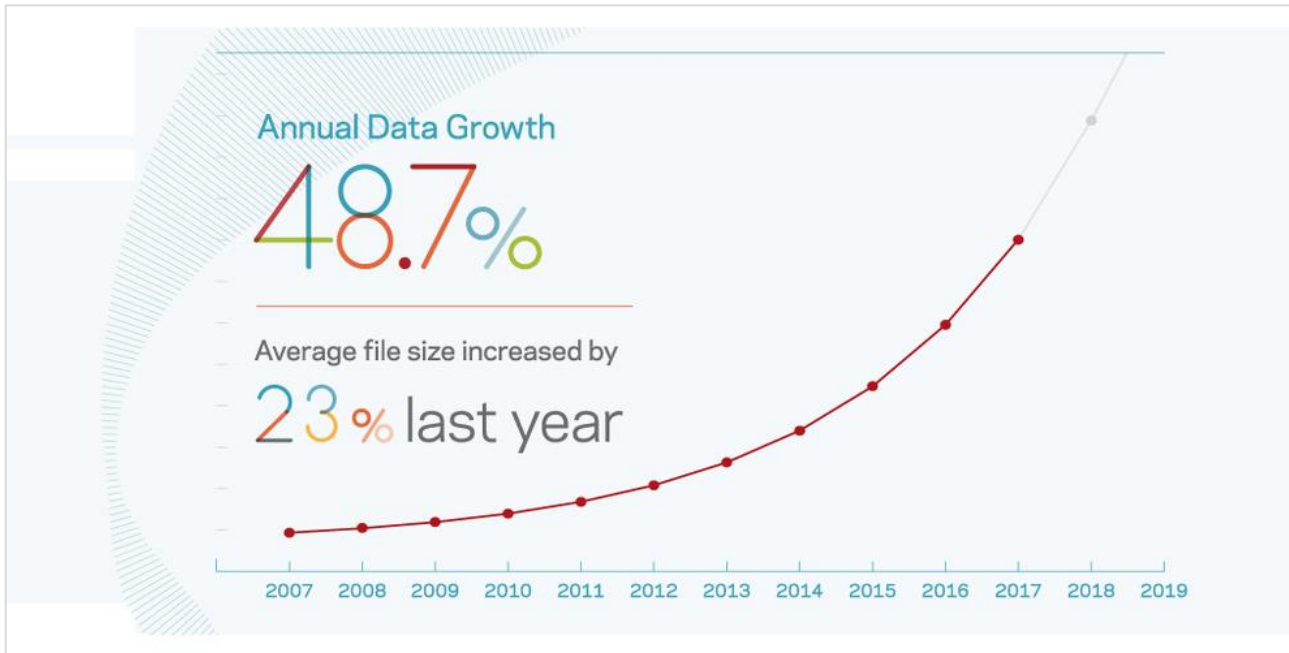


Pengenalan Data Mining

Made Satria Wibawa, M.Eng.
2020

Background

Ketersediaan data dengan jumlah masif



<https://vox.veritas.com/t5/Insights/Data-Genomics-Index-2017-Digital-Transformation-Continues-to/ba-p/844006>

EDITION: AS

ZDNet

VIDEOS EXECUTIVE GUIDES SECURITY CLOUD INNOVATION CXO HARDWARE MORE NEWSLETTERS

MUST READ: Where did the Moon come from? Scientists are using a supercomputer to find out

By 2025, nearly 30 percent of data generated will be real-time, IDC says

Both consumer preferences and enterprise needs will drive the creation of real-time data, according to a new white paper from IDC.

By Stephanie Condon for Between the Lines | November 27, 2018 -- 05:02 GMT (13:02 SGT) | Topic: Data Management

MORE FROM STEPHANIE CONDON

Cloud
Dell builds tighter VMware integrations for cloud, app modernization

Cloud
Oracle rolls out improvements across SaaS applications

Hardware

2-in-1 devices to carry PC market in 2018: IDC

WATCH NOW

<https://www.zdnet.com/article/by-2025-nearly-30-percent-of-data-generated-will-be-real-time-idc-says/>

Background

Kebutuhan Industri



Harvard Business Review

Data | Data Scientist: The Sexiest Job of the 21st Century

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

From the October 2012 Issue

Summary Save Share Print \$8.95 Buy Copies

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

Goldman, a PhD in physics from Stanford, was intrigued by the linking he did see going

