

# Overview of Engineering Analytics & Machine Learning (ECSE202)



# Subject Synopsis

This subject provides coverage of the knowledge of concepts and skills in the tools and processes used in data analytics of large amounts of engineering data. It encompasses the various stages of data analytics, from gathering the data, asking the right questions, to analyzing and interpreting data, identifying patterns and trends and making use of machine learning and predictive models to make intelligent and actionable recommendations for improvement in engineering systems.



# Subject Aims

This subject aims to equip students with the knowledge and skills to:

- apply the processes needed to perform analysis on engineering data.
- present results in the form of visualization.
- apply machine learning algorithms on engineering data.



# Subject Intended Learning Outcomes

1. Demonstrate basic application for data handling.
2. Perform data gathering and pre-processing.
3. Analyze data using visualization techniques.
4. Perform Machine Learning methods on a data set.

We will be using **Python** in this subject.

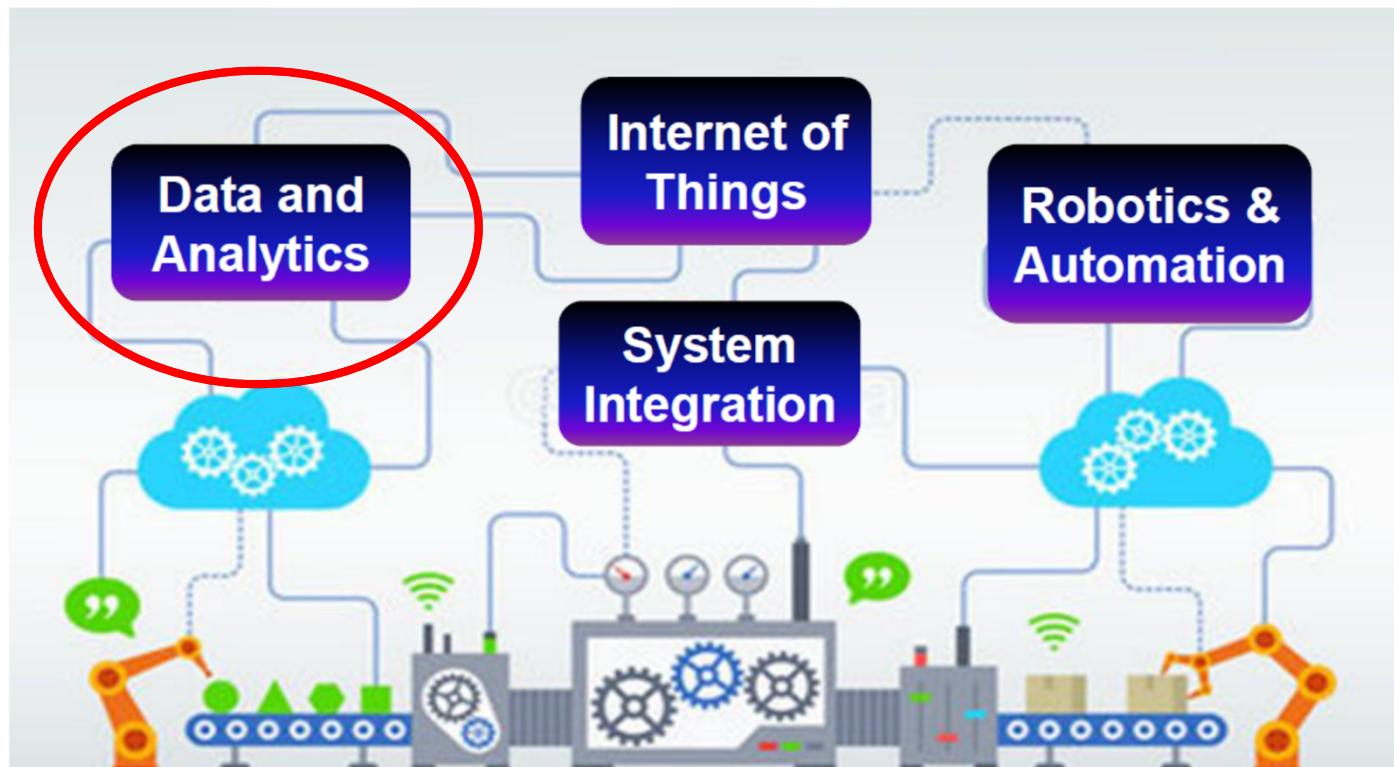


# Data Analytics Process & Motivation for Data Analytics



# Why Engineering Analytics?

Emerging Trends and Skills for the Electronics Industry



Source: Inputs from Electronics Workgroup and concurred by Electronics STC

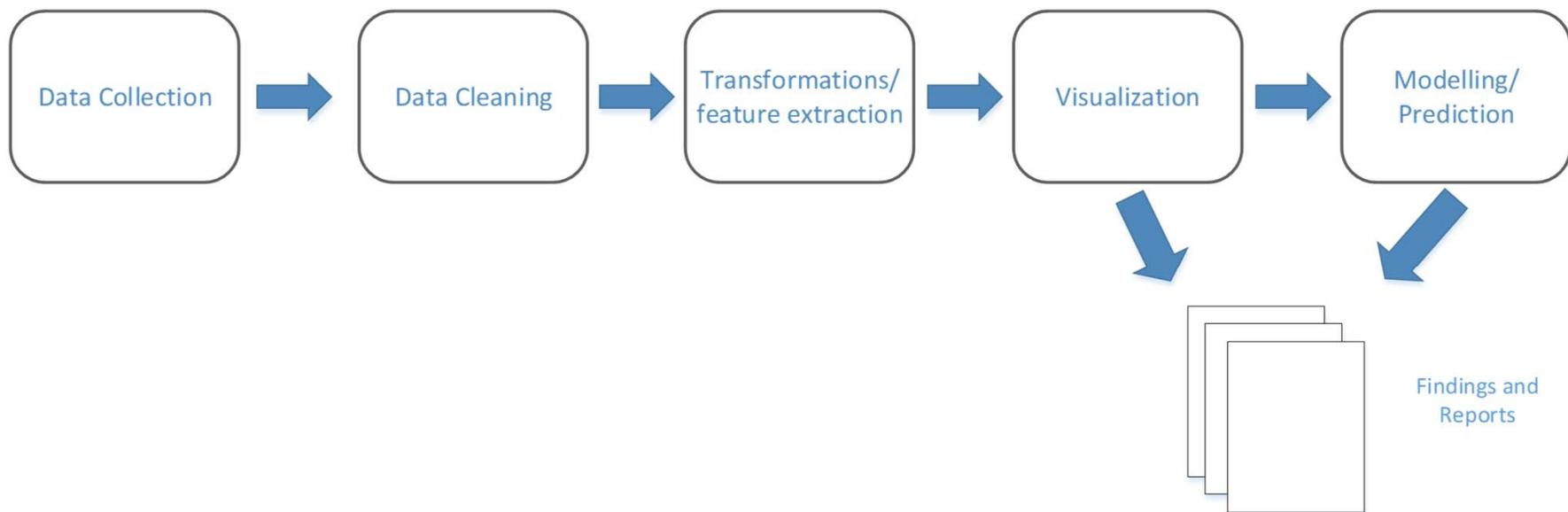
LEARNING FOR LIFE ADVANCING WITH SKILLS 24



# What is Data Analytics?

Data analytics is the process of **inspecting**, **cleansing**, **transforming** and **modelling** data with the goal of discovering useful information, suggesting conclusions and supporting decision making.

