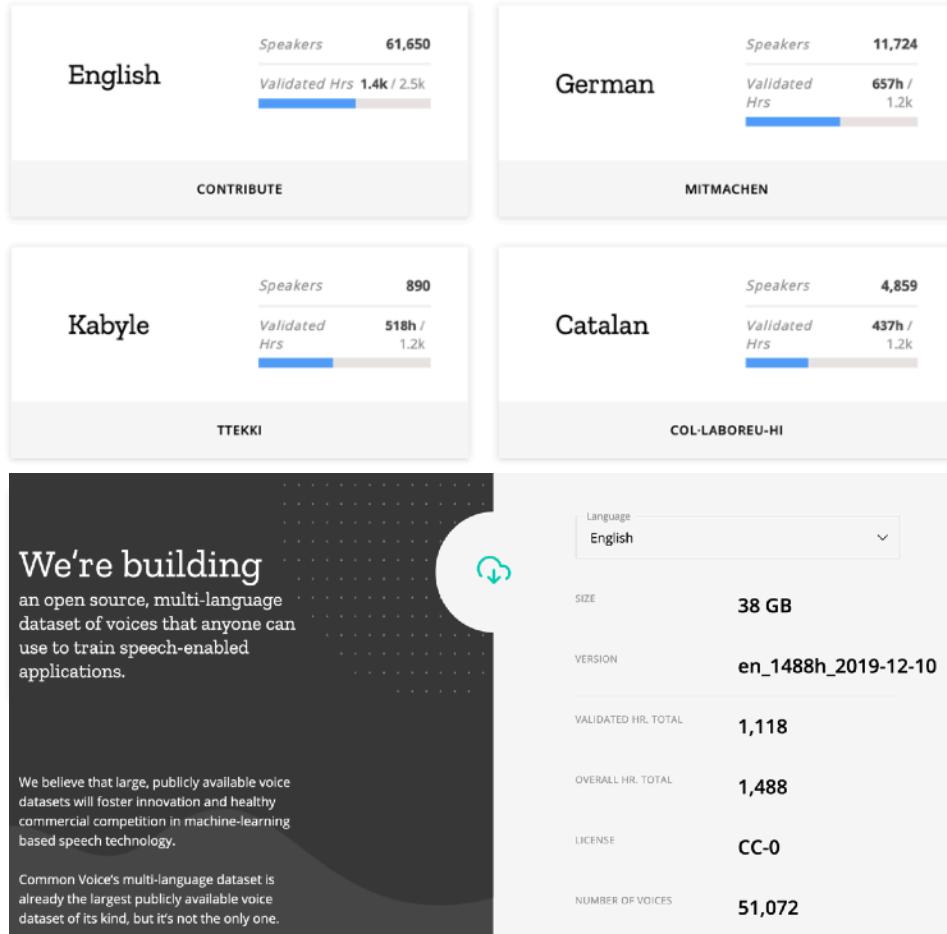
dar Pencemar Udara (ISPU) Jumat (13/9/2019) pukul dinyatakan memiliki kualitas udara dengan status Pekanbaru, Siak, Kampar, Rokan Hilir, dan Bengkalis.

elkan pada wilayah yang memiliki konsentrasi berukuran lebih kecil dari 10 mikron) lebih dari 300

@galuhsahid

Data Type #3

Sound



Data Type #4

Image



Synset: [people](#) has bounding box
Definition: (plural) any group of human beings (men or women or children) collectively; "old people"; "there were at least 200 people in the audience".



Synset: [homo_man](#), [human_being](#), [human](#) has bounding box
Definition: any living or extinct member of the family Hominidae characterized by superior intelligence, articulate speech, and erect carriage.



Synset: [child](#), [kid](#) has bounding box
Definition: a human offspring (son or daughter) of any age; "they had three children"; "they were able to send their kids to college".



Synset: [case](#), [display_case](#), [showcase](#), [vitrine](#) has bounding box
Definition: a glass container used to store and display items in a shop or museum or home.



Synset: [house](#) has bounding box
Definition: a dwelling that serves as living quarters for one or more families; "he has a house on Cape Cod"; "she felt she had to get out of the house".



Synset: [school](#), [schoolhouse](#) has bounding box
Definition: a building where young people receive education; "the school was built in 1932"; "he walked to school every morning".



Synset: [top](#) has bounding box
Definition: a garment (especially for women) that extends from the shoulders to the waist or hips; "he stared as she buttoned her top".



Synset: [woman](#), [adult](#), [female](#) has bounding box
Definition: an adult female person (as opposed to a man); "the woman kept house while the man hunted".



Synset: [support](#) has bounding box
Definition: any device that bears the weight of another thing; "there was no place to attach supports for a shelf".



Synset: [office](#), [business_office](#) has bounding box
Definition: place of business where professional or clerical duties are performed; "he rented an office in the new building".



Synset: [center](#), [centre](#) has bounding box

Where to get the data?

