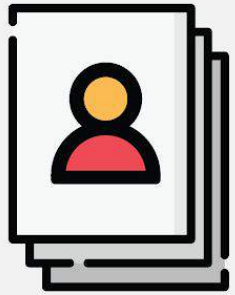
04 Supervised Learning from the Data Perspective

Topic Outline

- Types of Data
- Data in Different Use Cases
- Definition of Labelled Data
- How does a Machine Learn in Supervised Learning?

Types of Data

Numerical Data

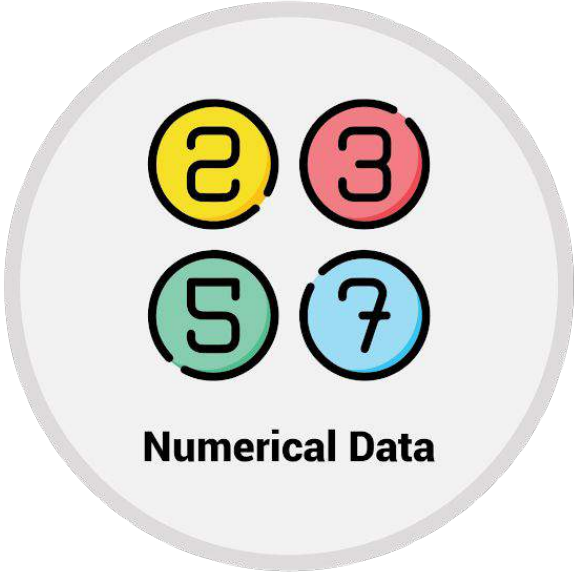


Images

Text

Sequential /
Time Series Data

Types of Data



Measurable attributes of an item. Normally stored as txt or csv format. People might refer to as csv data.

File ending format: csv, txt, xlsx, xlsm

Numerical Data Example

NovoExemplo - Bloco de notas

Ficheiro Editar Formatar Ver Ajuda

```
Exemplo
1  2  3
1  2  4
1  2  5
1  2  6
1  2  7
1  2  8
1  2  9
1  2  10
1  3  4
1  3  5
1  3  6
1  3  7
1  3  8
1  3  9
1  3  10
1  4  5
1  4  6
1  4  7
1  4  8
1  4  9
1  4  10
1  5  6
1  5  7
1  5  8
1  5  9
1  5  10
```

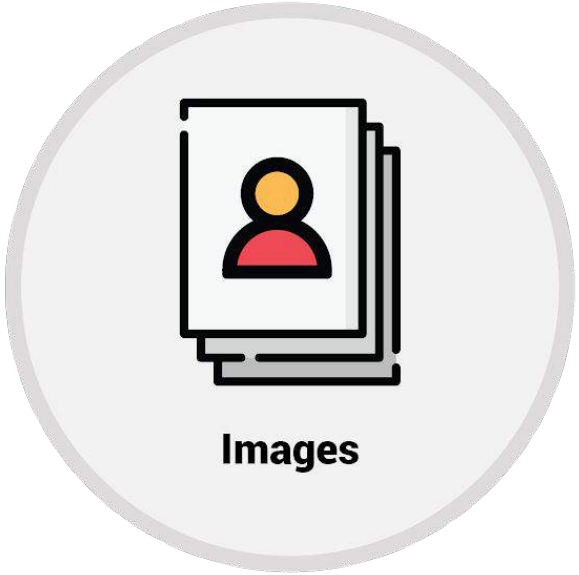
	A	B	C	D	E
1	Row I	Store	Product A	Product B	Product C
2	1	Store A	23	93	48
3	2	Store B	24	95	87
4	3	Store C	67	49	97
5	4	Store D	53	73	50
6	5	Store E	72	5	18
7	6	Store F	30	33	64
8	7	Store G	88	96	15
9	8	Store H	92	84	79
10	9	Store I	4	72	58
11	10	Store J	39	85	79
12	11	Store K	65	69	4
13	12	Store L	61	99	8
14	13	Store M	38	56	21
15	14	Store N	27	4	1
16	15	Store O	44	87	30
17	16	Store P	55	45	7
18	17	Store Q	23	13	11

csv_data_1.csv - Notepad

File Edit Format View Help

```
X,Y1,Y2,Y3
1.96,3.90,4.76,5.48
3.28,4.87,5.22,6.39
4.15,5.83,6.26,7.25
4.92,7.08,7.24,8.33
6.31,7.55,8.86,9.09
6.89,8.10,8.94,10.10
8.03,8.88,9.34,11.17
```

Types of Data



Abundant pictures about objects or scenes. Normally grey images is sufficient.

