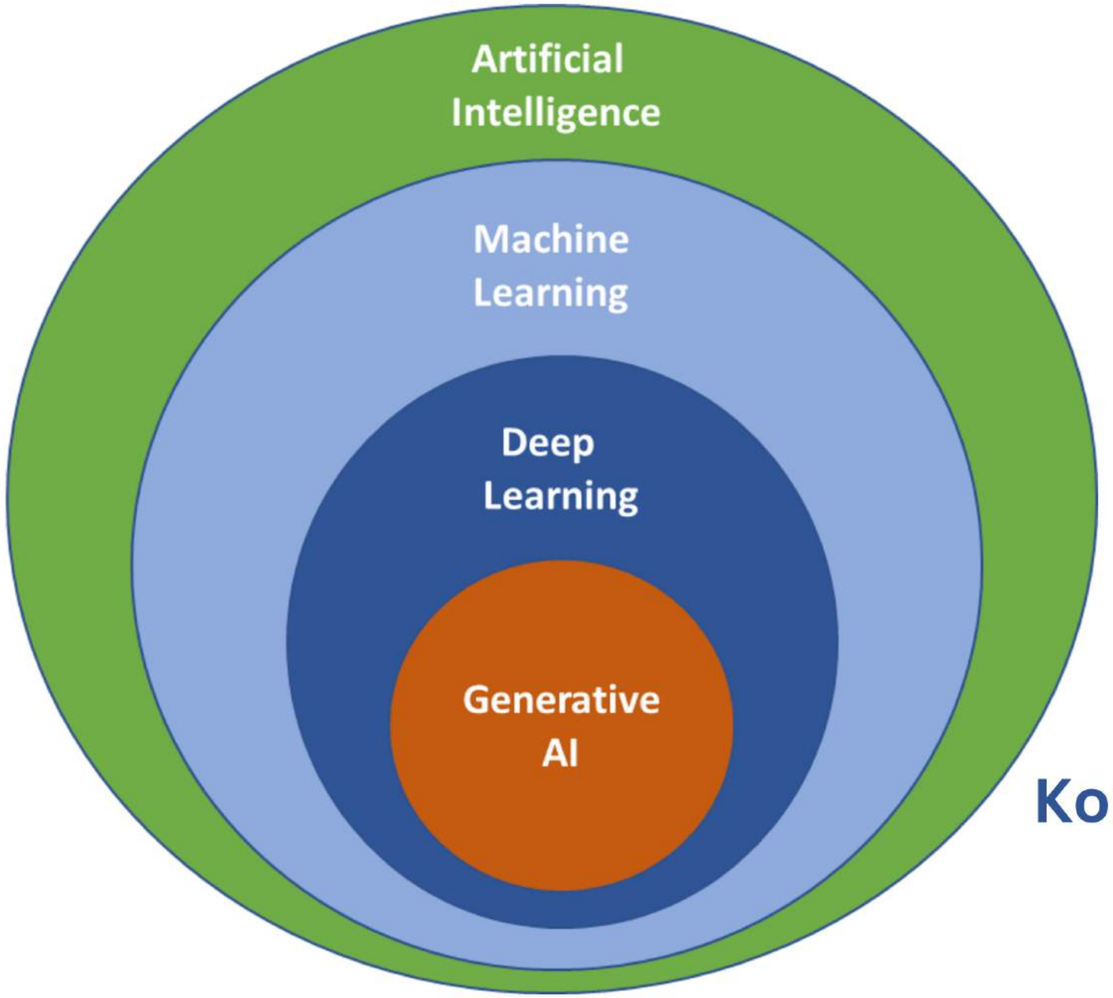
MACHINE LEARNING

Dhiya Salma Salsabila

PENGANTAR MACHINE LEARNING



Artificial Intelligence

Sistem yang dapat meniru kecerdasan manusia

Machine Learning

Sistem yang belajar dan membuat keputusan berdasarkan data tanpa program eksplisit

Deep Learning

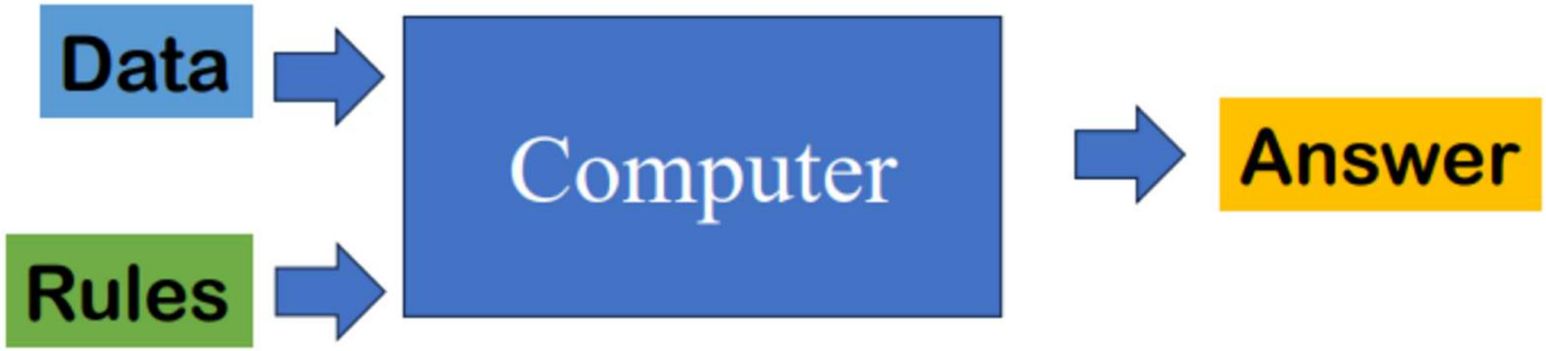
Konsep sistem berupa jaringan saraf tiruan untuk mempelajari pola dari data yang kompleks

