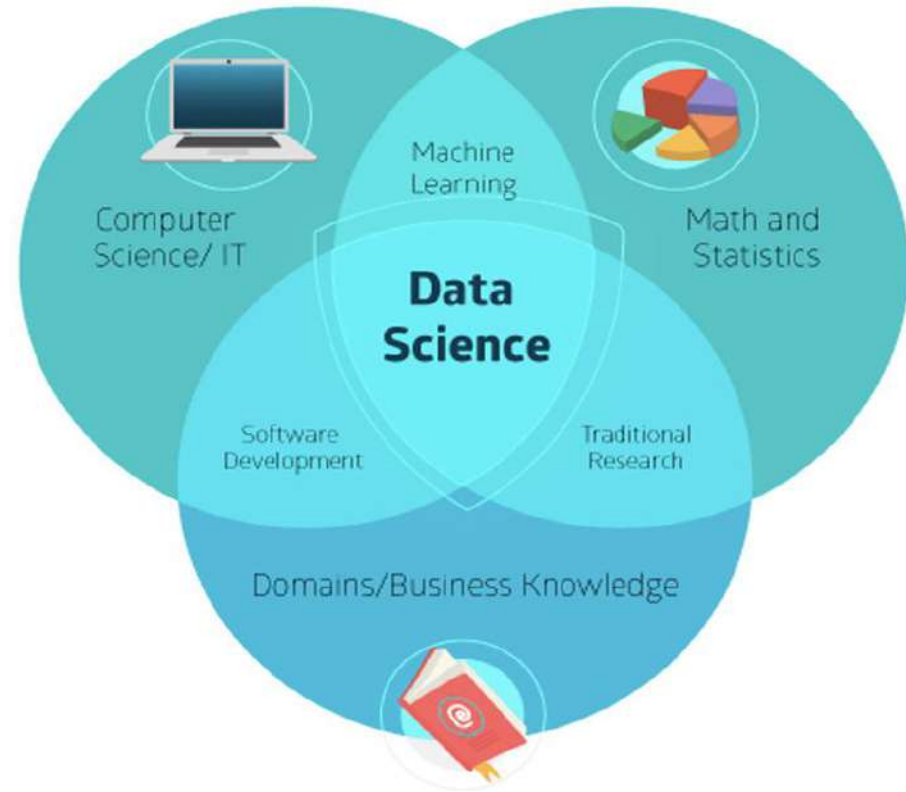


1. Pengantar Data Science

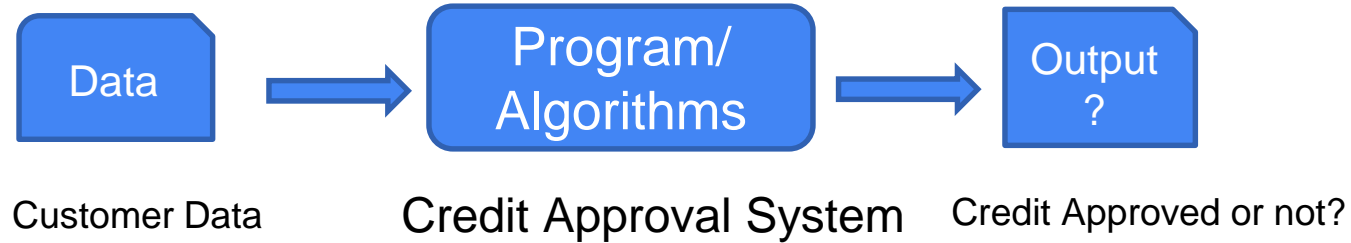
Data Science

Bidang keilmuan gabungan dari beberapa bidang yang menggunakan metoda dan proses saintifik, algoritma dan sistem untuk mendapatkan pengetahuan dan pemahaman mendalam (insights) dari data yang bersifat terstruktur maupun tidak terstruktur

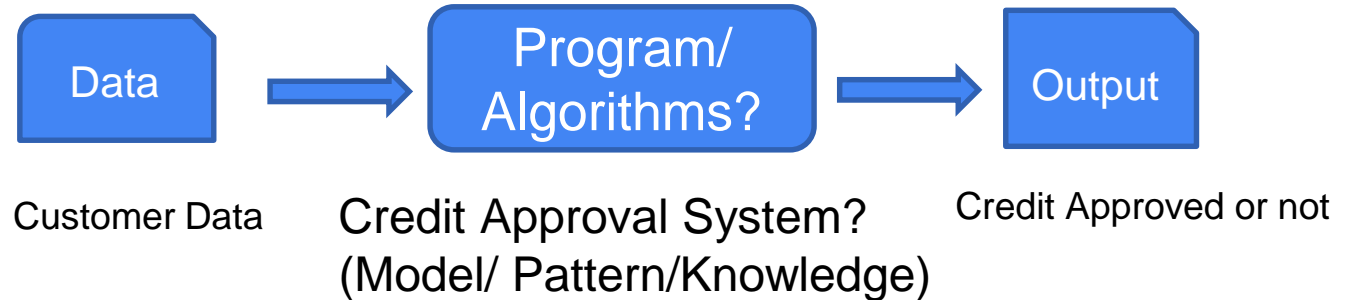


Data Science menggunakan Teknik Machine Learning

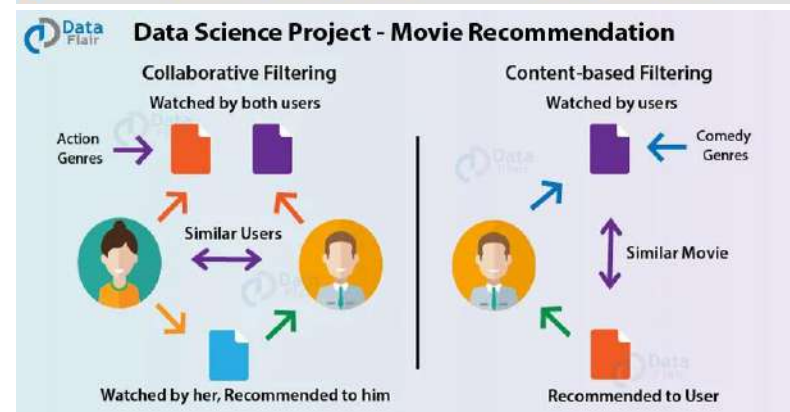
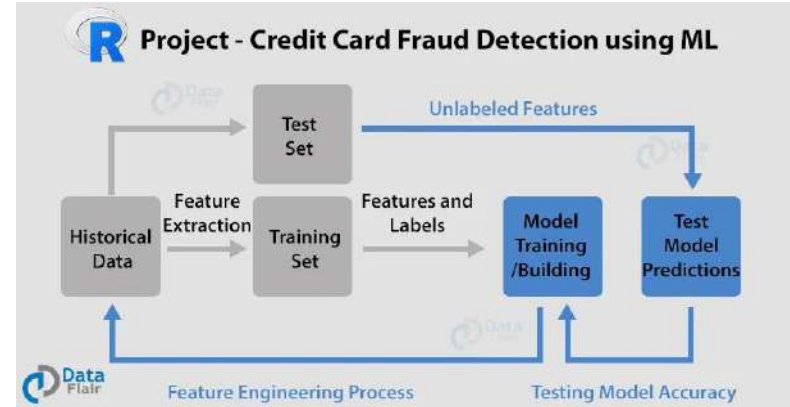
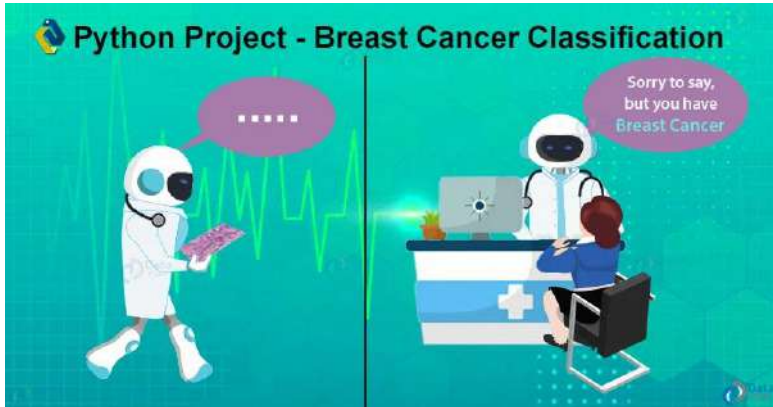
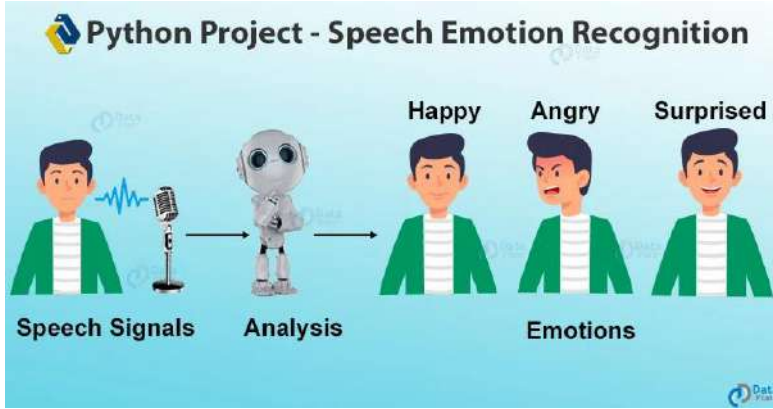
Conventional Program



