

Peta Jalan Ahli Al 2022

Achmad Benny Mutiara

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma

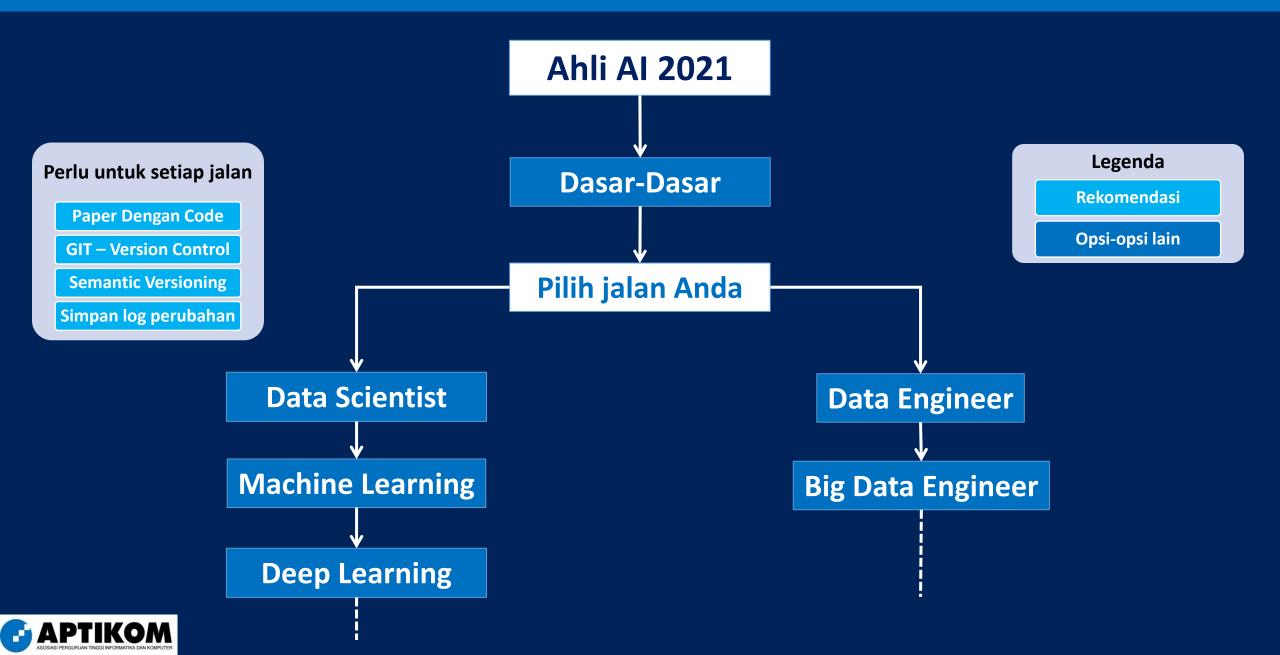
KETUM-APTIKOM

2023

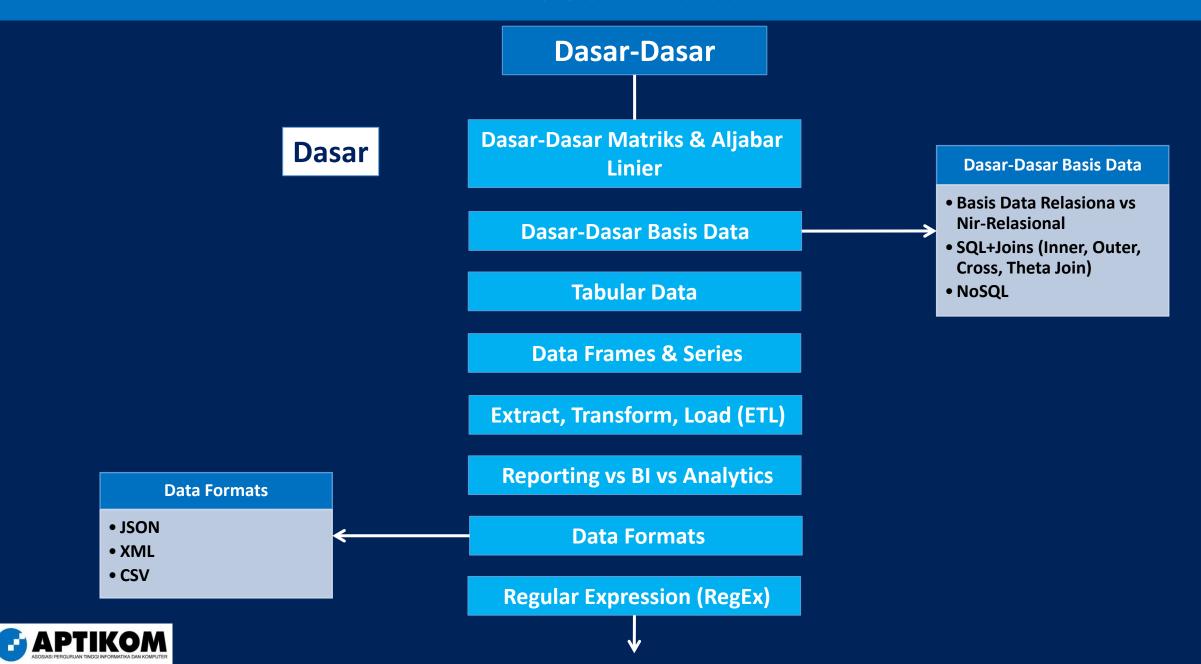
Pengantar



Pengantar

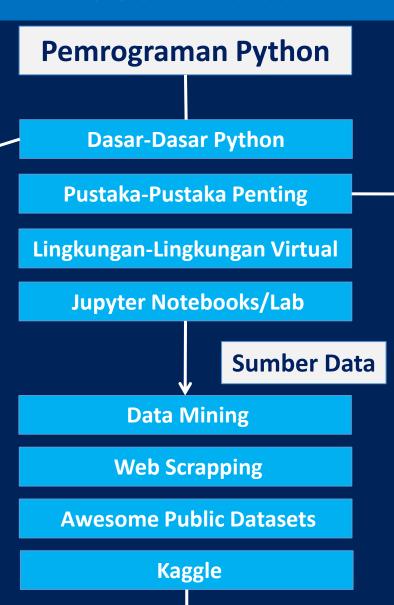






Dasar-Dasar Python

- Ekspresi
- Variabel
- Struktur Dat
- Fungsi
- Install Packages (via pip, conda atau sejenis
- Codestyle, mis. PEP8



Pustaka-Pustaka Penting

- Numpy
- Pandas



Exploratory Data Analysis/Data Munging /- Wrangling

Principal Component Analysis (PCA)

Dimensionality & Numerosity Reduction

Normalization

Data Scrubbing, Handling Missing Values

Unbiased Estimators

Binning sparse values

Feature Extraction

Denoising

Sampling



Data Scientist

Data Engineer

Peta Jalan Sains Data



Peta Jalan Sains Data

Teori Probabilitas

- Randomness, variabel acak dan sampel acak
- Distribusi probabilitas
- Probabilitas kondisional dan teorema Bayes
- Independensi statistik
- Variabel acak independen dan terdistribusi identik (IID)
- Fungsi distribusi kumulatif (cdf)
- Fungsi kerapatan probablitas (pdf)
- Fungsi massa probabilitas (pmf)

Important Laws

- Law of large numbers (LLM)
- Central limit theorem (CLT)

Estimasi

- Maximum likelihood estimation
- · Kernel density estimation

Pengujian Hipotesis

- p-Value
- Chi2 test
- F-test
- t-test

Data Scientist

Statistik

Teori Probabilitas

Distribusi Kontinyu

Distribusi Diskrit

Statistik Ringkasan (Summary Statistics)

Hukum Agung (Important Law)

Estimasi

Pengujian Hipotesis

Interval Kepercayaan (Confidence Interval (CI))

Metode Monte Carlo

Visualisasi

Distribusi Kontinyu

- Normal/Gaussian
- Uniform (kontinyu)
- Beta
- Dirichlet
- Eksponensial
- λ² (Chi-Squared)

