

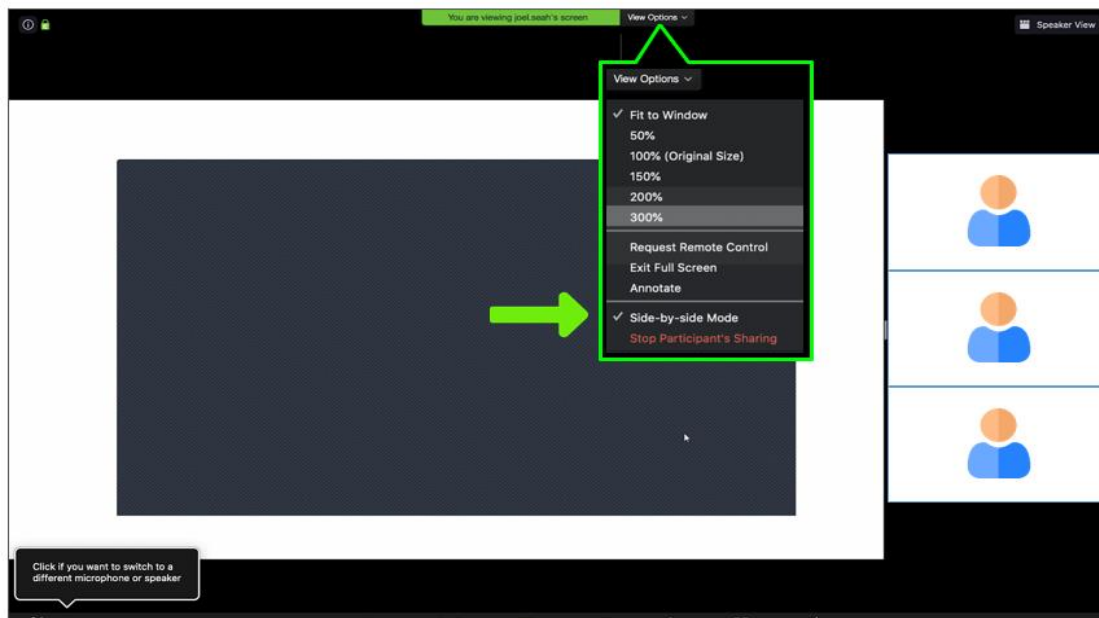


We'll Be Starting Shortly!

To help us run the workshop smoothly, kindly:

- Submit all questions using the Q&A function
- If you have an urgent request, please use the "Raise Hand" function

Using Zoom: Viewing Mode



Side-By-Side Mode

- When sharing screen (slide share)
- With small thumbnails of people on the sidebar

STEPS:

1. View Options
2. Side-By-Side Mode



Hello...

NILAM MUFIDAH

Data Team at Botika

Email : nilam19mufidah@gmail.com
Github : github.com/nilammufidah
LinkedIn : [linkedin.com/in/nilam-mufidah-8a50a8163/](https://www.linkedin.com/in/nilam-mufidah-8a50a8163/)





