

Extensive Data Mining Tanpa Coding di R

Aep Hidayatulloh

Sr. Data Analyst

aephidayatulloh.mail@gmail.com

<https://github.com/aephidayatulloh/rattle>



Perkenalkan...



Aep Hidayatulloh

Sr. Data Analyst @StarCore Analytics



telegram community
<https://t.me/GNURIndonesia>

Komunitas
R Indonesia



S1 Departemen Statistika
IPB University - tahun 2015



Statistician & App Developer - PT Ganesha Cipta Informatika [2013-2015]
Customer Value Management - PT Asuransi Jiwa Sequislife [2015-2017]
Sr. Data Analyst - StarCore Analytics [2017- Sekarang]



Email : aephidayatulloh.mail@gmail.com
GitHub : <https://github.com/aephidayatulloh>
RPubs : <https://rpubs.com/aephidayatulloh>



<https://bookdown.org/aepstk/introR>



<https://bookdown.org/aepstk/analisisdatabisnis>



Outline

- 01 ► Pengenalan R & Rattle
- 02 ► Bisnis & Data
- 03 ► Metode DM/ML
- 04 ► Let's Code Practice

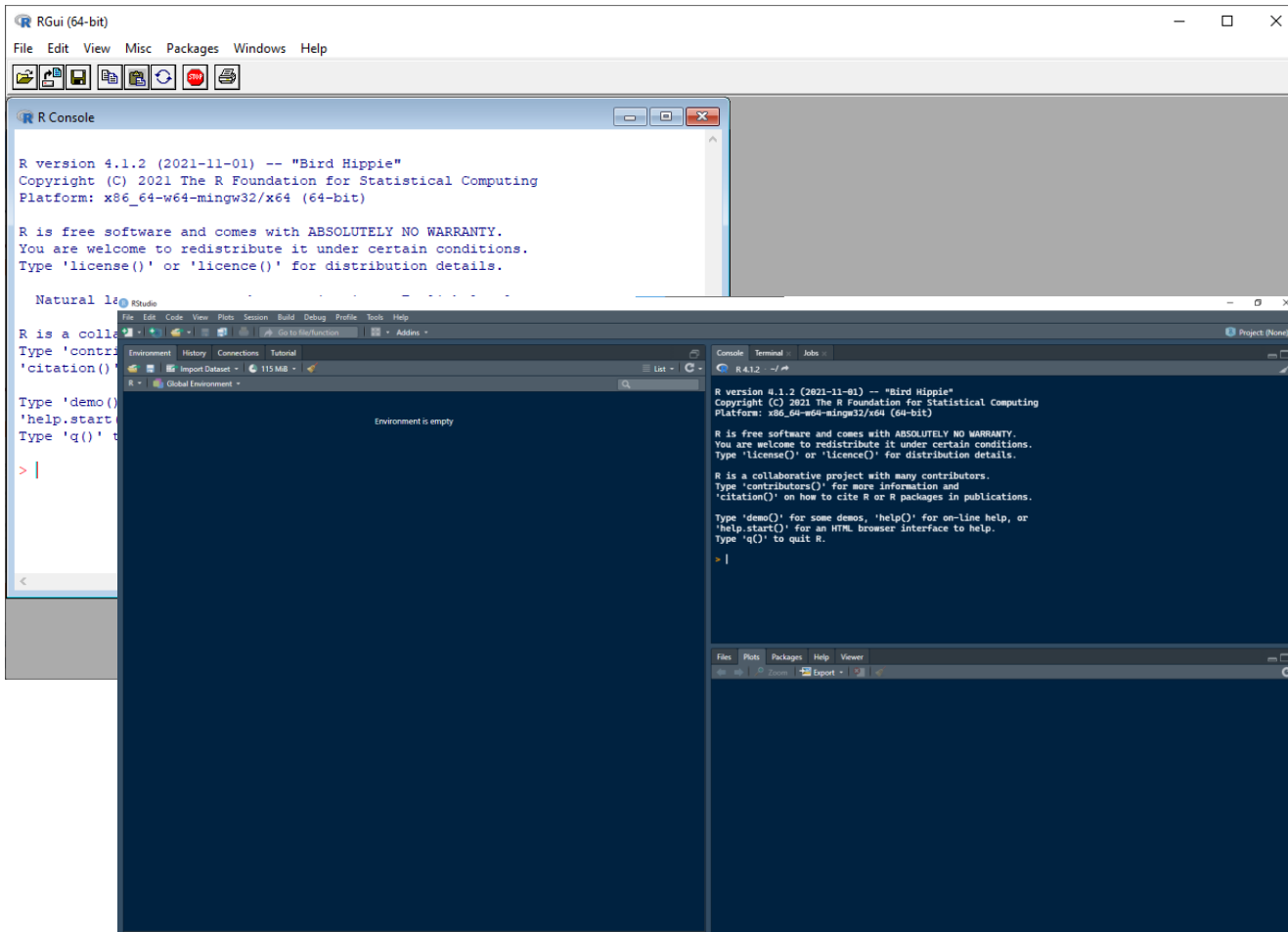




PART 01

Pengenalan R & Rattle

Program R



R GUI & RStudio

R adalah bahasa pemrograman dan sebuah tool untuk melakukan analisis statistika, data mining/machine learning, dan data science.

RStudio adalah Integrated Development Environment (IDE) untuk bahasa pemrograman R dan lainnya, seperti Python.

R Packages

Non GUI-based Packages

tidyverse

API untuk kebutuhan data science
(import, penyiapan dan
eksplorasi/visualisasi data)

tidymodels

API untuk machine learning yang
terintegrasi dengan tidyverse

ggplot2

Package yang terkenal untuk
eksplorasi dan visualisasi data dengan
mudah dan banyaknya kostumisasi
tampilan