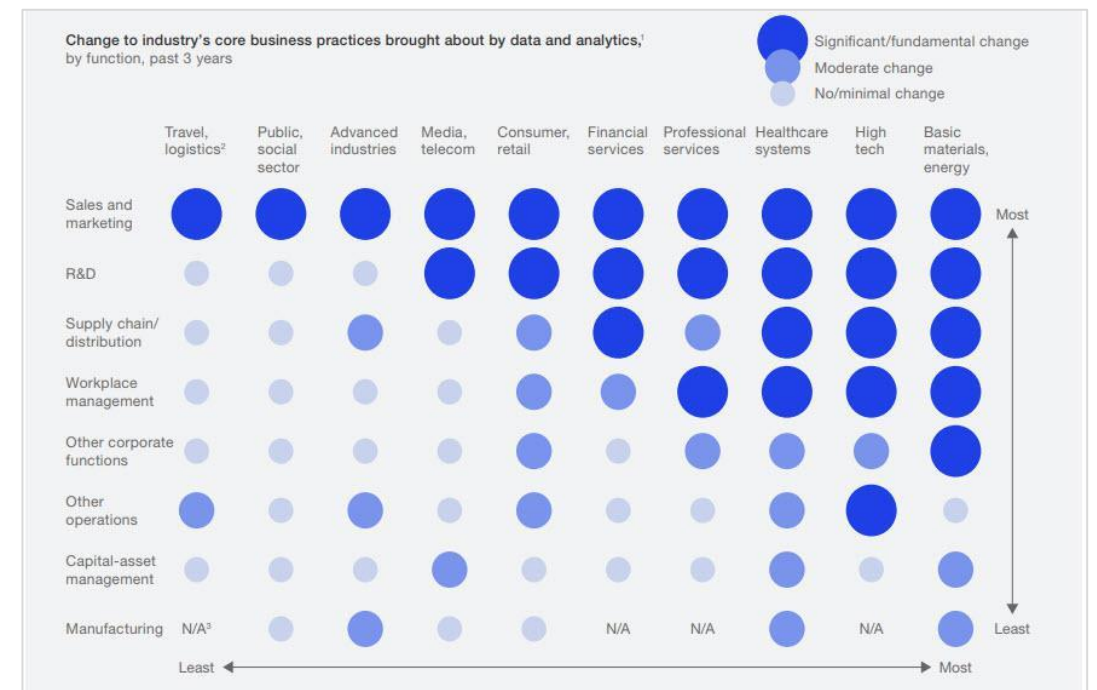
VIEW MORE FROM THE October 2012 Issue

Harvard Business Review

GETTING CONTROL OF BIG DATA

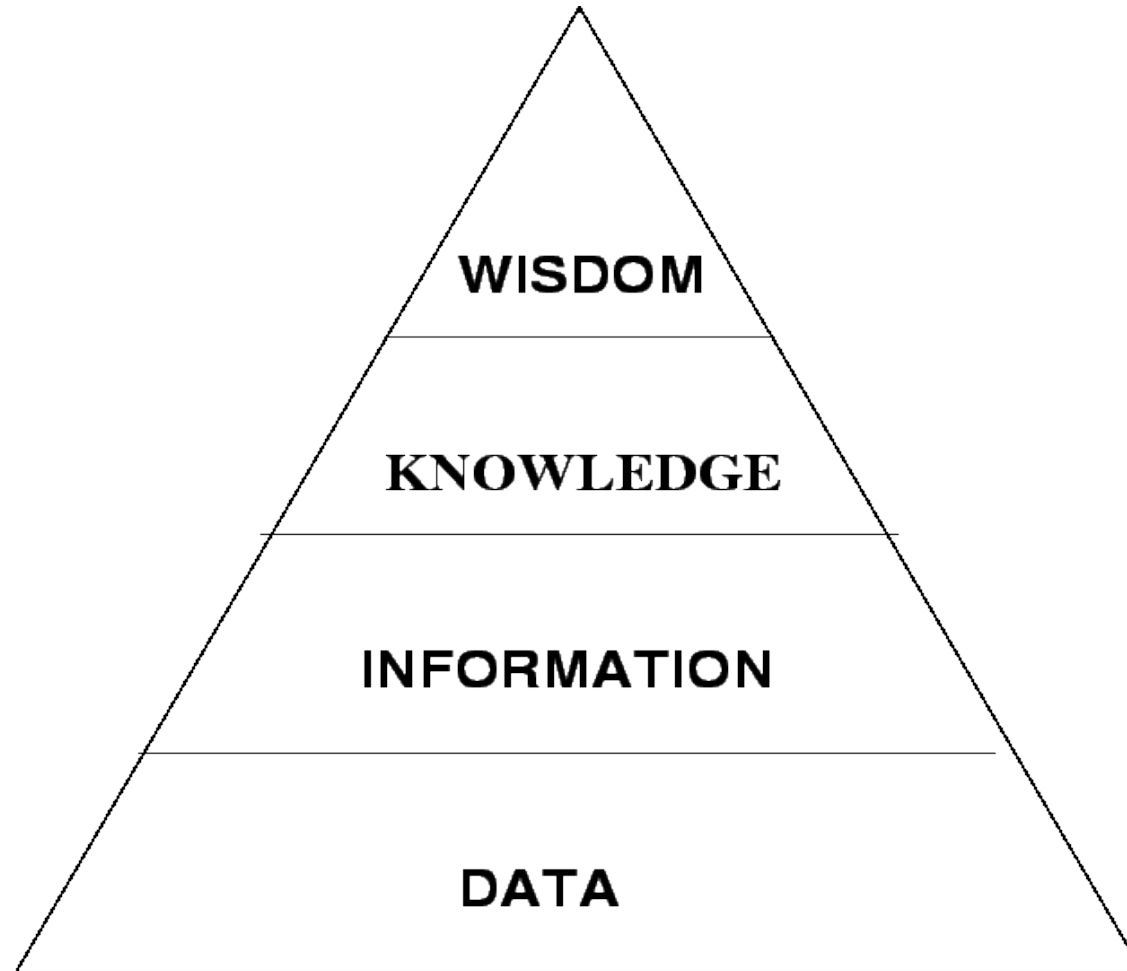
Explore The Archive

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>



<https://www.forbes.com/sites/louisolumbus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/#4c482c852926>

Data, Information, Knowledge, Wisdom



Data, Information, Knowledge, Wisdom

NIP	TGL	DATANG	PULANG
1103	02/12/2004	07:20	15:40
1142	02/12/2004	07:45	15:33
1156	02/12/2004	07:51	16:00
1173	02/12/2004	08:00	15:15
1180	02/12/2004	07:01	16:31
1183	02/12/2004	07:49	17:00

Data Kehadiran Pegawai

Data, **Information**, Knowledge, Wisdom

NIP	Masuk	Alpa	Cuti	Sakit	Telat
1103	22				
1142	18	2		2	
1156	10	1	11		
1173	12	5			5
1180	10			12	

Informasi Akumulasi Bulanan Kehadiran Pegawai

Data, Information, **Knowledge**, Wisdom

	Senin		Selasa	Rabu	Kamis	Jumat
Terlambat		7	0	1	0	5
Pulang Cepat		0	1	1	1	8
Izin		3	0	0	1	4
Alpa		1	0	2	0	2

Pola Kebiasaan Kehadiran Mingguan Pegawai

Data, Information, Knowledge, **Wisdom**

- Kebijakan **penataan jam kerja karyawan** khusus untuk hari senin dan jumat
- Peraturan jam kerja:
 - Hari Senin dimulai jam 10:00
 - Hari Jumat diakhiri jam 14:00
 - Sisa jam kerja dikompensasi ke hari lain

Karakteristik

- Data
 - Bersifat statis di alam
 - Menggambarkan fakta diskrit tentang peristiwa
 - Prasyarat untuk membentuk informasi
- Information
 - Kumpulan dari data yang telah diproses
 - Memiliki makna dan tujuan
- Knowledge
 - Pemahaman akan informasi berdasarkan tingkat kepentingan dan relevansi nya terhadap ruang lingkup masalah.
 - Integrasi dari proses persepsi manusia yang dapat membantu untuk menarik kesimpulan yang bermakna