Generative AI

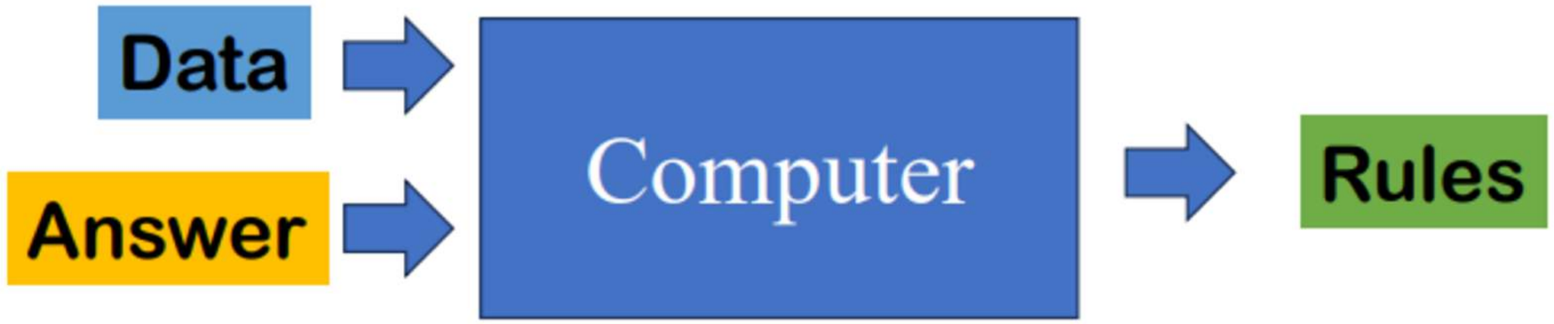
Sistem untuk membuat konten baru

DEFINISI MACHINE LEARNING

TRADITIONAL PROGRAMMING

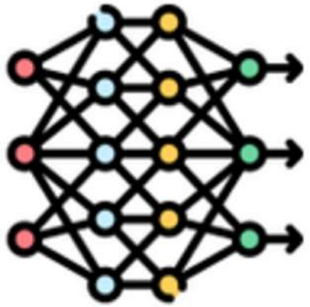


MACHINE LEARNING



TERMINOLOGI DALAM ML

Model



Model komputer yang dibuat berdasarkan data untuk membuat prediksi atau keputusan

Feature

Variabel yang digunakan untuk membuat prediksi

Data

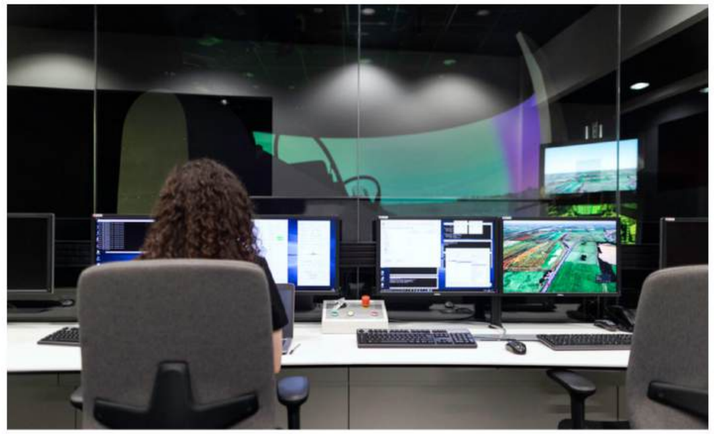
Data yang digunakan untuk melatih (*training*), validasi (*validation*), dan menguji model *(test)*