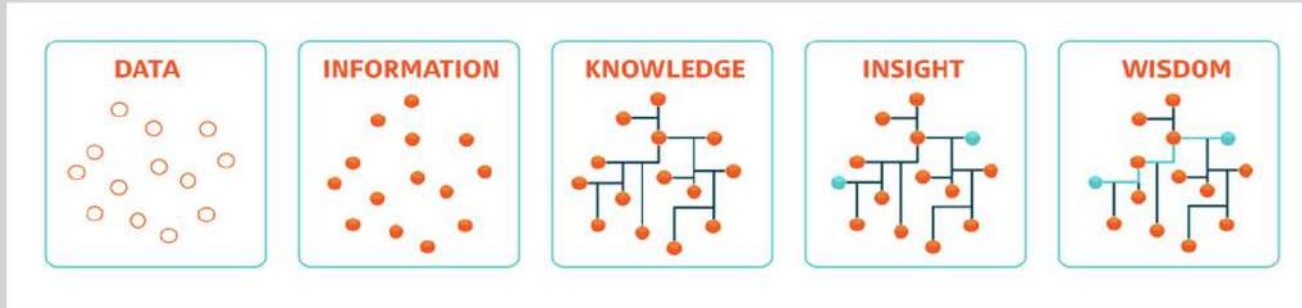
Machine Learning



(Contoh) Projek Data Science



Tujuan Data Science

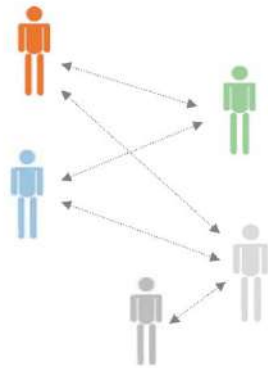
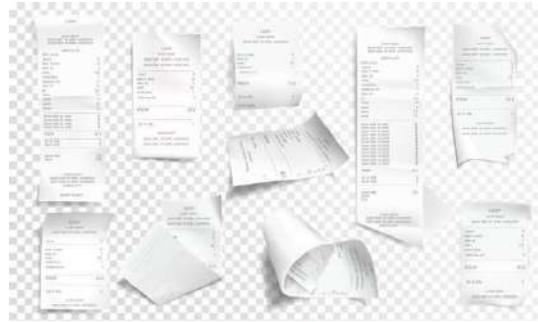


- Mendapatkan Pengetahuan dan *Insight*
- untuk
 - **Pengambilan Keputusan**
 - **Pembangunan Sistem AI** (Sistem Berbasis Pengetahuan)

Tujuan 1: Pengambilan Keputusan

- Data Retail suatu Minimarket
- Hasil analisis: Model Pembelian di Minimarket tsb.
 - Popok biasa dibeli bersama minuman bir
- Apa artinya?
- Berguna untuk apa? Penempatan barang di toko

Tujuan Data Science: Membantu Pengambilan Keputusan



“...90 days of point-of-sale data from Osco Drug stores - 1.2 million baskets...

...between 5pm and 7pm, customers tended to co-purchase beer and diapers.

...we have a correlation between beer, diapers and time...”

Increase co-purchases of products bought together, e.g. office desk supplies and the forgotten item

Cross-promotion using directional relationships, e.g. chips and dip

Price optimization, particularly during promotions

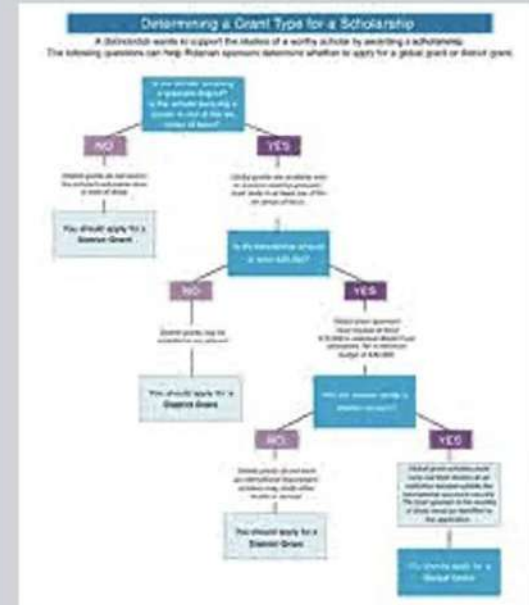
Inventory management, e.g. stocking the proper amount of the *dependent* product

Refine marketing, e.g. targeting segments based on their affinities



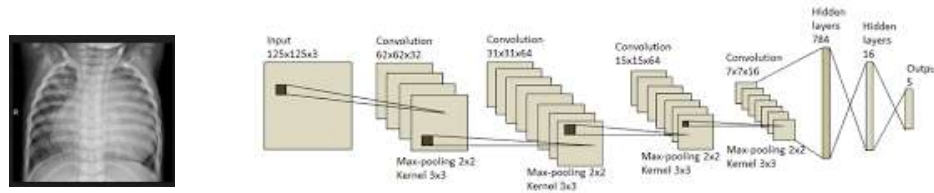
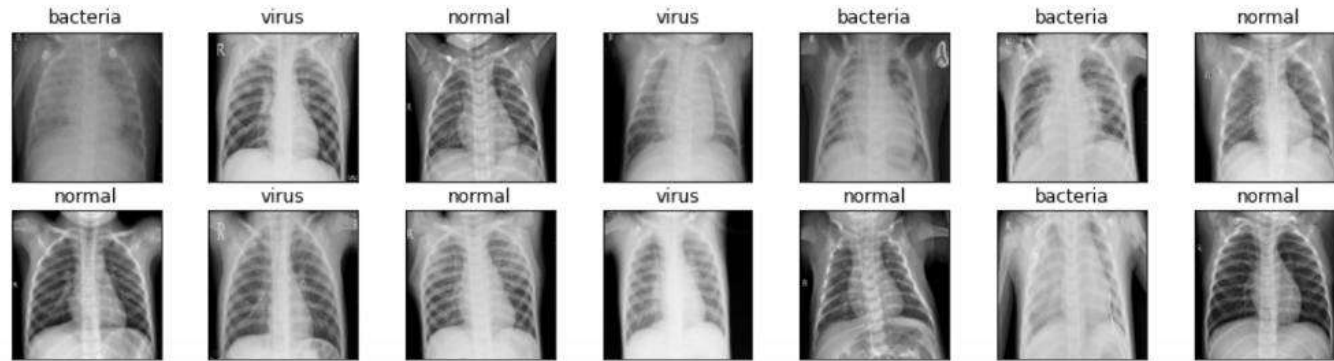
Tujuan 2: Pembuatan Aplikasi berbasis Artificial (1)

- Data Histori Pemberian Beasiswa
- Hasil analisis: Model Pemberian Beasiswa
- Pasang Model tersebut sebagai bagian dari Sistem Pemberian Beasiswa
 - Web based
 - Mobile Apps



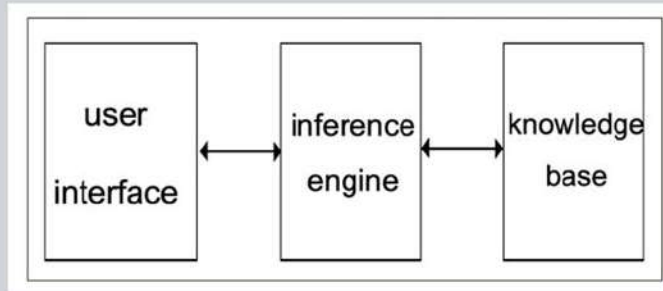
<https://rotary5630.org/decision-tree-scholar-grants/>

Tujuan Data Science: Pengambilan Keputusan Otomatis (Aplikasi Artificial Intelligence)



Normal, Virus, or Bacteria?