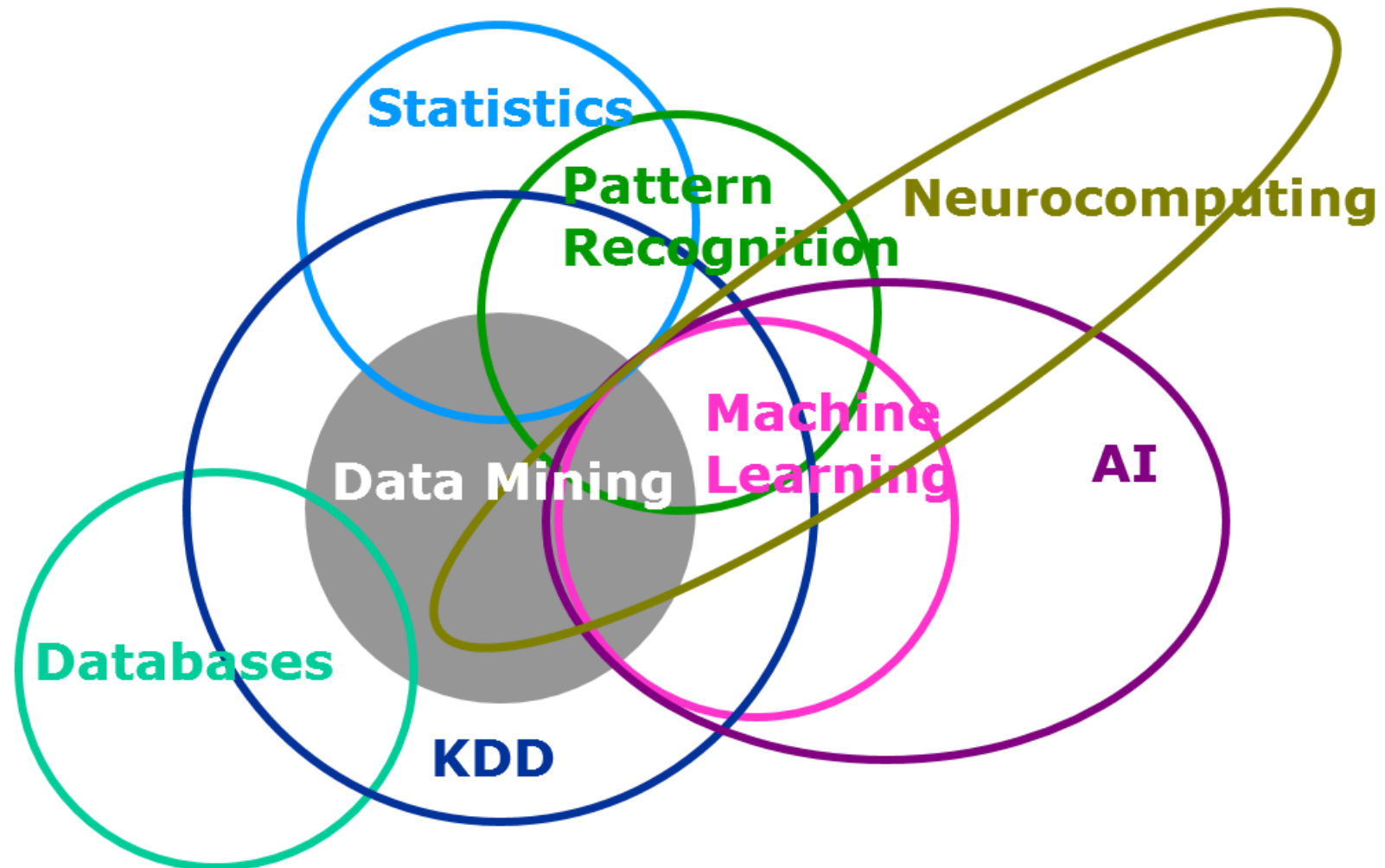
Definisi

- Disiplin ilmu yang mempelajari metode untuk mengekstrak **pengetahuan** atau **menemukan pola** dari suatu **data yang besar**
- Ekstraksi dari data ke pengetahuan:
- Data: **fakta yang terekam** dan tidak membawa arti
- Pengetahuan: pola, rumus, aturan atau model yang muncul dari data
- Istilah yang berhubungan dengan data mining:
 - Knowledge Discovery in Database (KDD)
 - Knowledge extraction
 - Pattern analysis
 - Information harvesting
 - Business intelligence