Label

Output yang diharapkan dalam supervised learning

PENGAMBILAN DATA

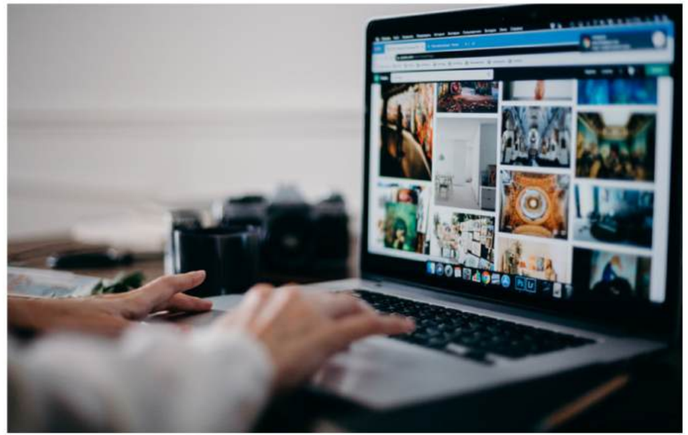
SIMULASI



EKSPERIMEN

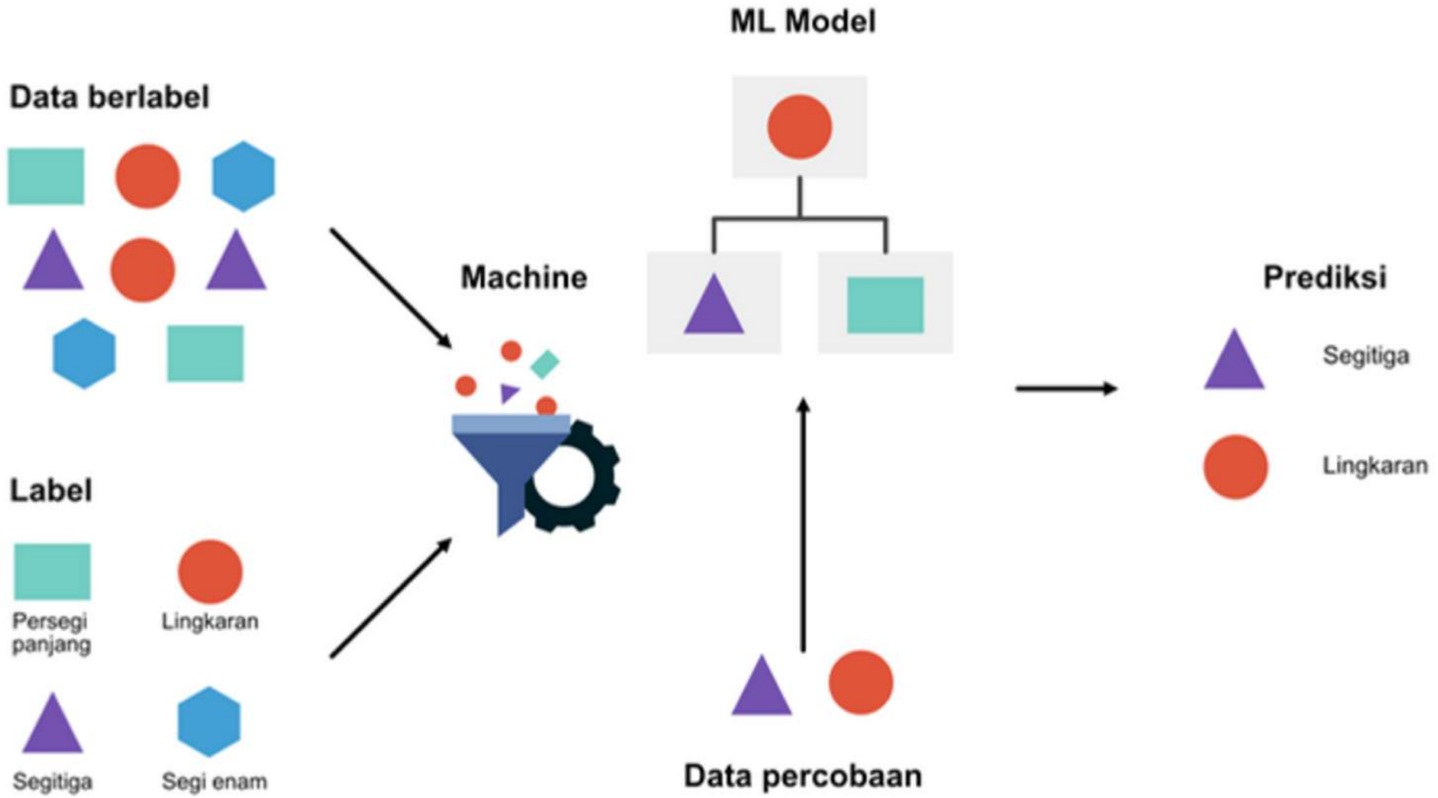
DATA LAPANGAN

INTERNET



JENIS TUGAS ML

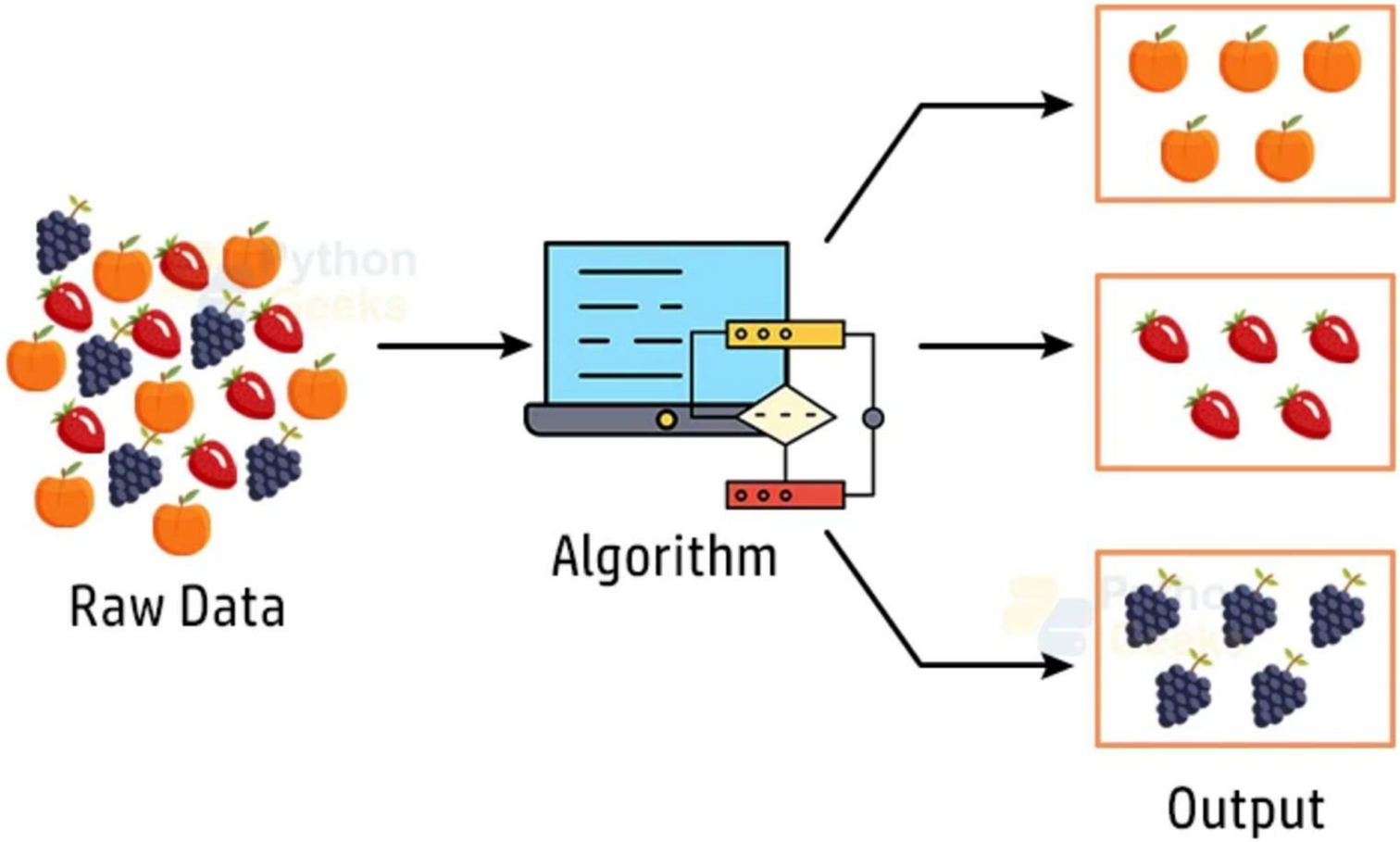
Klasifikasi



Memprediksi kelas atau kategori dari suatu kasus

JENIS TUGAS ML

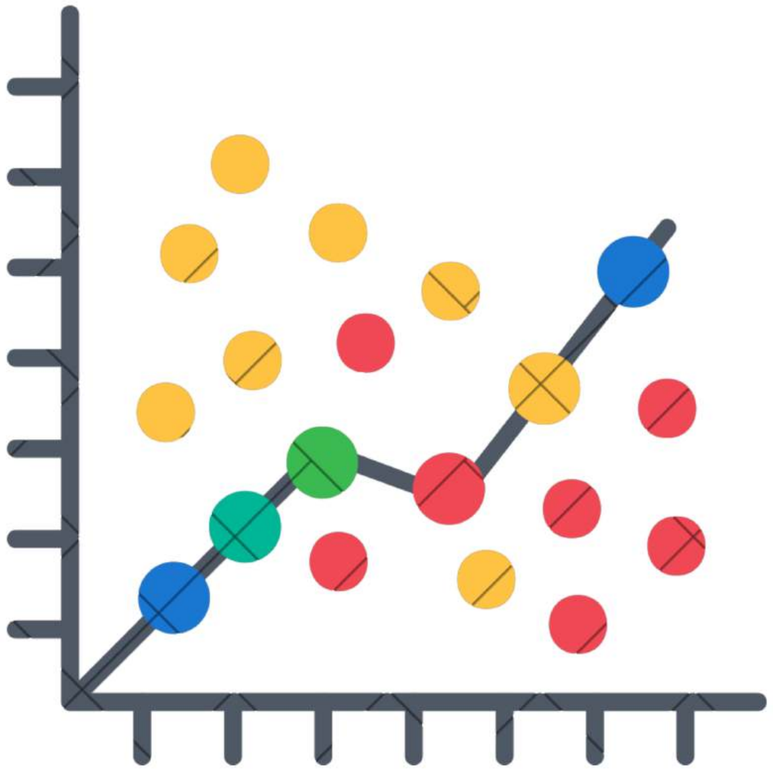
Clustering

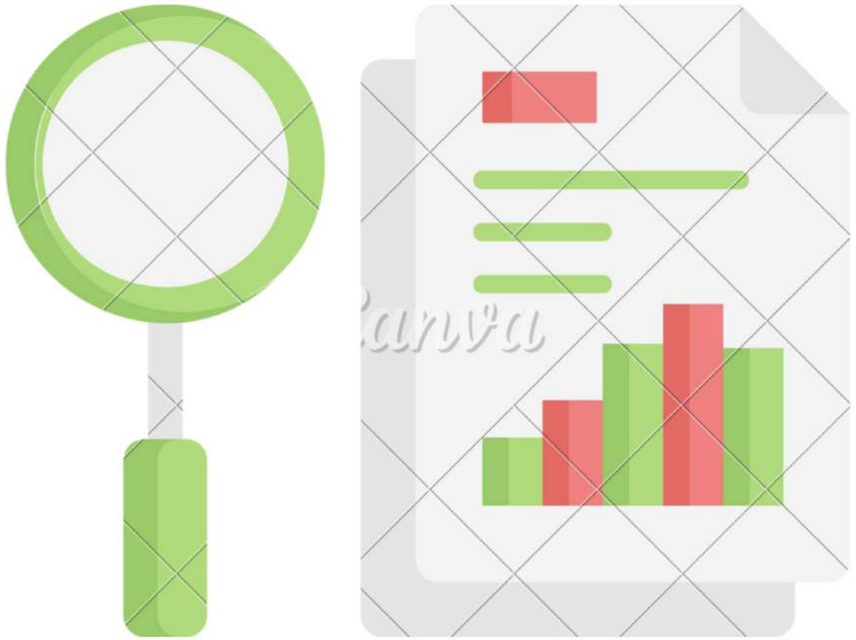


Mengelompokkan kasus berdasarkan kesamaan