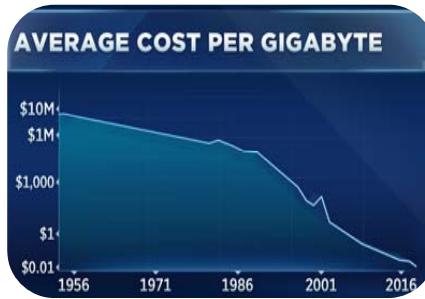
*Wikipedia*



# Data Analytics --- why now?



Easy access to high speed and cheap connection



Availability of cheap storage



Smart sensors and devices comes of age



Easily available and increasing computational power



Ever changing and rapidly business environment

# Types of Data Analytics

## Descriptive Analytics

- What happened?
- Uses historical data to give a perspective of what is going on

## Diagnostic Analytics

- Why did it happen?
- What and why it went wrong?

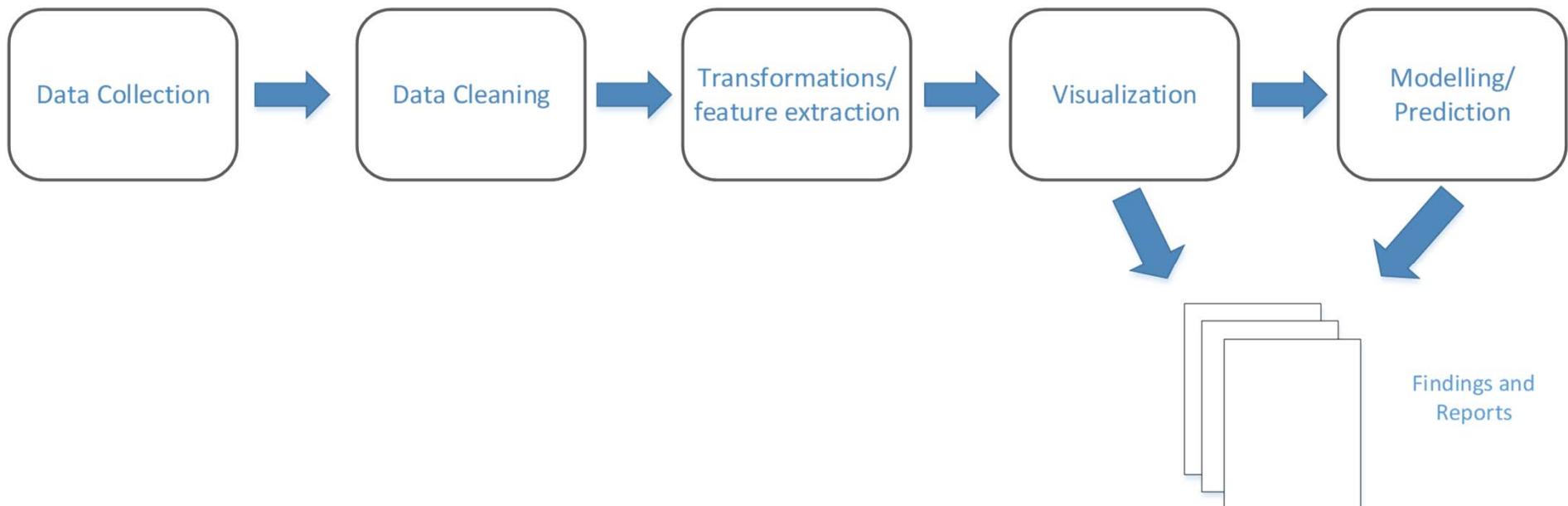
## Predictive Analytics

- What is going to happen?
- Building mathematical model using historical data

## Prescriptive Analytics

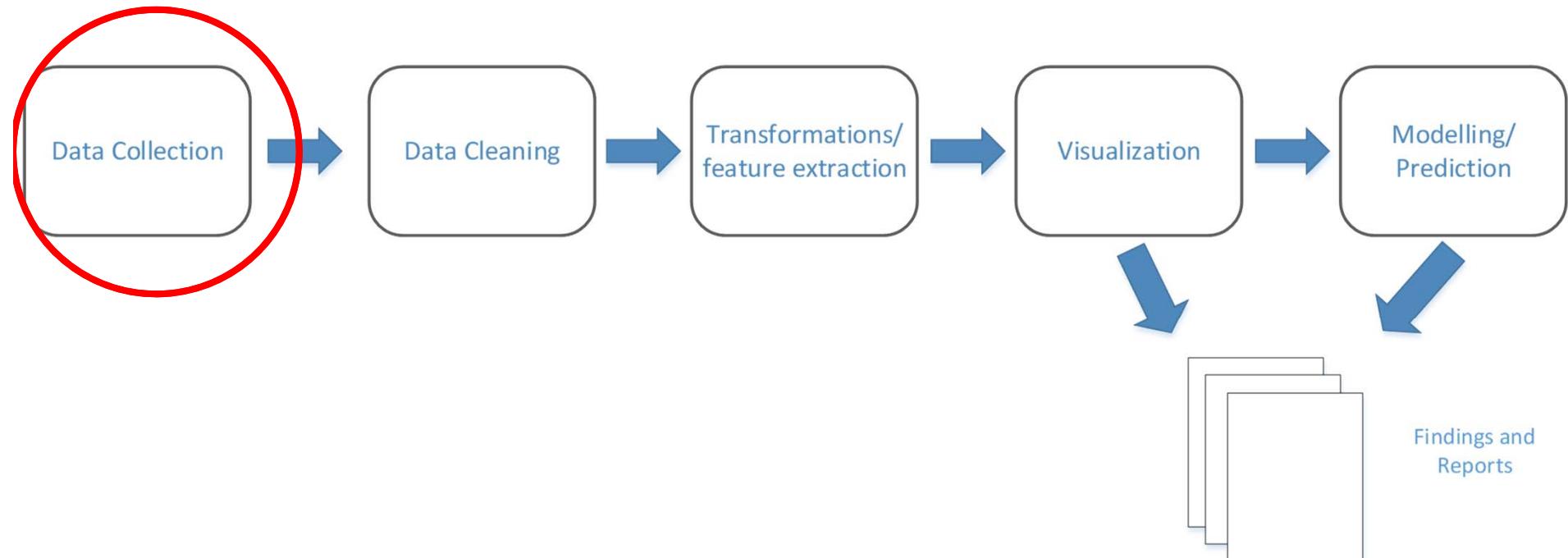
- What could we do about it?
- Find the best actions to achieve desired goal

# Data Analytics Process



The process can be iterative. After Visualization and Modelling, the data analyst can decide that more data needs to be collected or new features need to be extracted, and goes back to those stages.

# Data Collection



- It is the input of the Data Analytics process and solely responsible for the quality of the analysis.
- The data has to be reliable and quantifiable.

# Data Collection – Data Source

Different Types of Sensors...