- Use a **ready-to-use** dataset
- **Extract** the data by yourself
- **Collet and build your own** dataset from scratch

Data Source #1

Google

Research

<https://research.google/tools/datasets/>

TOOLS & DOWNLOADS ›

Datasets

In order to contribute to the broader research community, Google periodically releases data of interest to researchers in a wide range of computer science disciplines.

Search for datasets on the web with [Dataset Search](#).

Filters Sort by: Featured ▾ 81 datasets

Dataset type	Count
Image	11
Video	12
Audio	36
Text Annotation	27
Robotics	7
Other	6

Taskmaster-1
13,215 English task-based, annotated dialogs in six domains: ordering pizza, creating auto repair appointments, setting up ride service, ordering movie tickets, ordering coffee drinks and making restaurant reservations.

Taskmaster-2
Over 17,000 spoken, annotated dialogs in seven domains collected using a "Wizard of Oz" (human-in-the-loop) platform.

Coached Conversational Preference Elicitation
Wizard-of-Oz preference elicitation conversations in English between a user and an assistant about movie preferences, with annotated preference statements.

DiscoFuse
A dataset of 60 million examples for training sentence fusion models. The data has been collected from Wikipedia and from Sports articles.

Open Images Extended - Crowdsourced
Additional imagery sets to the main Open Images dataset, to improve its diversity (geographic, cultural, demographic, subject matter, etc). Currently composed of ~478K images contributed by users of the Crowdsource app.

Data Source #2

Google

Dataset

<https://datasetsearch.research.google.com>

The screenshot shows a Google search results page for the query "global temperatures". The top navigation bar includes a magnifying glass icon, the text "global temperatures", and standard search controls. Below the search bar, there are filters for "Last updated", "Download format", "Usage rights", "Topic", and "Free". A message indicates "100+ datasets found". The first result is a dataset titled "Global surface temperatures: BEST: Berkeley Earth Surface Temperatures" from climatedataguide.ucar.edu. It includes a purple circular icon with a white letter "U", a brief description, and a link to explore at climatedataguide.ucar.edu. The second result is "Climate Change: Earth Surface Temperature Data" from www.kaggle.com, featuring a blue circular icon with a white letter "B". The third result is "Global Climate Change Data" from data.world, featuring a blue circular icon with a white letter "D". The right side of the screen displays detailed information for the first dataset, including a "Time period covered" section (Dec 1700 - Jun 2019), a "Description" section (explaining the Berkeley Earth Surface Temperatures project), and a "Cited by" section (7 scholarly articles). Below this, there are additional filters and a second set of search results for "Germination Shifts of C3 and C4 Species under Simulated Global Warming Scenario" and "Consequences of Global Warming of 1.5 °C and 2 °C for Regional Temperature...".

@galuhsahid

Data Source #3

Kaggle Datasets

<https://www.kaggle.com/datasets>

The screenshot shows the Kaggle Datasets homepage. At the top, there's a navigation bar with links for Home, Compete, Data, Notebooks, Discuss, Courses, and More. A search bar is also at the top. Below the navigation, a section titled "Datasets" encourages users to help understand COVID-19. It features a "Get Started" button and a "View Contributions" button. A search bar below this section allows users to search for datasets. The main content area is titled "PUBLIC" and displays a list of datasets sorted by "Hottest". The datasets listed are:

- COVID-19 Open Research Dataset Challenge (CORD-19)** (Allen Institute For AI) - 5864 contributions, 10 hours ago, 2 GB, 8.8 rating, 68208 files (JSON, CSV, other), 10 tasks.
- UNCOVER COVID-19 Challenge** (Roche Data Science Coalition) - 476 contributions, 16 days ago, 123 MB, 8.2 rating, 84 files (CSV, other), 12 tasks.
- CoronaHack -Chest X-Ray-Dataset** (Praveen) - 80 contributions, a month ago, 1 GB, 9.4 rating, 5935 files (other, CSV).
- MeteoNet North-West France** (Katerpillar) - 28 contributions, 9 days ago, 11 GB, 8.2 rating, 447 files (other, CSV), 5 tasks.
- COVID-19 CT scans** (Larzel) - 76 contributions, 2 days ago, 1 GB, 10.0 rating, 81 files (other, CSV), 3 tasks.