GUI-based Packages

Rcmdr

GUI untuk bahasa pemrograman R
tanpa coding, fitur mirip SPSS/Minitab

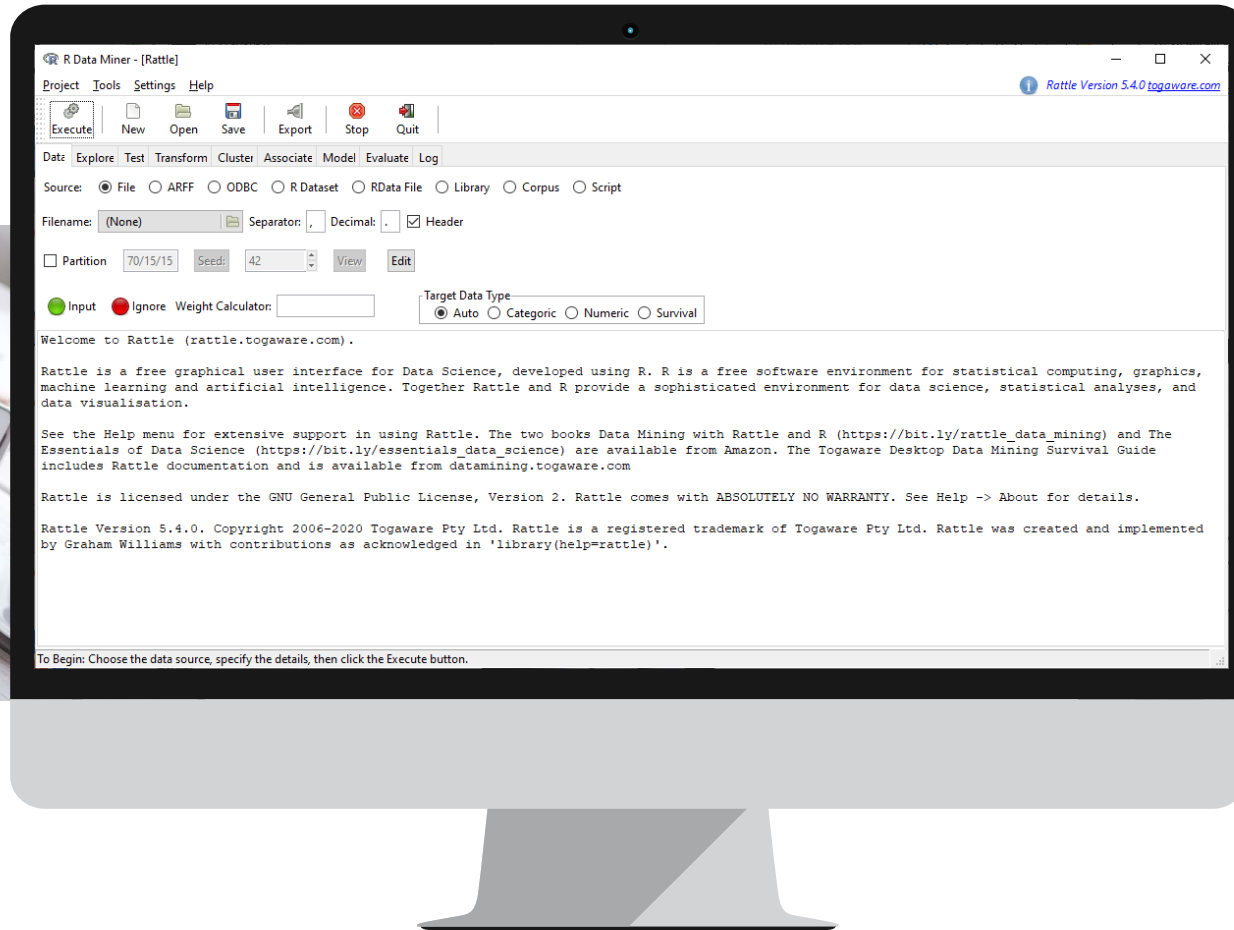
Rattle

GUI bahasa pemrograman R untuk
data mining/machine learning

esquisse

Eksplorasi dan visualisasi data dengan
cara interactive tanpa coding