Time Series, Unstructured, Noisy  
and prone to missing data



Structured and large

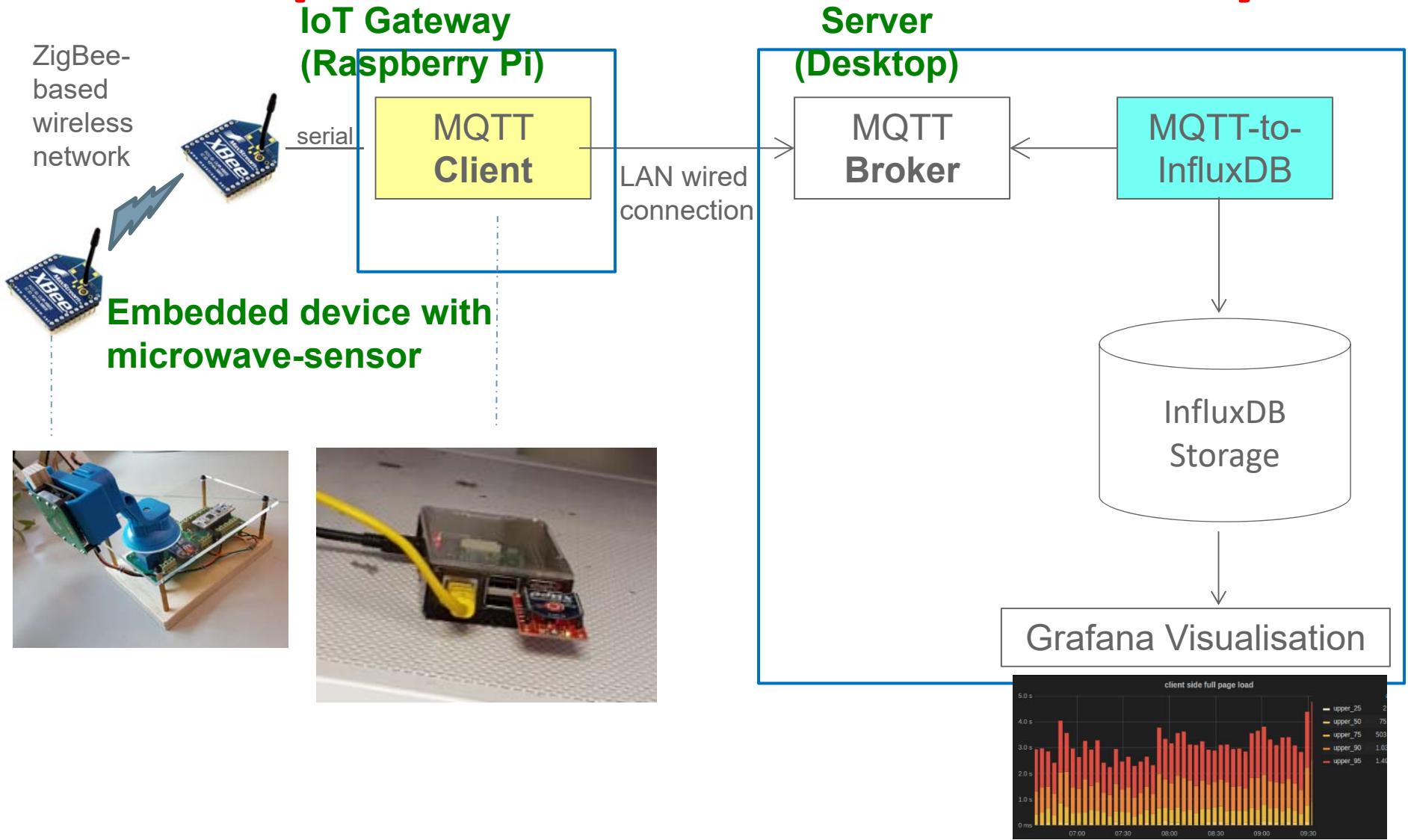


Structured/unstructured, involved using of  
API to access, real-time



Local, easy access but limited

# Example of Sensor Data: RemoteEye



# Data Cleaning

This is an important stage and is necessary because the data that is collected is often “dirty” due to:

- Missing data
- Unacceptable formats
- Erroneous values
- Data could be embedded inside text or other information and needs to be extracted
- Remove or compensate for outliers that skews the data unrealistically



# Data Cleaning

- Huge amount of effort is spent cleaning data to get it ready for analysis.
- It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data.
- Data preparation is not just the first step, but must be repeated over the course of analysis as new problems come to light or new data is collected.
- Data cleaning is sometimes painfully manual.

# Data Transformation / Feature Extraction

## Data Cleaning:

To address problems with the raw data, such as missing and erroneous values, data formatting issues etc.

## Data Transformation:

To enhance the data by changing the values and/or structure of the data and add additional features (feature extraction)



# Data Transformation / Feature Extraction

Some of the data transformation which will be covered in this subject include:

- Date Time processing
- Normalize
- Aggregate
- Generate New Attributes
- Select Attributes
- Filter Examples



# Applications of Data Analytics

- Industrial Internet of Things (IIOT) --- big data for manufacturing

<https://www.youtube.com/watch?v=sGXvZbo3l9I>

- Amazon Logistic Robot

<https://www.youtube.com/watch?v=cLVCGEmkJs0>



# Introduction to Machine Learning



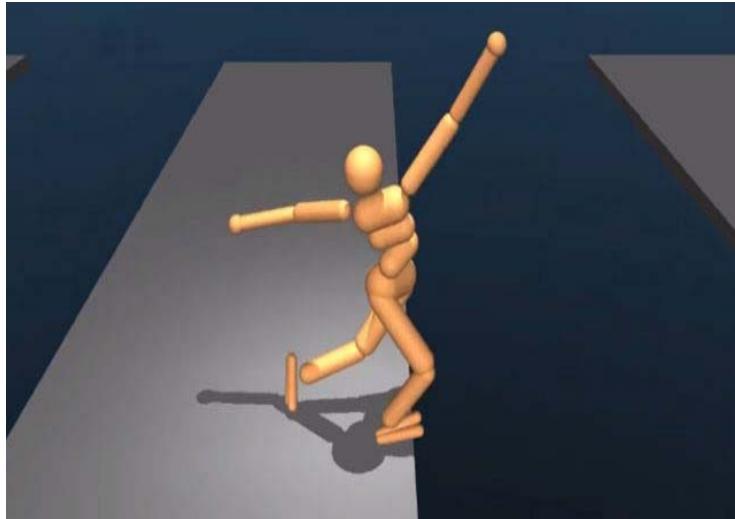
# FINANCIAL TIMES

Baidu offers open-source car software as lure for data

Move part of wider AI platform rollout to gain competitive edge with likes of Google



Source: <https://www.ft.com/content/55e79944-6168-11e7-91a7-502f7ee26895>

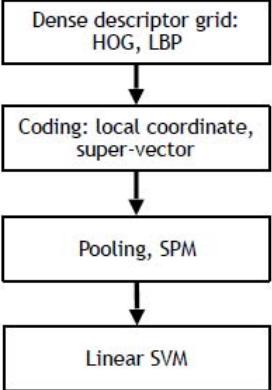


<https://www.youtube.com/watch?v=gn4nRCC9TwQ> Google AI walk

# IMAGENET Large Scale Visual Recognition Challenge

## Year 2010

NEC-UIUC

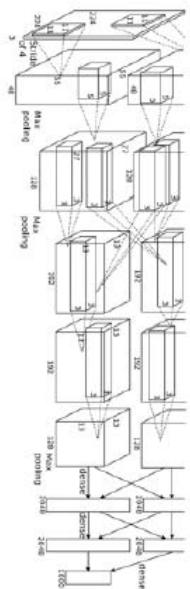


[Lin CVPR 2011]

[Lion image](#) by Swissfrog is licensed under CC BY 3.0

## Year 2012

SuperVision



[Krizhevsky NIPS 2012]

Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

## Year 2014

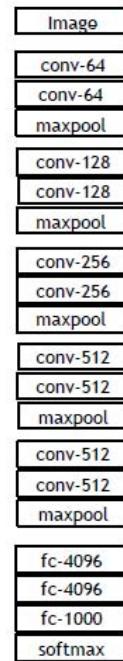
GoogLeNet

- Pooling
- Convolution
- Softmax
- Other



[Szegedy arxiv 2014]

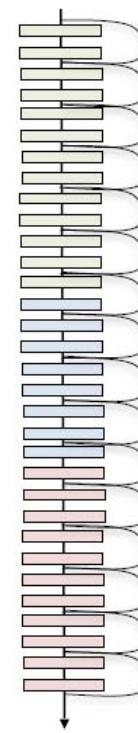
VGG



[Simonyan arxiv 2014]