Tujuan 2: Pembuatan Aplikasi berbasis Artificial (2)



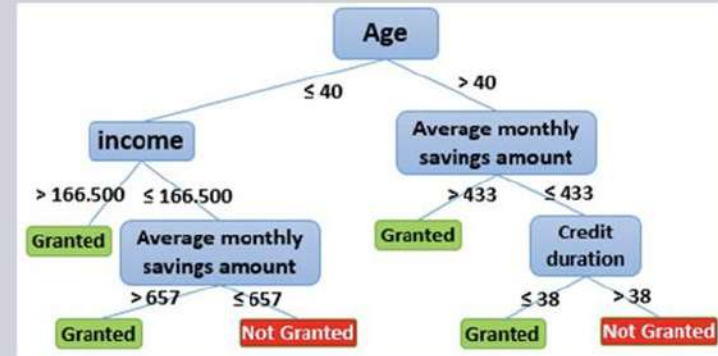
Data Permohonan



Aplikasi Persetujuan Pemberian Pinjaman



Terima/Tolak



Tujuan Tugas/ Task yang Biasa Dikembangkan

01

Descriptive:

Menjelaskan keadaan bisnis saat ini melalui data historis.

02

Diagnostic:

Menjelaskan mengapa suatu masalah terjadi dengan melihat data historis.

03

Predictive:

Memproyeksikan atau memprediksi hasil masa depan berdasarkan data historis.

04

Prescriptive:

Menggunakan hasil analitik prediktif dan pengetahuan lain dengan menyarankan upaya terbaik di masa depan.

Tujuan Tugas/ Task yang Biasa Dikembangkan

01

Descriptive:

Menjelaskan keadaan bisnis saat ini melalui data historis.

Macro-economic objectives	
1. Economic growth	High and sustainable economic growth.
2. Low Inflation	Low inflation. In UK target CPI 2% (±1)
3. Low Unemployment	Low unemployment, (less than 3% is often considered 'full employment')
4. Satisfactory balance of payments	Avoid 'large' deficit on the current account balance of payments.
5. Low government borrowing	For example: public sector debt less than 60%. Budget deficit less than 3%.
6. Stable exchange rate	Avoid destabilising devaluations/ appreciations
7. Minimise inequality	Avoid high inequality.
8. Protect environment	Important for very long run economic growth.

DATA KEUANGAN NEGARA DAN BISNIS



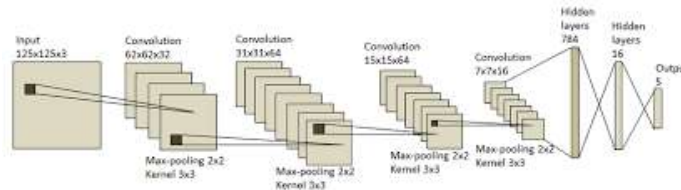
STATUS KEADAAN KEUANGAN SAAT INI (NOWCASTING)

Tujuan Tugas/ Task yang Biasa Dikembangkan

02

Diagnostic:

Menjelaskan mengapa suatu masalah terjadi dengan melihat data historis.



Normal, Virus, or Bacteria?

DATA GEJALA SEORANG PASIEN



PENYAKIT YANG DIDERITA

Tujuan Tugas/ Task yang Biasa Dikembangkan

03

Predictive:

Memproyeksikan atau memprediksi hasil masa depan berdasarkan data historis.

DATA KEUANGAN DAN AKUNTANSI ORGANISASI



PERKIRAAN PERUSAHAAN AKAN BANKRUT ATAU TIDAK (FORECASTING).

Tujuan Tugas/ Task yang Biasa Dikembangkan

04

Prescriptive:

Menggunakan hasil analitik prediktif dan pengetahuan lain dengan menyarankan upaya terbaik di masa depan.

DATA OPERASI MESIN



PERKIRAAN KERUSAKAN YANG TEJADI PADA MESIN TSB.



USULAN PERBAIKAN YANG HARUS DILAKUKAN

Jenis Tugas Analitiks yang bisqa Dilakukan

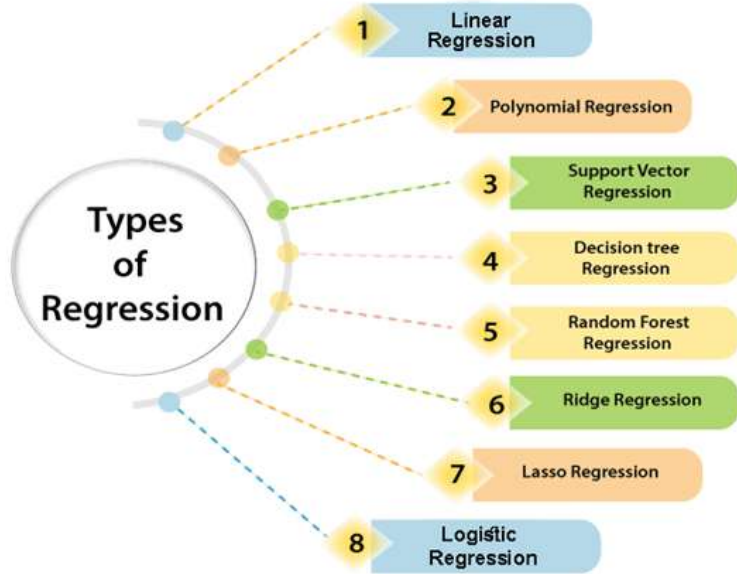
**Regresi /
Estimasi**

Klasifikasi

Klastering

Asosiasi

Regresi/ Estimasi



Uraian:

Memprediksi nilai kontinyu dari suatu kasus/ data

Contoh:

1. Memprediksi harga rumah berdasarkan karakteristik rumah tersebut
1. Memprediksi harga suatu saham besok hari
1. Menentukan dosis obat tertentu bagi pasien dan mengamati bagaimana efeknya pada
1. Memprediksi berapa lama waktu di Rumah sakit bagi seorang pasien

Klasifikasi

Uraian:

Memprediksi kelas atau kategori dari suatu kasus

Contoh:

1. Memprediksi apakah suatu pinjaman dapat dibayar atau tidak
1. Memprediksi apakah seorang pasien terjangkit Covid berdasar suara batuknya
1. Memprediksi apakah suatu perusahaan akan bangkrut



Klastering



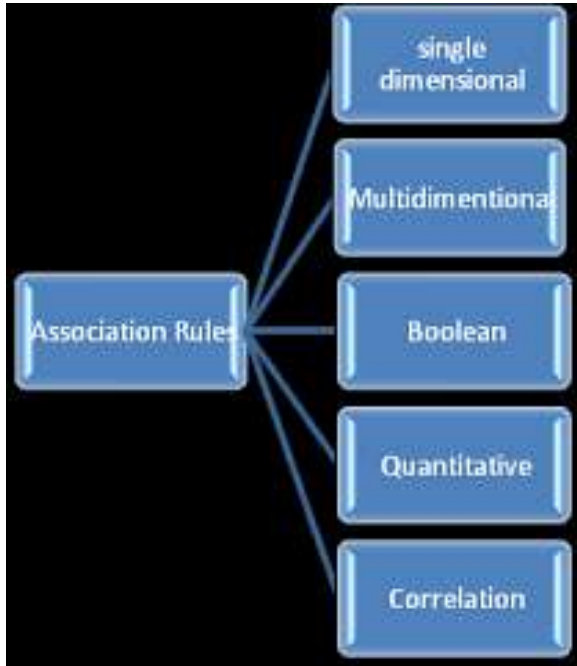
Uraian:

Mengelompokkan kasus berdasar kesamaan

Contoh:

1. Identifikasi kelompok Rumah Tangga berdasar pola pengeluaran bulanan
1. Segmentasi nasabah suatu bank

Asosiasi



Uraian:

Memprediksi suatu kumpulan item atau *event* yang akan terjadi secara bersamaan

Contoh:

1. Menemukan korelasi berbagai penyakit
1. Mencari kumpulan produk yang biasanya dibeli secara bersamaan

2. Metodologi Data Science

Mayoritas Proyek Pengembangan AI/DS Gagal

GARTNER
ESTIMATED

85%

of big data projects fail (2017). The initial estimation was 60% (GARTNER 2016)

THROUGH 2020

80%

of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization. (GARTNER 2018)

THROUGH 2022

20%

of analytic insights will deliver business outcomes. (GARTNER 2018)

EXECUTIVE
SURVEY

77%

respondents say that “business adoption” of big data and AI initiatives continues to represent a challenge for their organizations (NEWVANTAGE PARTNERS 2019)

<https://www.slideshare.net/PMI-Montreal/symposium-2019-gestion-de-projet-en-intelligence-artificielle>

Mayoritas Proyek Pengembangan AI/DS Gagal

- **PROBLEM** yang akan diselesaikan
 - Tidak Jelas; Problem salah; Over promising
- **DATA**
 - Tidak cukup (jumlah) atau tidak tepat (variabel)
 - Kualitas, tidak mencukupi
 - Tidak mengerti arti (semantic) data
 - Berbagai bias, hubungan antar variabel tidak dipikirkan (sampling, Fairness)
- **MODEL** yang dikembangkan
 - Terlalu kompleks; Tidak dimengerti
 - Metriks pengukuran tidak tepat
- **ALGORITHMS**
 - Terlalu sophisticated; Tidak dimengerti secara teknis
 - Tidak tepat
- **SUMBER DAYA MANUSIA**
 - One man show
 - Dukungan pemangku kepentingan kunci kurang

Perlu Metodologi Pengembangan

Pengembangan Sistem AI berdasar data

\neq

Data + Machine Learning (ML) Algorithms

Metodologi Pengembangan

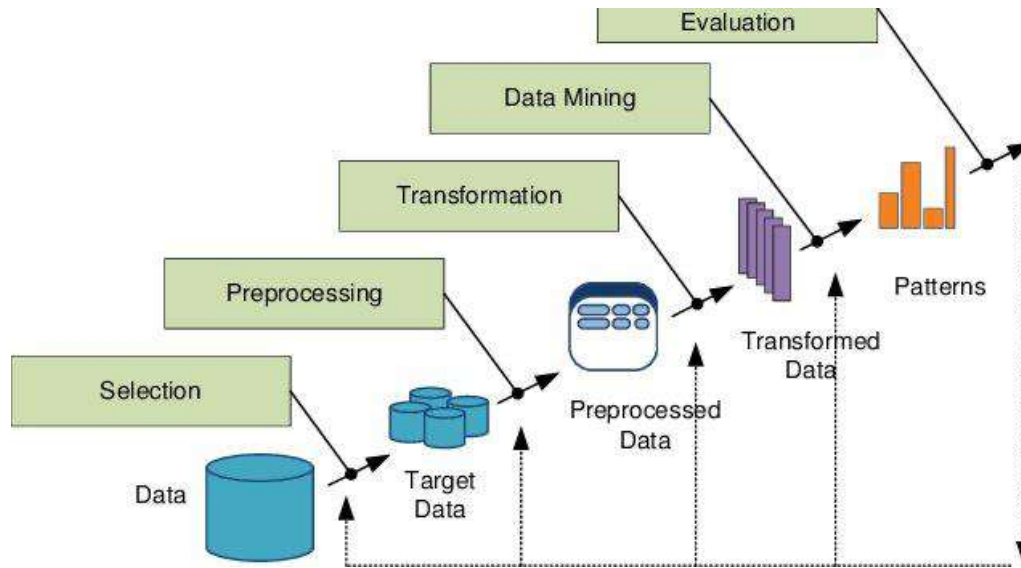
Metoda iterative yang dipakai untuk menyelesaikan masalah dengan menggunakan data dan data science melalui urutan langkah yang ditentukan

Jenis Metodologi

- Metodologi kegiatan Teknis
 - Kegiatan Data Science dianggap sebagai kegiatan teknis di bidang Teknologi Informasi ataupun Pengolahan data
- Metodologi kegiatan bisnis (dan teknis)
 - Kegiatan Data Science dianggap sebagai kegiatan bisnis yang terkait dengan penyelesaian suatu masalah organisasi menggunakan pendekatan data

Metodologi Teknis: Kegiatan DS/AI dianggap Kegiatan Teknikal

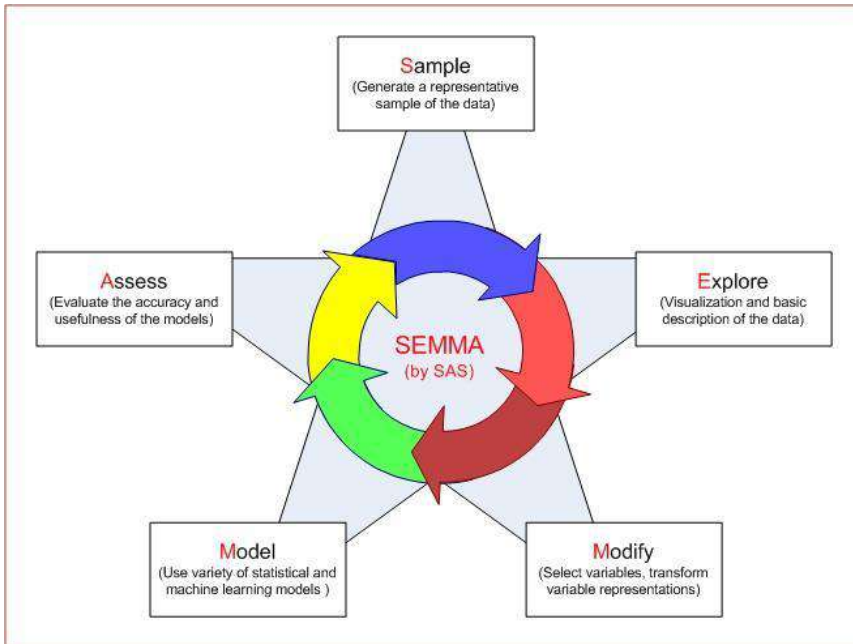
- Knowledge Discovery and Data Mining



<https://www.kdnuggets.com/gspubs/ai-mag-kdd-overview-1996-Fayyad.pdf>

Metodologi Teknis: Kegiatan DS/AI dianggap Kegiatan Teknikal

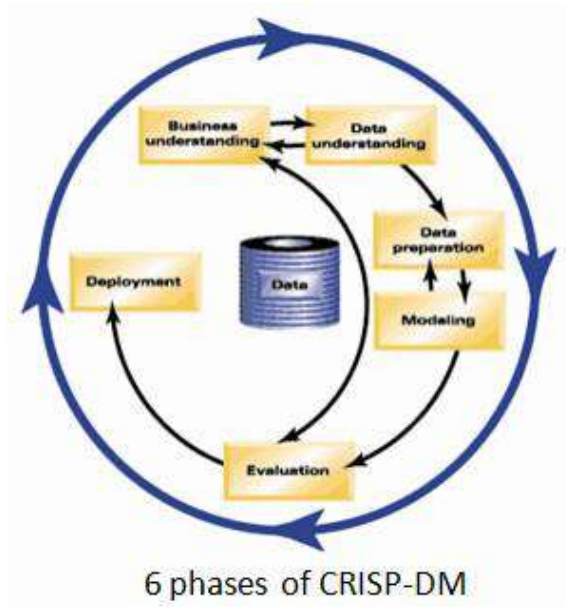
- SEMMA dari SAS Institute



<https://documentation.sas.com/?docsetId=emref&docsetTarget=n061bzurmej4j3n1jnjbijm1a2.htm&docsetVersion=14.3&locale=en>

Metodologi Lengkap: Kegiatan DS/AI dianggap Kegiatan Bisnis: Masalah Bisnis menjadi Masalah DS/AI