JENIS TUGAS ML

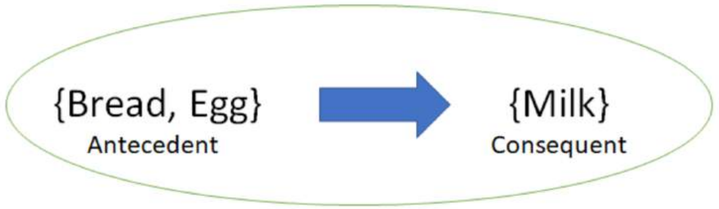
Regresi





Memprediksi nilai dari suatu kasus

JENIS TUGAS ML



Itemset = {Bread, Egg, Milk}

Asosiasi

Memprediksi kumpulan item/event yang akan terjadi secara bersamaan

TOOLS PENGEMBANGAN ML

Programming

Menggunakan bahasa pemrograman untuk membuat model.
Contoh : Python, R

No Code/Minimal Code

Menggunakan program drag and drop untuk membuat model
Contoh : Rapid Miner, Orange, WEKA

LLM

Menggunakan Generative AI untuk membuat model
Contoh : blackbox.ai, chatGPT

TIPE TIPE MACHINE LEARNING

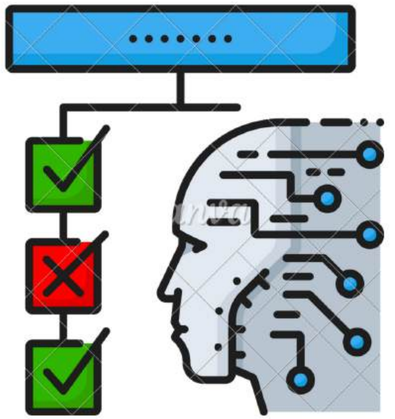
Supervised

- Belajar dengan “guru”, data training di beri label
- Memprediksi luaran yang diberikkan

Unsupervised

- Belajar tanpa “guru”, data training tanpa label
- Menentukan pola dari data

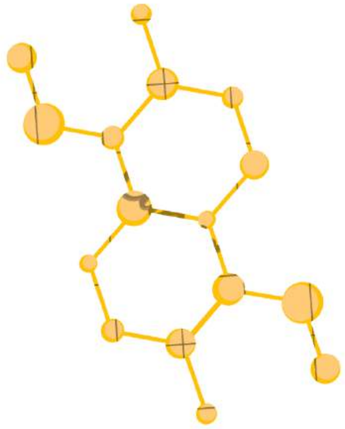
Reinforcement

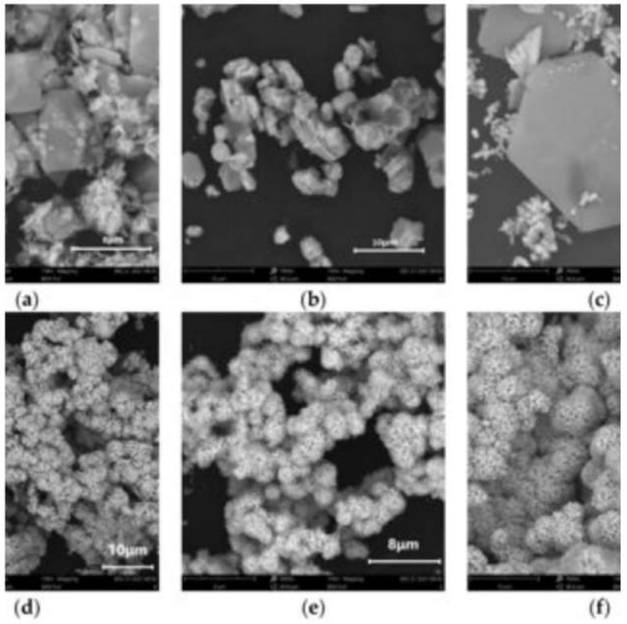


- Belajar dari pengalaman dan interaksi dengan lingkungan
- Memaksimalkan “reward”