File ending format: png, jpeg, jpg, bmp, tif

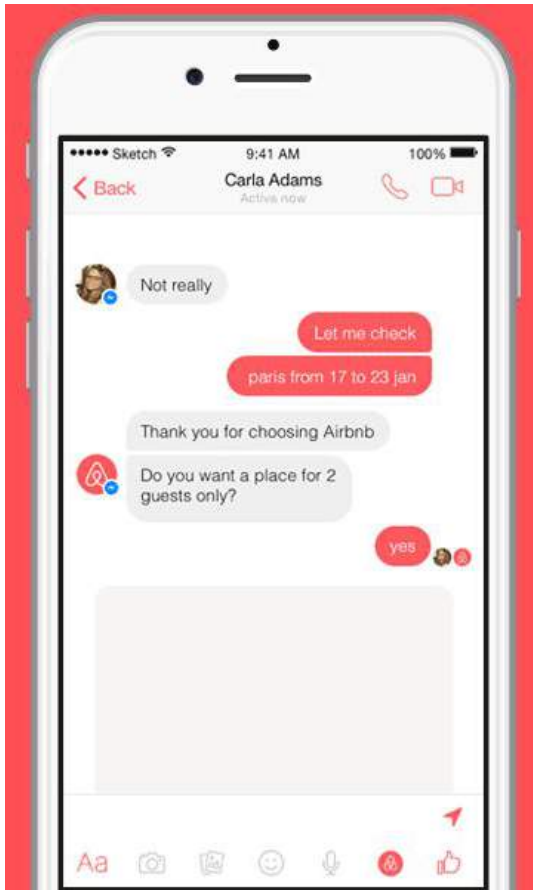
Images Data Example

Types of Data

Language-based
unstructured data.
Example: English,Malay,
and Spanish.

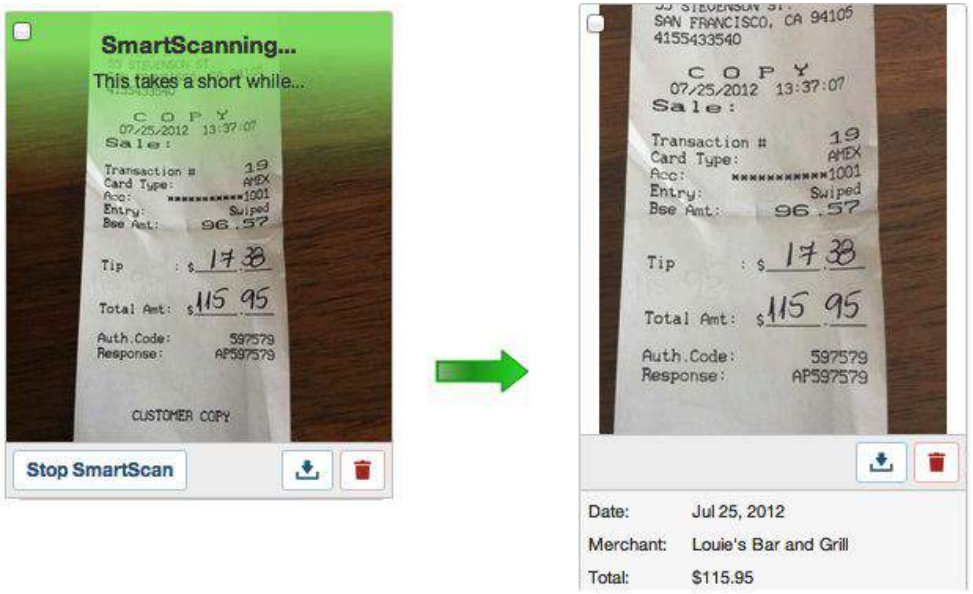
File ending format: csv, txt, doc, docx

Textual Data Example



Chatbot

Translation



Optical Character Recognition

https://gifer.com/en/ARix
https://mc.ai/parsing-in-tongues-neural-machine-translation/
https://www.pcworld.com/article/228342/expensify_receipt_scanning.html