Rattle GUI



Install RGtk2 & Rattle

```
install.packages("RGtk2")
```

```
install.packages("rattle",  
dependencies = "Suggests")
```

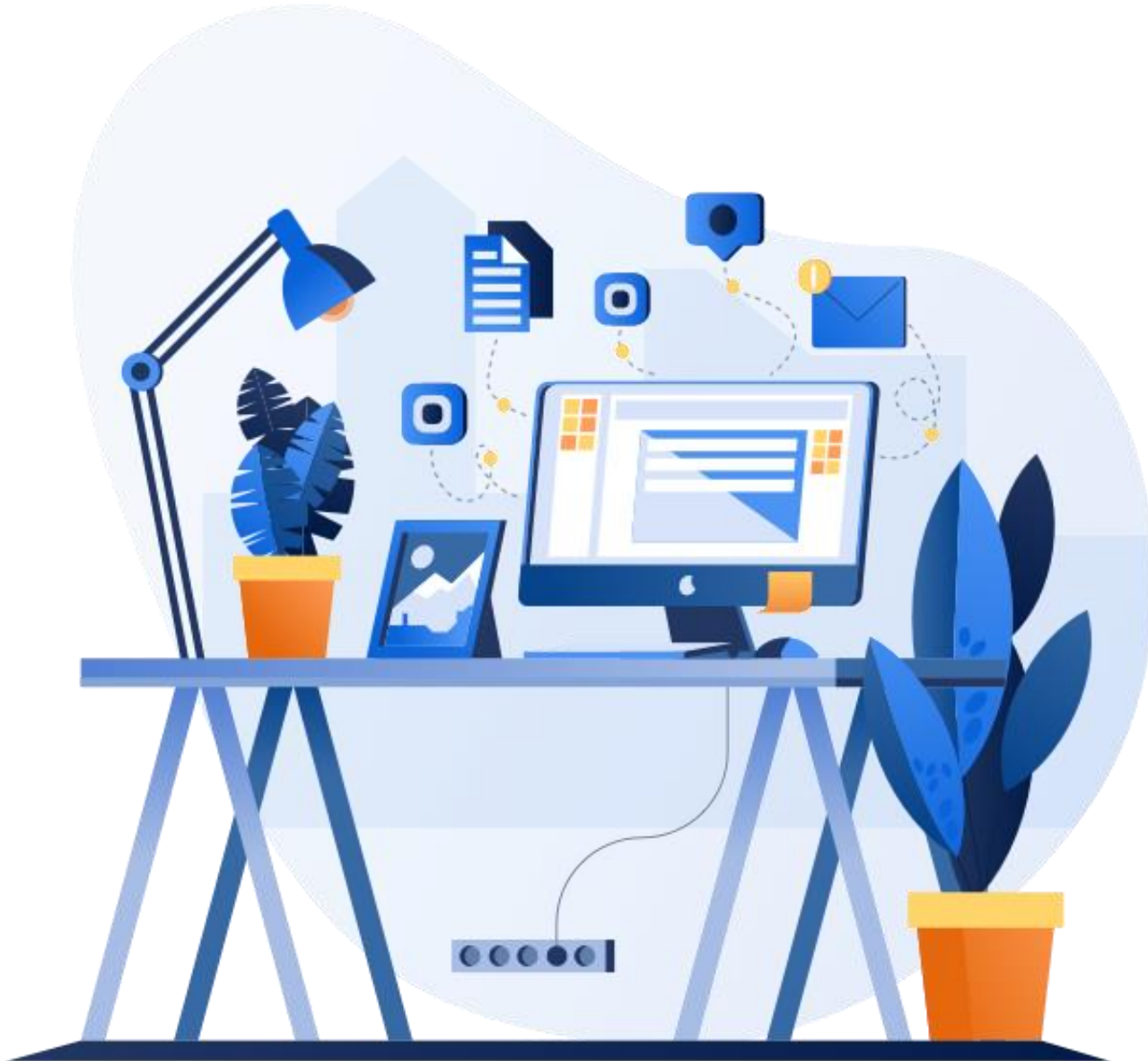
Rattle: R Analytical Tool To Learn Easily
A Graphical User Interface for Data Mining using R