@galuhsahid

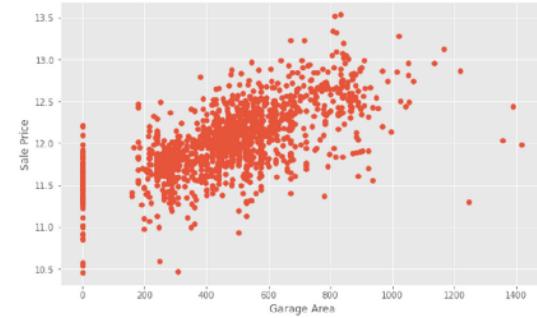
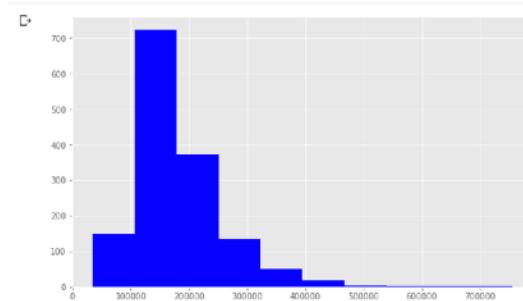
Getting to know your data

Exploratory Data Analysis

- Analyze your data to summarize their main characteristics
- Examples include: check for basic statistics (e.g. mean, median), missing data, outliers

```
df[ "SalePrice" ].describe()

count      1460.000000
mean     180921.195890
std      79442.502883
min     34900.000000
25%    129975.000000
50%    163000.000000
75%    214000.000000
max     755000.000000
Name: SalePrice, dtype: float64
```



Preparing your data

Feature engineering

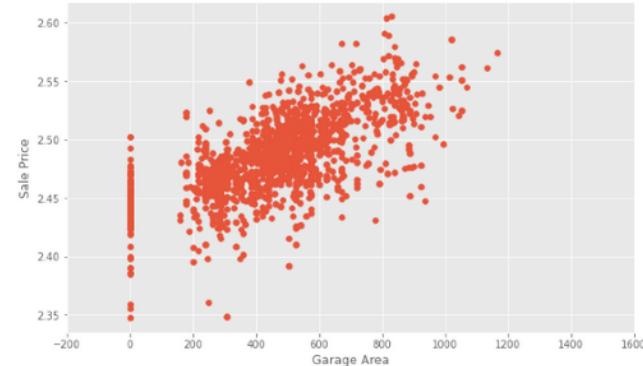
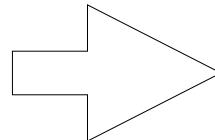
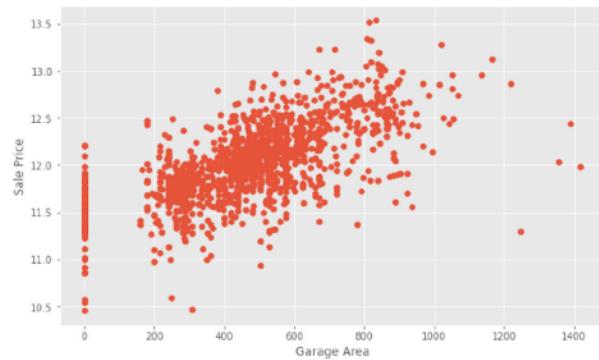
- Handling categorical data

	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour
Id								
1	60	RL	65.0	8450	Pave	NaN	Reg	Low
2	20	RL	80.0	9600	Pave	NaN	Reg	Low
3	60	RL	68.0	11250	Pave	NaN	IR1	Low
4	70	RL	60.0	9550	Pave	NaN	IR1	Low
5	60	RL	84.0	14260	Pave	NaN	IR1	Low

Preparing your data

Feature engineering

- Handling outliers



Step #3

Train your model

Step #3

Train your model

You need to know:

- What is a **feature**
- What is a **model**
- How does the **training** process work
- How **loss** helps our model to get better
- How **evaluation metrics** help us know if our model is good enough

Features

Examples of features for house price prediction

A	B	C	D	E	F	G	H	I
Id	LotArea	Utilities	LotConfig	HouseStyle	YearBuilt	Neighborhood	RoofStyle	SalePrice
1	8450	AllPub	Inside	2Story	2003	CollgCr	Gable	208500
2	9600	AllPub	FR2	1Story	1976	Veenker	Gable	181500
3	11250	AllPub	Inside	2Story	2001	CollgCr	Gable	223500
4	9550	AllPub	Corner	2Story	1915	Crawfor	Gable	140000
5	14260	AllPub	FR2	2Story	2000	NoRidge	Gable	250000
6	14115	AllPub	Inside	1.5Fin	1993	Mitchel	Gable	143000
7	10084	AllPub	Inside	1Story	2004	Somerst	Gable	307000
8	10382	AllPub	Corner	2Story	1973	NWAmes	Gable	200000
9	6120	AllPub	Inside	1.5Fin	1931	OldTown	Gable	129900
10	7420	AllPub	Corner	1.5Unf	1939	BrkSide	Gable	118000
11	11200	AllPub	Inside	1Story	1965	Sawyer	Hip	129500
12	11924	AllPub	Inside	2Story	2005	NridgHt	Hip	345000
13	12968	AllPub	Inside	1Story	1962	Sawyer	Hip	144000
14	10652	AllPub	Inside	1Story	2006	CollgCr	Gable	279500
15	10920	AllPub	Corner	1Story	1960	NAmes	Hip	157000
16	6120	AllPub	Corner	1.5Unf	1929	BrkSide	Gable	132000
17	11241	AllPub	CulDSac	1Story	1970	NAmes	Gable	149000
18	10791	AllPub	Inside	1Story	1967	Sawyer	Gable	90000
19	13695	AllPub	Inside	1Story	2004	SawyerW	Gable	159000
20	7560	AllPub	Inside	1Story	1958	NAmes	Hip	139000
21	14215	AllPub	Corner	2Story	2005	NridgHt	Gable	325300

Features

Examples of features for house price prediction

We want to predict this...