Distribusi Diskrit

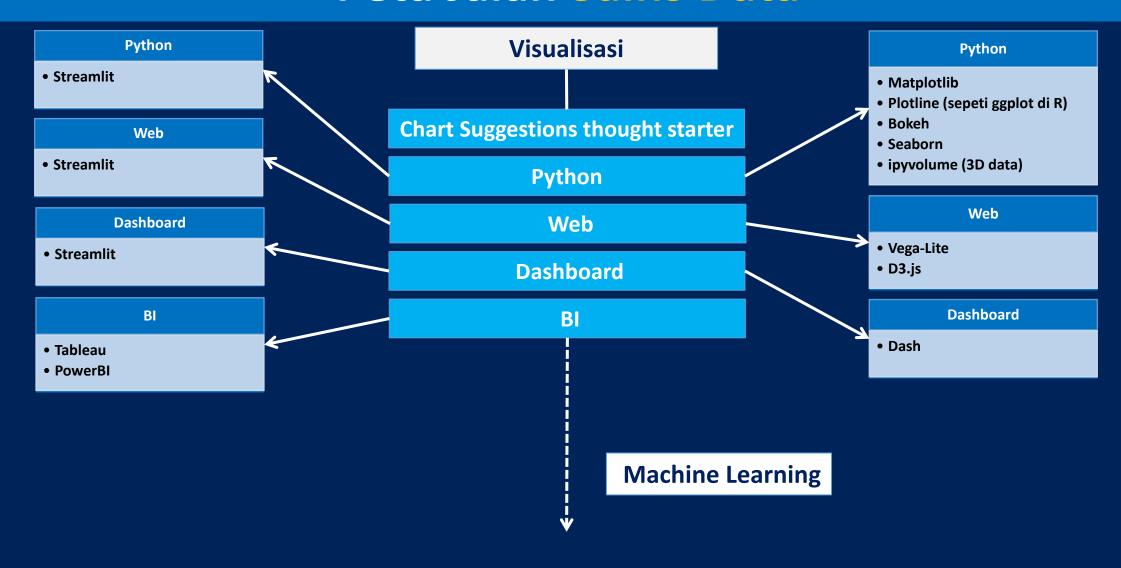
- Uniform (diskrit)
- Binomial
- Multinomial
- Hypergeometrik
- Poisson
- Gemotrik

Summary Statistics

- Expectation dan mean
- Variance dan standard deviation
- Covariance dan correlation
- Median dan quartile
- Interquartile range
- Percentile/quantile
- Mode



