



Pengenalan Machine Learning/AI dengan Python

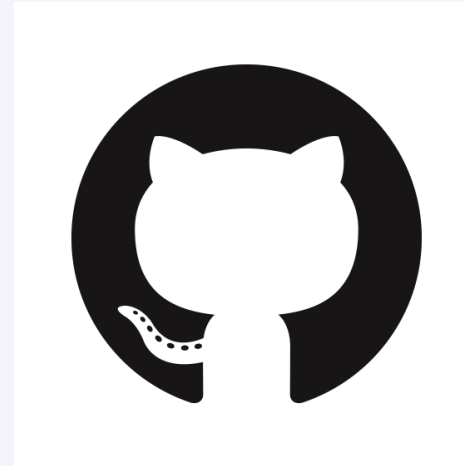
Debi Tomika



**BADAN PUSAT STATISTIK
KABUPATEN KUANTAN SINGINGI**

Materi

- Semua Materi Tersedia di GitHub dan Laci
 - File presentasi
 - Jupyter notebooks
 - Raw data dapat diperoleh dari Fungsi DLS BPS Provinsi Riau



[https://github.com/debitomika/pengenalan ML](https://github.com/debitomika/pengenalan_ML)

or

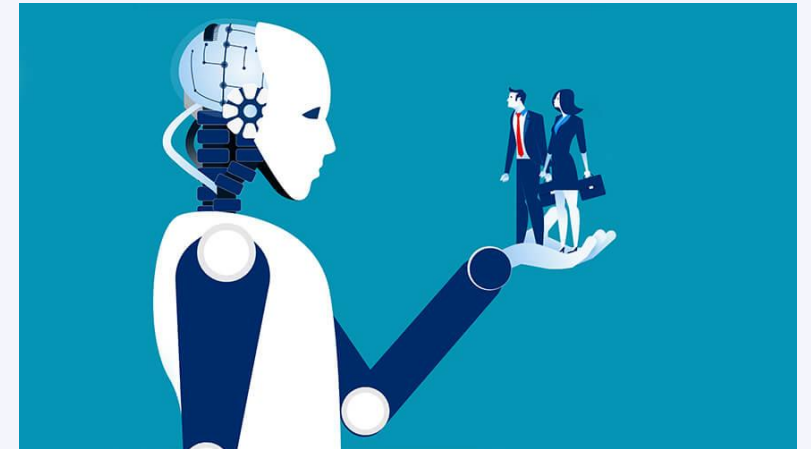
[http://s.bps.go.id/pengenalan ml](http://s.bps.go.id/pengenalan_ml)

Kenapa Insan BPS Harus Menguasai Machine Learning?

- Mengutip Juknis Prakom Poin I.C.47 **Melakukan Implementasi *Data Mining*** (Angka Kredit 0,6, Batasan Penilaian 1 (satu) kali per bulan)
- “*Machine Learning* sangat penting bagi suatu organisasi dalam membuat keputusan yang lebih baik dengan menemukan pola dan tren dalam data, meringkas data, dan mengambil informasi yang relevan.”
- Melakukan implementasi *data mining* adalah kegiatan atau proses menemukan pola tersembunyi atau hubungan yang tidak terduga/sebelumnya tidak diketahui dalam data dengan memanfaatkan teknik *data mining* yaitu *machine learning*, *artificial intelligence* (AI), dan/atau statistik

Pengenalan

- Machine learning membuat langkah besar
 - data sets yang semakin besar dan berkualitas
 - Peningkatan kemampuan komputasi komputer
 - Algoritma yang semakin canggih
- Dapat diaplikasikan dalam berbagai bidang
 - Komersial
 - Ilmiah/Sains
- Hubungannya dengan artificial intelligence
 - Akan tetapi, AI \neq machine learning