## Year 2015

MSRA

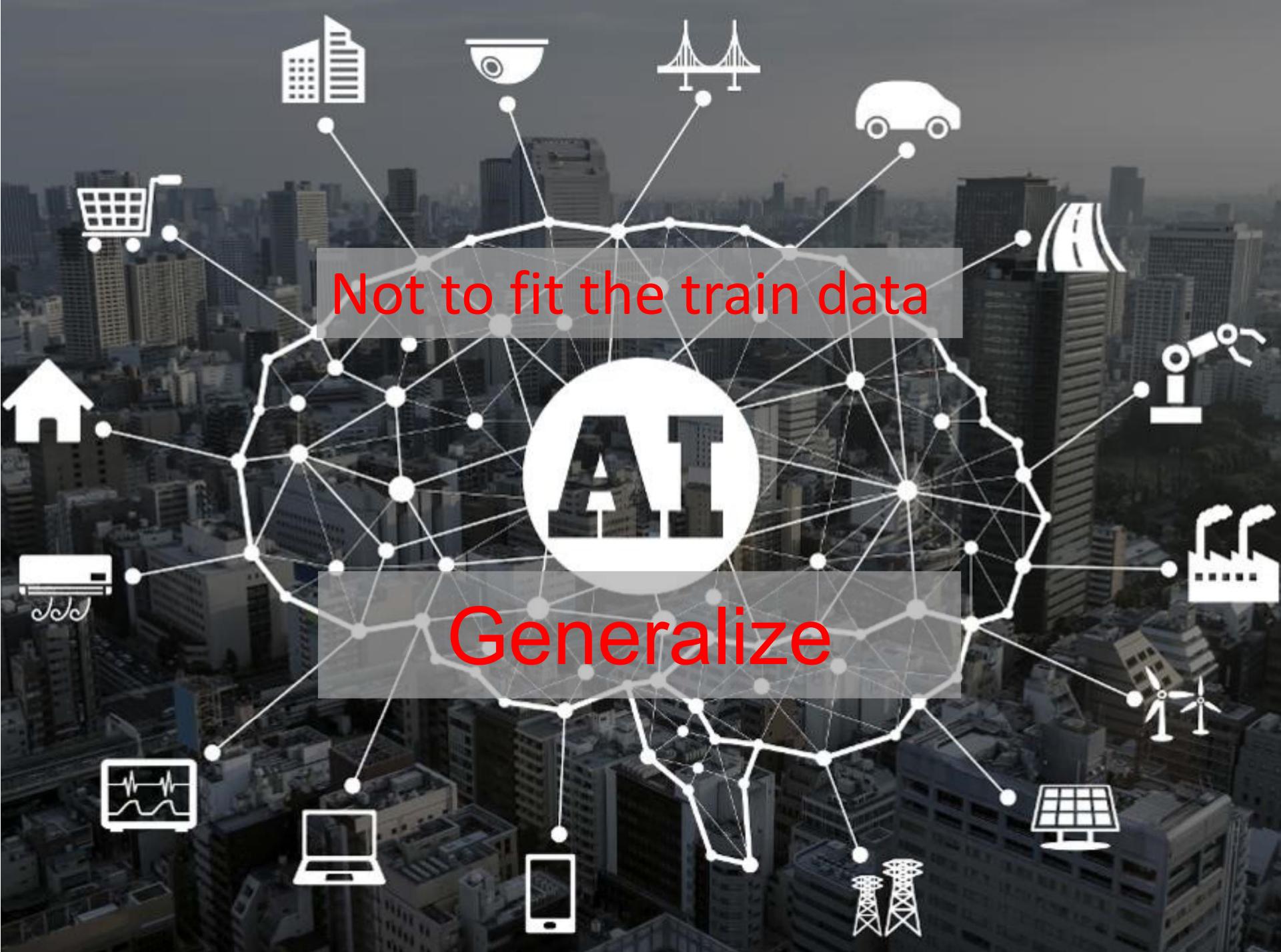


[He ICCV 2015]

# Human Learning



- Show sample of examples for learning
  - Not for the kid to be fixed to the material
  - Able to apply
  - Able to extrapolate
  - Able to adapt to similar cases
  - Able to innovate
-



Not to fit the train data

Generalize

# Different type of machine learning methods

	<b>Supervised Learning</b>	<b>Unsupervised Learning</b>
Characteristic	<ul style="list-style-type: none"><li>• Known number of classes</li><li>• Based on a training set with correct label</li><li>• Use to predict/classify future observations</li></ul>	<ul style="list-style-type: none"><li>• Number of classes unknown</li><li>• No prior knowledge</li><li>• Used to understand/explore data</li></ul>
Algorithms	<ul style="list-style-type: none"><li>• Linear Regression</li><li>• Logistic Regression</li><li>• SVM</li><li>• Decision Tree</li><li>• Deep Learning Neural Network</li></ul>	<ul style="list-style-type: none"><li>• PCA</li><li>• K-means</li><li>• Hidden Markov Model</li></ul>

