Bootcamp Dibimbing.id

Day 17: Machine Learning With R

Table of Contents

Materi Hari Ini:

1. “Probability” vs “Odds”
2. Intuition of Logistic Regression
3. Logistic Regression Hands-On in R
4. Intuition of Naive Bayes
5. Naive Bayes Hands-On in R
6. Intuition of Decision Trees
7. Decision Tress Hands-On in R
8. General Machine Learning Theories

R-Studio and R

Links to Download and Install RStudio & R

1

R

https://cran.r-project.org/bin/windows/base/R-4.1.0-win.exe

2

RStudio

https://download1.rstudio.org/desktop/windows/RStudio-1.4.1717.exe

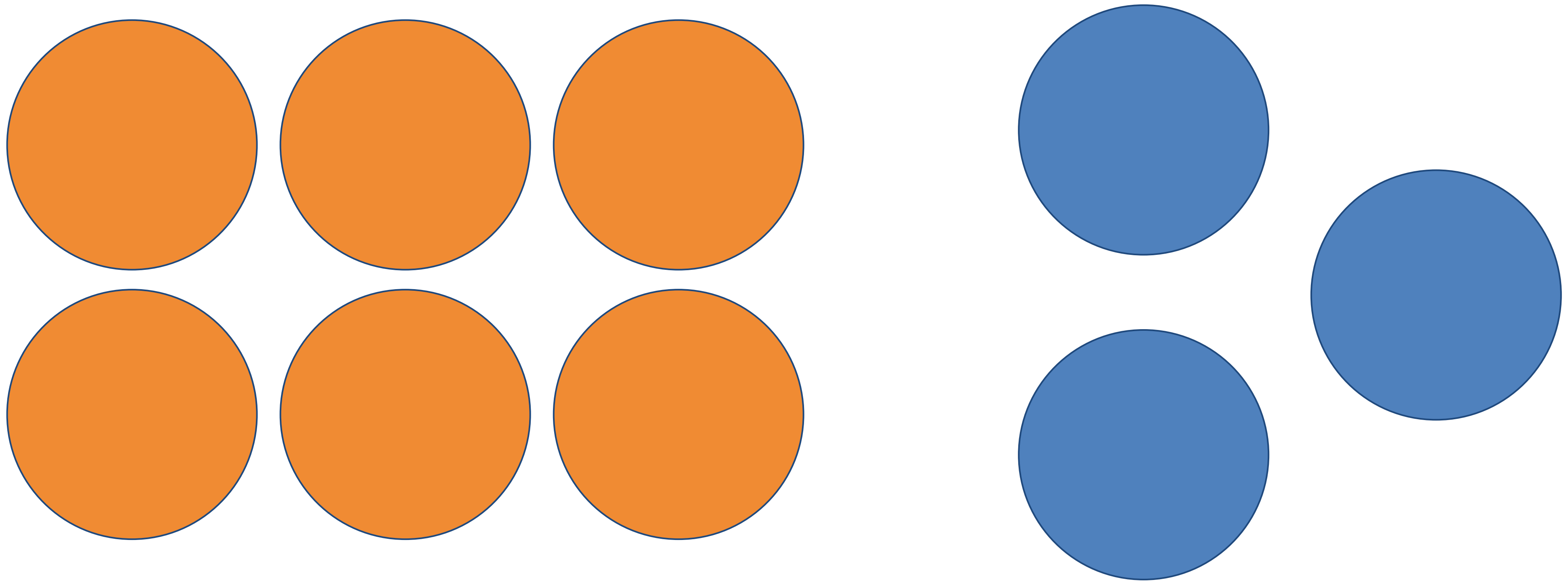
Probability and Odds

Apa bedanya “probability” dan “odds”?

1. Probability = Rasio kemungkinan terjadinya suatu event vs kemungkinan seluruh outcome.

2. Odds = Rasio kemungkinan terjadinya suatu event vs kemungkinan tidak terjadinya event tersebut.

Probability and Odds: Example



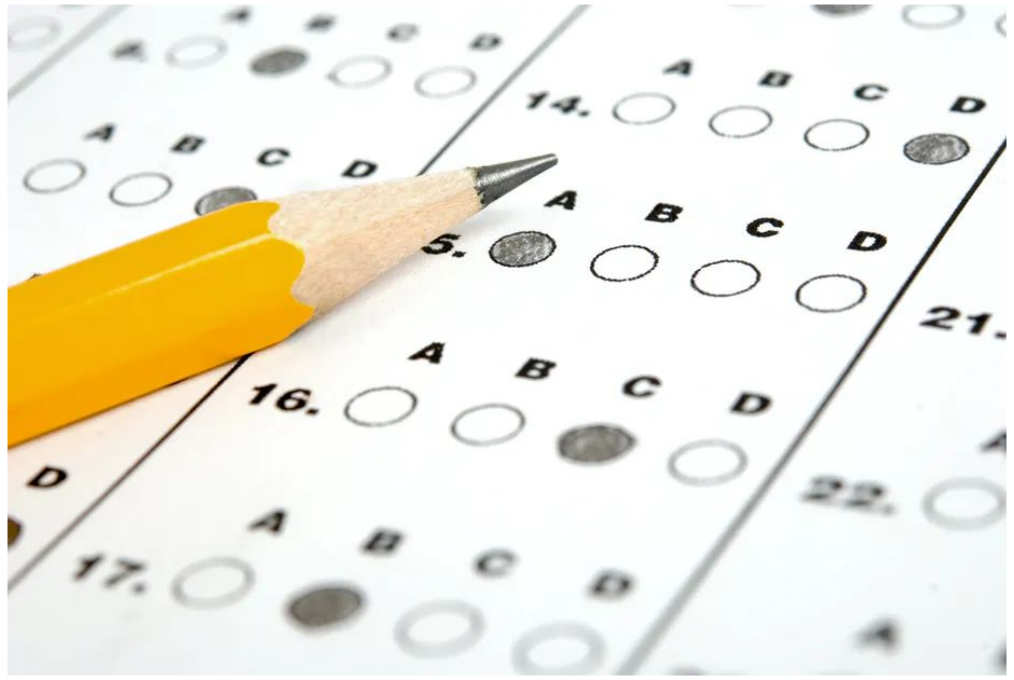
“Probability” of taking a blue ball is 33%.

The “odds” of taking a blue ball is 3 to 6, or 1 to 2. In fraction, it’s 3/6 = ½.

Probability and Odds: Riddle

3 out of 4 people passed the test.

A. What is the ‘probability’ of passing the test?
B. What is the ‘odds’ of you passing the test?



Probability and Odds: Answer

3 out of 4 people passed the test.

A. What is the ‘probability’ of a student passing the test?
   ¾ = 75%

B. What is the ‘odds’ of passing the test?
   3 to 1. In fraction mode, it’s 3/1 = 3.

Probability and Odds: Answer

3 out of 4 people passed the test.

$$\text{Probability of passing} = \frac{\text{(odds of passing)}}{\text{(odds of passing)} + 1}$$

Probability of passing = 3/(3+1) = ¾ = 0.75

Log Odds

Let's go back to the 'test' example.