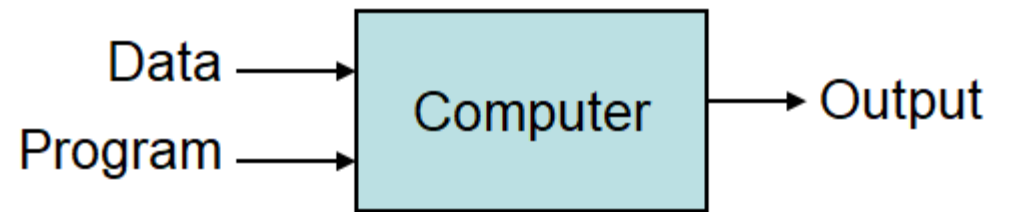


Definisi Machine learning

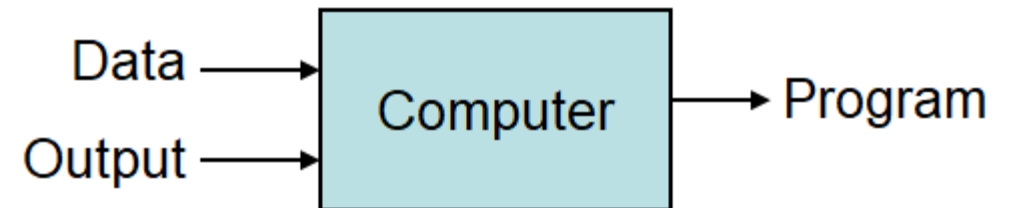
Machine learning adalah sebuah cabang dari artificial intelligence berdasarkan gagasan bahwa sebuah sistem dapat belajar dari data, mengenali pola dan membuat keputusan dengan intervensi manusia yang minimal.

“A computer program that can learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, so that its performance at tasks in **T**, as measured by **P**, improves with experience **E**.”