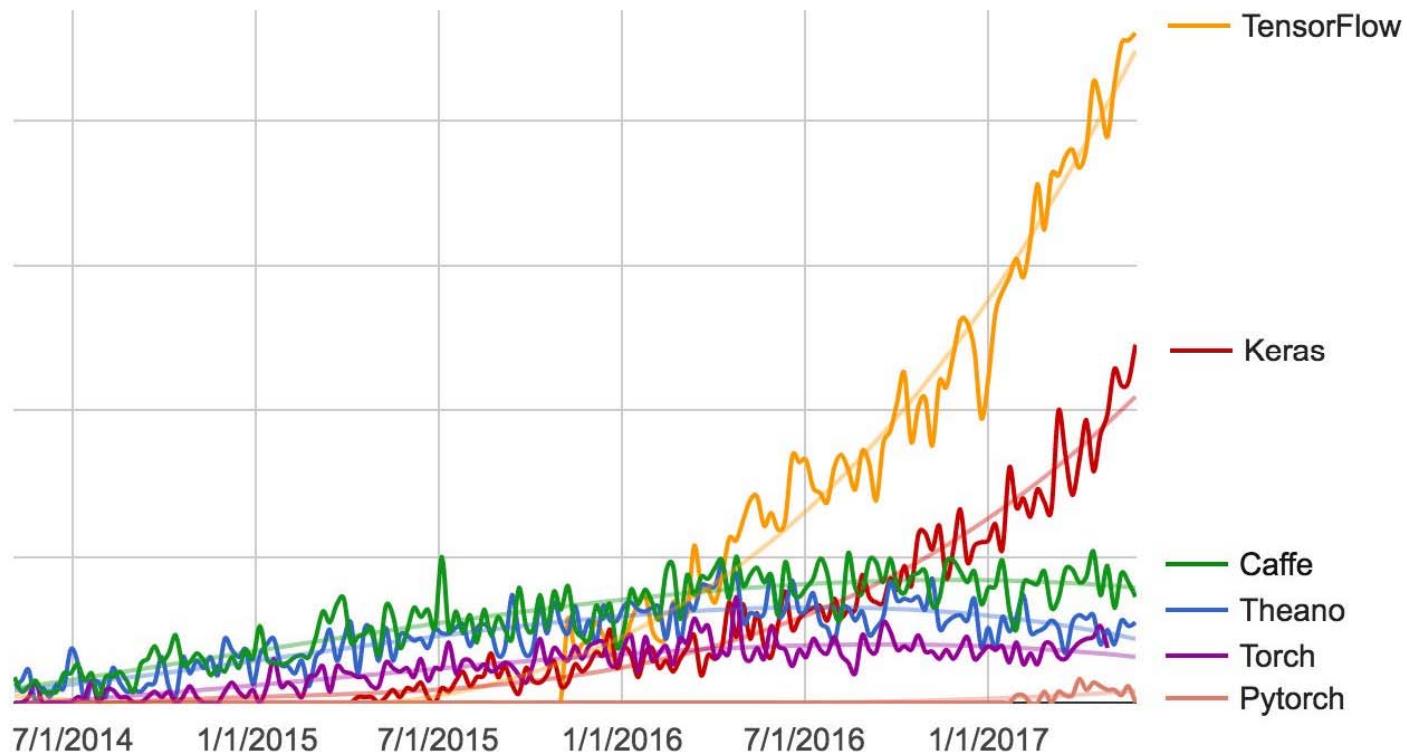
---

# Introduction to Python for Data Science

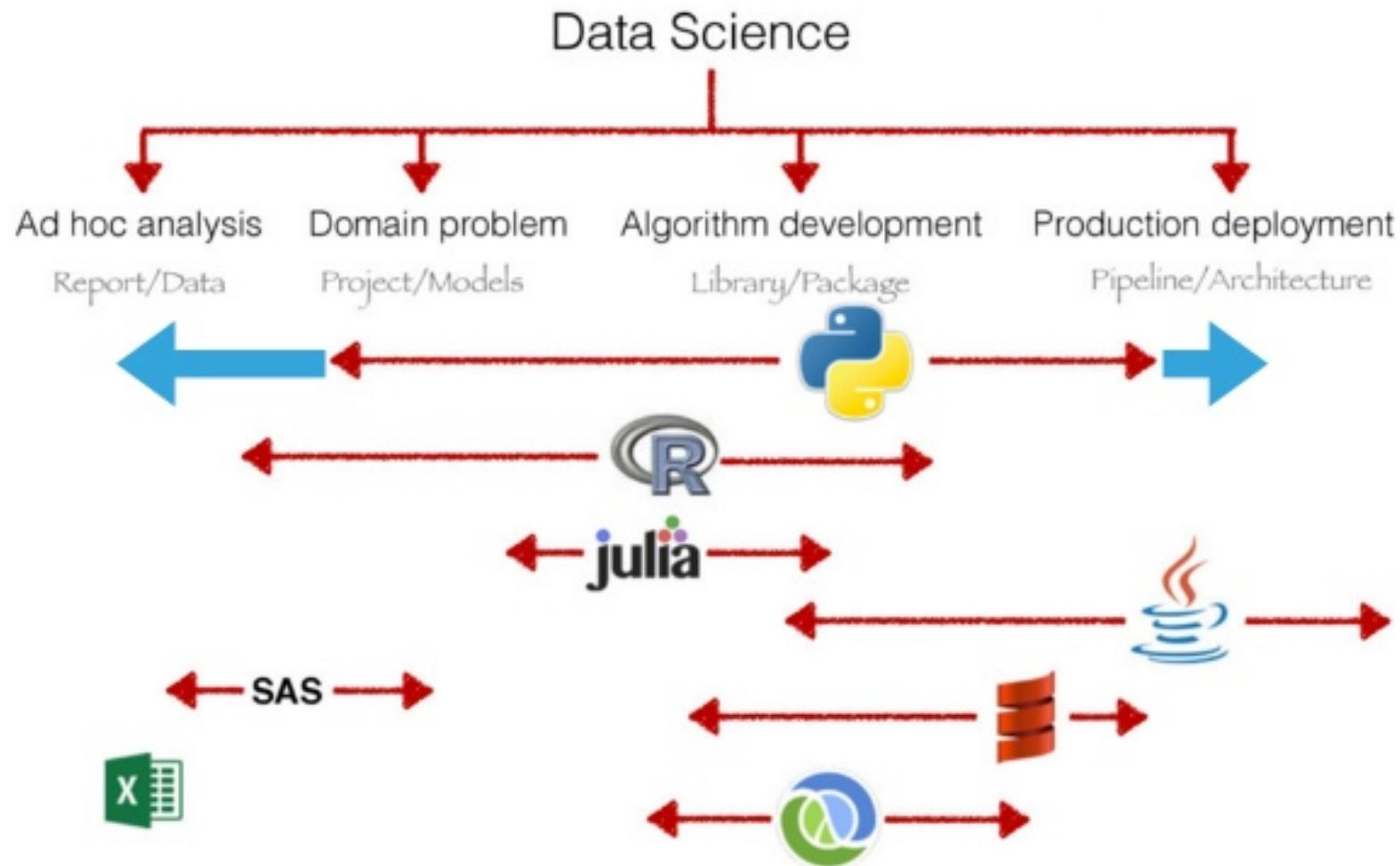


# Why Python

Deep learning framework search interest



# Why Python



Source: <https://www.slideshare.net/icaromedeiros/why-python-is-better-for-data-science>

# Anaconda Virtual Environment

 **ANACONDA DISTRIBUTION**  
Most Trusted Distribution for Data Science

**ANACONDA NAVIGATOR**  
Desktop Portal to Data Science

**ANACONDA PROJECT**  
Portable Data Science Encapsulation

**DATA SCIENCE LIBRARIES**

Data Science IDEs	Analytics & Scientific Computing	Visualization	Machine Learning
 	  	 	 
