# Ilhem Bekkar

ESI student – 3rd year

Main interest:
- Machine learning and deep learning problems
- Optimization
- Hardware performance

Samsung Innovation campus Alumni

# Contents

**01** Introduction
- ML definition
- ML algorithms

**02** ML project pipeline
- Data gathering
- Data preprocessing
- Data modeling

**03** Practice
- Understand data
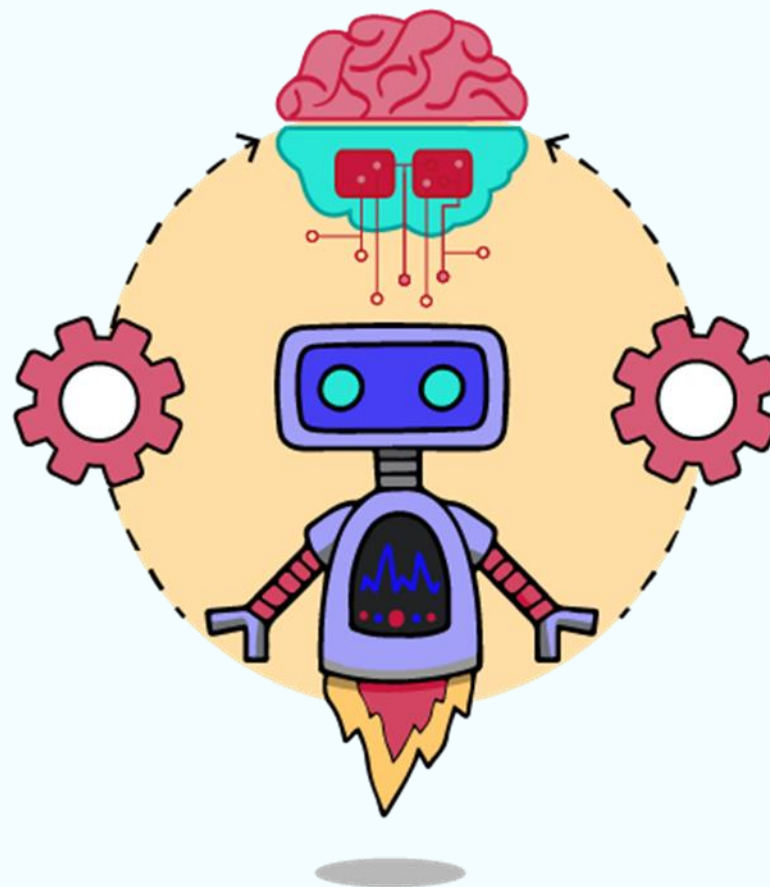- Data preprocessing
- Model building

# Contents

# What is Machine learning

Machine learning is a category of algorithm that allows systems to explore data in order to perform a specific task effectively without being explicitly programmed. Instead, it relies on patterns and relationships.
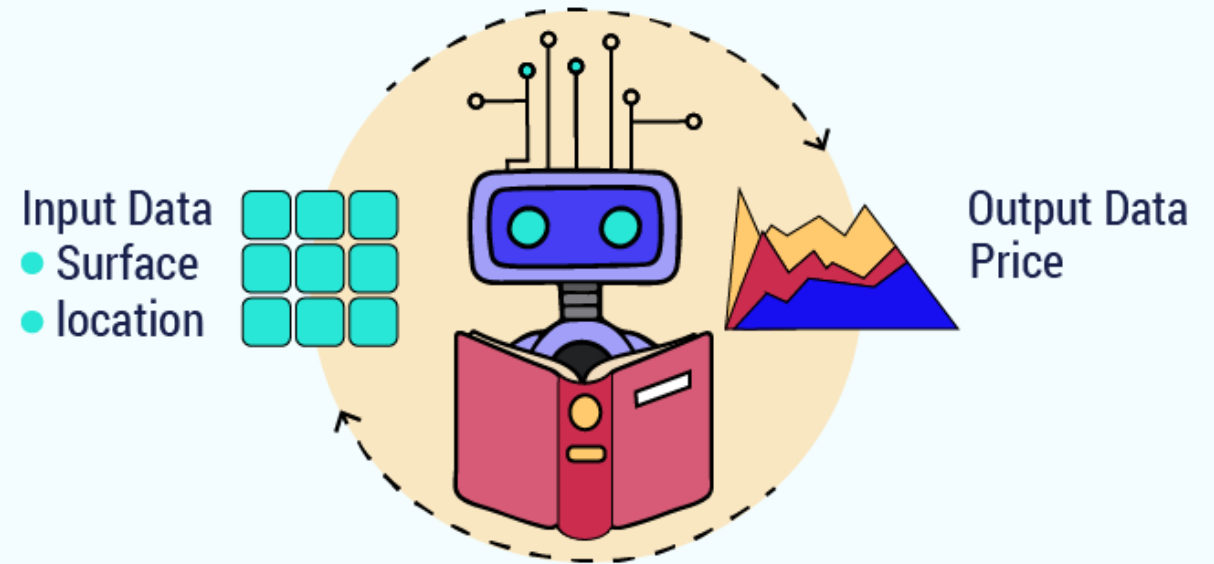


Machine Learning

# In machine learning:

- The programmer's job is to introduce to the Machine Learning algorithm historic data.
- The Machine Learning system will learn patterns from this historic data.