If we face a difficult test, the odds of passing the test becomes smaller, for example 1 to 10. (1/10 = 0.1)

If we face an easy test, the odds of passing the test becomes bigger, for example 10 to 1. (10).

Log Odds

Let’s imagine a **very difficult** test.
A test that no matter how hard we learn, we can never pass it. The odds of passing is..**0.**

Let’s imagine a **very easy** test. The odds of passing is 10000:1 = 10000.

Let’s imagine a **test** that is **1:1** to our **skill**.
The odds of passing is 1:1 = 1.

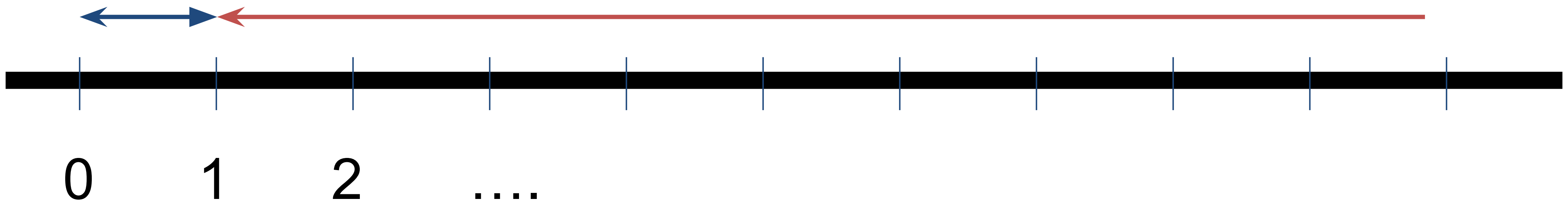
Log Odds

This is the problem of using odds.

If the **odds** are **against** us, the **odds** ranges from 0 to 1.

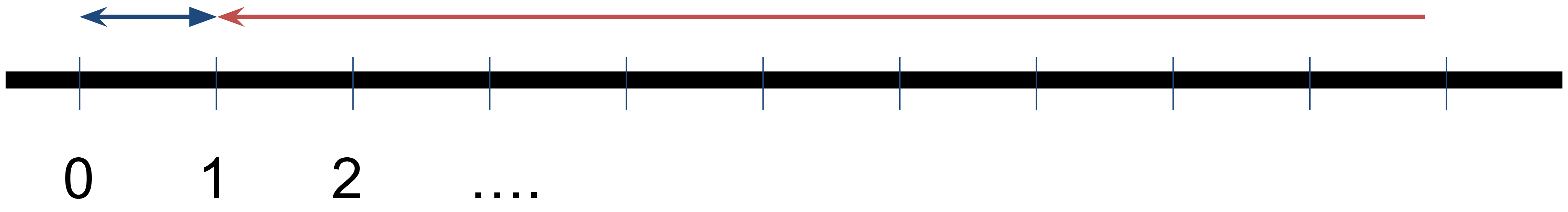
If the **odds** are **for** us (we are advantageous), the **odds** range from 1 to **infinite**.

It's simply **not balanced.**



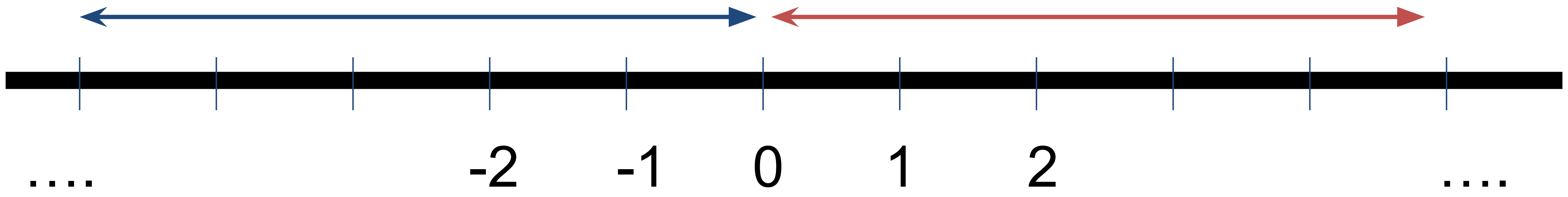
Log Odds

That’s why we try to make it more ‘symmetrical’ by applying **logarithm** to it. which **logarithm**? Usually the **natural logarithm.**



Log Odds

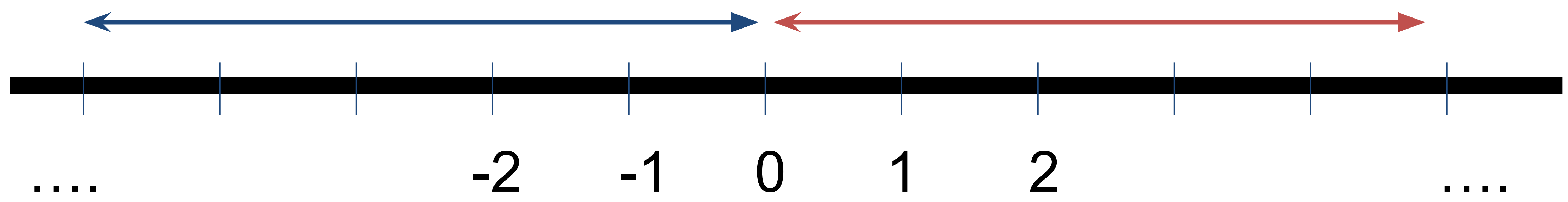
- In an **impossibly difficult** test, our **log odds** is log(0) = **- infinite**
- In a **difficult** test, our **logg odds** is log(1/10) = -2.3
- In a **1:1** test, our **log odds** is log(1) = 0
- In an **easy** test, our **logg odds** is log(10) = 2.3
- In a **very easy** test, our log odds is log(10000) = 9.2