Types of Data

Sequential / Time Series Data

Aforementioned data with the addition of time element.

Sequential / Time Series Data

Video (image frames stacked together)

Stock (numbered data stacked over time)

https://srirangatarun.wordpress.com/2018/07/09/video-frame-prediction-with-keras/
https://medium.com/@alon.lek/how-not-to-predict-stock-prices-2bf9da9ba91d

Data in Different Use Cases

Data types have to be aligned with use cases.

01 Medical	02 Finance	03 Construction	04 Education
Discovery of disease through image detection approach (*MRI, CT, and other image data*)	Analyse trend of graph automatically (*stock graph*)	Locate worker that violate the construction site rules (*workers image, workers ppe image*)	Auto generate set of examination question (*past examination questions*)
Discovery of disease through segmentation approach (*MRI, CT, and other image data*)	Algorithmic trading (*transaction data, customer data*)	Automatic construction cost estimating (*construction cost data, staff experience data*)	Automatic grading (*answer schema*)

Data in Different Use Cases

Data types have to be aligned with use cases.

05 Microchip

Locate chips inside the main board and define its types (*images of chip*)

Detect damaged chip during quality inspection (*image of damaged chip*)

06 Real Estate

Optical Character Recognition

1. *Computer Text ROI Generation*
2. *Computer Texts Labelling*
3. *Handwriting Texts Labelling*
4. *Handwriting Texts Generation*

07 Petroleum

Monitor the traffic of ships and condition of weather (*map of ship route, weather forecasting record*)

Automatic fuel quality inspection (*fuel data*)

08 Sports

Evaluate the performance of athlete (*image of people playing sports*)

Sports media content (*sports data*)

Data in Different Use Cases

Data types have to be aligned with use cases.

09 Agriculture

Monitoring the activities of agriculture.
(*image of agriculture activities*)

Monitor the growth of weed seedlings
(*soil data, weed images*)

10 Airlines

Face Detection
(*Asian*)

Gender Annotation
(*Face*)

Emotion Detection
(*Face*)

Gesture Detection
(*Hand signals*)

11 Faces

Detecting faces for devices unlock, attendance record, security screening
(*image of human faces*)

Automatic face editing
(*human faces images, skin images*)

12 Food Industry

Monitoring workers activities
(*image of normal factory activities*)

Diet plan recommendations
(*customer data, food data, food images*)

Data in Different Use Cases

Data types have to be aligned with use cases.