In the future, Machine Learning systems will use these patterns to make predictions on new data



## Machine Learning System

Input Data
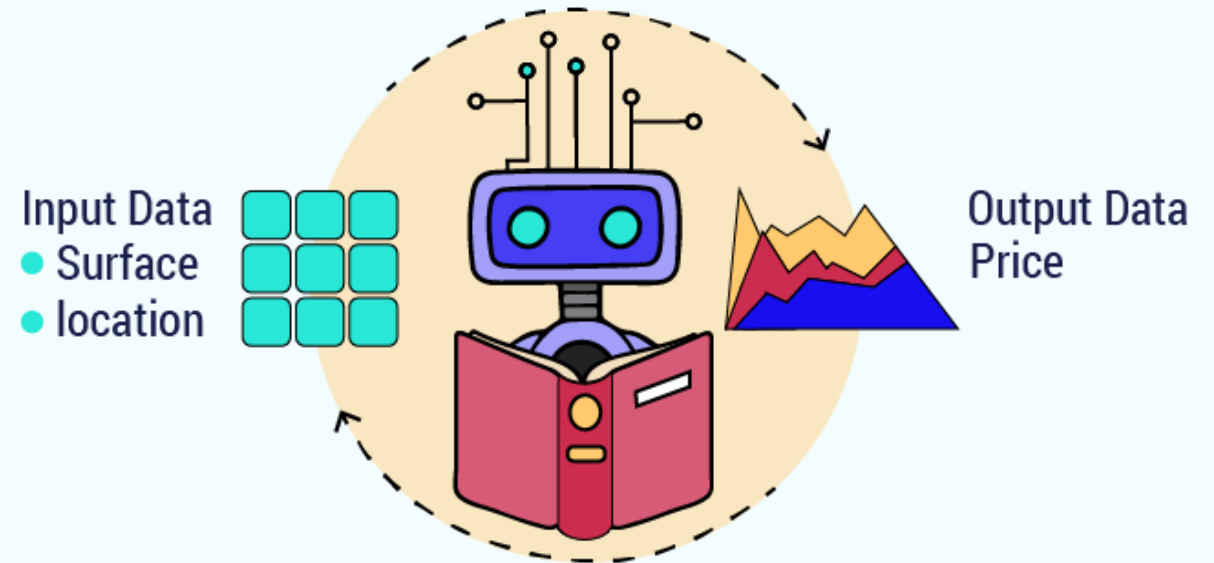- Surface
- location

Output Data
Price

# Machine learning exemple

Let's say we want to predict the price of a new house based on its surface and location.

During the learning phase, our Machine Learning algorithms' job is to find the relationship between the house characteristics and the price.