A	B	C	D	E	F	G	H	I
Id	LotArea	Utilities	LotConfig	HouseStyle	YearBuilt	Neighborhood	RoofStyle	SalePrice
1	8450	AllPub	Inside	2Story	2003	CollgCr	Gable	208500
2	9600	AllPub	FR2	1Story	1976	Veenker	Gable	181500
3	11250	AllPub	Inside	2Story	2001	CollgCr	Gable	223500
4	9550	AllPub	Corner	2Story	1915	Crawfor	Gable	140000
5	14260	AllPub	FR2	2Story	2000	NoRidge	Gable	250000
6	14115	AllPub	Inside	1.5Fin	1993	Mitchel	Gable	143000
7	10084	AllPub	Inside	1Story	2004	Somerst	Gable	307000
8	10382	AllPub	Corner	2Story	1973	NWAmes	Gable	200000
9	6120	AllPub	Inside	1.5Fin	1931	OldTown	Gable	129900
10	7420	AllPub	Corner	1.5Unf	1939	BrkSide	Gable	118000
11	11200	AllPub	Inside	1Story	1965	Sawyer	Hip	129500
12	11924	AllPub	Inside	2Story	2005	NridgHt	Hip	345000
13	12968	AllPub	Inside	1Story	1962	Sawyer	Hip	144000
14	10652	AllPub	Inside	1Story	2006	CollgCr	Gable	279500
15	10920	AllPub	Corner	1Story	1960	NAmes	Hip	157000
16	6120	AllPub	Corner	1.5Unf	1929	BrkSide	Gable	132000
17	11241	AllPub	CulDSac	1Story	1970	NAmes	Gable	149000
18	10791	AllPub	Inside	1Story	1967	Sawyer	Gable	90000
19	13695	AllPub	Inside	1Story	2004	SawyerW	Gable	159000
20	7560	AllPub	Inside	1Story	1958	NAmes	Hip	139000
21	14215	AllPub	Corner	2Story	2005	NridgHt	Gable	325300

Features

Examples of features for house price prediction

We want to predict this...