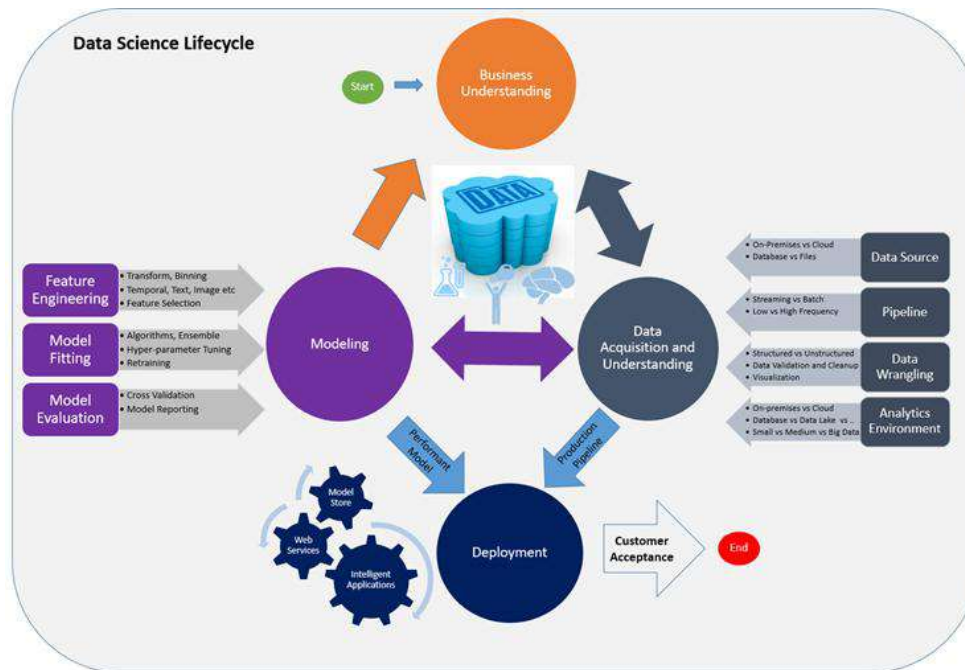
- CRISP-DM



<https://documentation.sas.com/?docsetId=emref&docsetTarget=n061bzurmej4j3n1jnj8bbjim1a2.htm&docsetVersion=14.3&locale=en>

Metodologi Lengkap: Kegiatan DS/AI dianggap Kegiatan Bisnis: Masalah Bisnis menjadi Masalah DS/AI

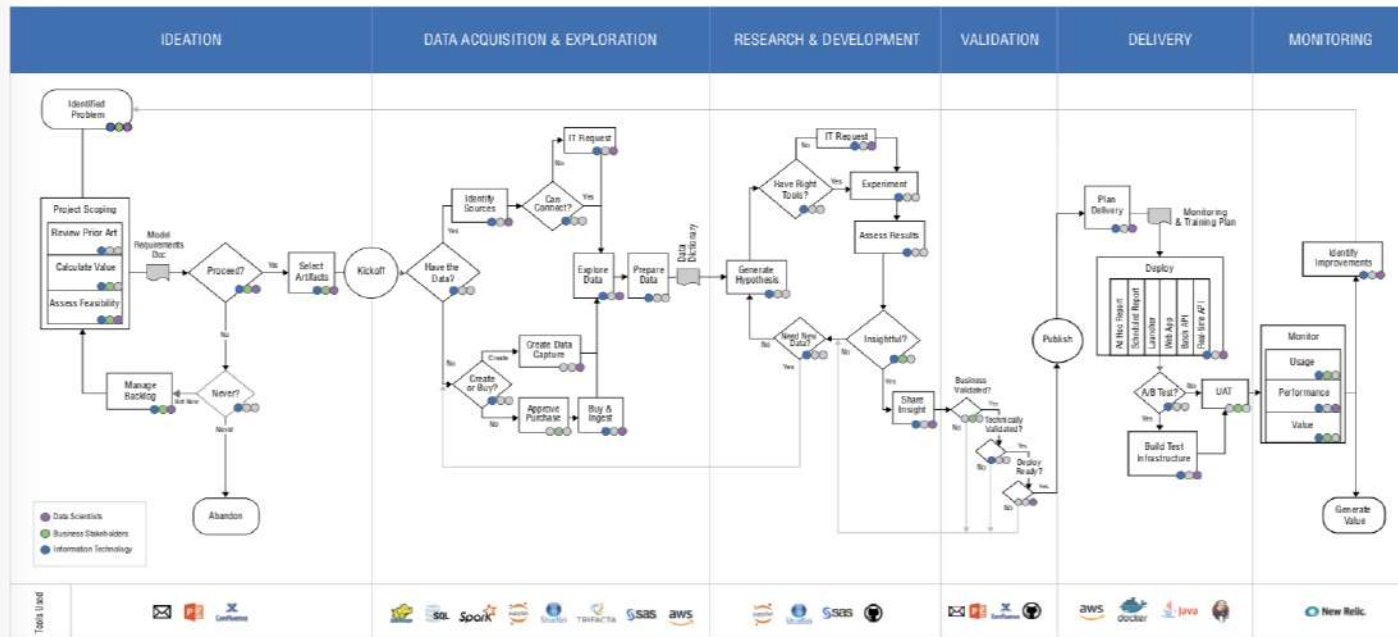
- Microsoft's Team Data Science Process



<https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>

Metodologi Lengkap: Kegiatan DS/AI dianggap Kegiatan Bisnis: Masalah Bisnis menjadi Masalah DS/AI

- Domino DataLab Methodology



Bagaimana di Indonesia?

Standard Kompetensi Kerja Nasional: KepMen Ketenagakerjaan No 299 thn 2020



MENTERI KETENAGAKERJAAN
REPUBLIK INDONESIA

KEPUTUSAN MENTERI KETENAGAKERJAAN
REPUBLIK INDONESIA
NOMOR 299 TAHUN 2020
TENTANG

PENETAPAN STANDAR KOMPETENSI KERJA NASIONAL INDONESIA
KATEGORI INFORMASI DAN KOMUNIKASI GOLONGAN POKOK AKTIVITAS
PEMROGRAMAN, KONSULTASI KOMPUTER DAN KEGIATAN YANG
BERHUBUNGAN DENGAN ITU (YBDI) BIDANG KEAHLIAN *ARTIFICIAL
INTELLIGENCE* SUBBIDANG *DATA SCIENCE*

TUJUAN UTAMA	FUNGSI KUNCI	FUNGSI UTAMA	FUNGSI DASAR
Menemukan pengetahuan, <i>insight</i> atau pola yang bermanfaat dari data untuk berbagai keperluan (orang mengambil keputusan atau sistem memproses lebih lanjut)	Menganalisis Kebutuhan (Requirements) Organisasi	<i>Business Understanding</i>	<ol style="list-style-type: none"> Menentukan objektif bisnis Menentukan tujuan teknis Membuat rencana proyek
		<i>Data Understanding</i>	<ol style="list-style-type: none"> Mengumpulkan data Menelaah data Memvalidasi data
	Mengembangkan model	<i>Data Preparation</i>	<ol style="list-style-type: none"> Memilah data Membersihkan data Mengkonstruksi data Menentukan Label Data Mengintegrasikan data
		<i>Modeling</i>	<ol style="list-style-type: none"> Membangun skenario pengujian Membangun model
		<i>Model Evaluation</i>	<ol style="list-style-type: none"> Mengevaluasi hasil pemodelan Melakukan review proses pemodelan
	Menggunakan model yang dihasilkan	<i>Deployment</i>	<ol style="list-style-type: none"> Membuat rencana deployment model Melakukan deployment model Melakukan rencana pemeliharaan Melakukan pemeliharaan
		<i>Evaluation</i>	<ol style="list-style-type: none"> Melakukan review proyek Membuat laporan akhir proyek

3. Peran Associate Data Scientist

Siapa yang terlibat dalam Kegiatan Data Science

01

Data Scientist

Mengembangkan model terbaik dari data untuk menjawab permasalahan bisnis

02

Data Engineer

Menyiapkan (big) data untuk diolah/ dimodelkan

03

Data Analyst

Menganalisis/ mencari insight dari data (dan menampilkannya dalam dashboard)

04

Project/ Product Manager

Mengelola projek/ produk berbasis data.