This search is based on millions of data sets containing house prices based on their surface and location
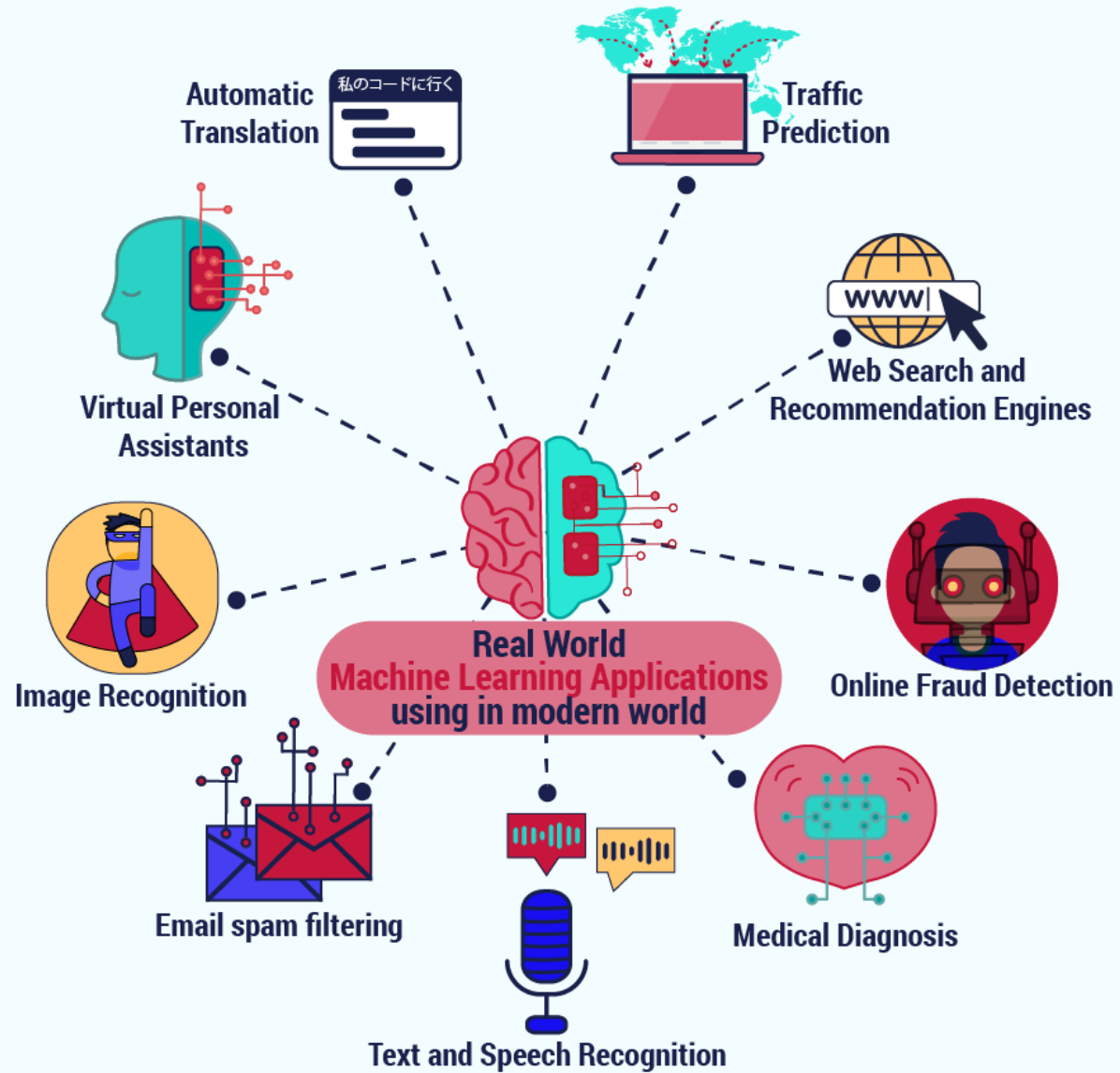


## Machine Learning System

Input Data
- Surface
- location

Output Data
Price

# Contents

Automatic Translation

Traffic Prediction

Virtual Personal Assistants

Web Search and Recommendation Engines

Image Recognition

Real World
Machine Learning Applications
using in modern world

Online Fraud Detection

Email spam filtering

Medical Diagnosis

Text and Speech Recognition

# Machine Learning process

**Step 1: Clarify the problem and constraints**

Step 2: Data Collection

Step 3: Data Exploration

Step 4: Data Cleaning

Step 5: Feature Engineering

Step 6: Model Selection

Step 7: Model Training & Evaluation

# Machine Learning process

Step 1: Clarify the problem and constraints

**Step 2: Data Collection**

Step 3: Data Exploration

Step 4: Data Cleaning

Step 5: Feature Engineering

Step 6: Model Selection

Step 7: Model Training & Evaluation

# Machine Learning process

Step 1: Clarify the problem and constraints
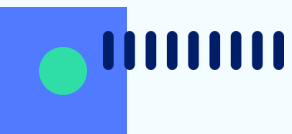
Step 2: Data Collection