<https://rattle.togaware.com>

Load rattle package

```
library(rattle)
```

```
rattle()
```

PART 02

Bisnis & Data

Bisnis

Retail Company

Sebuah perusahaan retail **MyMart** sudah mengumpulkan data dari **4970 pelanggan**, dengan melakukan promo **cashback 50K untuk pembelian $\geq 800K$** .

* **Note:** Data yang diambil diantaranya adalah *visit* dan *spending* dalam 1 bulan di bulan **1**, **2**, dan **3**, yang menunjukkan bulan **November**, **Oktober**, dan **September**.

MyMart mencatat pelanggan yang mengakhiri periode promo ini dengan mengambil promo tersebut atau tidak. Pada akhirnya, sebuah model akan dibangun untuk secara efektif memilih pelanggan terbaik untuk penawaran promo berikutnya.



Variable

Variable	Deskripsi
member_id	ID pelanggan
gender	Jenis kelamin pelanggan
visit_last_1mo	Banyaknya visit selama bulan 1 (November)
visit_last_2mo	Banyaknya visit selama bulan 2 (Oktober)
visit_last_3mo	Banyaknya visit selama bulan 3 (September)
spending_last_1mo	spending amount selama bulan 1 (November)
spending_last_2mo	spending amount selama bulan 2 (Oktober)
spending_last_3mo	spending amount selama bulan 3 (September)
age	Usia pelanggan saat diberi promo
monthly_income	Rata-rata penghasilan per bulan tiap pelanggan
marital_status	Status pernikahan pelanggan
payment_channel	Channel pembayaran yang paling sering digunakan
groceries	Apakah pelanggan membeli produk groceries selama 3 bulan terakhir (1: ya; 0: tidak)
toiletries	Apakah pelanggan membeli produk toiletries selama 3 bulan terakhir (1: ya; 0: tidak)
food	Apakah pelanggan membeli produk food selama 3 bulan terakhir (1: ya; 0: tidak)
electronic	Apakah pelanggan membeli produk electronic selama 3 bulan terakhir (1: ya; 0: tidak)
clothes	Apakah pelanggan membeli produk clothes selama 3 bulan terakhir (1: ya; 0: tidak)
home_appliances	Apakah pelanggan membeli produk home appliances selama 3 bulan terakhir (1: ya; 0: tidak)
recency_last_visit	Berapa hari sejak visit terakhir pelanggan? (recency)
response	Apakah pelanggan mengambil promo? (1: ya; 0: tidak)