PEMANFAATAN UTAMA ML DALAM BIDANG TEKNIK

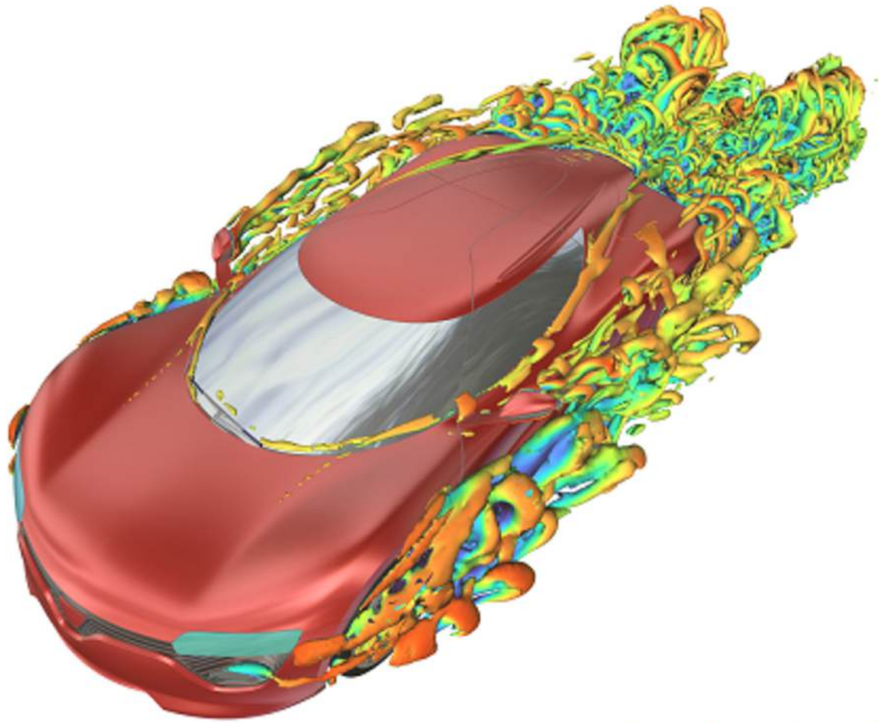
PERANCANGAN MATERIAL





OPTIMASI PROSES/DESAIN

cadence



PEMANFAATAN UTAMA ML DALAM BIDANG TEKNIK

INSPEKSI VISUAL

PREDIKSI MAINTENANCE

PEMANFAATAN ML DALAM NUKLIR

Journal of Environmental Radioactivity 241 (2022) 106772

Contents lists available at ScienceDirect

Journal of Environmental Radioactivity

journal homepage: www.elsevier.com/locate/jenvrad

Machine learning analysis of ^{137}Cs contamination of terrestrial plants after the Fukushima accident using the random forest algorithm

Igor Shuryak

Center for Radiological Research, Columbia University Irving Medical Center, 630 West 168th Street, VC-11-234/5, New York, NY, 10032, USA

ARTICLE INFO

Keywords:
Radioactive contamination
^{137}Cs
Radionuclides
Plants
Trees
Modeling
Machine learning
Random forests

ABSTRACT

Radioactive contamination of terrestrial plants was extensively investigated and quantitatively modeled after the Fukushima nuclear power plant accident. This phenomenon, which is important for ecosystem functioning and protection of human health, is influenced by multiple factors, including plant species, time after the accident, and climate. Machine learning algorithms such as random forests (RF) have a record of strong performance on large multi-dimensional data sets, but, to our knowledge, combined data on post-Fukushima plant contamination with radionuclides were not yet subjected to a machine learning analysis. Here we performed such analysis on two large published data sets: (1) ^{137}Cs activity concentrations in four common Japanese forest tree species. (2) Plant/soil ^{137}Cs concentration ratios in multiple perennial plant species. The goal was to show the usefulness of machine learning for identifying and quantifying the main trends of ^{137}Cs contamination in terrestrial plants. Each data set was split randomly into training and testing parts, RF was fitted and tuned on the training parts, and its performance was assessed on the testing parts by three metrics: coefficient of determination (R^2), root mean squared error, and mean absolute error. Synthetic noise variables and the Boruta algorithm were used in a customized procedure to identify the most important predictor variables, which consistently outperformed random noise. Good agreement between observations and RF predictions (*e.g.* R^2~0.9 on testing data) was obtained on both data sets. The effects of the most important predictors (*e.g.* time after the accident, ^{137}Cs land contamination level, and plant species) and interactions between them were quantified by partial dependence plots. These results of machine learning analyses of large data collections can help to complement previous modeling efforts, and to clarify the patterns of ^{137}Cs contamination of plants after the Fukushima accident.