 	 	 	 
<i>...and many more!</i>			

**CONDA®**  
Data Science Package & Environment Manager

# Jupyter Notebook

The screenshot shows the Jupyter Notebook interface with a title bar "jupyter spectrogram (autosaved)". The menu bar includes File, Edit, View, Insert, Cell, Kernel, and Help. Below the menu is a toolbar with various icons for file operations. The main area contains a section titled "Simple spectral analysis" with the following text:  
An illustration of the [Discrete Fourier Transform](#)  
$$X_k = \sum_{n=0}^{N-1} x_n \exp^{-\frac{2\pi i}{N} kn} \quad k = 0, \dots, N-1$$

```
In [2]: from scipy.io import wavfile  
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view its spectral structure using matplotlib's builtin specgram routine:

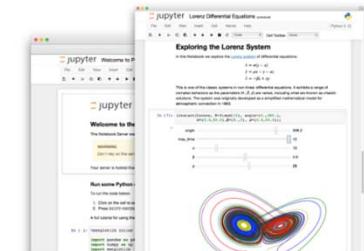
```
In [5]: fig, (ax1, ax2) = plt.subplots(1,2,figsize=(16,5))  
ax1.plot(x); ax1.set_title('Raw audio signal')  
ax2.specgram(x); ax2.set_title('Spectrogram');
```

Two plots are displayed: a line plot titled "Raw audio signal" showing a noisy waveform, and a spectrogram titled "Spectrogram" showing frequency components over time.



## Jupyter Notebook

The Jupyter Notebook is a web-based interactive computing platform that allows users to author data- and code-driven narratives that combine live code, equations, narrative text, visualizations, interactive dashboards and other media.

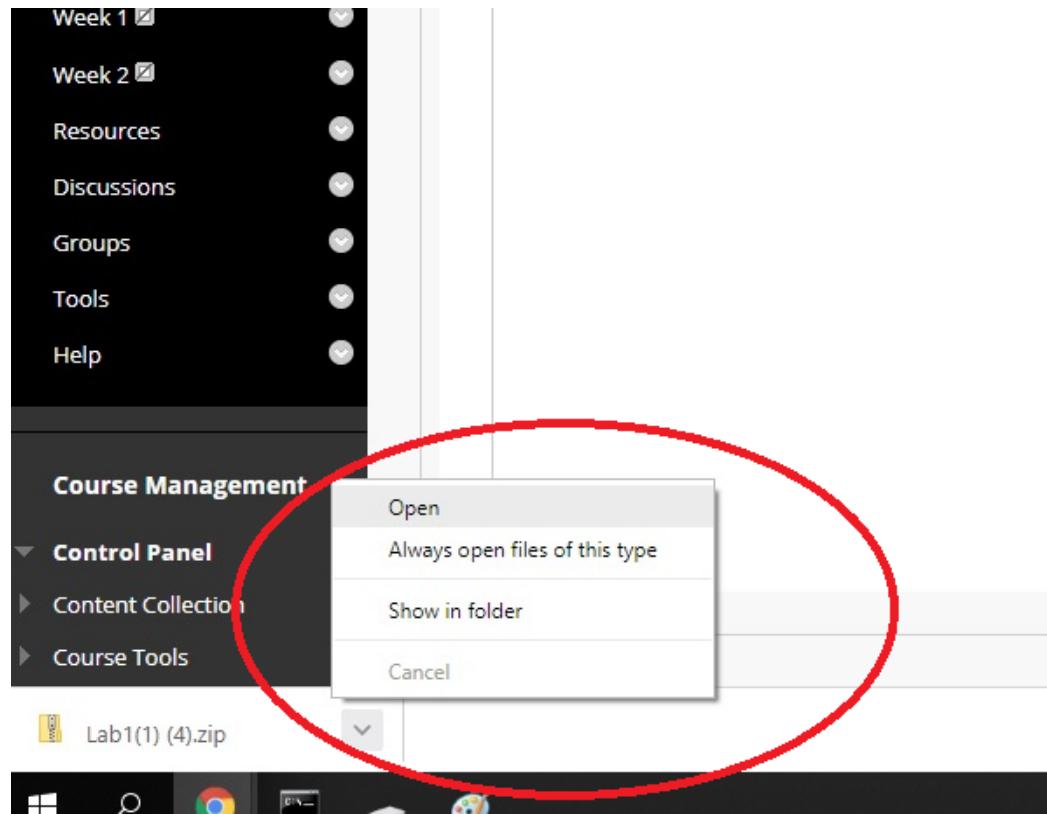