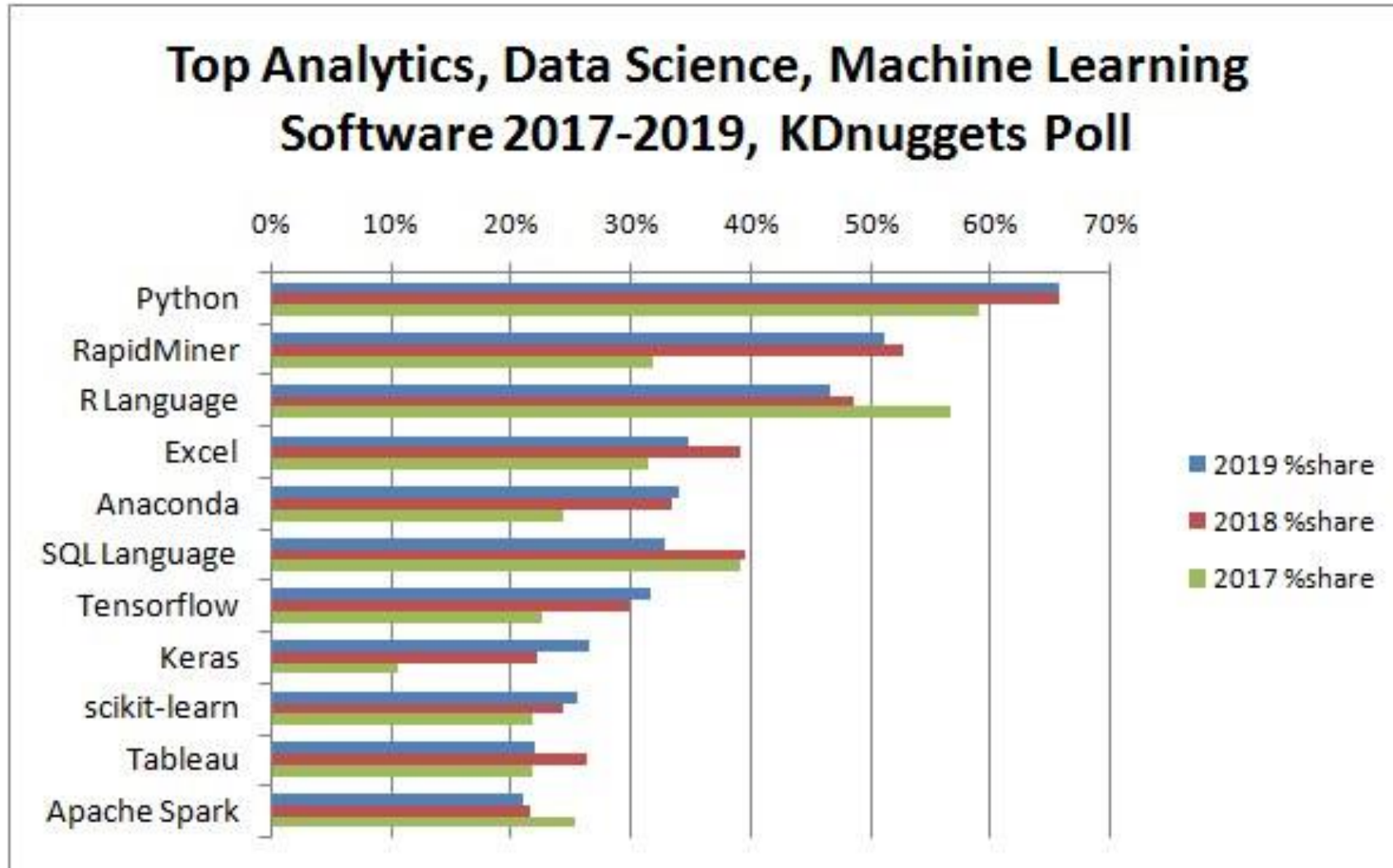
Definisi

- Melakukan **ekstraksi** untuk mendapatkan **informasi penting** yang sifatnya **implisit** dan **sebelumnya tidak diketahui**, dari suatu **data** (Witten et al., 2011)
- Kegiatan yang meliputi **pengumpulan**, pemakaian **data historis** untuk menemukan **keteraturan, pola dan hubungan** dalam set data berukuran **besar** (Santosa, 2007)
- Extraction of interesting (**non-trivial, implicit, previously unknown and potentially useful**) **patterns or knowledge** from huge amount of data (Han et al., 2011)

Multi Disiplin Ilmu



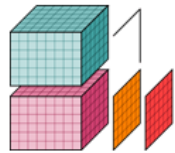
Tools



Tools



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



xarray



scikit-image
image processing in python



IP[y]:
IPython

silahkan install tools
mandiri berdasarkan
panduan