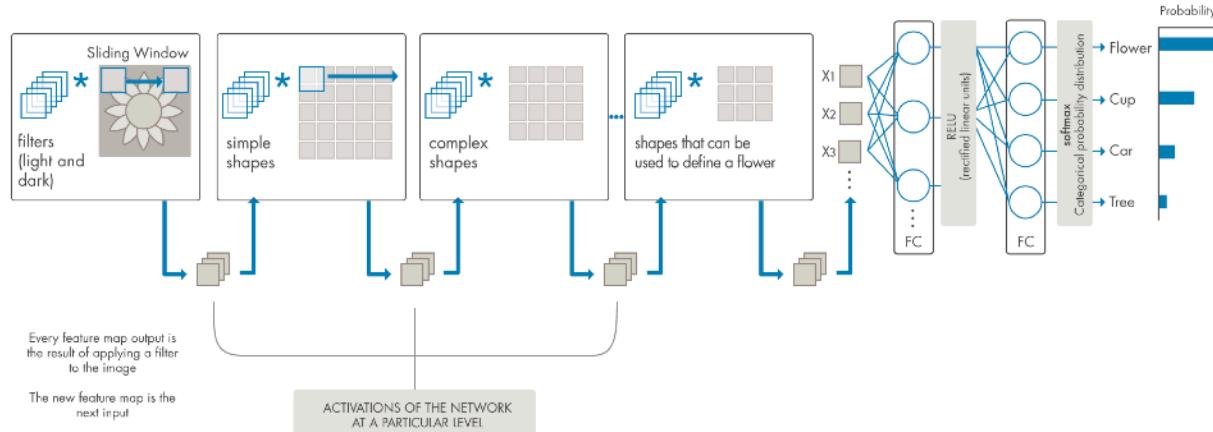
...using these features

A	B	C	D	E	F	G	H	I
Id	LotArea	Utilities	LotConfig	HouseStyle	YearBuilt	Neighborhood	RoofStyle	SalePrice
1	8450	AllPub	Inside	2Story	2003	CollCr	Gable	208500
2	9600	AllPub	FR2	1Story	1976	Veenker	Gable	181500
3	11250	AllPub	Inside	2Story	2001	CollCr	Gable	223500
4	9550	AllPub	Corner	2Story	1915	Crawfor	Gable	140000
5	14260	AllPub	FR2	2Story	2000	NoRidge	Gable	250000
6	14115	AllPub	Inside	1.5Fin	1993	Mitchel	Gable	143000
7	10084	AllPub	Inside	1Story	2004	Somerst	Gable	307000
8	10382	AllPub	Corner	2Story	1973	NWAmes	Gable	200000
9	6120	AllPub	Inside	1.5Fin	1931	OldTown	Gable	129900
10	7420	AllPub	Corner	1.5Unf	1939	BrkSide	Gable	118000
11	11200	AllPub	Inside	1Story	1965	Sawyer	Hip	129500
12	11924	AllPub	Inside	2Story	2005	NridgHt	Hip	345000
13	12968	AllPub	Inside	1Story	1962	Sawyer	Hip	144000
14	10652	AllPub	Inside	1Story	2006	CollCr	Gable	279500
15	10920	AllPub	Corner	1Story	1960	NAmes	Hip	157000
16	6120	AllPub	Corner	1.5Unf	1929	BrkSide	Gable	132000
17	11241	AllPub	CulDSac	1Story	1970	NAmes	Gable	149000
18	10791	AllPub	Inside	1Story	1967	Sawyer	Gable	90000
19	13695	AllPub	Inside	1Story	2004	SawyerW	Gable	159000
20	7560	AllPub	Inside	1Story	1958	NAmes	Hip	139000
21	14215	AllPub	Corner	2Story	2005	NridgHt	Gable	325300

What are the features for an image classification task?

Features

What are the features for an image classification task?



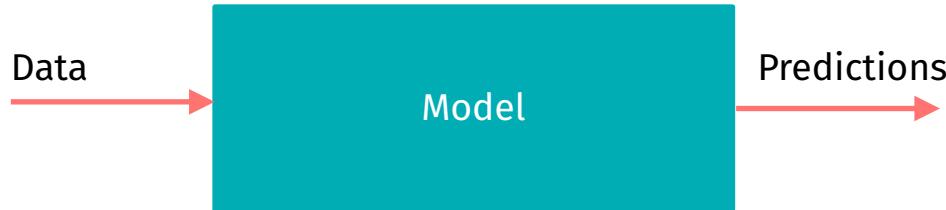
What is a model?

- A model maps examples to predicted labels
- It is defined by **weights that are learned during the training process**
- Once trained, you can use it to **make predictions about data that it has never seen before**

What is a model?

- There are many algorithms that you can use:
 - Linear regression
 - Logistic regression
 - Decision tree
 - Support Vector Machine (SVM)
 - Naive Bayes
 - kNN
 - ...

The training process



- Iteration 1: **2*number of floors + 3*area size = predicted house price**

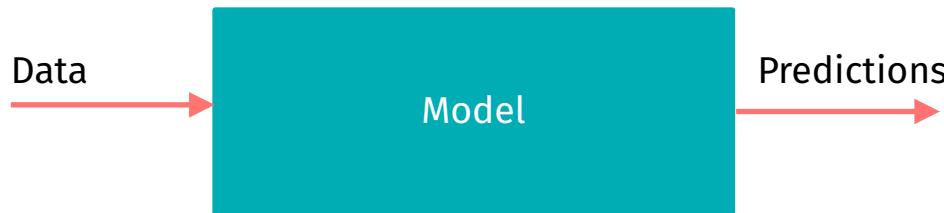
House #1:

predicted: 200 million

actual: 500 million

difference: 300 million

The training process



- Iteration 1: **2***number of floors + **3***area size = predicted house price
- Iteration 2: **4***number of floors + **6***area size = predicted house price

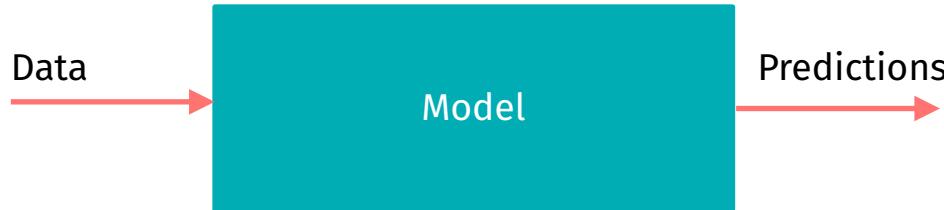
House #1:

predicted: 400 million

actual: 500 million

difference: 100 million

The training process



- Iteration 1: **2***number of floors + **3***area size = predicted house price
- Iteration 2: **4***number of floors + **6***area size = predicted house price

House #1:

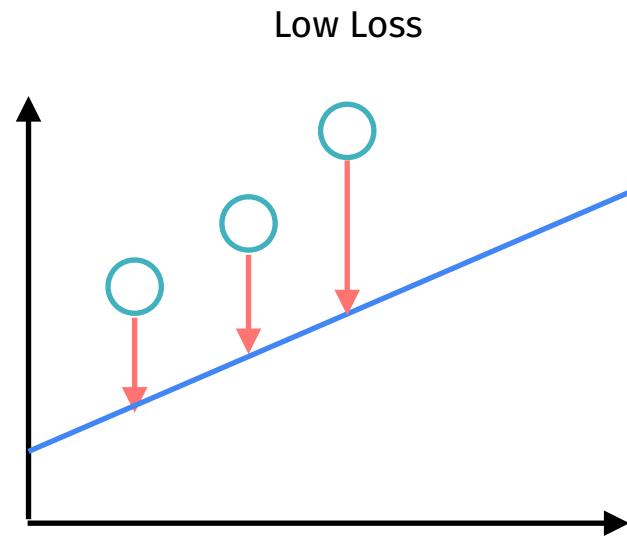
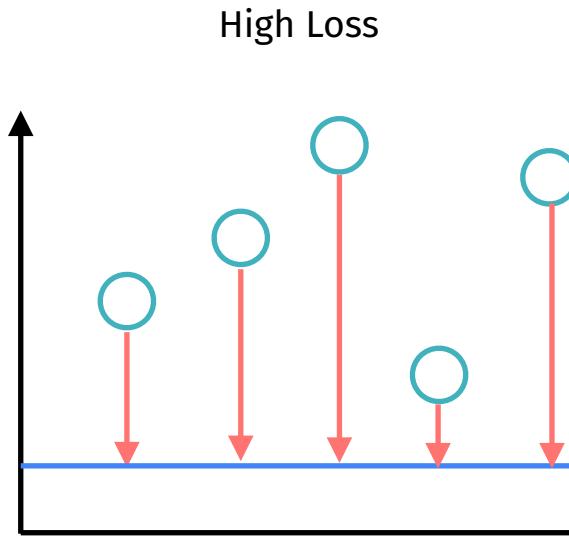
predicted: 400 million

actual: 500 million

difference: 100 million

Our model does not get smart right away - it needs to be “trained”

How loss helps our model get better



- Arrows represent loss
- Blue lines represent predictions

How evaluation metrics help us know that our model is good enough

- Evaluation metrics:
- Accuracy
- Mean Absolute Error
- Root Mean Squared Error
- ... and more

	Actual Spam	Actual Not Spam
Predicted Spam	15	10
Predicted Not Spam	5	30

Accuracy:

(Correctly classified spam emails + correctly classified not spam emails)/total emails = $(15 + 30) / (15+10+5+30) = 75\%$

Step #4

Use the model to make predictions

Tools & resources

Programming languages

- Python or R is usually the go-to programming language
- However, you can now train your own machine learning models using JavaScript thanks to TensorFlow.js

Libraries

- Data manipulation: numpy, pandas
- NLP: NLTK, spaCy
- Image processing: PIL, OpenCV
- Machine learning: scikit-learn, TensorFlow, TensorFlow Lite

Tools & resources

Teachable Machine

Teachable Machine

Train a computer to recognize your own images, sounds, & poses.

A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.

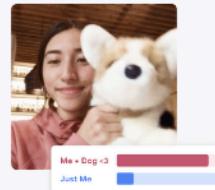
Get Started

TensorFlow.js p5.js Coral Node.js

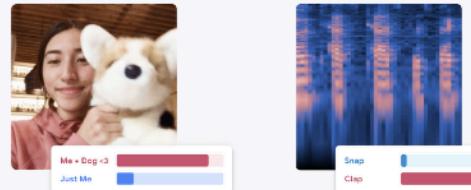
The screenshot shows a user interface for training a machine learning model. At the top, there's a blue button labeled "Get Started". Below it, there are icons for TensorFlow.js, p5.js, Coral, Node.js, and a flag icon. In the center, there's a progress bar with two categories: "Snap" at 48% and "Clap" at 0%. A dotted line leads from this bar to a small audio waveform icon. The background is white with some light gray shadows.

What can I use to teach it?

Teachable Machine is flexible – use files or capture examples live. It's respectful of the way you work. You can even choose to use it entirely on-device, without any webcam or microphone data leaving your computer.

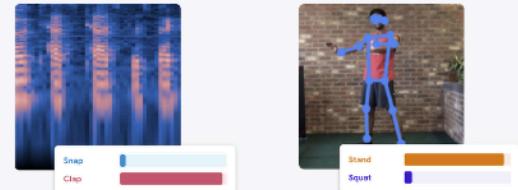


About



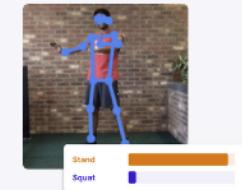
Images

Teach a model to classify images using files or your webcam.



Sounds

Teach a model to classify audio by recording short sound samples. (WAV/MP3/etc file support coming soon.)



Poses

Teach a model to classify body positions using files or striking poses in your webcam.

How do I use it?