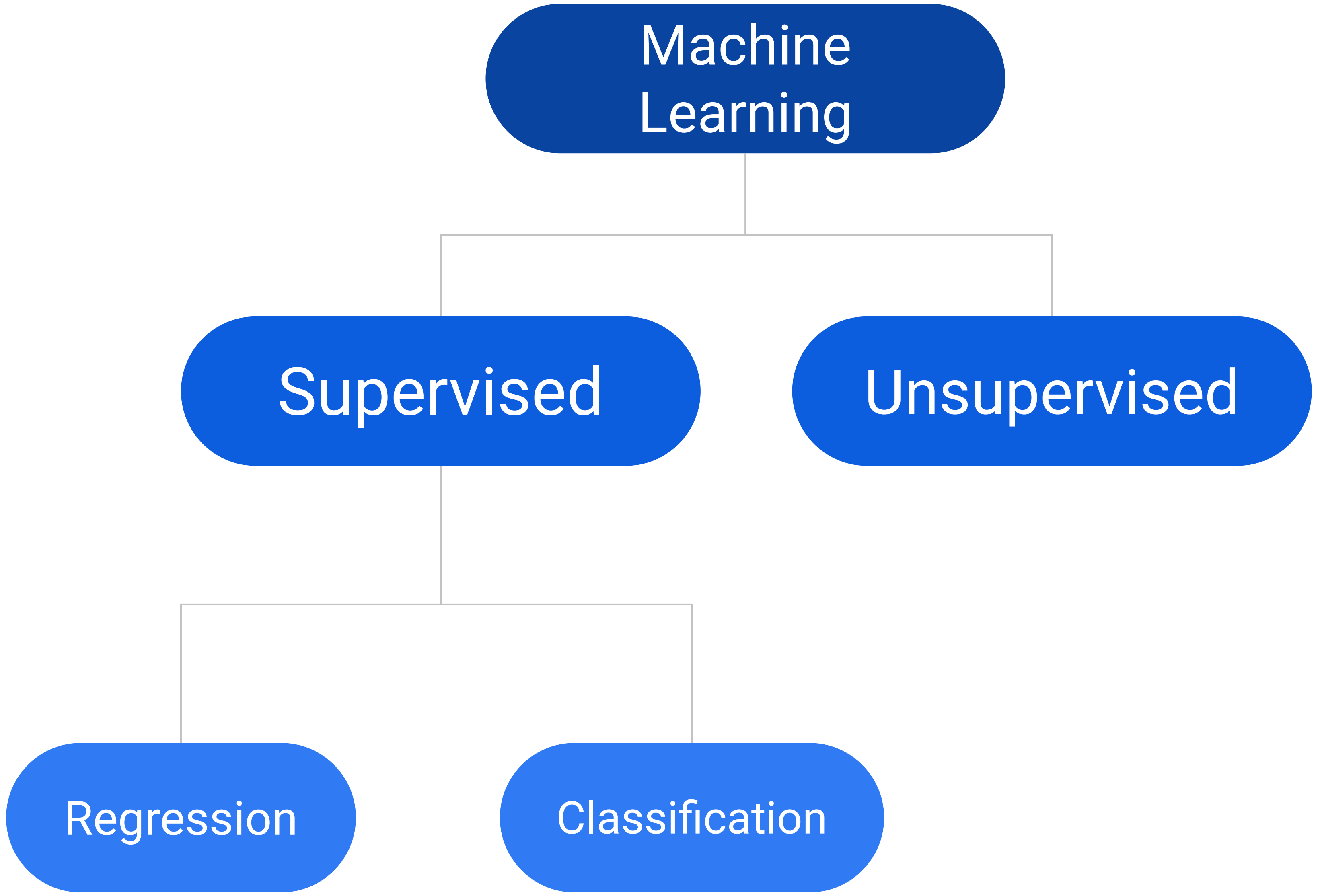
Log Odds

It’s more symmetrical as the value ranges from **-infinite** to **infinite**, and when the odds are 1:1, it sits **at the middle 0**.

This is why we use **log odds** in **logistic regression** and other **Machine Learning** algorithms.



Types of Machine Learning



Machine Learning

Machine Learning: Analogy

Bayangkan sebuah ruang kelas SD.

Lagi belajar matematika, topiknya penjumlahan.

“Anak-anak,

1+2 = 3,
2+3 = 5,
4+6 = 10,
8+6 = 14,

maka…. 7+8 = ?”

Ini adalah Unsupervised/Supervised Learning?
Regression/Classification?

Machine Learning

Machine Learning: Analogy

Bayangkan sebuah ruang kelas SD.

Lagi pelajaran Sains, dan guru memberi contoh-contoh makhluk hidup dan ‘tipe’ mereka.

“Harimau itu karnivora,
Gajah itu herbivora,
Hiu itu karnivora,
Jerapah itu herbivora,

Buaya itu herbivora/karnivora?”

Ini termasuk Unsupervised/Supervised Learning?
Regression/Classification?

Machine Learning

Machine Learning: Analogy

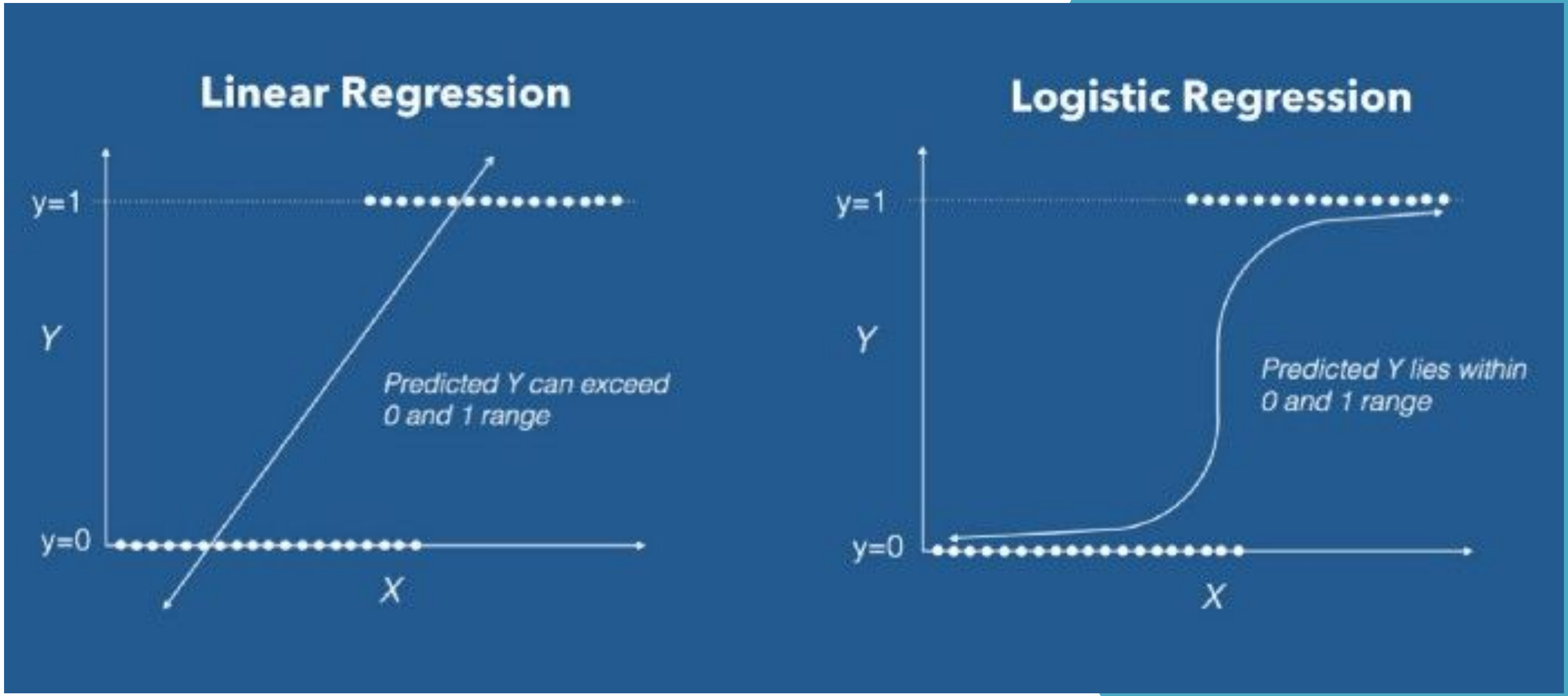
Bayangkan sebuah ruang kelas SD.