Text Analytics and Text Classification

Agenda

- Crawling data from Twitter
- Data Preparation
- Text Classification



Crawling and Scraping Data



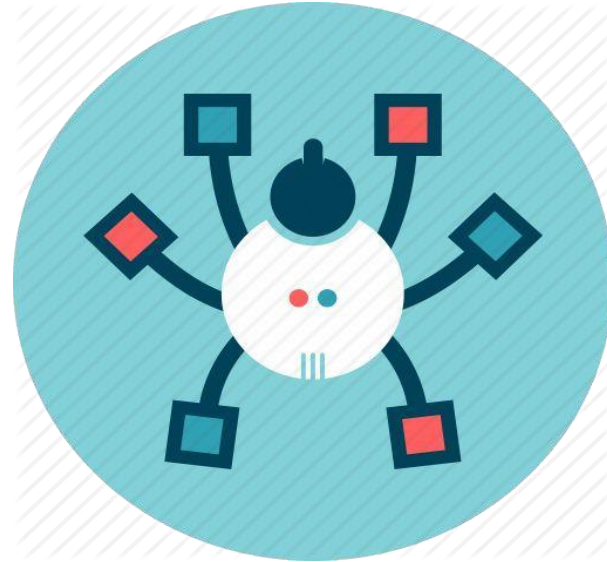
Scraping Data

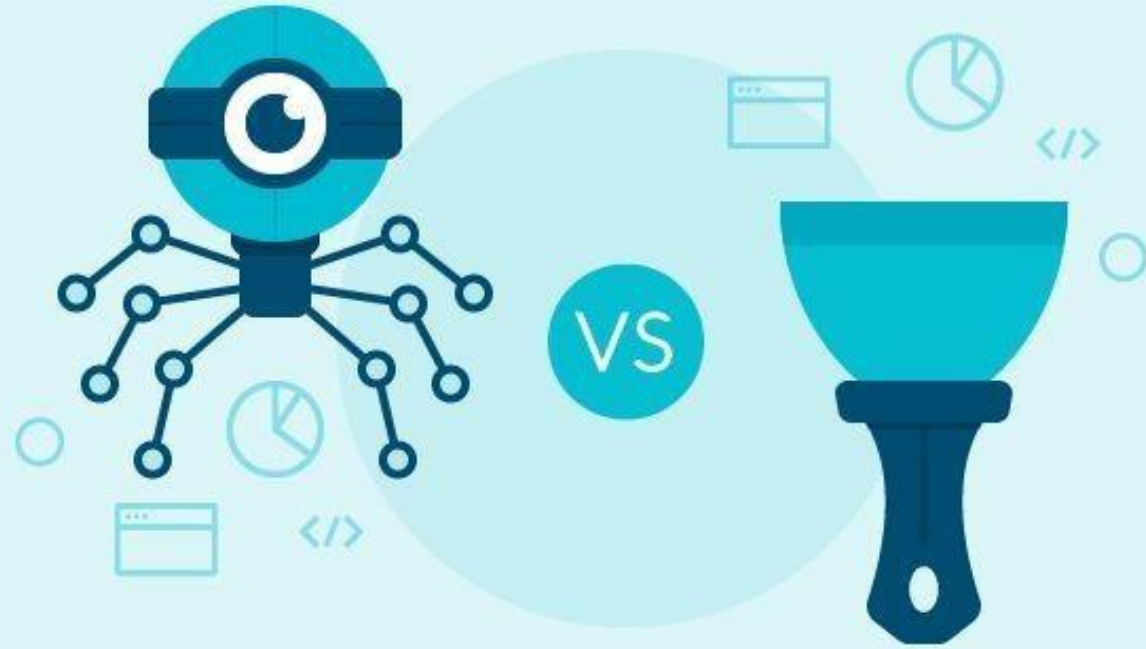
Data scraping is the finding of data and then scraping it. It pulls data directly from a page. This doesn't pull exclusively from the web, it can be taken from anywhere that data exist.



Crawling Data

Web crawling is digging deep into the nooks and crevices of the world wide web to retrieve the stuff you missed ins spring cleaning. Think about spiders or bots, scavenging through the web to find whatever is relevant to your quest





Crawling = Scraping ?

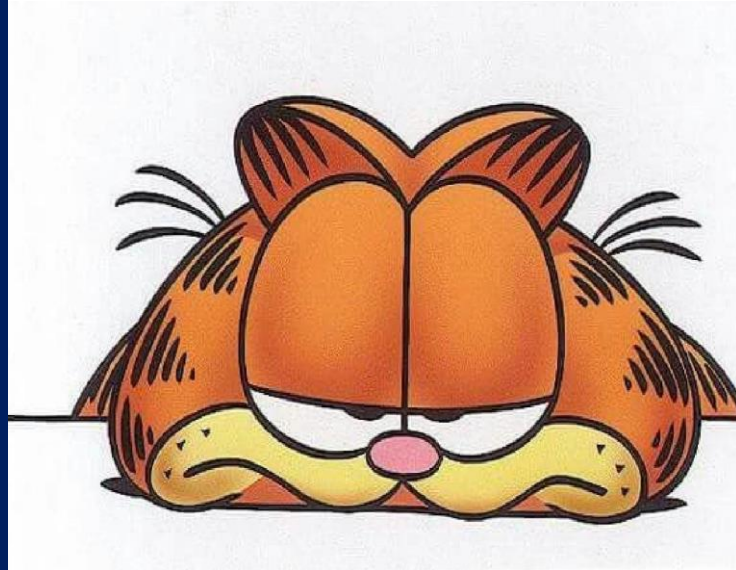
Data Preparation



Spend almost
80 % of our
time

But

It is
IMPORTANT



I feel bored !!!