05

Domain Expert

Memberi arahan tentang domain permasalahan

06

IT People

Menyiapkan infrastruktur IT (terutama deployment)

Kompetensi (Minimal) bagi seorang Data Scientist

TUGAS Menganalisis dan mengembangkan model pengetahuan terbaik dari data

LEVEL KKNI 7

KODE UNIT	JUDUL UNIT KOMPETENSI
J.62DMI00.001.1	Menentukan Objektif Bisnis
J.62DMI00.002.1	Menentukan Tujuan Teknis DS
J.62DMI00.005.1	Menelaah Data
J.62DMI00.006.1	Memvalidasi Data
J.62DMI00.007.1	Menentukan Objek Data
J.62DMI00.008.1	Membersihkan Data
J.62DMI00.009.1	Mengkonstruksi Data
J.62DMI00.012.1	Membangun Skenario Model
J.62DMI00.013.1	Membangun Model
J.62DMI00.014.1	Mengevaluasi Hasil Pemodelan
J.62DMI00.015.1	Melakukan Proses Review Pemodelan

Kompetensi (Minimal) bagi seorang Associate Data Scientist

TUGAS Membantu Data Scientist dalam *dari* mengembangkan pengetahuan, pola, dan *insight dari data*

LEVEL KKNI 6

KODE UNIT	JUDUL UNIT KOMPETENSI
J.62DMI00.004.1	Mengumpulkan Data
J.62DMI00.005.1	Menelaah Data
J.62DMI00.006.1	Memvalidasi Data
J.62DMI00.007.1	Menentukan Objek Data
J.62DMI00.008.1	Membersihkan Data
J.62DMI00.009.1	Mengkonstruksi Data
J.62DMI00.010.1	Menentukan Label data
J.62DMI00.013.1	Membangun Model
J.62DMI00.014.1	Mengevaluasi Hasil Pemodelan

4. Langkah Pengembangan

1. Business Understanding: Menentukan Masalah Bisnis

Kasus: Kegagalan Kredit



Problem:

Bagaimana menurunkan NPL suatu bank

Pertanyaan:

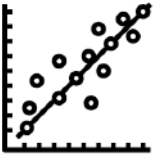
Bagaimana memperbaiki perhitungan Credit score

Measurable outcomes:

% Penurunan kredit gagal bayar

1. Business Understanding: Menentukan Tugas Analytics

Apa Tugas Analitiks yang perlu diselesaikan untuk menjawab permasalahan bisnis?

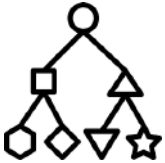


A. Regresi/Estimasi: Memprediksi nilai kontinyu dari kasus

- Prediksi harga rumah berdasar karakteristik tertentu
- Prediksi harga saham besok

1. Business Understanding: Menentukan Tugas Analytics

Apa Tugas Analitiks yang perlu diselesaikan untuk menjawab permasalahan bisnis?



B. Klasifikasi: Memprediksi kelas/ kategori dari kasus

- Prediksi kolektibilitas suatu pinjaman
- Prediksi kebangkrutan suatu perusahaan di tahun depan

1. Business Understanding: Menentukan Tugas Analytics

Apa Tugas Analitiks yang perlu diselesaikan untuk menjawab permasalahan bisnis?



C. Klustering: Mengelompokkan kasus berdasar kemiripan

- Segmentasi nasabah perbankan
- Pengelompokkan pasien yang mirip kasusnya

1. Business Understanding: Menentukan Tugas Analytics

Apa Tugas Analitiks yang perlu diselesaikan untuk menjawab permasalahan bisnis?



D. Asosiasi: Memprediksi kumpulan item/ kejadian yang biasa terjadi bersama

- Mencari barang jualan yang biasa dibeli bersama
- Menyusun portofolio saham

1. Business Understanding: Menentukan Tugas Analytics

Pengukuran Performansi tergantung Jenis Task Analytics

Metriks Performansi: Ukuran keberhasilan dari proses data science yang dilakukan

Contoh: Root Mean Squared Error (RMSE)

R-Square

Jackard Index

Log-loss

Precision

Recall

F1-Score

1. Business Understanding: Menentukan Tugas Analytics

Kasus: Kegagalan Kredit

Apa Tugas Analitiks yang perlu diselesaikan untuk menjawab permasalahan bisnis tersebut?



Problem:

Bagaimana menurunkan NPL suatu bank

Pertanyaan:

Bagaimana memperbaiki perhitungan Credit score

Tugas Analitik:

Klasifikasi

Performance Metrics:

F1-Score

1. Business Understanding: Menentukan Kebutuhan Data

Data apa yang diperlukan?
Dari mana bisa diperoleh?

Struktur Data: Bagaimana deskripsi data (atribut) yang diperlukan

Jumlah Data: Berapa banyak (record) data yang diperlukan

Sumber Data: Darimana data bisa diperoleh? Apakah sudah tersedia?

- Internal: Sistem Informasi/ ERP, Excel, dokumen
- Eksternal: Web API, Web Scraping
- Dataset via public data
- Dataset via open data

1. Business Understanding: Merencanakan Manajemen Proyek

Bagaimana rencana pelaksanaan proyeknya?

Cost Benefit Analysis: Apakah menguntungkan untuk melakukannya?

Situation Assessment: Analisa keadaan organisasi

Project Plan: Scope (WBS), Time, Schedule, Tim Pengembang

2. Data Understanding :

Mengenali/ mendalami data yang dimiliki

01

Mengumpulkan Data

Mengumpulkan Data yang Diperlukan

Jumlah Data (Baris dan Kolom)
Deskripsi data

02

Menelaah data

Menganalisa data secara eksploratif

Karakteristik atribut/ fitur
Keterkaitan antar data

03

Memvalidasi Data

Menilai kesesuaian kualitas data dengan masalah yang akan dipecahkan

Kualitas Data

Python Libraries

Scientific Computing

Pandas (Data structure and tools)

Numpy (Array and matrices)

Scipy (Integrals, solving differential equations, optimization)

Visualization

Matplotlib (plots & graphs, most popular)

Seaborn (plots : heat maps, time series, violin plots)

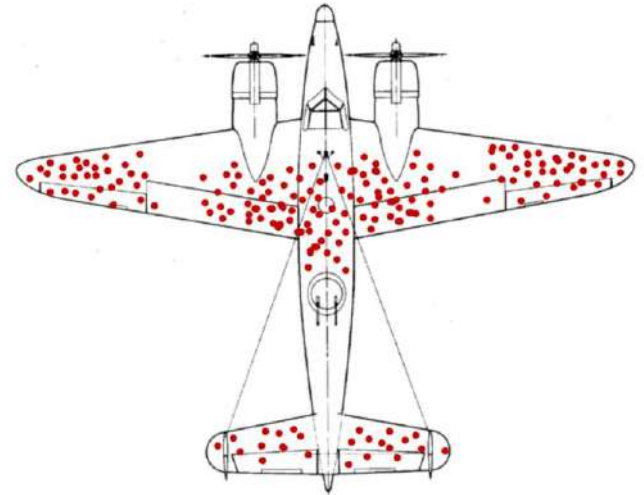
Algorithmic

Scikit-learn (Machine learning : regression, classification, etc)

Statsmodels (Explore data, estimate statistical models, perform statistical test)

2. Data Understanding : Mengapa Perlu Mengenal/ mendalami data yang dimiliki

- The United States armed forces faced a dilemma during the war, because returning bomber planes were riddled with bullet holes and they needed better ways to protect them
- “Where should they put it?”
- When they plotted out the damage these planes were incurring, it was spread out, but largely concentrated around the tail, body and wings.
- Should they upgrade these sections?