Data

Input/Prediktor

Target

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	member_id	gender	visit_last_1mo	visit_last_2mo	visit_last_3mo	spending_last_1mo	spending_last_2mo	spending_last_3mo	age	monthly_income	marital_status	payment_channel	groceries	toiletries	food	electronic	clothes	home_appliances	recency_last_visit	response
2	237398	Female	4	6	6	207164	170482	157592	25	3172700	Married	Cash	1	1	1	0	0	1	27	0
3	502009	Female	6	6	2	792577	821678	859218	32	4141600	Married	Cash	0	1	0	0	0	1	5	0
4	11841	Female	3	3	2	1019203	1058981	1036460	24	5292900	Single	Debit Card	0	1	1	1	0	1	3	0
5	645389	Female	3	2	1	260765	281803	276175	28	2606600	Single	Credit Card	0	0	1	0	0	0	27	0
6	225416	Female	3	4	3	956641	899804	814255	27	4520100	Single	Debit Card	0	1	1	0	0	1	1	0
7	909605	Female	3	2	4	454295	415457	413672	24	7301300	Married	Debit Card	0	1	1	0	1	0	2	1
8	52795	Female	0	0	1	0	0	447018	31	5302000	Married	Cash	1	1	1	0	0	0	37	0
9	536887	Male	1	2	2	577327	534312	470981	32	9366900	Married	Cash	0	1	1	0	0	0	19	1
10	609899	Female	7	4	3	661529	651732	652627	30	13895200	Married	Debit Card	0	0	1	0	0	0	11	1
11	599263	Female	2	1	0	461273	485364	0	27	9550000	Married	Cash	0	0	0	0	0	0	8	0
12	394378	Female	6	5	4	570724	570862	495474	27	6132800	Married	Cash	0	0	1	0	0	0	28	1
13	440873	Female	5	2	6	198611	203043	195667	36	5959400	Married	Debit Card	0	1	1	0	0	0	16	0
14	311224	Female	7	9	7	604219	654103	615189	28	3062700	Married	Credit Card	0	0	1	0	0	1	4	1
15	534320	Male	3	3	4	237511	270097	279177	43	3661200	Married	Cash	0	1	1	0	0	1	15	1
16	622228	Female	4	5	2	305992	330652	335845	32	5961800	Married	Credit Card	1	1	1	0	0	0	14	0
17	813652	Female	1	1	1	505503	505677	594101	22	6753500	Single	Cash	0	1	1	0	0	0	3	0
18	302442	Male	1	1	5	306538	314632	294242	33	3217300	Married	Debit Card	0	0	1	0	0	0	28	0
19	566543	Female	3	2	4	549395	531798	447470	29	2395800	Single	Debit Card	0	1	1	0	0	0	29	1
20	172166	Female	5	5	1	578749	558831	576726	38	4459100	Married	Cash	0	1	1	0	0	1	21	1
21	437415	Female	3	4	6	299722	301667	281176	37	5226500	Married	Credit Card	0	1	1	0	0	0	25	0
22	569305	Female	4	3	0	510433	610902	0	23	4455800	Single	Credit Card	0	0	1	0	0	0	11	0
23	11937	Female	1	1	4	396000	472502	506685	30	6354300	Married	Cash	0	1	1	0	0	0	28	1
24	193893	Male	0	2	6	0	221906	234459	25	5463700	Married	Cash	0	1	1	0	0	0	50	0
25	312080	Male	1	3	6	219443	218097	243365	30	7844700	Single	Credit Card	1	0	1	0	1	1	15	1
26	745951	Male	3	4	1	442846	416872	434817	29	3914300	Married	Debit Card	1	1	1	1	0	0	21	1