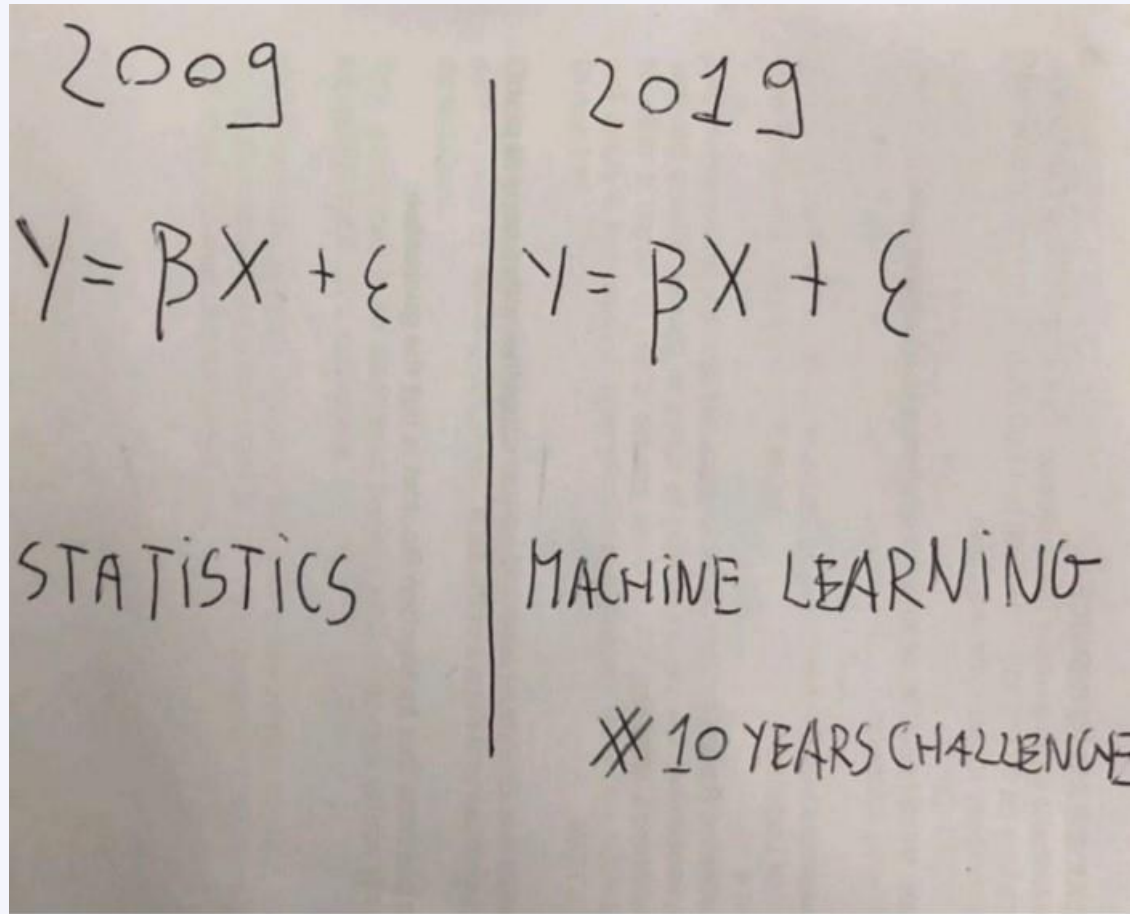
Traditional Programming



Machine Learning



Machine Learning or Statistics?



Machine Learning or Statistics?

Mengatakan machine learning adalah sebutan untuk statistik di masa kini, sama halnya dengan mengatatakan

- Physics is just glorified mathematics.
- Zoology is just glorified stamp collection.
- Architecture is just glorified sand-castle construction.

Machine Learning or Statistics?

- “Perbedaan utama antara machine learning dan statistik adalah tujuannya. Model *machine learning* dirancang untuk membuat prediksi seakurat mungkin. Model statistik dirancang untuk inferensi tentang hubungan antar variabel. ”

Machine Learning or Statistics?

- ML theory from fields like mathematics & statistics
- ML algorithms from fields like optimization, matrix algebra, calculus
- ML implementations from computer science & engineering concepts (e.g. kernel tricks, feature hashing)

Pembagian Machine learning

- Supervised learning
 - regresi: memprediksi nilai numerik
 - klasifikasi: memprediksi nilai kategorik
- Unsupervised learning
 - clustering: mengolompokkan data menurut “jarak”
 - association: menemukan kejadian yang sering terjadi bersamaan
 - link prediction: menemukan hubungan dalam data
 - data reduction: memproyeksikan variabel
- Reinforcement learning