# Lab Setup

The screenshot shows a learning management system interface for the course "ECSE202 ENGINEERING ANALYTICS & MACHINE LEARNING". The left sidebar lists various sections: Home Page, About Subject, Week 1 (circled in red), Week 2, Resources, Discussions, Groups, Tools, and Help. The main content area is titled "Week 1" and contains two items: "ENGANLYML Seminar1.pdf" and "Lab1.zip" (also circled in red). Both items have a "Enabled: Statistics Tracking" status indicator.

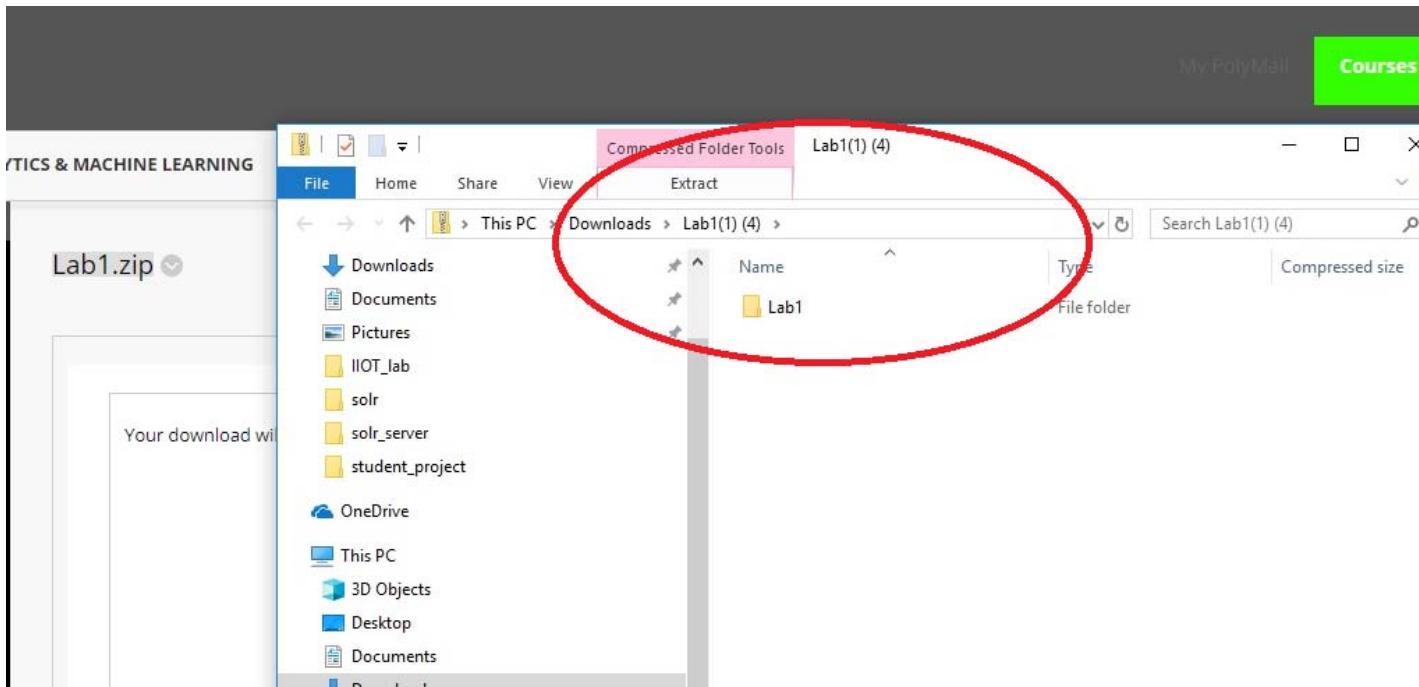
- Go to Polymall
  - Access the current week folder e.g. Week 1
  - Click on the Lab1.zip link to download the files
-

# Lab Setup



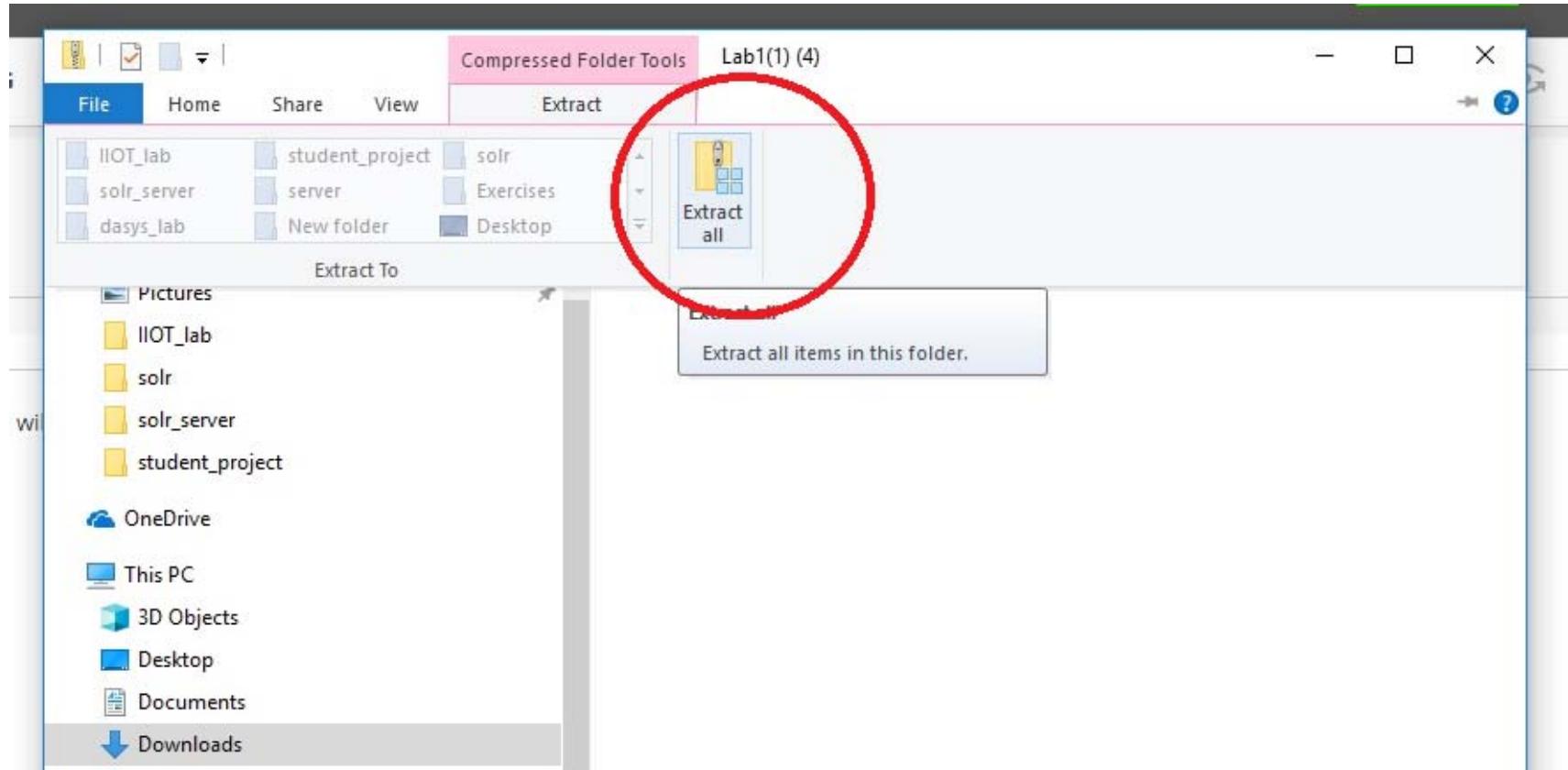
- When file download completed, select open to open the zip file
  - Due to user right setting of lab machine it may not work if you choose to save the file
-

# Lab Setup



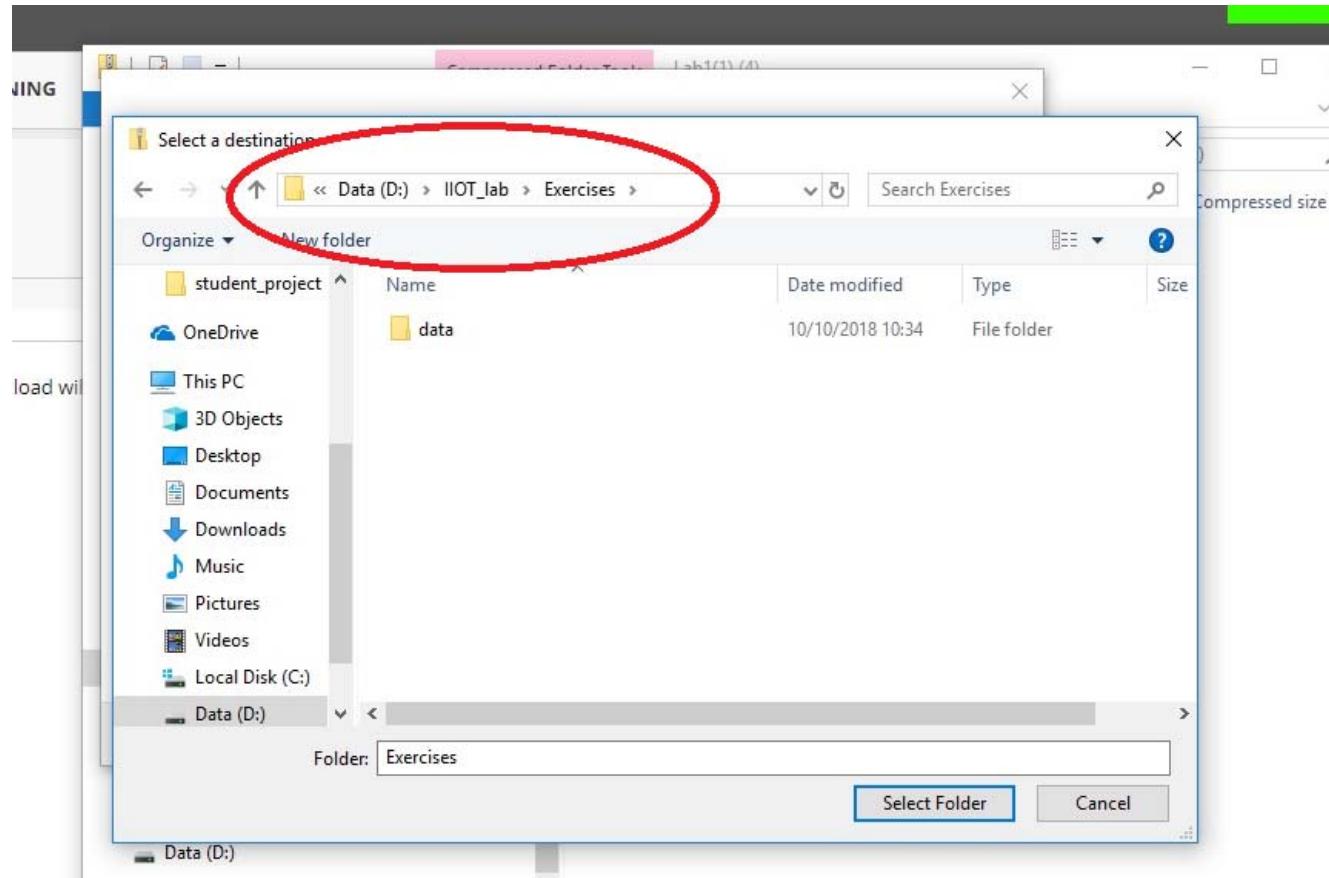
- An explorer would be activate and show you the content of the zip file which usually is a folder with name such as Lab1, Lab2 and etc
-

# Lab Setup



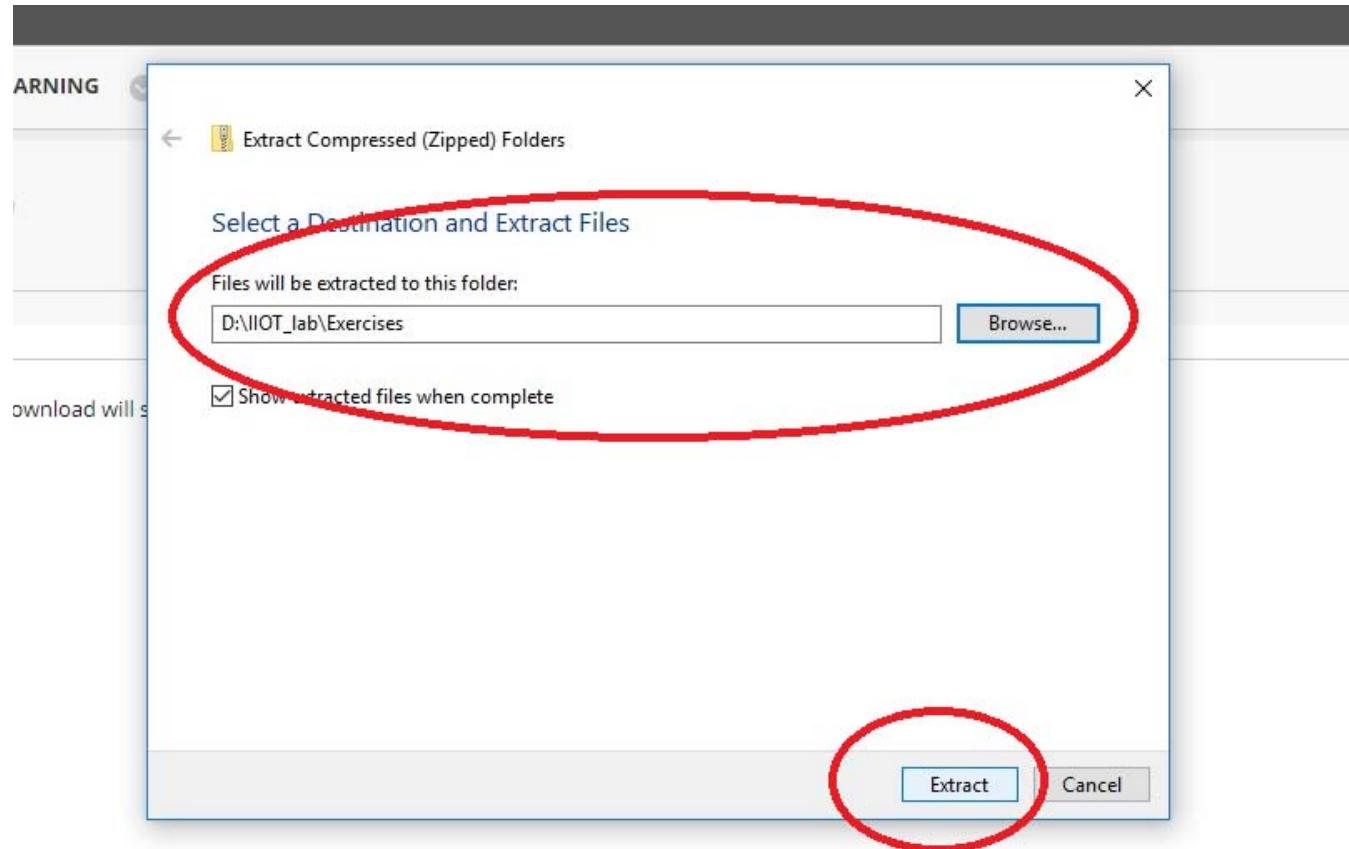
- Select Extract All to uncompressed the zip file
-

# Lab Setup



- Navigate to d:\IIOT\_lab\Exercise to extract the content to this folder
  - Due to user right setting, other folder may not work.
-

# Lab Setup



- Finally ensure you are extracting of the right folder and click Extract.
-

# Lab Setup