2. Data Understanding : Mengumpulkan Data

Mengumpulkan Data yang Diperlukan

Jumlah Data: Berapa banyak yang dapat diperoleh

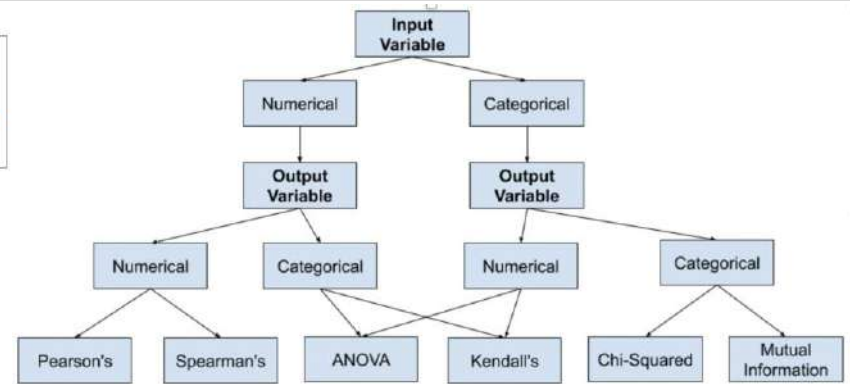
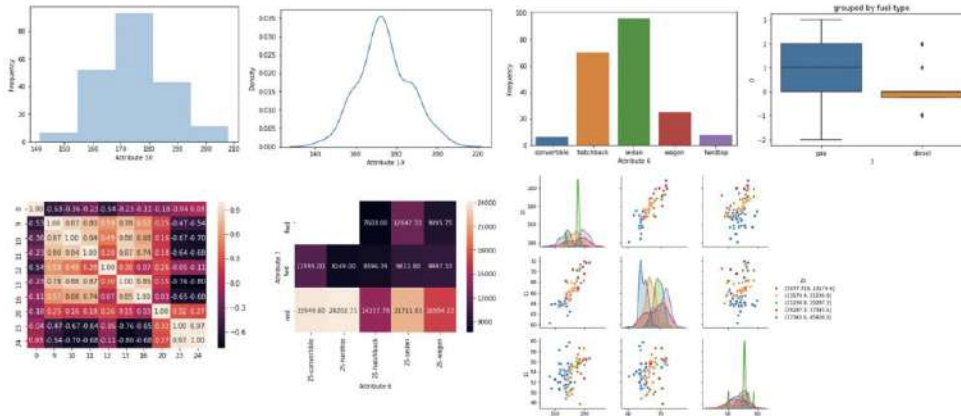
Deskripsi Data: Penjelasan arti atribut/ fitur

2. Data Understanding : Menelaah Data

Menganalisa data secara eksploratif (EDA)

Karakteristik Atribut: Deskripsi data (atribut) yang diperoleh

Keterkaitan antar Data: Analisis statistik korelasi, Anova, Chi-Squared,...



Copyright © MachineLearningMastery.com

2. Data Understanding : Memvalidasi Data

Menilai kesesuaian kualitas data dengan masalah yang akan dipecahkan

Laporan Kualitas Data:

- Ukuran Data (Atribut/ fitur dan Jumlah record)
- Deskripsi statistical atribut
- Relasi antar atribut (dan label)
- Visualisasi data

3. Data Preparation :

Memperbaiki kualitas data untuk Pemodelan

01

Memilih dan memilah data

Memilih data yang akan dipergunakan

Rekord terpakai
Atribut terpakai

02

Membersihkan Data

Meminimalkan noise (tidak lengkap, salah)

Data lengkap
Data yang diperbaiki
Data Pecilan

03

Mengkonstruksi data

Menambahkan fitur dan transformasi data

Fitur tambahan (Feature Engineering)
Transformasi data (standardisasi, transformasi)

04

Integrasi Data

Menggabungkan data

Gabungan data

4. Modeling : Mengembangkan Model (Pengetahuan)

01

Membangun Skenario Pemodelan

Membuat strategi pencarian model terbaik

- Pemilihan Algoritma Machine Learning (ML)
- Pembagian Data
- Penentuan Langkah Eksperimen

02

Membangun model

Mengembangkan model dengan Teknik ML

- Eksekusi Algoritma
- Pengaturan Parameter
- Pengukuran Performance Metrics

4. Modeling : Membangun Skenario Pemodelan

Membuat strategi pencarian model terbaik

Pemilihan Algoritma Machine Learning (ML)
Pembagian Data
Penentuan Langkah Eksperimen

4. Modeling : Membangun Skenario Pemodelan

Membuat strategi pencarian model terbaik

A. Memilih Algoritma: Disesuaikan dengan Tugas Analytics yang dipilih

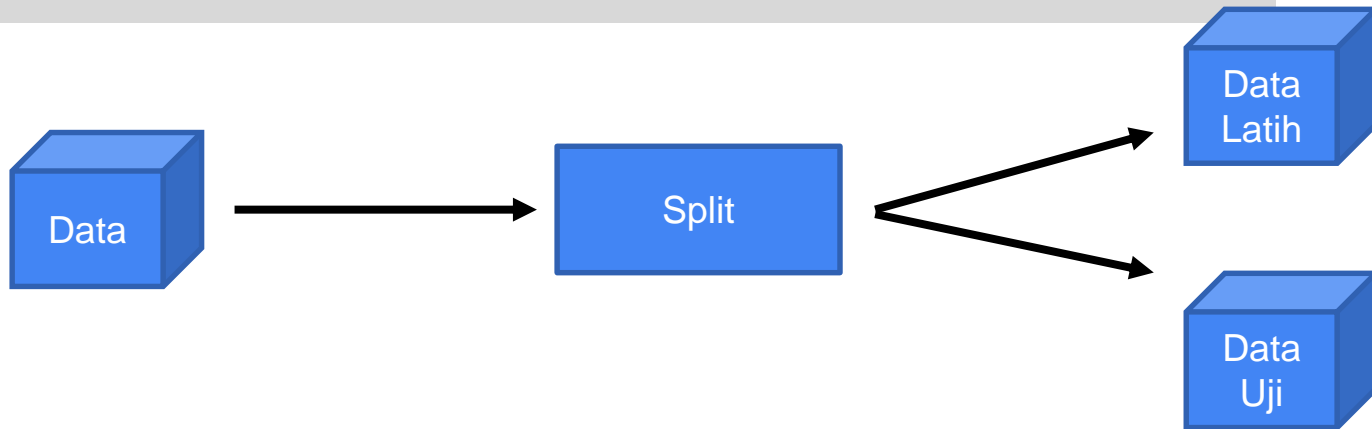
1. k-Nearest Neighbor (k-NN)
2. Naïve Bayes
3. Regression Techniques
4. Support Vector Machines (SVMs)
5. Decision Trees
6. Random Forests
7. Deep Learning Algorithms
8. ...

4. Modeling : Membangun Skenario Pemodelan

Membuat strategi pencarian model terbaik

B. Membagi data: Sesuai dengan ketersediaan data

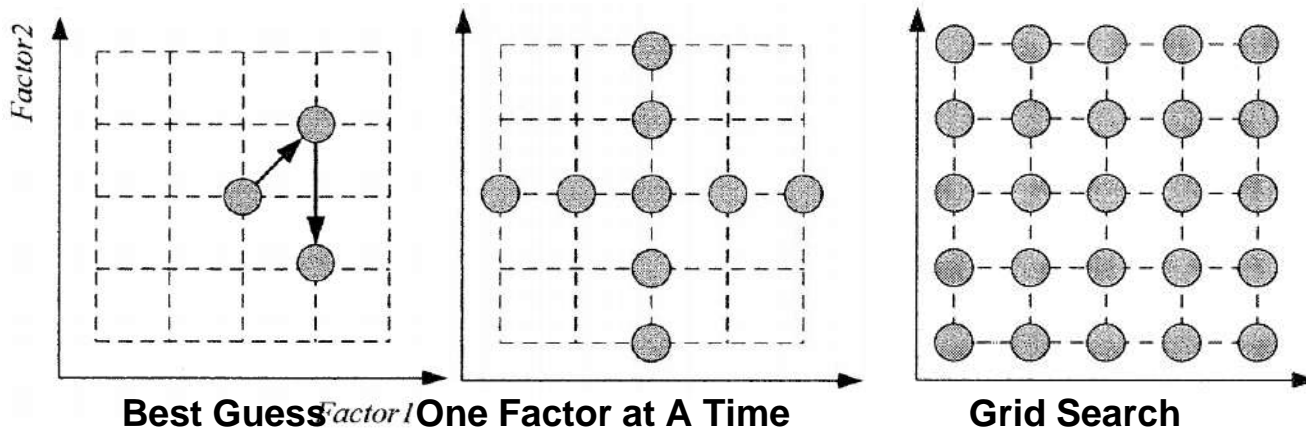
1. Data Latih: Untuk mengembangkan model
2. Data Uji: Untuk Mengukur performansi model