1 Gather

Gather and group your examples into classes, or categories, that you want the computer to learn.

[Tutorial: Gather samples](#)



2 Train

Train your model, then instantly test it out to see whether it can correctly classify new examples.

[Tutorial: Train your model](#)



3 Export

Export your model for your projects: sites, apps, and more. You can download your model or host it online for free.

[Tutorial: Export your model](#)

@galuhsahid

Tools & resources

TensorFlow Hub

TensorFlow Hub is a library for reusable machine learning modules.

TensorFlow Hub is a library for the publication, discovery, and consumption of reusable parts of machine learning models. A **module** is a self-contained piece of a TensorFlow graph, along with its weights and assets, that can be reused across different tasks in a process known as transfer learning. Transfer learning can:

- Train a model with a smaller dataset,
- Improve generalization, and
- Speed up training.

[Browse modules on tfhub.dev](#)

The screenshot shows the TensorFlow Hub homepage. At the top, there's a navigation bar with links for Install, Learn, API, Resources (which is underlined), Community, and Why TensorFlow. Below the navigation is a header "Hub" with sub-links Overview, Guide, API, and Modules. The main content area has a heading "TensorFlow Hub is a library for reusable machine learning modules." followed by a detailed description of what TensorFlow Hub is and how it works. It includes a bulleted list of benefits and a link to browse modules on tfhub.dev. To the right, there's a code snippet for installing TensorFlow Hub via pip, followed by three cards: "Introducing TensorFlow Hub" (with a video thumbnail), "TensorFlow Hub at the Dev Summit" (with a video thumbnail), and "TensorFlow Hub on GitHub" (with a GitHub logo).

← mobilebert

Problem domain
Text embedding

Architecture
Transformer

Publisher
Tensorflow

Dataset
SQuAD

Language: English
Fine-tuned: No
License: Apache2.0

Last updated: 2020-04-23

Model formats

.JS (mobilebert/1) **TFLite (v1, default/1)**

Description

Pre-trained model optimized to work with TensorFlow.js
Please refer to the complete documentation here: <https://www.npmjs.com/package/@tensorflow-models/qna>

[Copy URL for TF.js](#) [Download Assets](#)

Overview

This model is a compressed version of BERT - MobileBERT - that runs 4x faster and has 4x smaller model size. BERT is a deep bidirectional transformer trained on Wikipedia and the BookCorpus. The details are described in the paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" [1].

MobileBERT is further trained on SQuAD, or Stanford Question Answering Dataset which is a reading comprehension dataset consisting of articles from Wikipedia and a set of question-answer pairs for each article.

The MobileBERT Q&A model takes a passage and a question as input, then returns a segment of the passage that most likely answers the question. It requires semi-complex pre-processing including tokenization and post-processing steps that are described in the BERT paper and implemented in the sample app.
To use the WordPiece tokenizer for pre-processing, you need the [vocabulary file](#).

@galuhsaид

Demo #1: Predicting house prices

<https://bit.ly/intro-ml-nusantech>

Demo #2: Image classification

Image classification

The screenshot shows a Teachable Machine project interface. At the top, there's a file navigation bar with options like "New File", "Share", and "Markdown". Below the file list, there's a preview window showing a person's hand making a rock gesture. The main area displays a bar chart with three categories: "Rock" at 32%, "Paper" at 1%, and "Scissors" at 67%. Below the chart, there's a note: "This machine learning model was made using Teachable Machine." and a link: "Learn how we made it without coding."