```
D:\IIOT_lab\Exercises>cd Lab1

D:\IIOT_lab\Exercises\Lab1>DIR
Volume in drive D is Data
Volume Serial Number is AAE0-D7A2

Directory of D:\IIOT_lab\Exercises\Lab1

11/10/2018  09:48 AM    <DIR>      .
11/10/2018  09:48 AM    <DIR>      ..
11/10/2018  09:48 AM            35,478 01_Lab_Introduction_
11/10/2018  09:48 AM            390,623 01_Lab_Introduction_
11/10/2018  09:48 AM    <DIR>      images
11/10/2018  09:48 AM            203 test.py
                           3 File(s)   426,304 bytes
                           3 Dir(s)  95,649,259,520 bytes free

D:\IIOT_lab\Exercises\Lab1>activate py36
Restore environment to a previous revision
```

- Open a Command Prompt Terminal
  - Navigate to the folder you said the lab content
  - Run “activate py36” to activate the anaconda environment for our lab
-

# Lab Setup

```
D:\IIOT_lab\Exercises>cd Lab1

D:\IIOT_lab\Exercises\Lab1>DIR
Volume in drive D is Data
Volume Serial Number is AAE0-D7A2

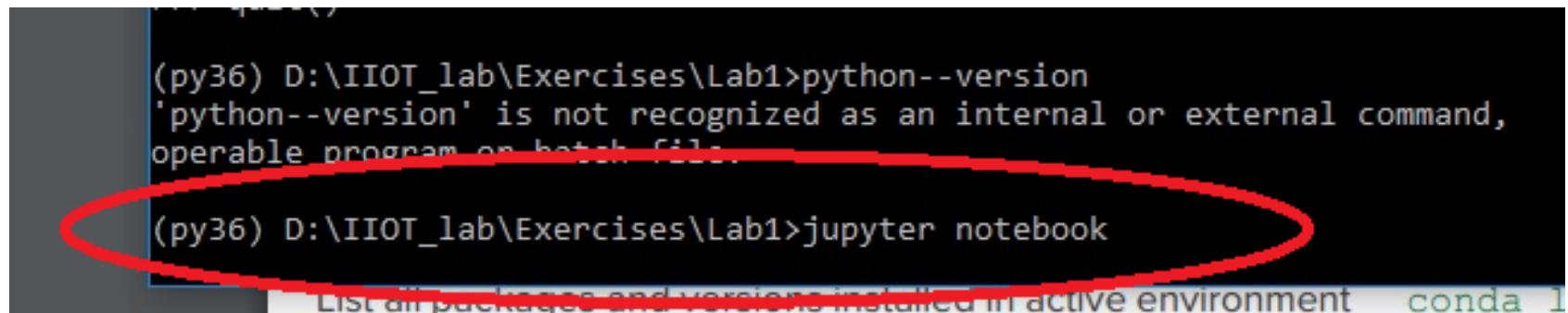
Directory of D:\IIOT_lab\Exercises\Lab1

11/10/2018  09:48 AM    <DIR>      .
11/10/2018  09:48 AM    <DIR>      ..
11/10/2018  09:48 AM            35,478 01_Lab_Introduction_
11/10/2018  09:48 AM            390,623 01_Lab_Introduction_
11/10/2018  09:48 AM    <DIR>      images
11/10/2018  09:48 AM            203 test.py
                           3 File(s)   426,304 bytes
                           3 Dir(s)  95,649,259,520 bytes free

D:\IIOT_lab\Exercises\Lab1>activate py36
Restore environment to a previous revision
```

- Open a Command Prompt Terminal
  - Navigate to the folder you said the lab content
  - Run “activate py36” to activate the anaconda environment for our lab
-

# Lab Setup



```
(py36) D:\IIOT_lab\Exercises\Lab1>python--version
'python--version' is not recognized as an internal or external command,
operable program or batch file.

(py36) D:\IIOT_lab\Exercises\Lab1>jupyter notebook

List all packages and versions installed in active environment  conda 1
```

- If py36 is successfully activated, you would see (py36) beside your command prompt
  - To quit py36 environment, you may use the command “deactivate”
  - Run the command “jupyter notebook” to activate the Jupyter environment for our lab
-

# Lab Setup



The screenshot shows the Jupyter Notebook interface. At the top, there's a navigation bar with the Jupyter logo, 'jupyter' text, and 'Files', 'Running', and 'Clusters' tabs. On the right of the bar are 'Quit' and 'Logout' buttons. Below the bar, a message says 'Select items to perform actions on them.' To the right of this message are 'Upload', 'New', and a refresh icon. A file list table follows, with columns for selection, name, last modified, and file size. The table contains the following data:

	Name	Last Modified	File size
<input type="checkbox"/> 0	/		
<input type="checkbox"/>	images	3 minutes ago	
<input type="checkbox"/>	01_Lab_Introduction_to_Python.ipynb	3 minutes ago	35.5 kB
<input type="checkbox"/>	01_Lab_Introduction_to_Python.pdf	3 minutes ago	391 kB
<input type="checkbox"/>	test.py	3 minutes ago	203 B

- If successful, the default browser would be activate to start with the Jupyter Notebook navigator.
  - You may click on the 01\_Lab\_Introduction\_to\_Python.ipynb to start the lab
-

# Reference