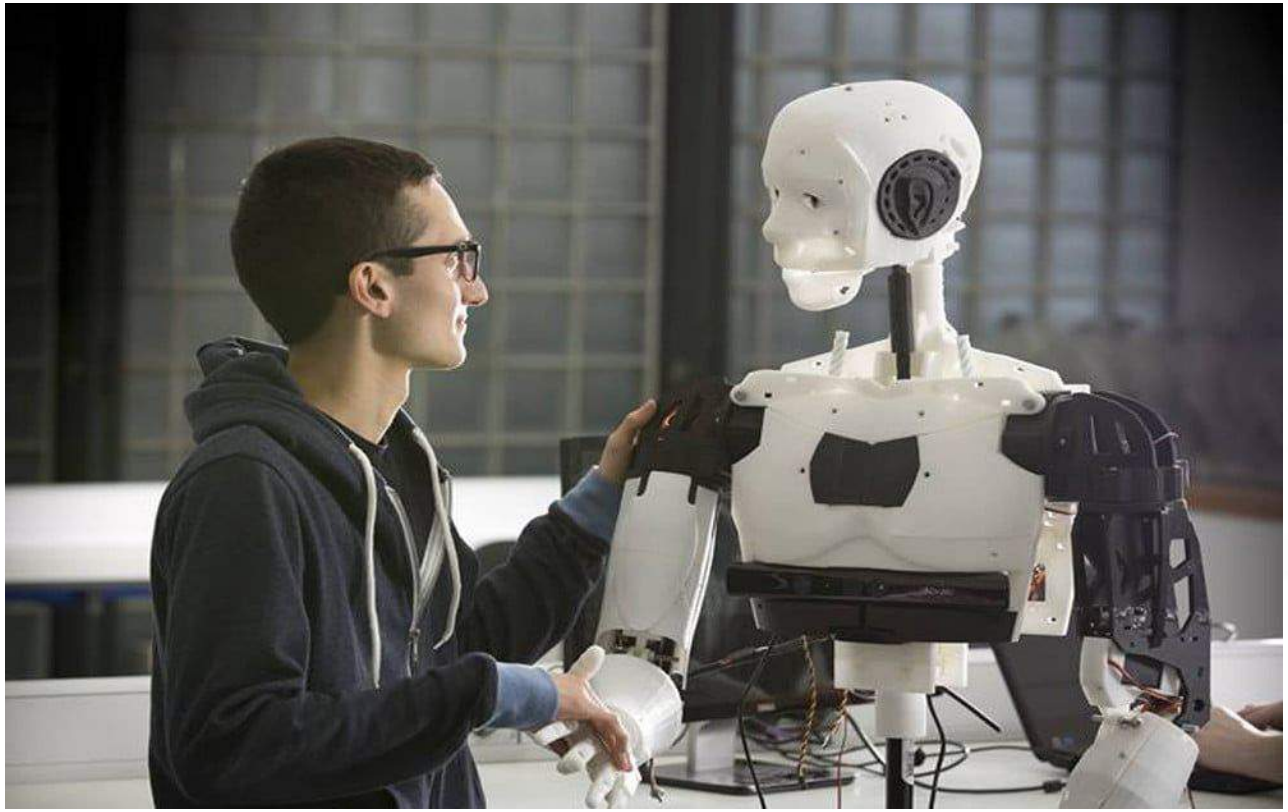
13 Manufacturing	14 Mining
Defect Detection (*Quality Inspection for Manufacturing Sector images*)	Monitoring workers activities and their safety (*image of normal workers activities*)
Predictive cost maintenance (*machines images and data*)	Automatic self driving trucks (*historical mining data, road images*)

What is Labelled Data?

Definition of Labelled Data

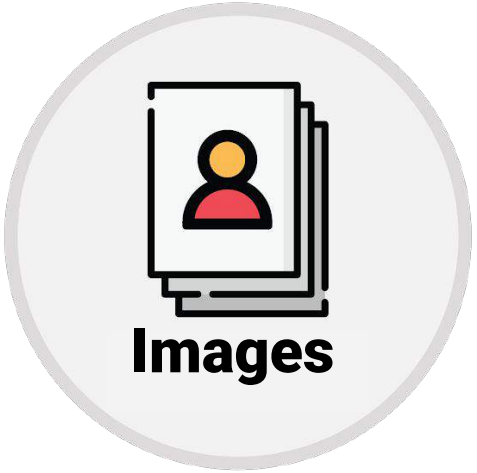
- **Data (of various forms) get marked up to show the object of interest**
- The object of interest is the target for the ML / DL algorithm to identify
    - Commonly comes with certain attributes, properties, characteristics