Peta Jalan Sains Data

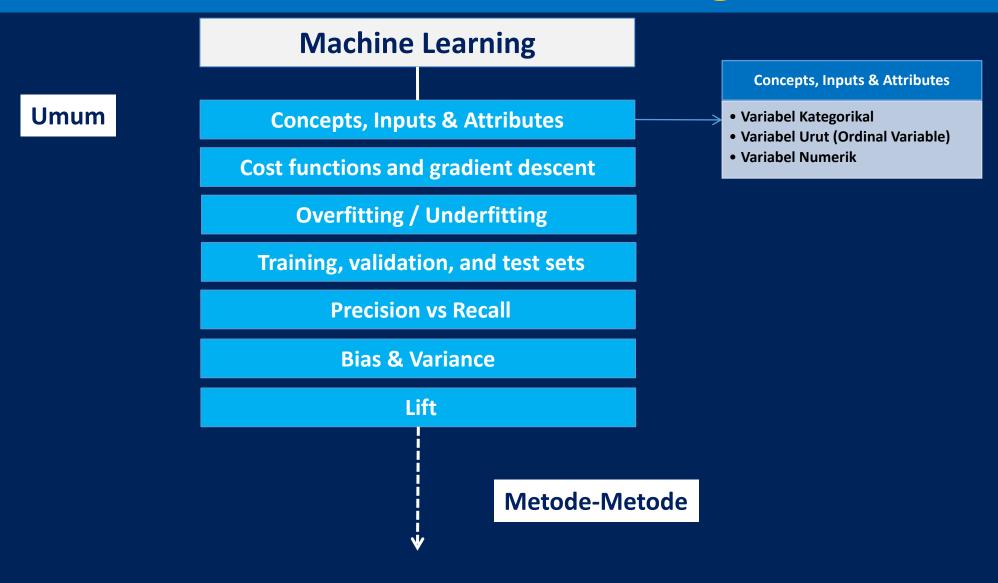




Peta Jalan Machine Learning

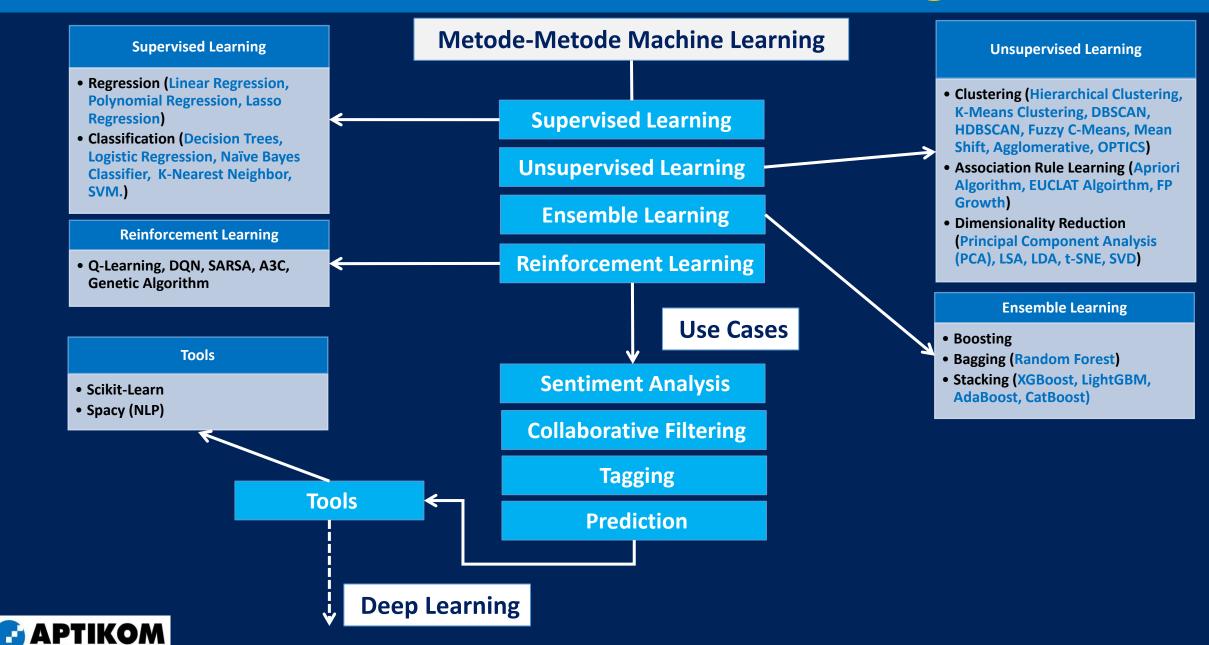


Peta Jalan Machine Learning





Peta Jalan Machine Learning





Paper-Paper

Neural Network

Deep Learning

Deep Learning Papers Reading Roadmap

Papers with code

Papers with code - state of the art

Understanding Neural Networks

Lost Functions

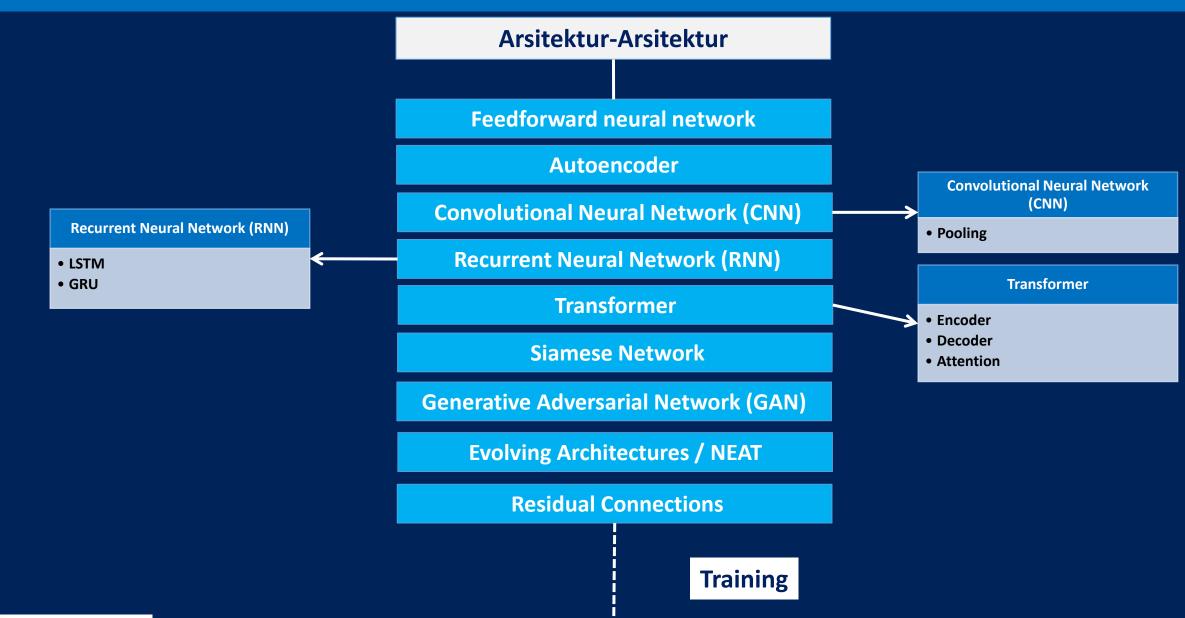
Activation Functions

Weight Initialization

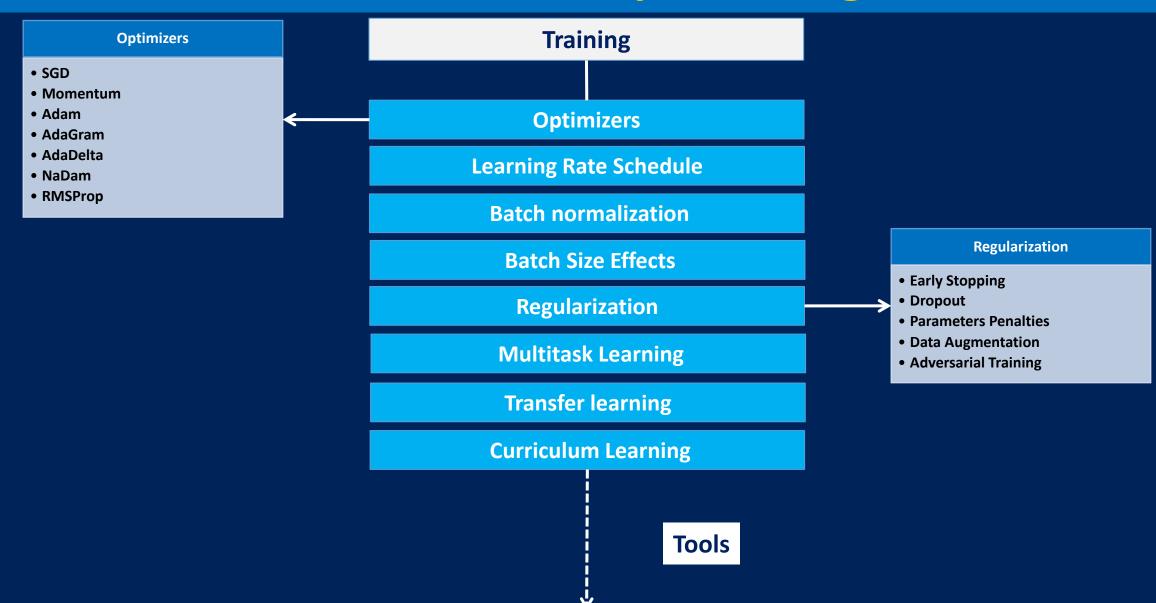
Vanishing / Exploding Gradient Problem

Arsitektur-Arsitektur