Sumber

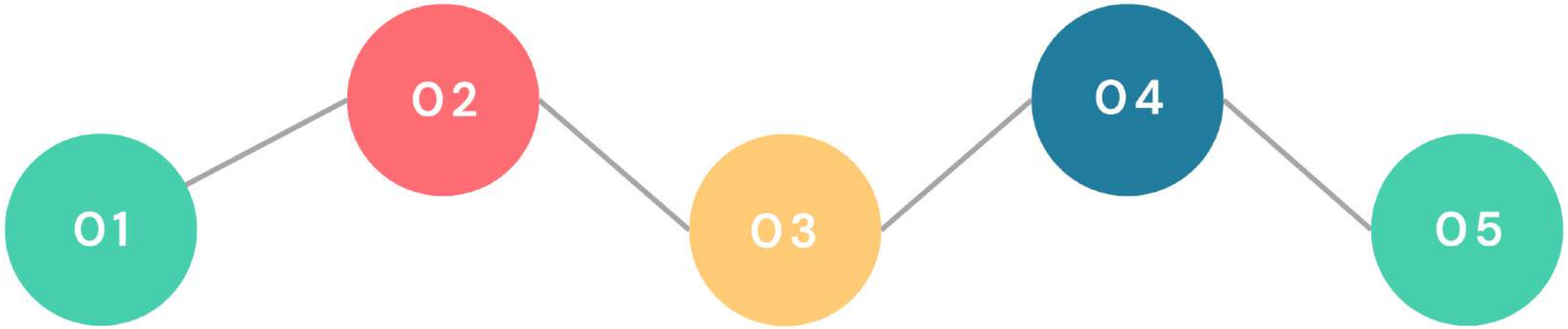
LANGKAH KERJA MACHINE LEARNING

Data Preparation

- Memilah data
- Membersihkan data
- Mengkonstruksi data
- Menentukan label data

Model Evaluation

- Mengevaluasi hasil pemodelan
- Melakukan review proses pemodelan



Data Understanding

- Mengumpulkan data
- Mempelajari data
- Memvalidasi data

Modeling

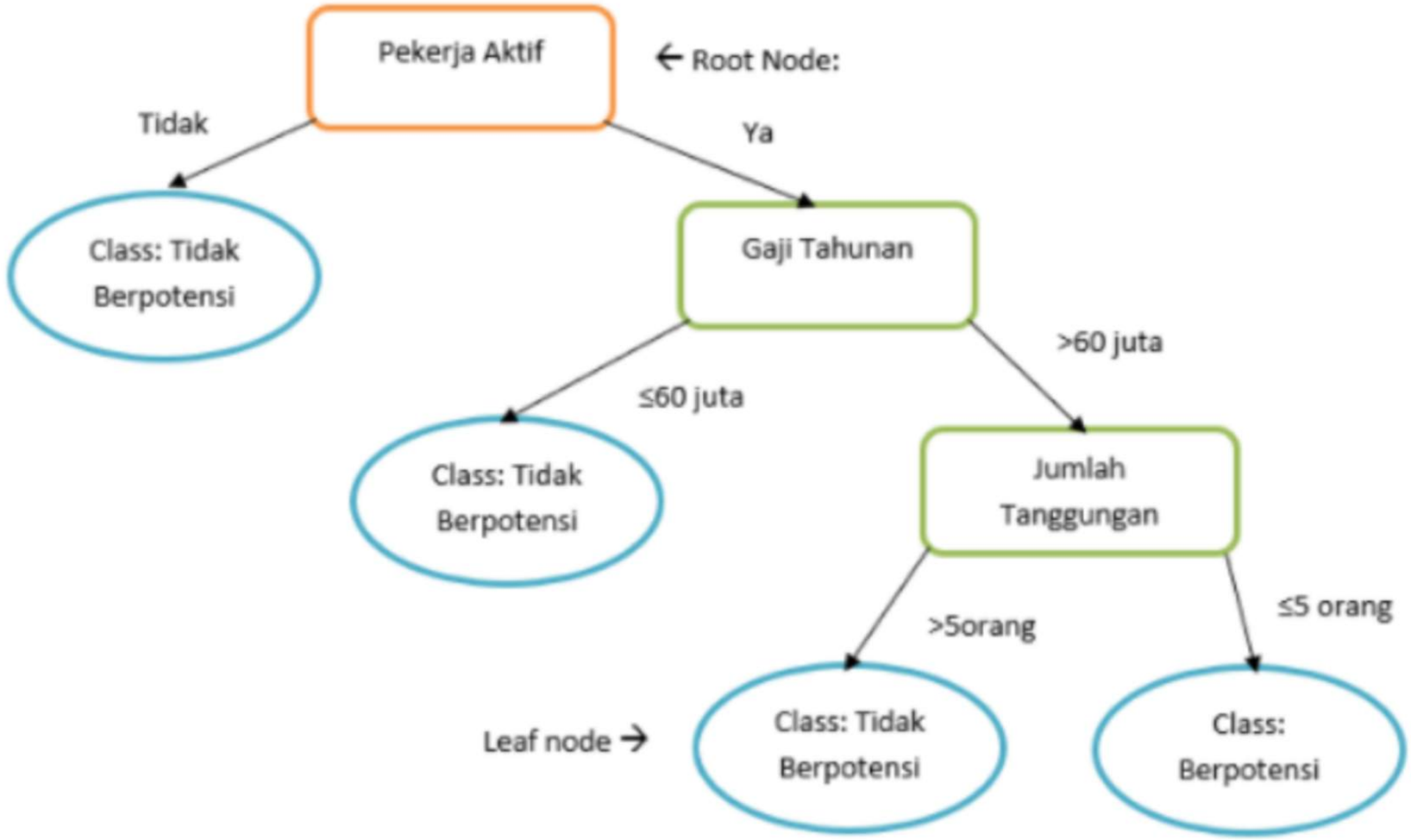
- Membangun skenario pengujian
- Membangun model