**Step 3: Data Exploration**

Step 4: Data Cleaning

Step 5: Feature Engineering

Step 6: Model Selection

Step 7: Model Training & Evaluation

| DISTANCE (MI) | RATING |
|---|---|
| 1.5 | 3.6 |
| | |

# Machine Learning process

Step 1: Clarify the problem and constraints

Step 2: Data Collection

Step 3: Data Exploration

**Step 4: Data Cleaning**

Step 5: Feature Engineering

Step 6: Model Selection
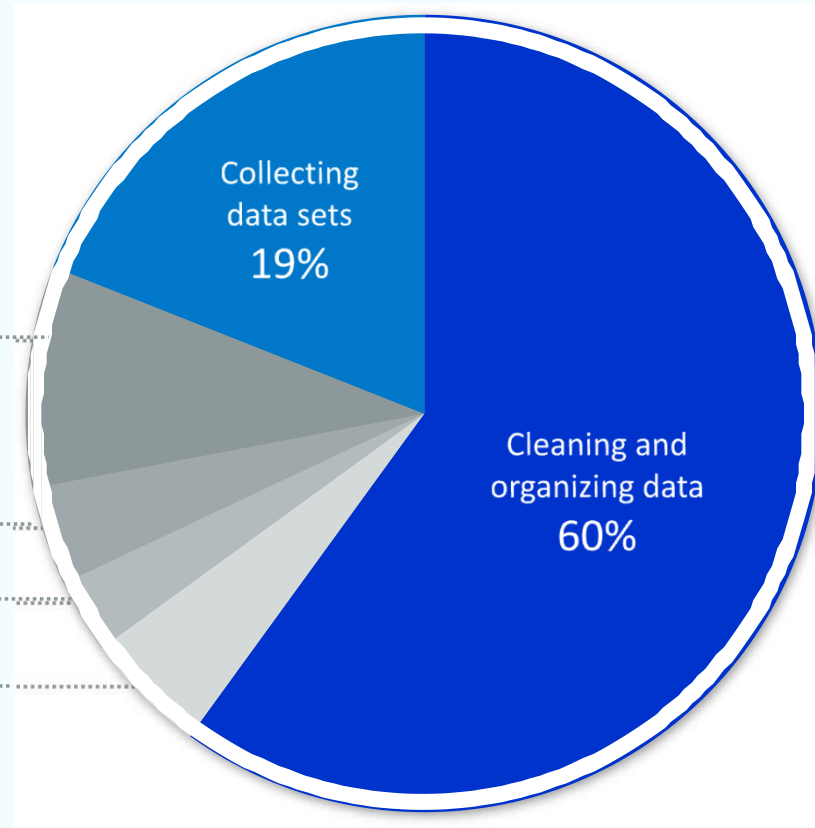
Step 7: Model Training & Evaluation

# What do data scientists spend the most time doing?



Mining data for patterns 9%

Refining algorithms 4%

Building training sets 3%

Other 5%

Collecting data sets 19%

Cleaning and organizing data 60%

> **"** A well-known fact in the data science community : **"**
> data scientists spend 60% of their time cleaning data.

```
import pandas as pd
df = pd.read_csv('train.csv', encoding = "ISO-8859-1")
df.head()
```

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 3 | Braund, Mr. Owen Harris | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikkinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 | S |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN | S |

If you have a question

# sa9ssi

# Machine Learning process

Step 1: Clarify the problem and constraints
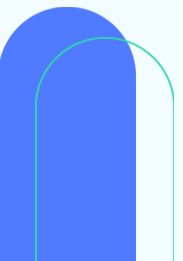
Step 2: Data Collection

Step 3: Data Exploration

Step 4: Data Cleaning

**Step 5: Feature Engineering**

Step 6: Model Selection

Step 7: Model Training & Evaluation

> "Feature engineering is the process of transforming the raw data to a representative dataset to your model in the best way to get the best results."

- Feature preprocessing

| DIST (MI) | RATING | PRICE ($) |
|-----------|--------|-----------|
| 0.8 | 2.7 | 147 |
| 1.5 | 3.6 | 136 |
| ... | ... | ... |
| 19.4 | 4.8 | 209 |

> **Feature engineering is the process of transforming the raw data to a representative dataset to your model in the best way to get the best results.**

- Feature selection

FEATURES

TARGET

| DISTANCE (MI) | RATING |
| --- | --- |

| PRICE ($) |
| --- |

# Machine Learning process

Step 1: Clarify the problem and constraints
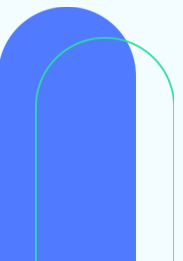
Step 2: Data Collection

Step 3: Data Exploration

Step 4: Data Cleaning

Step 5: Feature Engineering

**Step 6: Model Selection**

Step 7: Model Training & Evaluation

# Machine Learning process

Step 1: Clarify the problem and constraints

Step 2: Data Collection

Step 3: Data Exploration

Step 4: Data Cleaning

Step 5: Feature Engineering

Step 6: Model Selection

**Step 7: Model Training & Evaluation**

TRAINING DATA

TEST DATA

TESTING

FEATURES

MODEL

TARGET

# Let's practice ! ✈

# Let's practice ! ✈

# Let's practice ! ✈

# Home work

🔗 Deadline | 17.05.2020 at 20:00

# Thanks !

Good night !

MACHINE LEARNING BOOTCAMP

Icon

You can resize these icons keeping the quality.

You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.

# Break Section

1 mn

# Sa7a Ftorkom