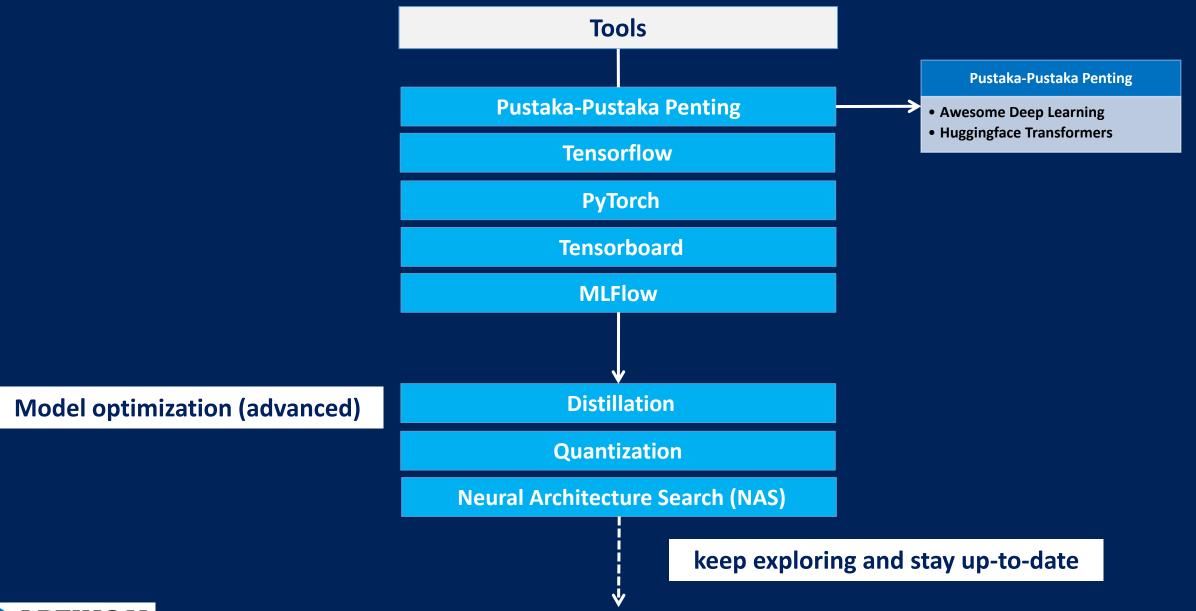




🔂 APTIKOM



🔂 APTIKOM





Peta Jalan Data Engineer



Peta Jalan Data Engineer

Data Engineer

Summary of Data Formats

Data Discovery

Data Source & Acquisition

Data Integration

Data Fusion

Transformation & Enrichment

Data Survey

OpenRefine

How much Data

Using ETL

Data Lake vs Data Warehouse

Dockerize your Python Application

keep exploring and stay up-todate



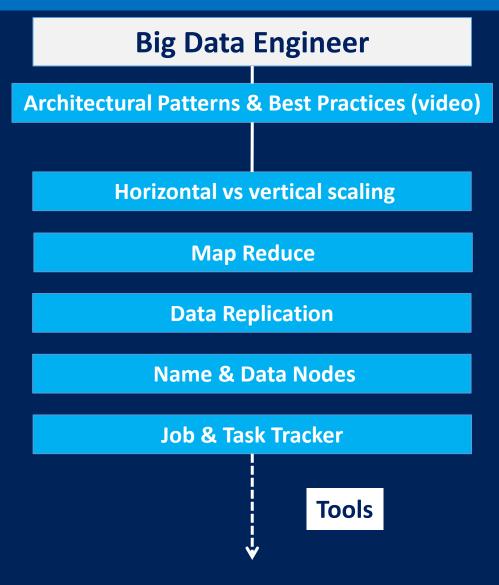
Peta Jalan Big Data Engineer



Peta Jalan Big Data Engineer

Big Data Architectures

Prinsip-Prinsip





Peta Jalan Big Data Engineer

Tools

Onnx
OpenVINO
MLFlow
Kafka & KSQL
Databases
Scalability
Cloud Services

Awesome Production ML

Check the Awesome Big Data List Hadoop (large data) Spark (in memory) RAPIDS (on GPU) Flume, Scribe: For Unstruct Data **Data Warehouse with Hive Elastic (EKL) Stack Avro** Flink Dask

Hadoop (large data)

- HDFS
- Loading data with Sqoop and Pig
- Storm: Hadoop Realtime

Cloud Services

Databases

• AWS SageMaker

• MongoDB, Neo4j

• Cassandra

- Google ML Engine
- Microsoft Azure Machine Learning Studio

keep exploring and stay up-todate