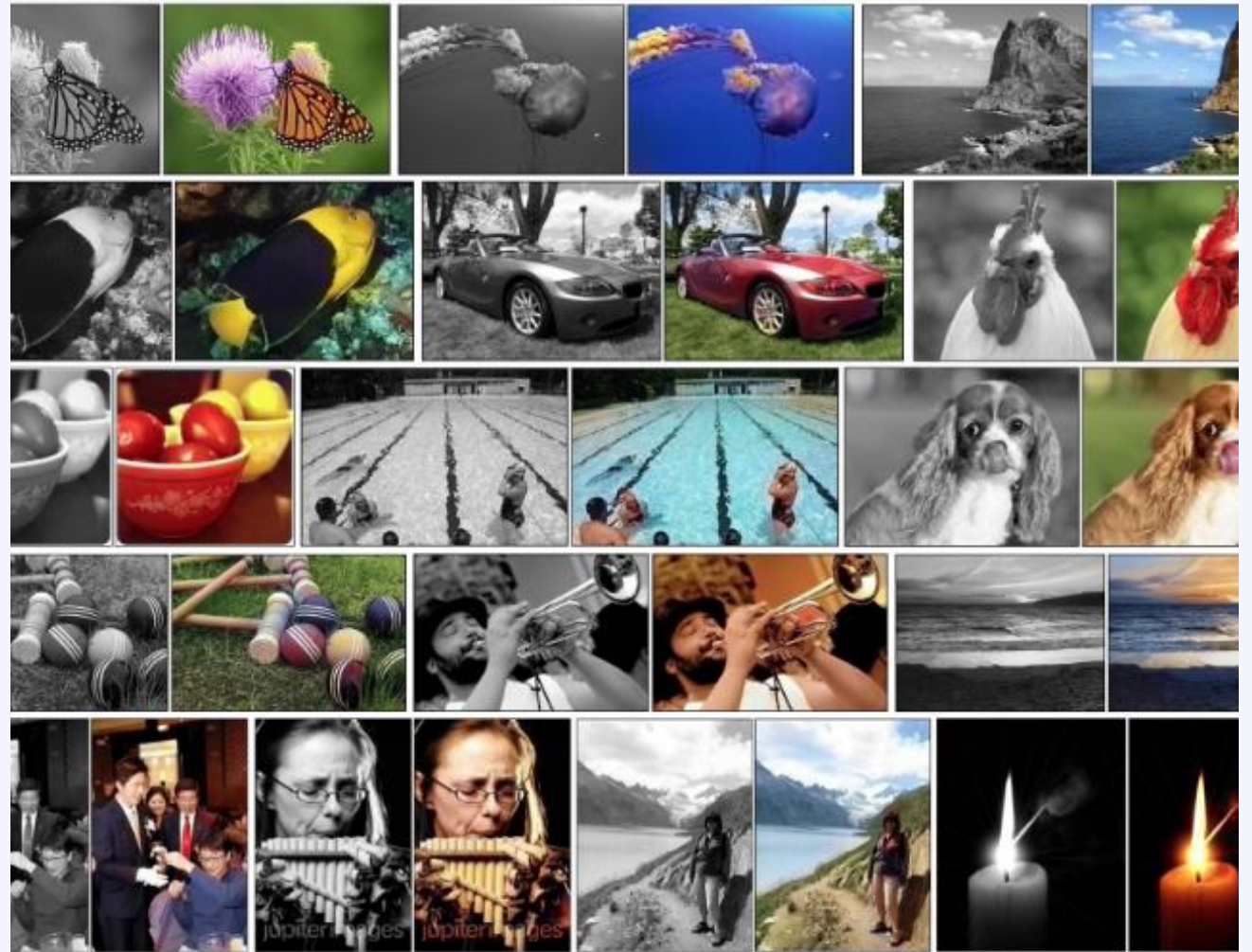
Tips Belajar Machine learning

- Mengutip professor David Perkins dari Harvard University
- Cara tercepat mempelajari *machine learning* adalah dengan contoh kasus