Deployment

- Membuat rencana deployment model
- Melakukan deployment model

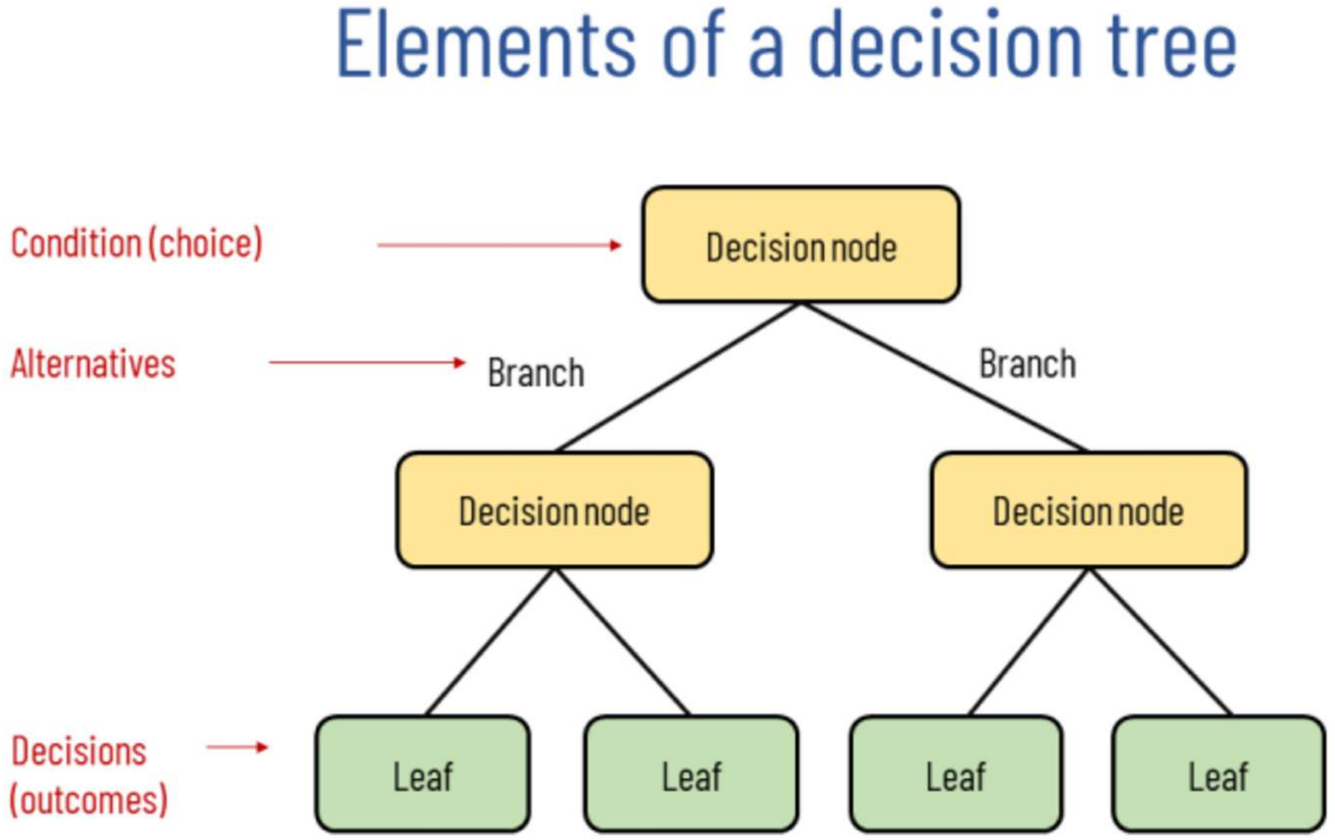
Decision Tree (DT)

Metode pembelajaran berbasis pohon yang digunakan untuk tugas klasifikasi dan regresi.
Metode ini bekerja dengan membagi dataset menjadi subset yang lebih kecil secara bertahap berdasarkan fitur-fitur yang relevan



PARAMETER DALAM DT

- **Criterion: Membantu pohon memilih fitur terbaik untuk memisahkan data.**
- **Max Depth: Menentukan kedalaman maksimum dari tree.**
- **Min Samples Split: Jumlah minimum sampel yang diperlukan untuk membagi node.**
- **Min Samples Leaf: Jumlah minimum sampel yang diperlukan untuk menjadi sebuah leaf node.**
- **Max Features: Menentukan jumlah fitur maksimum yang dipertimbangkan untuk membuat split di setiap node.**
- **Splitter: Menentukan strategi pembagian pada setiap node: best: Membagi berdasarkan fitur terbaik, random: Membagi berdasarkan fitur yang dipilih secara acak.**



THANK YOU