What wrong with our data?



- Have different pattern
- Duplicates
- Missing value
- Unexpected data values
- Doesn't fit with our need

**Dig a little deeper to
your data**

**Knowing your data
Cleaning
Manipulating Data**

**So, what
should we do ?**



Text Analytics





Type of Algorithms

Supervised Learning

Let's make it simple

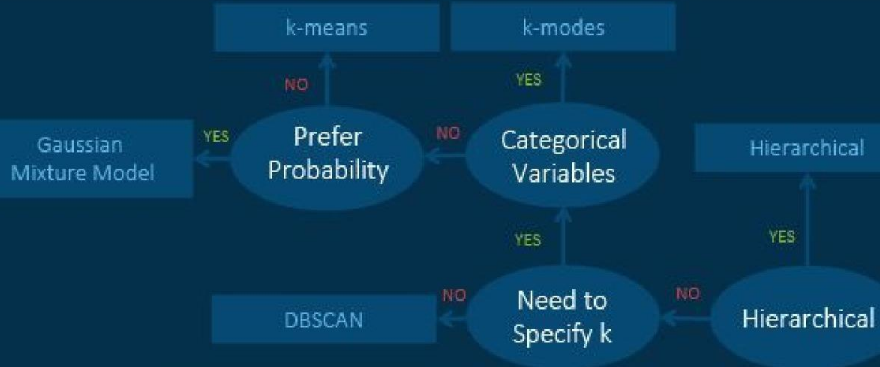
Semi-Supervised learning

**Look at the label
& Output**

Unsupervised Learning

Machine Learning Algorithms Cheat Sheet

Unsupervised Learning: Clustering

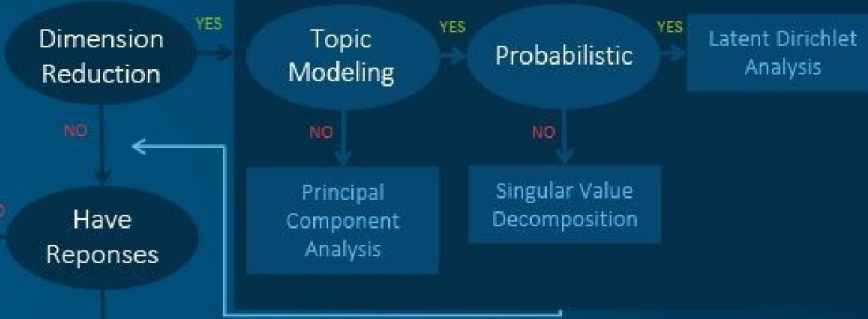


Supervised Learning: Classification



START

Unsupervised Learning: Dimension Reduction

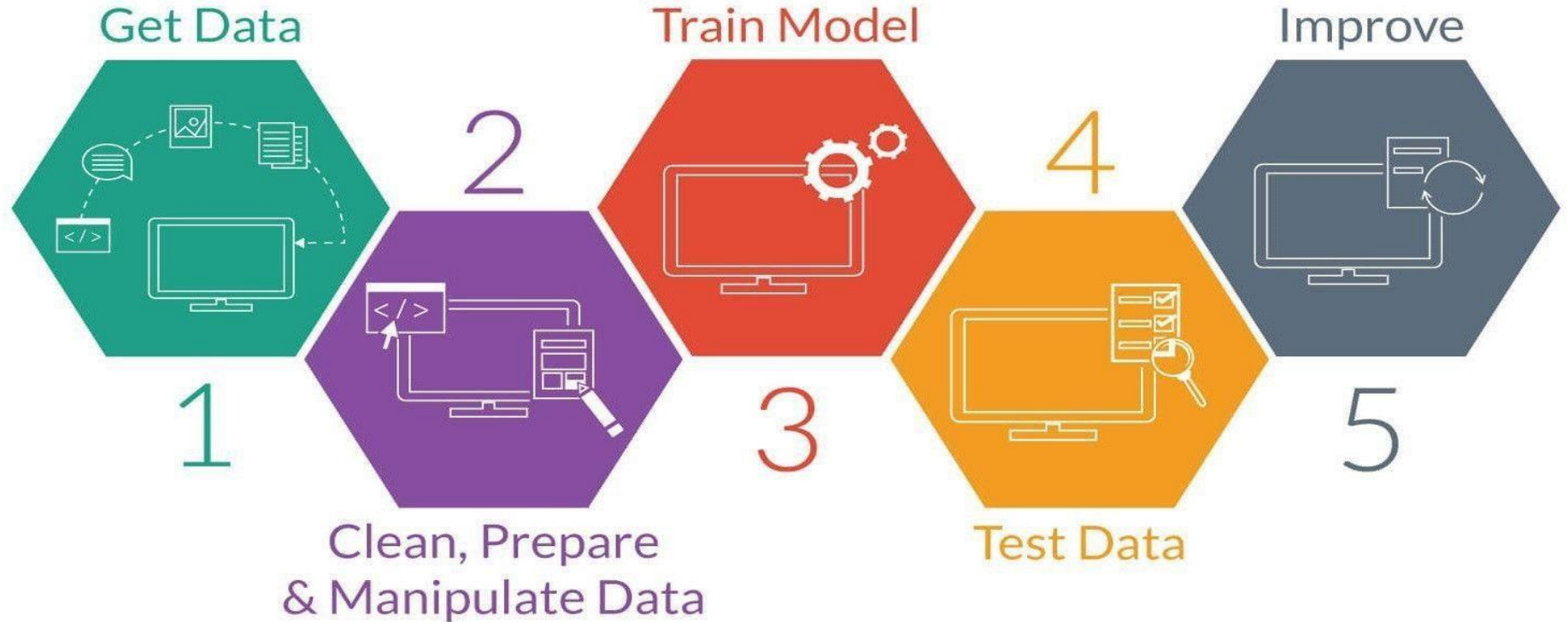


Supervised Learning: Regression





Our Steps





Time to Praticice

[Github.com/nilammufidah/botika-SCL](https://github.com/nilammufidah/botika-SCL)



Reference



- <https://www.edureka.co/blog/what-is-big-data/>
- <https://www.edureka.co/blog/big-data-tutorial>
- <https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/>
- <https://www.datahen.com/blog/data-scraping-vs-data-crawling/>
- <https://www.talend.com/resources/what-is-data-preparation/>
- <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>
- <https://towardsdatascience.com/cleaning-and-preparing-data-in-python-494a9d51a878>
- <https://www.kaggle.com/timoboz/google-trends-data>
- <https://www.kdnuggets.com/2017/06/which-machine-learning-algorithm.html>