Lagi pelajaran olahraga, dan mereka diminta membuat kelompok secara bebas. Dari 20 murid, menjadi 3 kelompok.

Ini termasuk Unsupervised/Supervised Learning? Regression/Classification?

Machine Learning

Logistic vs Linear Regression



Logistic Regression

Even though the name is logistic ‘**regression**’, this can be interpreted ‘**classification**’ algorithm, because the output value is only either ‘0’ or ‘1’.

For example:

- Predicting whether a loan is accepted or not based on how ‘rich’ a customer is.

Input will be the wealth of the customer.
Output will be a **value ranging from 0 to 1**, meaning the probability of a loan is accepted.

0 means the **loan is not accepted**
1 means the **loan is accepted**.

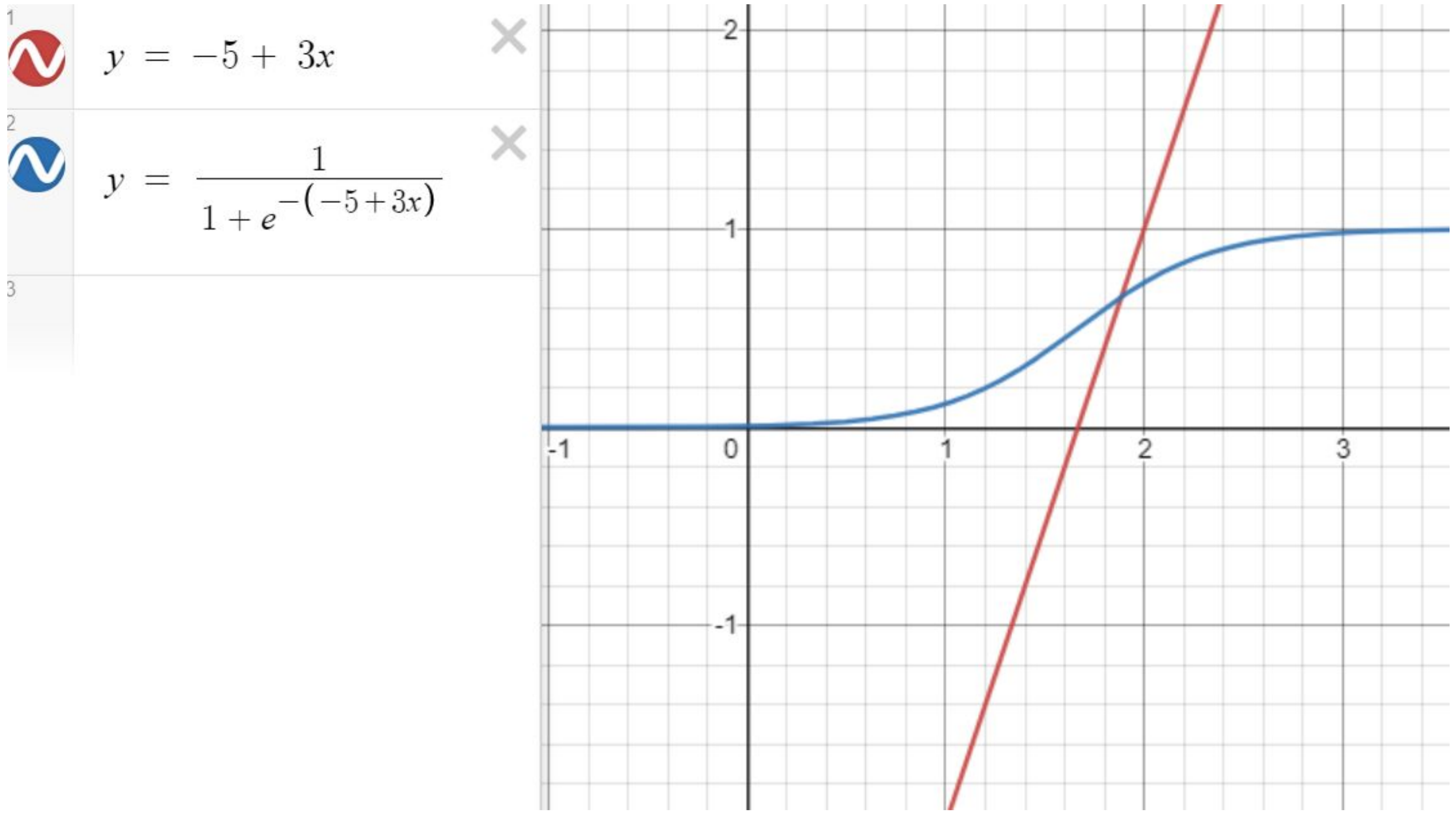
If the output is > 0.5, loan is accepted
If the output is < 0.5, loan is not accepted

Logistic
Regressio

Linear to Logistic Regression

Red Line: Linear Regression, value can ‘theoretically’ be of **any real number**.