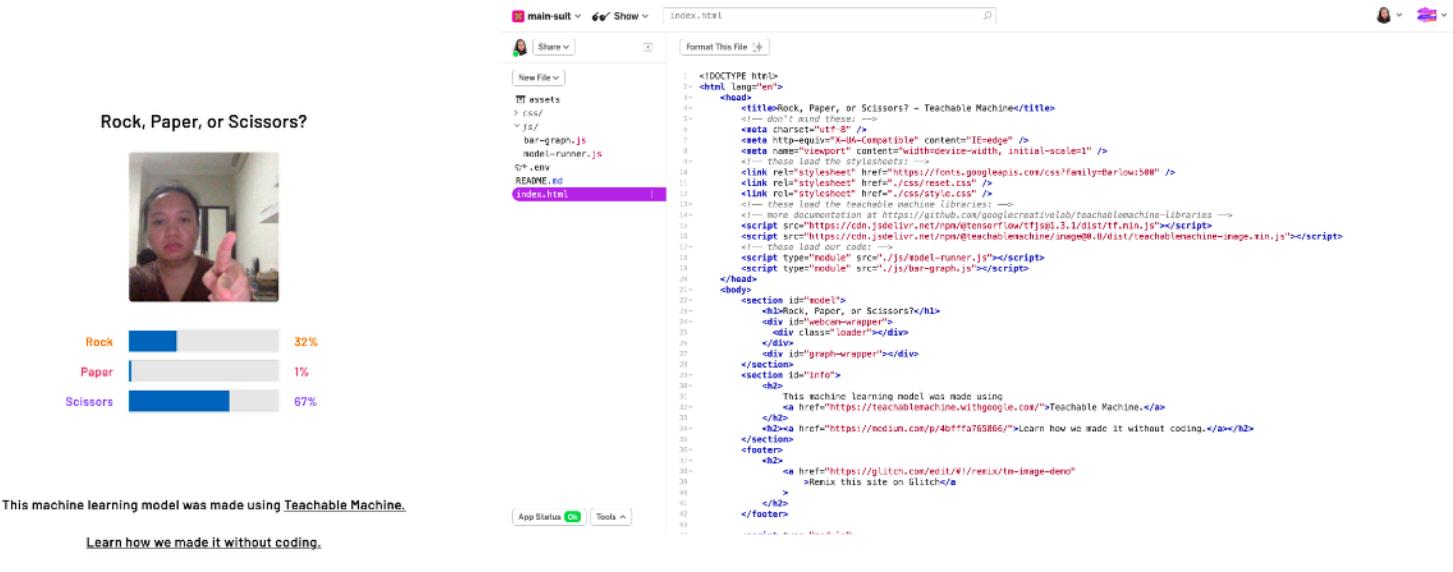
The screenshot shows the Teachable Machine training interface. It features three separate sections for "Rock", "Paper", and "Scissors", each with a grid of sample images and "Upload" buttons. A "Training" button is visible. To the right, there's a "Preview" section showing a video feed and a bar chart with the same 32%, 1%, and 67% results as the main interface. A note at the bottom says: "Kalau sudah, klik "Train". Jangan tinggalkan tab-mu selama proses pelatihan berlangsung, ya!"

Mengekspor model
Sebelum mengekspor model, kamu bisa mencoba-coba modelmu terlebih dahulu.

<https://main-suit.glitch.me>

@galuhsahid

Image classification



The screenshot shows a web application built with Teachable Machine. At the top, there's a header with the title "Rock, Paper, or Scissors?". Below the header is a video feed of a person's face, which is used as input for the model. To the right of the video feed is a bar chart showing the classification results:

Category	Percentage
Rock	32%
Paper	1%
Scissors	67%

At the bottom left, there's a note: "This machine learning model was made using [Teachable Machine](#)." Below that is another link: "Learn how we made it without coding." On the right side of the interface, there's a code editor window titled "index.html" showing the source code for the application.

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <title>Rock, Paper, or Scissors? - Teachable Machine</title>
    <meta charset="utf-8" />
    <meta http-equiv="X-UA-Compatible" content="IE=edge" />
    <meta name="viewport" content="width=device-width, initial-scale=1" />
    <!-- These load the stylesheets -->
    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Barlow:500" />
    <link rel="stylesheet" href="css/reset.css" />
    <link rel="stylesheet" href="css/style.css" />
    <!-- These load the teachable machine libraries -->
    <script src="https://cdn.jsdelivr.net/npm@teachablemachine@1.1/dist/tf.min.js"></script>
    <script src="https://cdn.jsdelivr.net/npm@teachablemachine@image@0.0.1/dist/teachablemachine-image.min.js"></script>
    <!-- These load our code! -->
    <script type="module" src="js/model-runner.js"></script>
    <script type="module" src="js/bar-graph.js"></script>
  </head>
  <body>
    <section id="model">
      <h1>Rock, Paper, or Scissors!</h1>
      <div id="webcam-wrapper">
        <div class="loader"></div>
        <div id="graph-wrapper"></div>
      </div>
      <div id="info">
        <p>This machine learning model was made using<br>
          <a href="https://teachablemachine.withgoogle.com/">Teachable Machine.</a>
        </p>
        <h2><a href="https://medium.com/p/4bffff7e5866/">Learn how we made it without coding.</a></h2>
      </div>
      <div id="glitch">
        <h3><a href="https://glitch.com/edit/#!/remix/tn-image-demo">Remix this site on Glitch</a></h3>
      </div>
    </section>
    <footer>
      <div>
        <img alt="Teachable Machine logo" />
        <small>A project by Google AI Research</small>
      </div>
    </footer>
  </body>
</html>
```

<https://main-suit.glitch.me>

References

More machine learning

- On building ML projects: [First Steps Towards Your First Machine Learning Project](#)
- On ML with JavaScript: [Machine Learning on the Web](#)
- On ML with TensorFlow: [A Whirlwind Tour of Machine Learning with TensorFlow](#)

Learning resources

- [Deep Learning with Python](#) (book) by François Chollet
- [Machine Learning Glossary](#)
- [Machine Learning Crash Course](#)
- [TensorFlow Tutorials](#)
- Teachable Machine Tutorials ([1](#), [2](#), [3](#))
- [But what is a neural network?](#) (video)

Thank you!

@galuhsahid