Regression

Mewarnai gambar hitam-putih secara otomatis

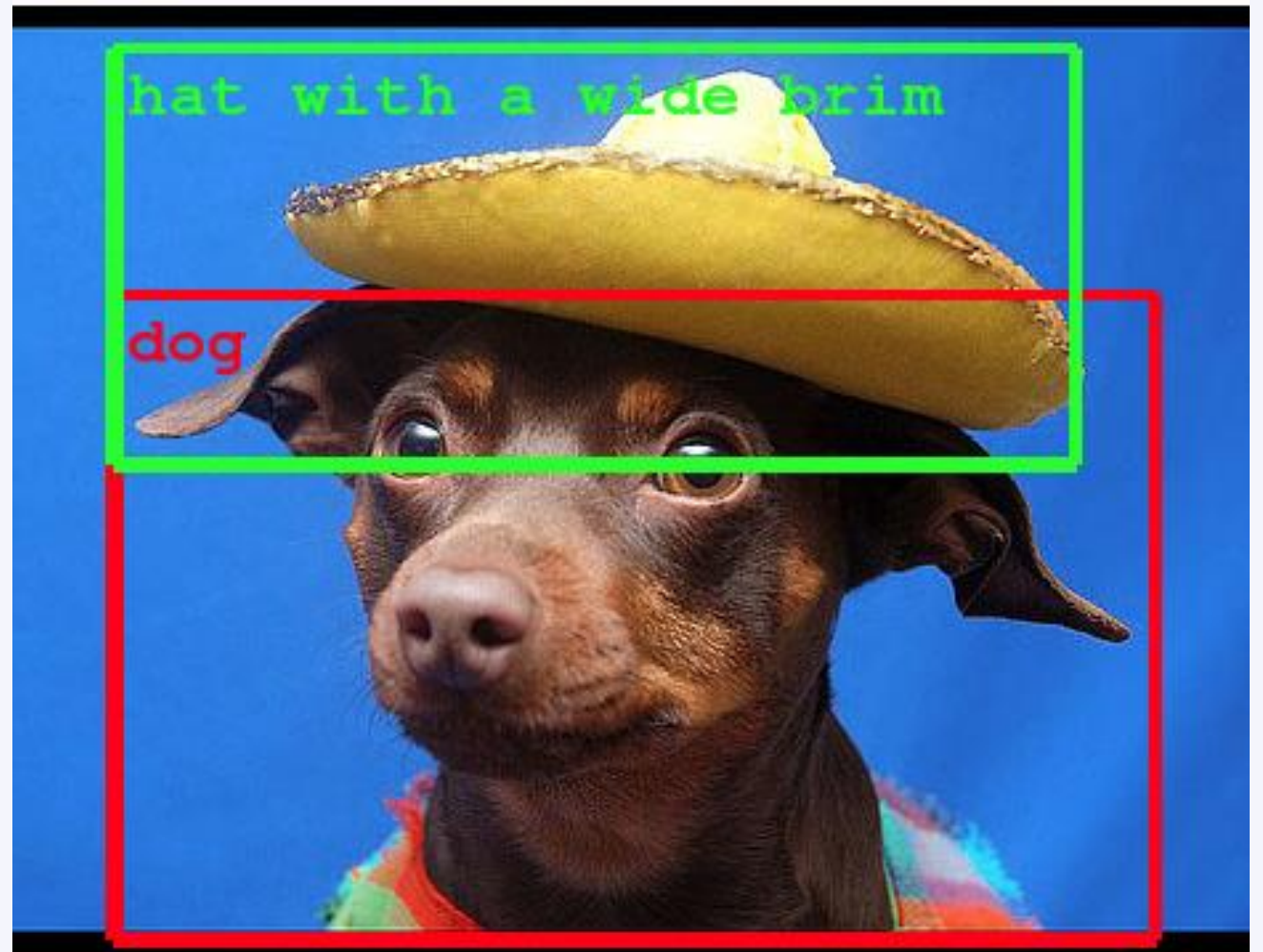
<https://tinyclouds.org/colorize/>



Classification

Pengenalan objek

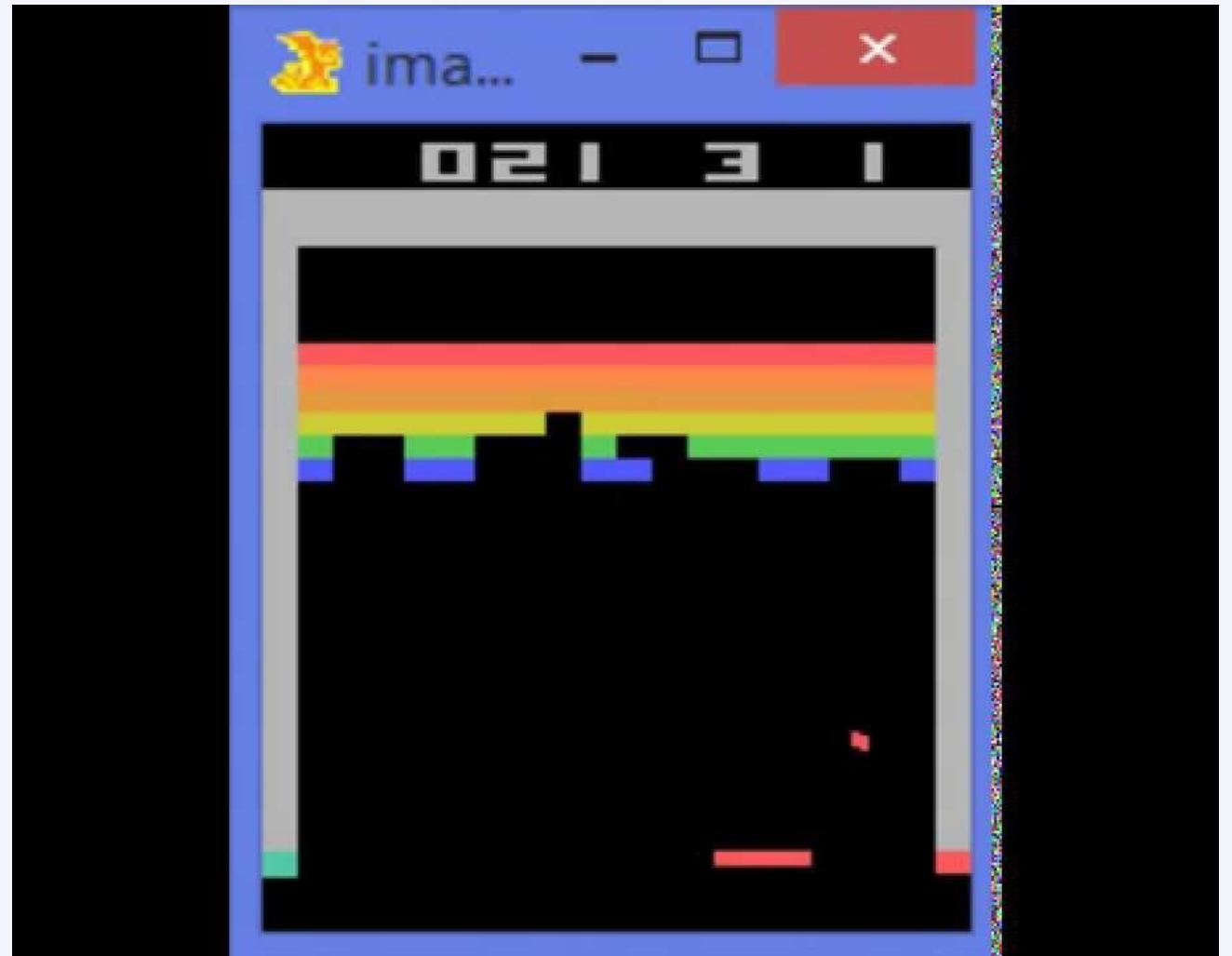
<https://ai.googleblog.com/2014/09/building-deeper-understanding-of-images.html>