Blue Line: Logistic Regression

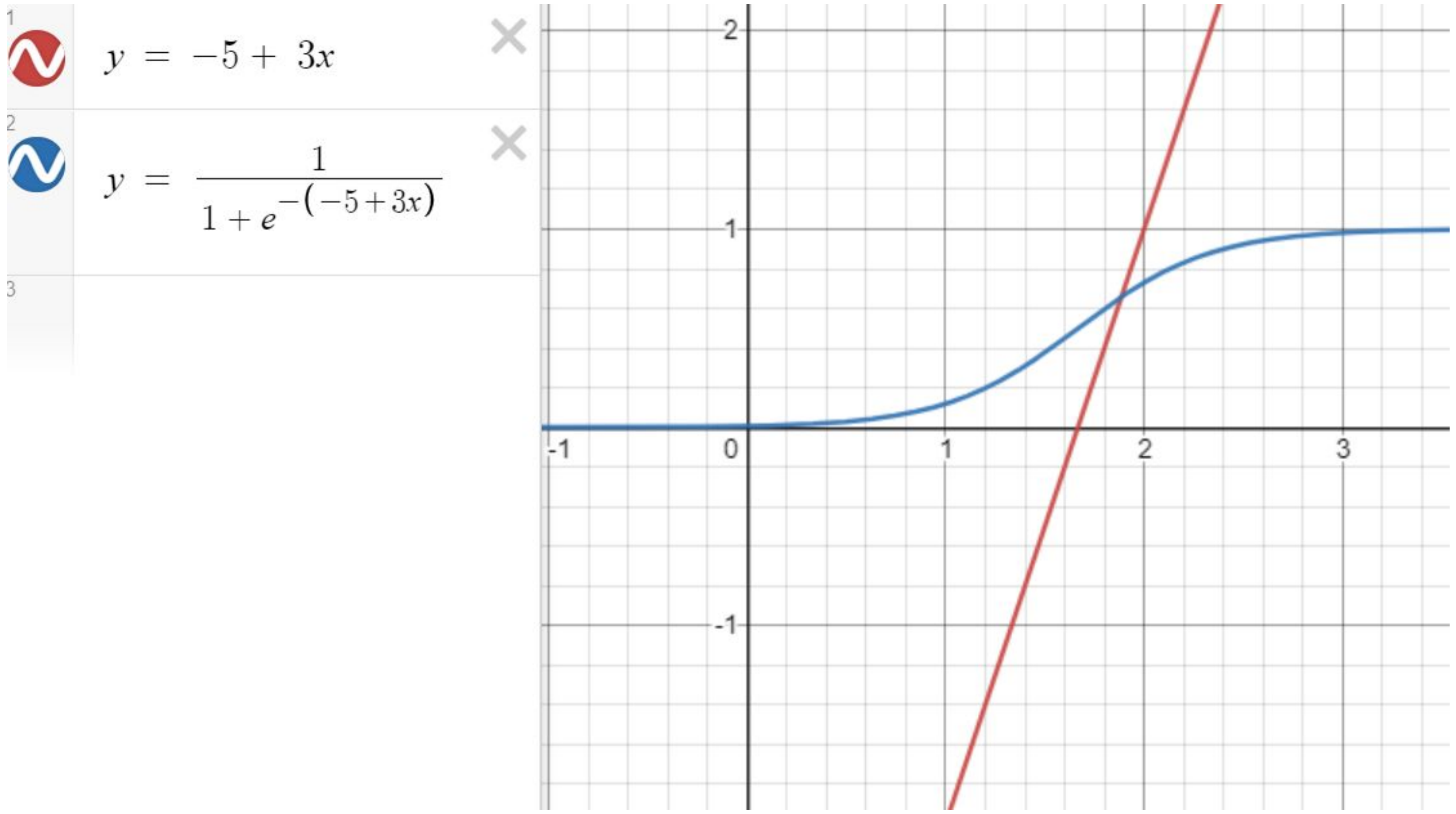


Logistic Regression

Linear to Logistic Regression

Red Line: Linear Regression, value can ‘theoretically’ be of **any real number**.

Blue Line: Logistic Regression



Logistic Regression

Logistic Regression Hands On

Dataset Download:
https://www.kaggle.com/ronitf/heart-disease-uci

R Script:
‘logistic_regression_heart_disease.R’

Logistic Regression

Naive Bayes

Naive Bayes is a term that is collectively used for **classification algorithms** that are based on **Bayes Theorem.**

Naive Bayes terdiri dari dua kata, **Naive** dan **Bayes**. **Bayes** berarti menggunakan prinsip **Bayes Theorem**, sedangkan **Naive** berarti diasumsikan bahwa semua variabel input adalah **independent** satu sama lain.

Naive Bayes

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

P(A|B) = Peluang kejadian A terjadi jika diketahui B (kejadian B benar)

P(B|A) = Peluang kejadian B terjadi jika diketahui A benar

Naive Bayes

Naive Bayes Example

Data di slide berikut menunjukkan apakah seseorang akan pergi atau tidak.

Data dikumpulkan selama 14 hari, dan berisi tentang keadaan cuaca serta suhu pada hari tersebut.

Karena di iklim tropis, diasumsikan bahwa suhu dan cuaca saling bebas (bisa saja berawan tapi panas, berawan tapi dingin, terik berangin sehingga sejuk, dll).

Naive Bayes

Naive Bayes Example

	Cuaca	Suhu	Pergi?		Cuaca	Suhu	Pergi?
1	Terik	Sejuk	Ya	8	Hujan	Sejuk	Ya
2	Terik	Panas	Ya	9	Hujan	Sejuk	Ya
3	Berawan	Sejuk	Ya	10	Terik	Panas	Tidak
4	Berawan	Panas	Ya	11	Terik	Panas	Tidak
5	Berawan	Dingin	Ya	12	Terik	Sejuk	Tidak
6	Berawan	Dingin	Ya	13	Hujan	Sejuk	Tidak
7	Hujan	Dingin	Ya	14	Hujan	Dingin	Tidak

Naive Bayes

Naive Bayes Example

Pertanyaan:

Jika hari ini Terik dan Sejuk, apakah orang ini akan pergi?

Naive Bayes

Naive Bayes Example

Langkah pertama: buat tabel untuk variabel ‘Cuaca’ dan ‘Suhu’ seperti berikut

Cuaca	Ya	Tidak	P(Ya)	P(Tidak)
Terik	2	3	2/9	3/5
Berawan	4	0	4/9	0
Hujan	3	2	3/9	2/5
Total	9	5	100%	100%

Suhu	Ya	Tidak	P(Ya)	P(Tidak)
Panas	2	2	2/9	2/5
Sejuk	4	2	4/9	2/5
Dingin	3	1	3/9	1/5
Total	9	5	100%	100%

Naive Bayes

Naive Bayes Example

Langkah Kedua: Hitung probabilitas orang tersebut Pergi (Ya) atau Tidak Pergi (Tidak)

$$P(Ya|Terik, Sejuk) = \frac{P(Terik|Ya) * P(Sejuk|Ya) * P(Ya)}{P(Terik\ Sejuk)}$$

$$P(Tidak|Terik, Sejuk) = \frac{P(Terik|Tidak) * P(Sejuk|Tidak) * P(Tidak)}{P(Terik\ Sejuk)}$$

Naive Bayes

Naive Bayes Example

Langkah Kedua: Hitung probabilitas orang tersebut Pergi (Ya) atau Tidak Pergi (Tidak)

$$P(Ya|Terik, Sejuk) = \frac{(2/9) * (4/9) * (9/14)}{P(Terik\ Sejuk)}$$

$$P(Tidak|Terik, Sejuk) = \frac{(3/5) * (2/5) * (5/14)}{P(Terik\ Sejuk)}$$

Naive Bayes

Naive Bayes Example

Langkah Kedua: Hitung probabilitas orang tersebut Pergi (Ya) atau Tidak Pergi (Tidak)

P(Ya|Terik, Sejuk) α (2/9) * (4/9) * (9/14) = 0.0635

P(Tidak|Terik, Sejuk) α (3/5)* (2/5)* (5/14) = 0.0857142

Naive Bayes

Naive Bayes Example

Langkah Terakhir:

Karena nilai P(Tidak|Terik, Sejuk) lebih besar dari P(Ya|Terik, Sejuk), maka kemungkinan besar, orang tersebut tidak akan pergi hari ini.

Bagaimana jika hari ini Terik dan Panas?