PART 03

Metode

Metode Data Mining/ Machine Learning

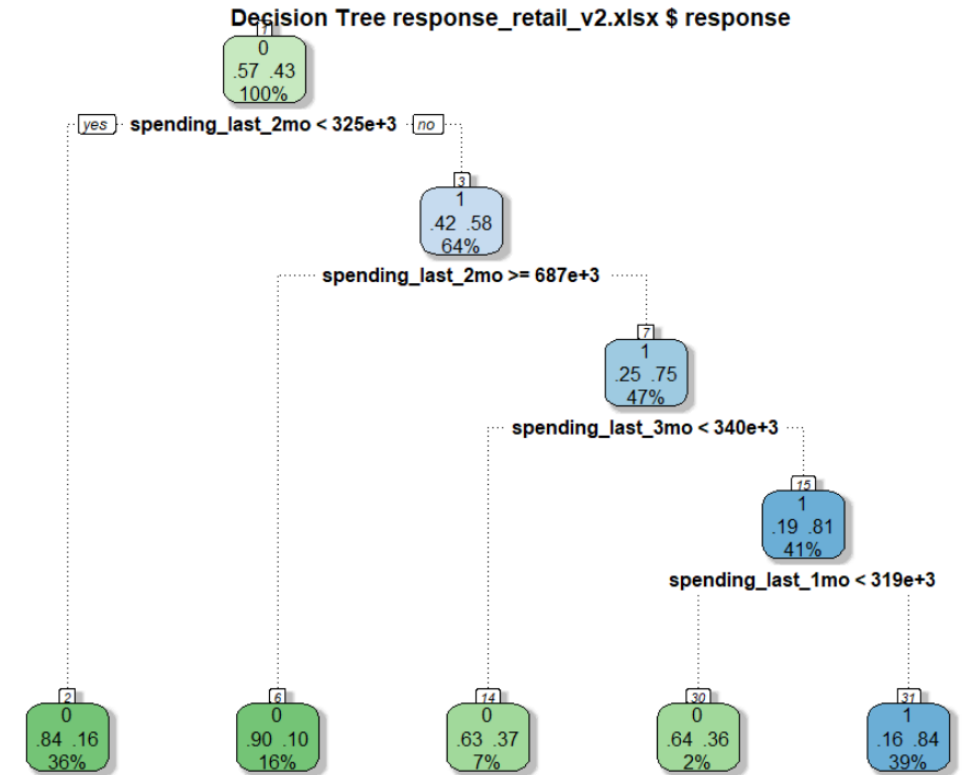
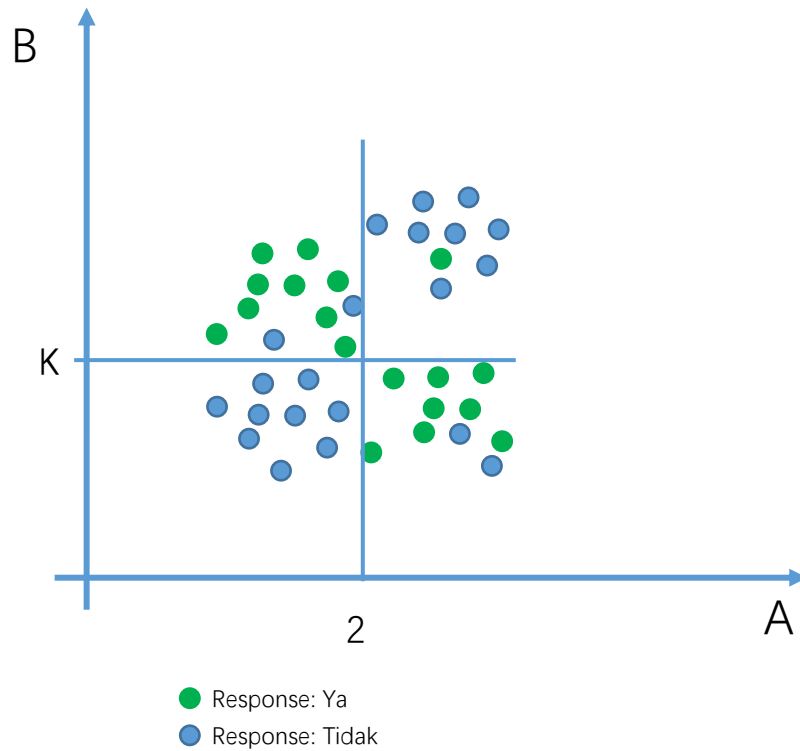