Reinforcement learning

Belajar bermain game

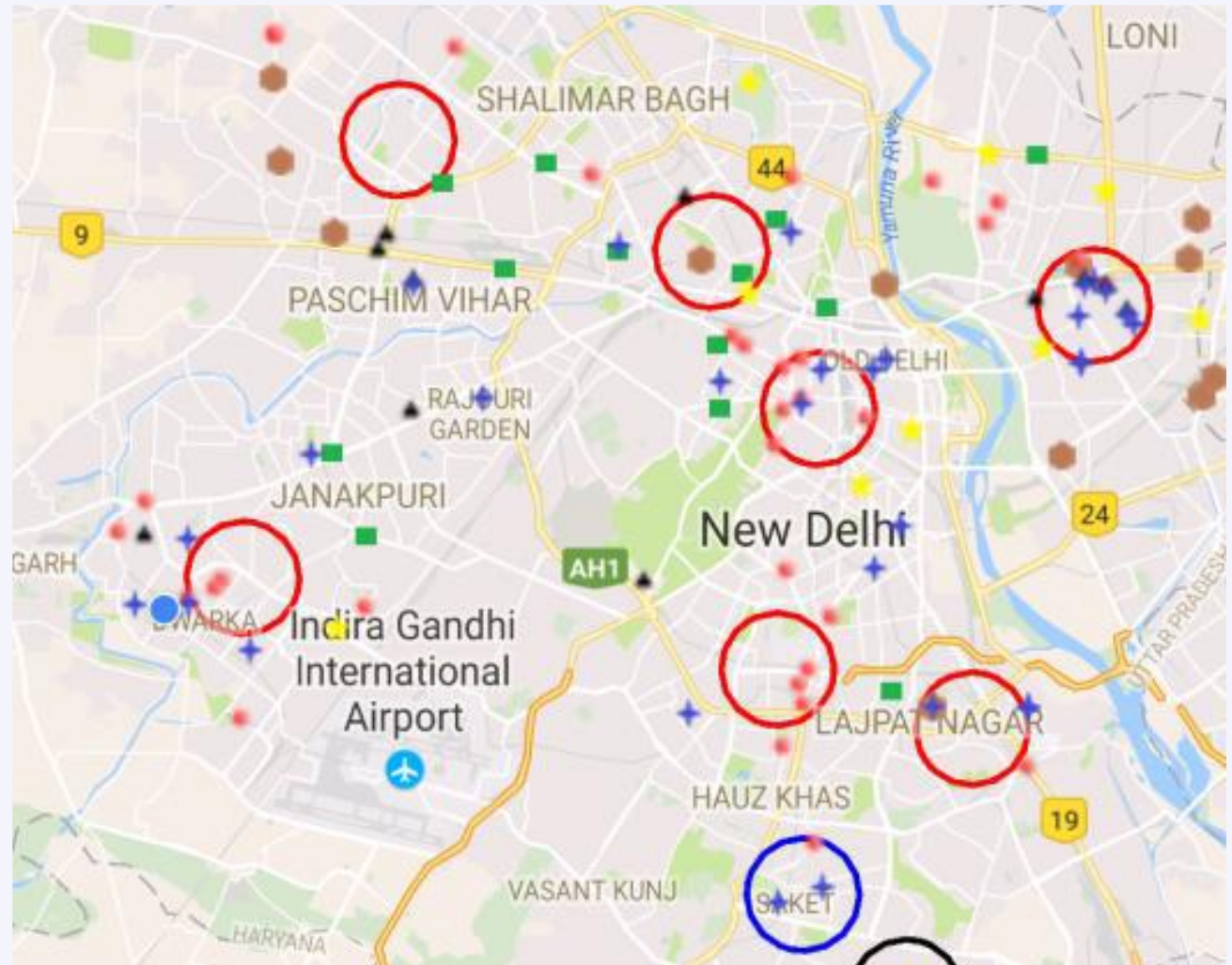
<https://www.youtube.com/watch?v=V1eYniJ0Rnk>