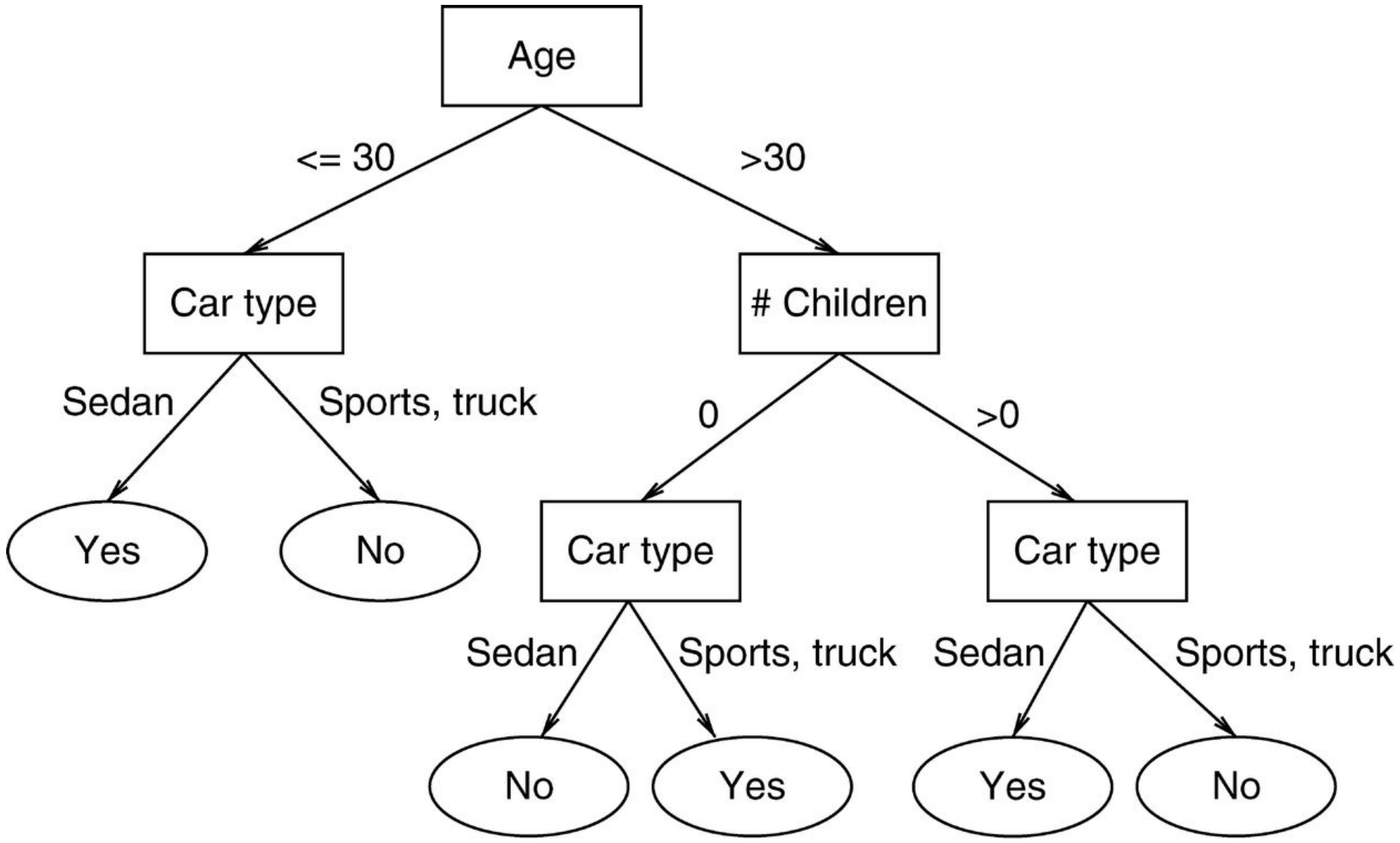
Naive Bayes

Hands-On Naive Bayes in R

Dataset: Iris (sudah dalam paket Caret)

What is Decision Tree?

- Machine Learning Algorithm that constructs "rules" that divide the data into several "decisions" after one another, so it looks like a "tree".



Decision Tree

How does Decision Tree Work?

- Decision Tree algorithm attempts to divide the data so it can achieve a ‘pure’ leaf with the least amount of ‘branch’

- We need 2 metrics to decide how to split our data:
  - **Entropy**
  - **Information Gain**

Decision Tree

How does Decision Tree Work?

- Decision Tree algorithm attempts to divide the data so it can achieve a 'pure' leaf with the least amount of 'branch'

- We need 2 metrics to decide how to split our data:
  - **Entropy**
  - **Information Gain**

Decision Tree

How does Decision Tree Work?

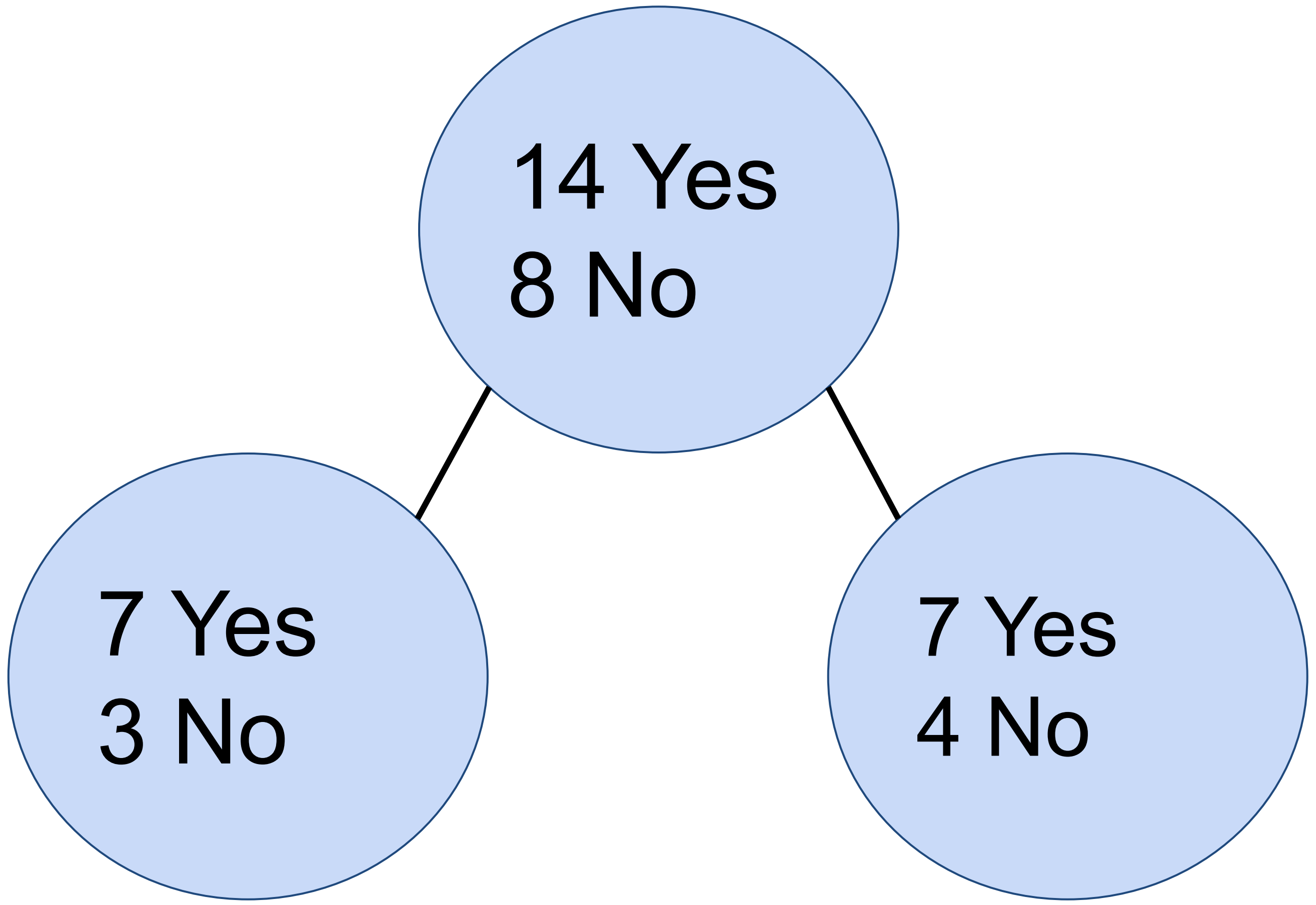
- Entropy Formula:

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

Decision Tree

How does Decision Tree Work?

- Entropy Formula:

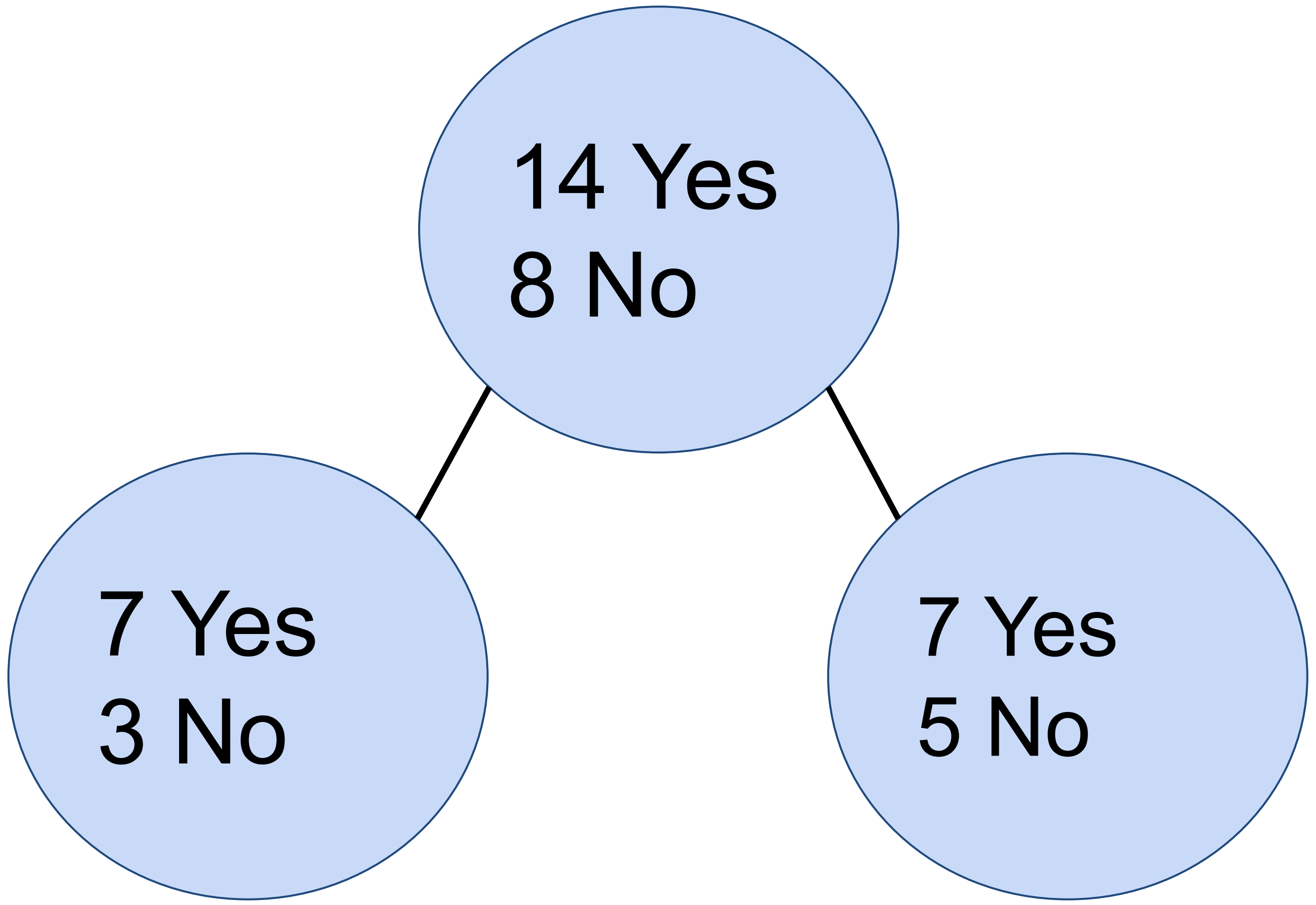


Entropy: $- (3/10) * \log_2 (3/10) - (7/10) * \log_2 (7/10) = 0.88$

Decision Tree

How does Decision Tree Work?

- Entropy Formula:



Entropy = ???

Decision Tree

How does Decision Tree Work?

Low-entropy nodes are more preferrable than **high-entropy** nodes.

If a node as only 1 class member in it (e.g. 10 yes and 0 no), it has **low entropy**.