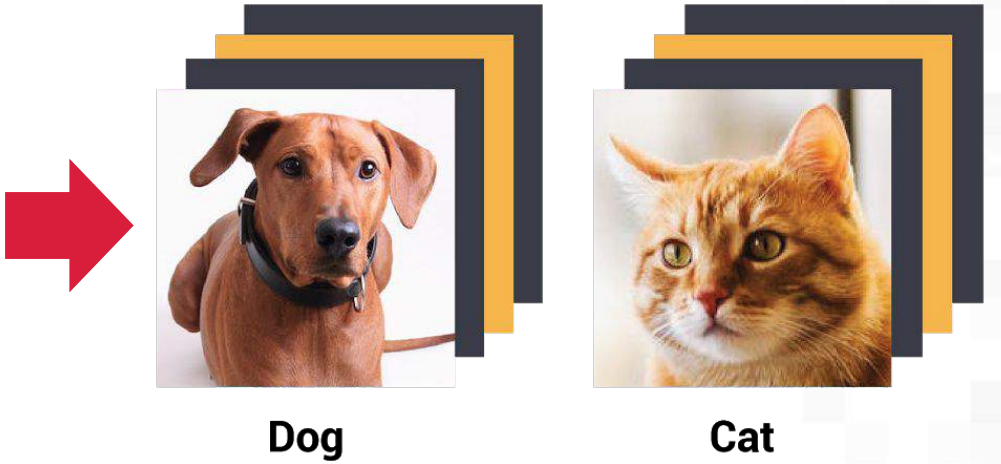
Think of it as:
Human highlighting the target of interest for algorithm to learn.



https://www.digitaltrends.com/cool-tech/research-shows-humans-like-robots-more-if-they-have-flaws/

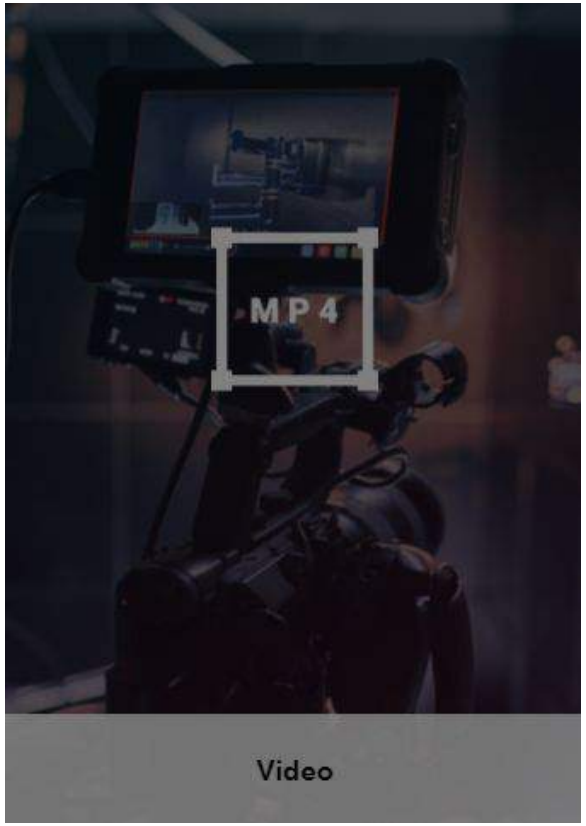
Most data does not come labelled...





Other Types of Annotated Data

Tabular



Video

Voice

How does a Machine Learn in Supervised Learning?

The Machine Learning/ Deep Learning Workflow

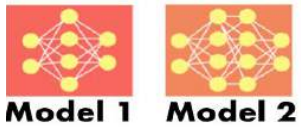
Validate data

Validate data healthiness

Preprocess data

Clean and preprocess data

Train models



Build models depends on

- Problem Domain
- Amount of Data

Decide whether to build

- Classification
- Regression
- Forecasting

Evaluate models

Monitor the performance of the model (Train the model if necessary)

Perform new hypothesis testing (Repeat the process from beginning again)

Deploy models

Apply the trained model to predict new data

Make prediction

Access model serving on cloud / on premise to get predictions

Monitor prediction

Monitor the production environment to ensure relevance to get accurate predictions

Topic Summary

- Each **use case** mandates the usage of **appropriate** and relevant **data types**.
- Most data are not labelled, so a **human** may facilitate the computer or **algorithm to understand** the data through **data annotation**.
- A **complete AI implementation with machine learning** would generally progress in the following **workflow** - *Validate data, preprocess data, train models, evaluate models, deploy models, make predictions and monitor predictions*.