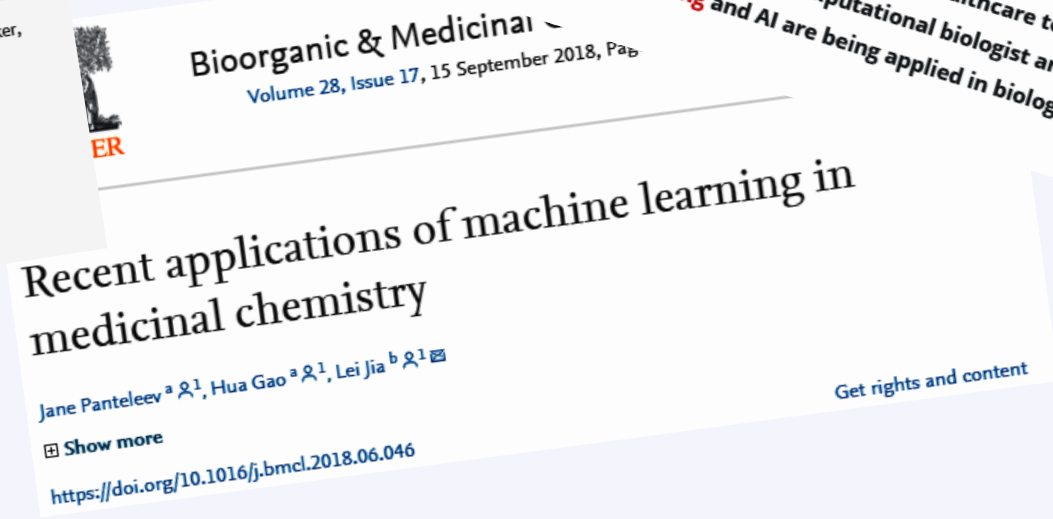
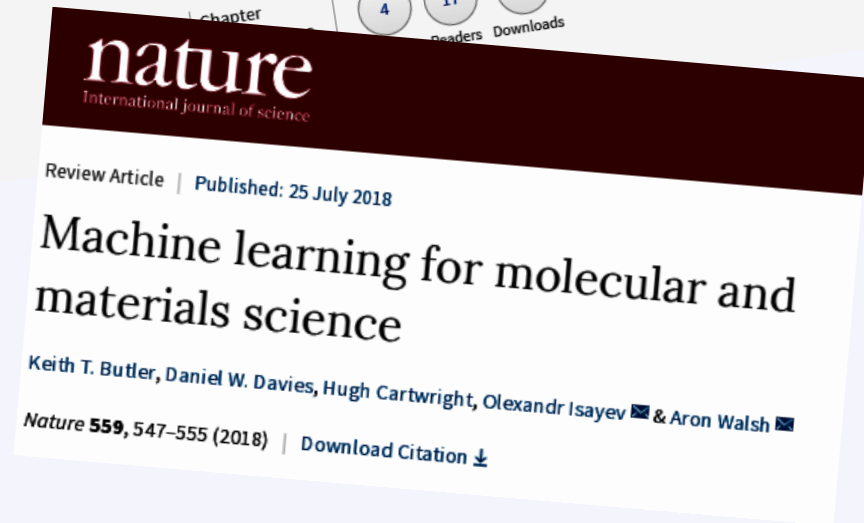
Clustering