- <https://medium.com/@pytyagi/work-flow-in-machine-learning-project-327eddb946b4>

Image



- https://www.123rf.com/photo_10823919_illustration-of-a-kid-carrying-a-box-ful-of-toys-for-donation-or-organizing.html
- https://www.123rf.com/photo_11197789_illustration-of-kids-buying-souvenirs-from-a-toy-store.html?fromid=dGF5OG1ESHhVa0w0UVZodEpvcHJNZz09
- <https://www.xenonstack.com/images/blog/10-vs-of-big-data.png>
- <https://www.uihere.com/free-cliparts/cartoon-clip-art-thinking-woman-1601359>
- <https://www.uihere.com/free-cliparts/stock-photography-thought-illustration-thinking-child-1388176>
- <https://www.uihere.com/free-cliparts/big-data-analytics-information-internet-of-things-ppt-element-1330286>
- <https://www.uihere.com/free-cliparts/school-cartoon-royalty-free-illustration-2-children-holding-up-their-hands-in-front-of-the-seat-1329114>
- <https://www.uihere.com/free-cliparts/idea-drawing-icon-vector-bulb-1267305>
- <https://id.pinterest.com/pin/541980136404391230>
- <https://depositphotos.com/stock-photos/dig.html?filter=all&qview=49849527>
- <https://depositphotos.com/vector-images/detective-cartoon-vector.html?qview=47423759>
- <http://www.dataception.in/blog/machine-learning/>
- <https://www.efilecabinet.com/grow-your-business-immediately-with-efilecabinet/im-scared-that-i-have-way-too-much-data-to-migrate-and-even-if-i-could-it-would-take-forever-to-organize/>



THANK YOU

Find us on :



botika.online



botikaonline



BOTIKA



Botika

Your Feedback Matters!



bit.ly/3hmJ3Nr