Target: Numerik
Regression

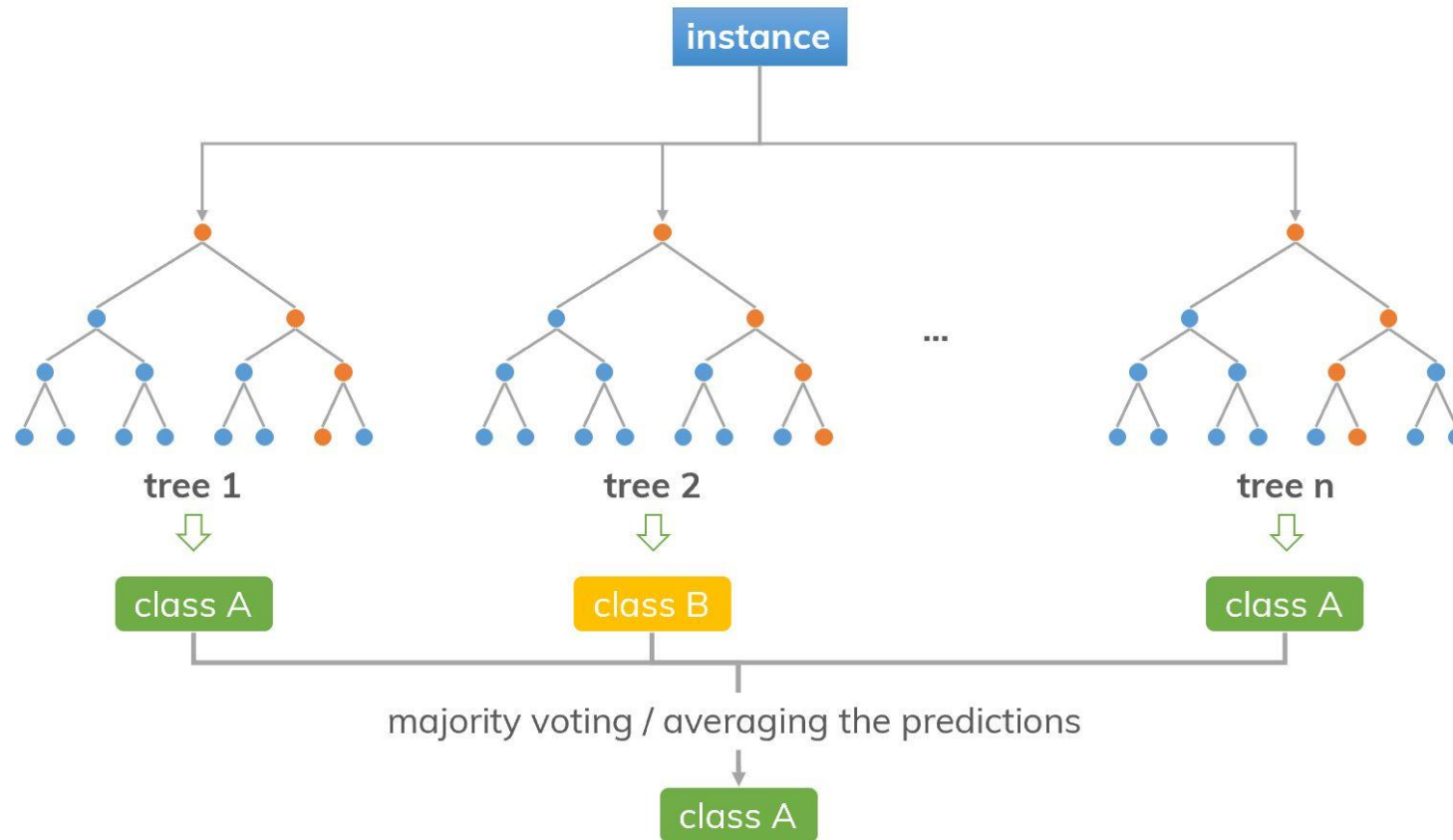


Target: Kategorik
Classification

Decision Tree



Random Forest





PART 04

Let`s Code Practice

Terima kasih