Prediksi tindak kejahatan menggunakan k-means clustering

<http://www.grdjournals.com/uploads/article/GRDJE/V02/I05/0176/GRDJEV02I050176.pdf>

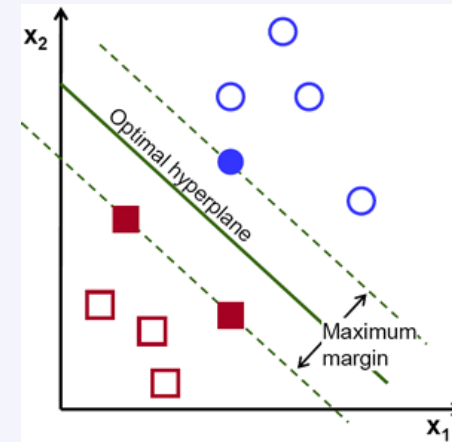


Aplikasi di bidang sains



Algoritma Machine learning

- Regression:
Ridge regression, Support Vector Machines, Random Forest, Multilayer Neural Networks, Deep Neural Networks, ...
- Classification:
Naive Base, , Support Vector Machines, Random Forest, Multilayer Neural Networks, Deep Neural Networks, ...
- Clustering:
k-Means, Hierarchical Clustering, ...



Frameworks

- Bahasa pemrograman

- Python

- R

- C++

- Anaconda

- Banyak libraries

- scikit-learn

- PyTorch

- TensorFlow

- Keras

- ...

classic machine learning

deep learning frameworks

Anaconda

- Anaconda adalah paket distribusi Python dari **Continuum Analytics** yang berisi paket Python ditambah beberapa paket tambahan untuk keperluan pemrograman data science, matematika hingga teknik dalam satu distribusi *platform* yang user friendly. File instalasi Anaconda dapat [diunduh di tautan ini](#).



Where packages, notebooks, projects and environments are shared.

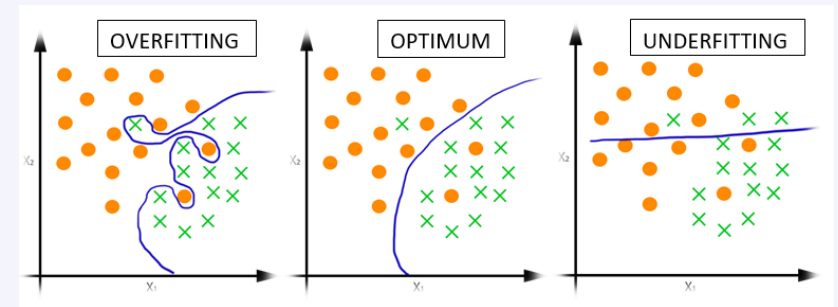
Jupyter Notebook

- Jupyter Notebook merupakan tool yang populer untuk mengolah data di python. Jupyter Notebook memungkinkan untuk mengintegrasikan antara kode dengan output di dalam satu dokumen secara interaktif.



Supervised learning: methodology

- Select model, e.g., random forest, (deep) neural network, ...
- Train model, i.e., determine parameters
 - Data: input + output
 - training data → determine model parameters
 - validation data → yardstick to avoid overfitting
- Test model
 - Data: input + output
 - testing data → final scoring of the model
- Production
 - Data: input → predict output



Contoh Aplikasi Machine Learning Pada Data BPS

- Klasifikasi rumah tangga miskin di provinsi Riau
- Data Susenas Maret 2020 Provinsi Riau
- Artikel terkait akurasi, presisi, recall, dan f1 score
<https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>



Demo

Klasifikasi rumah tangga miskin di provinsi Riau berdasarkan data Susenas Maret 2020 dengan menggunakan python



**BADAN PUSAT STATISTIK
KABUPATEN KUANTAN SINGINGI**

Materi Tambahan Terkait Data Science

- Materi terkait data science, berupa ebook, paper, course, dan video tutorial, Dapat di download di https://s.bps.go.id/DATASCIENCE_MISC



**BADAN PUSAT STATISTIK
KABUPATEN KUANTAN SINGINGI**



**Sensus
Penduduk
2020**

**Thank
you!**