4. Modeling : Membangun Skenario Pemodelan

Membuat strategi pencarian model terbaik

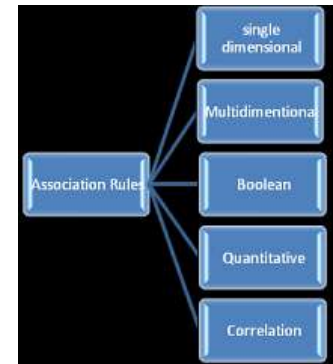
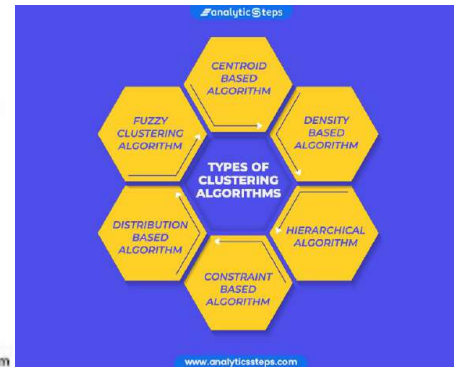
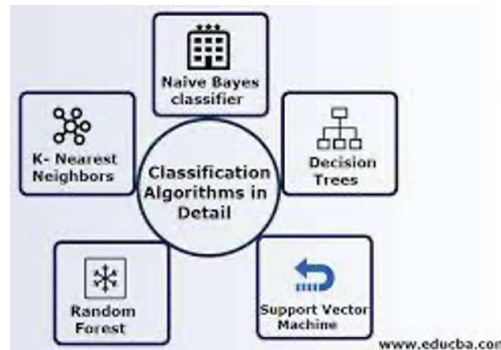
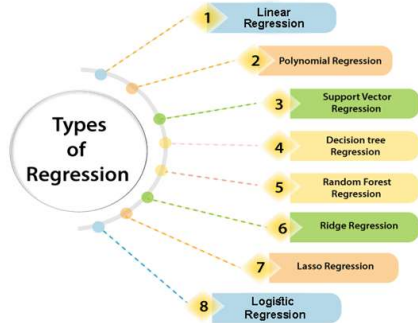
C. Menentukan Langkah Eksperimen: Untuk mendapatkan model terbaik secara efisien dan efektif



4. Modeling : Membangun model

Mengembangkan model dengan Teknik ML

Pemilihan Algoritma Machine Learning (ML)
Pembagian Data
Penentuan Langkah Eksperimen

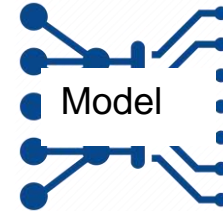
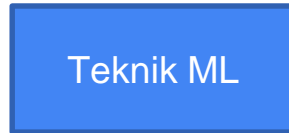
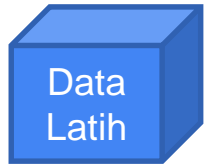


Tidak ada Algoritma yang SELALU TERBAIK untuk setiap dataset! Coba beberapa algoritma!!

4. Modeling : Membangun model

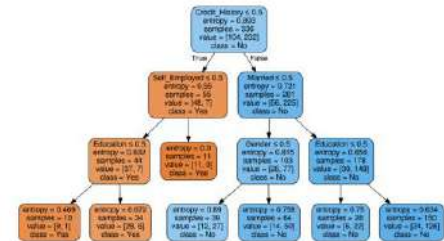
Mengembangkan model dengan Teknik ML

A. Proses Pelatihan : Untuk mendapatkan model



Variable	Type	Definition
BAD	Num	BAD: 1 = a applicant defat (not on loan or seriously delinquent), 0 = applicant paid loan
LOAN	Num	LOAN: Amount of the loan request
MORTDUE	Num	MORTDUE: Amount due on existing mortgage
VALUE	Num	VALUE: Value of current property
REASON	Char	REASON: DebtCon = debt cons or inflation; Remaing = home improvement
JOB	Char	JOB: Occupational categories
YOE	Num	YOE: Year(s) at present job
DELNOG	Num	DELNOG: Number of major derogatory reports
DELNO	Num	DELNO: Number of delinquent credit lines
CLAGE	Num	CLAGE: Age of oldest credit line in months
INNO	Num	INNO: Number of recent credit inquiries
CIND	Num	CIND: Number of credit lines
DEBTINC	Num	DEBTINC: Debt-to-income ratio

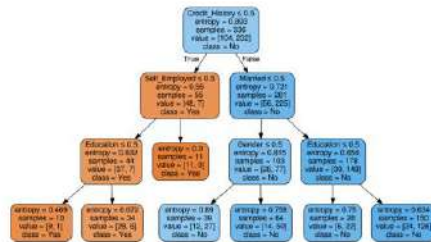
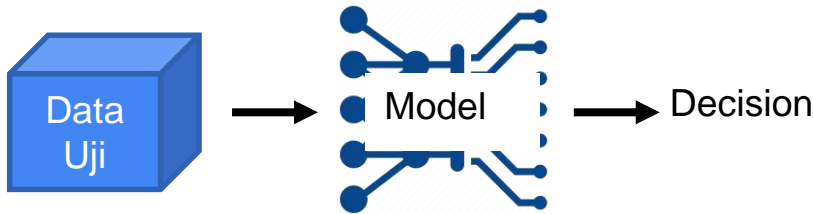
1. k-Nearest Neighbor (k-NN)
2. Naïve Bayes
3. Regression Techniques
4. Support Vector Machines (SVMs)
5. Decision Trees
6. Random Forests
7. Deep Learning Algorithms
8. ...



4. Modeling : Membangun model

Mengembangkan model dengan Teknik ML

B. Proses Pengujian : Untuk mengukur Performansi



TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

	p' (Predicted)	n' (Predicted)
p (Actual)	True Positive	False Negative
n (Actual)	False Positive	True Negative

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

5. Model Evaluation

Mengevaluasi Performansi Model Yang Dihasilkan

01

Mengevaluasi Model

Mengukur performansi model

Performansi Capaian vs Target
Memilih Model terbaik

02

Mengevaluasi Proses

Menilai apakah proses sudah maksimal

Review Proses untuk mencari
batasan atau kekurangan model



6. Deployment

Memasang model ke dalam aplikasi atau menggunakannya

01

Membuat rencana deployment

Mengukur performansi model

Rencana pemasangan model menjadi suatu sistem

02

Melakukan deployment model

Menilai apakah proses sudah maksimal

Pemasangan model ke dalam sistem operasional

03

Membuat rencana pemeliharaan

Mengukur performansi model

Memonitor performansi sistem dan membuat rencana pemeliharaan

04

Melakukan pemeliharaan model

Menilai apakah proses sudah maksimal

Melakukan perawatan atau membuat model baru

7. Evaluation

Mengevaluasi semua kegiatan proyek DATA SCIENCE

01

Melakukan review proyek DS

Lesson learned

Umpan balik dan lesson learned selama melakukan proyek

02

Membuat laporan akhir proyek DS

Laporan pelaksanaan an pencapaian

Laporan akhir proyek

5. Langkah Berikutnya?

Materi yang diberikan merupakan materi dasar untuk seorang Associate Data Scientist. Untuk mengembangkan kemampuannya masih banyak hal yang perlu dikembangkan, misalkan:

- 1. Keseimbangan data**
- 2. Kelengkapan data**
- 3. Privacy**
- 4. Risiko penggunaan Data Science**
- 5. Fairness**
- 6. Explanation Data Science**

Referensi

- Standard Kompetensi Kerja Nasional Indonesia Bidang AI sub bidang Data Science
 - <https://skkni.kemnaker.go.id/tentang-skkni/dokumen>
- CRISP-DM
 - <http://crisp-dm.eu/>
- IBM Data Science Methodology
 - <https://www.slideshare.net/JohnBRollinsPhD/foundational-methodology-for-data-science>
- Microsoft Methodology
 - <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>
- Domino Methodology
 - <https://www.dominodatalab.com/>

Terima Kasih