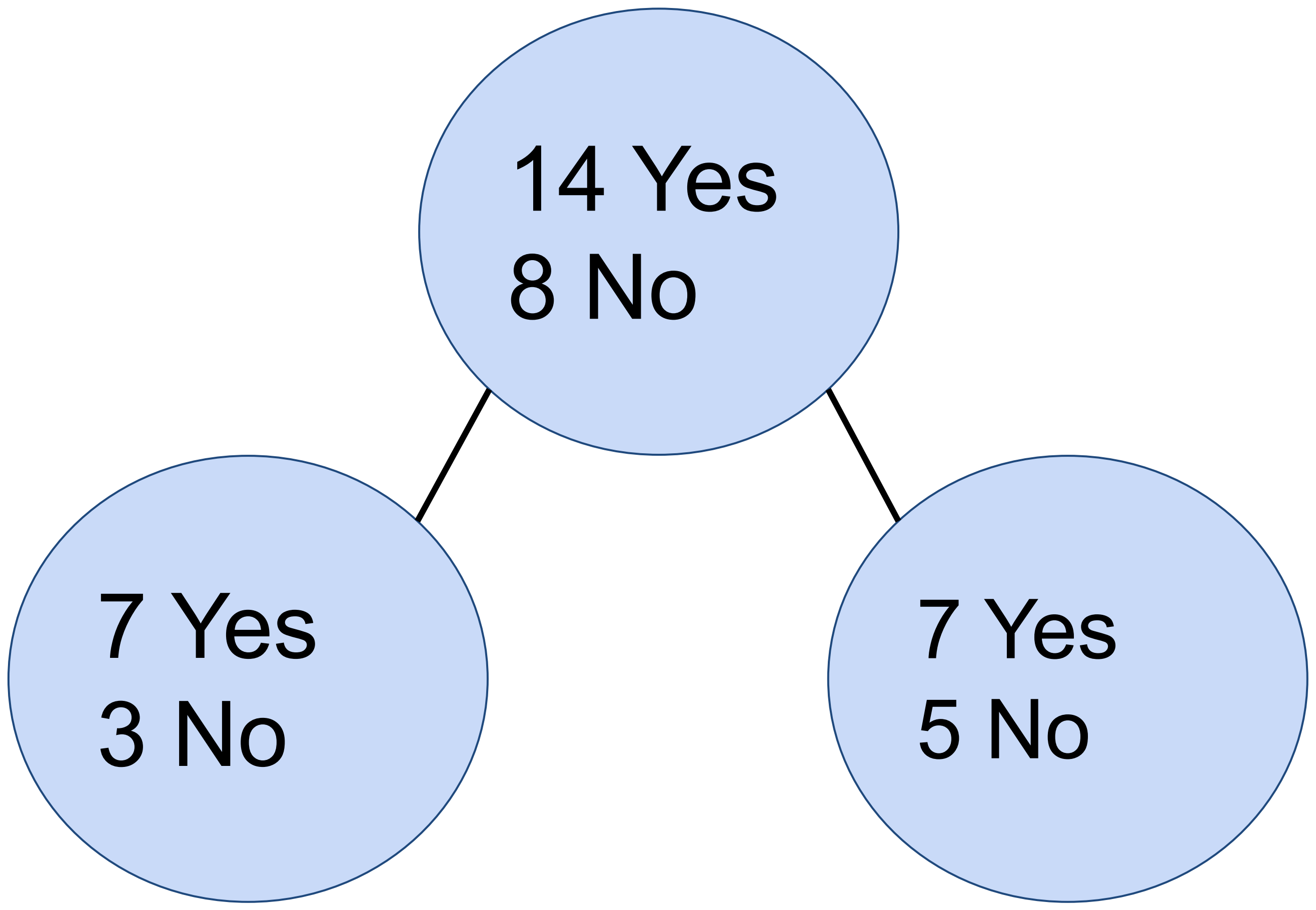
If a node has equal class member in it (e.g. 5 yes and 5 no), it has **high entropy**, and this means that the branch is practically **not ideal** as it cannot "divide" the data well enough.

Thus, we need to calculate **entropy for all nodes**, and choose the division structure in which we "**reduce the entropy as fast as possible"**.

That's why we need **Information Gain**.

Decision Tree

How does Decision Tree Work?

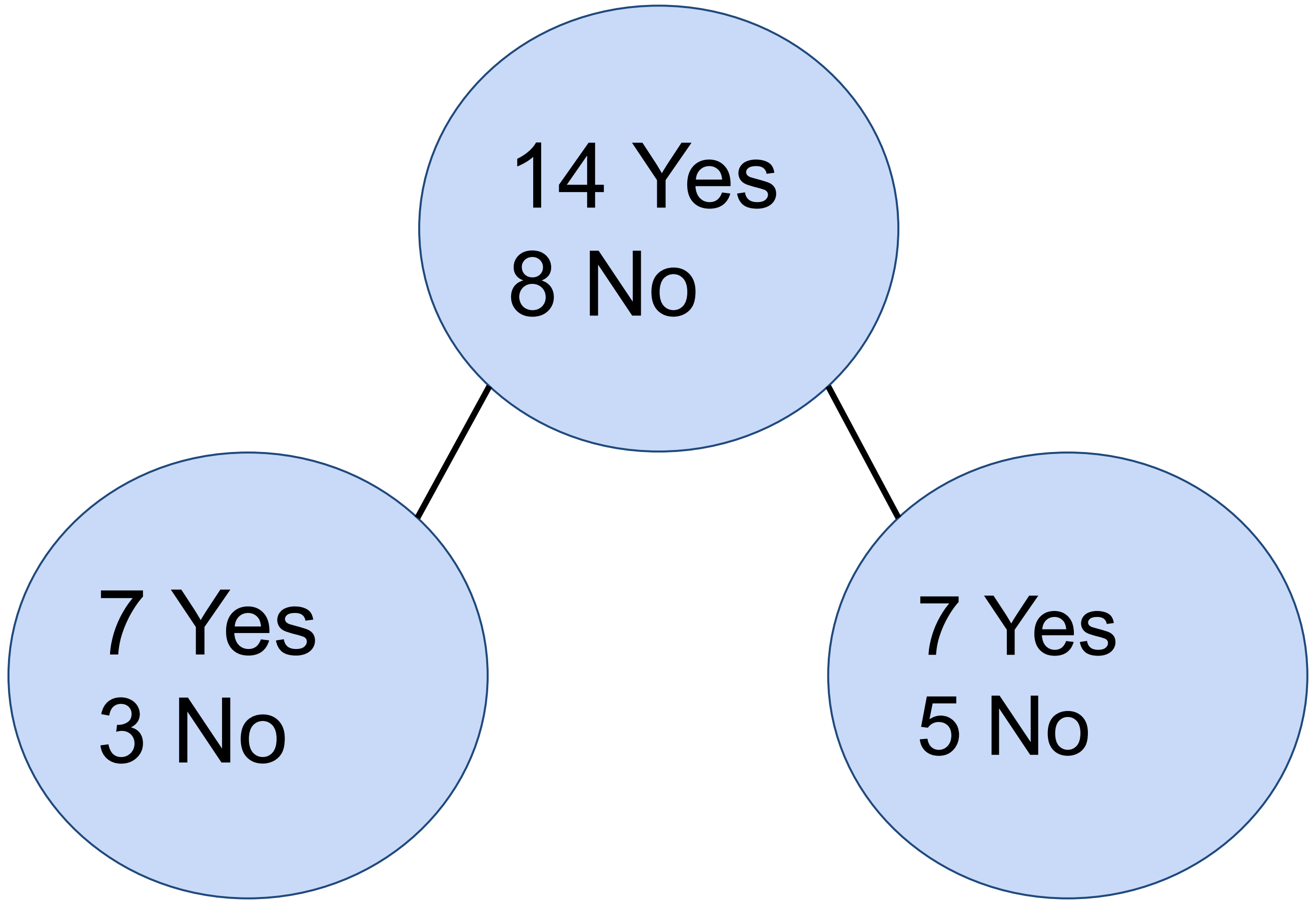


Information Gain Calculation:

Entropy of Parent Node
- **Weighted Entropy of Child Node 1**
- **Weighted Entropy of Child Node 2**

Decision Tree

How does Decision Tree Work?

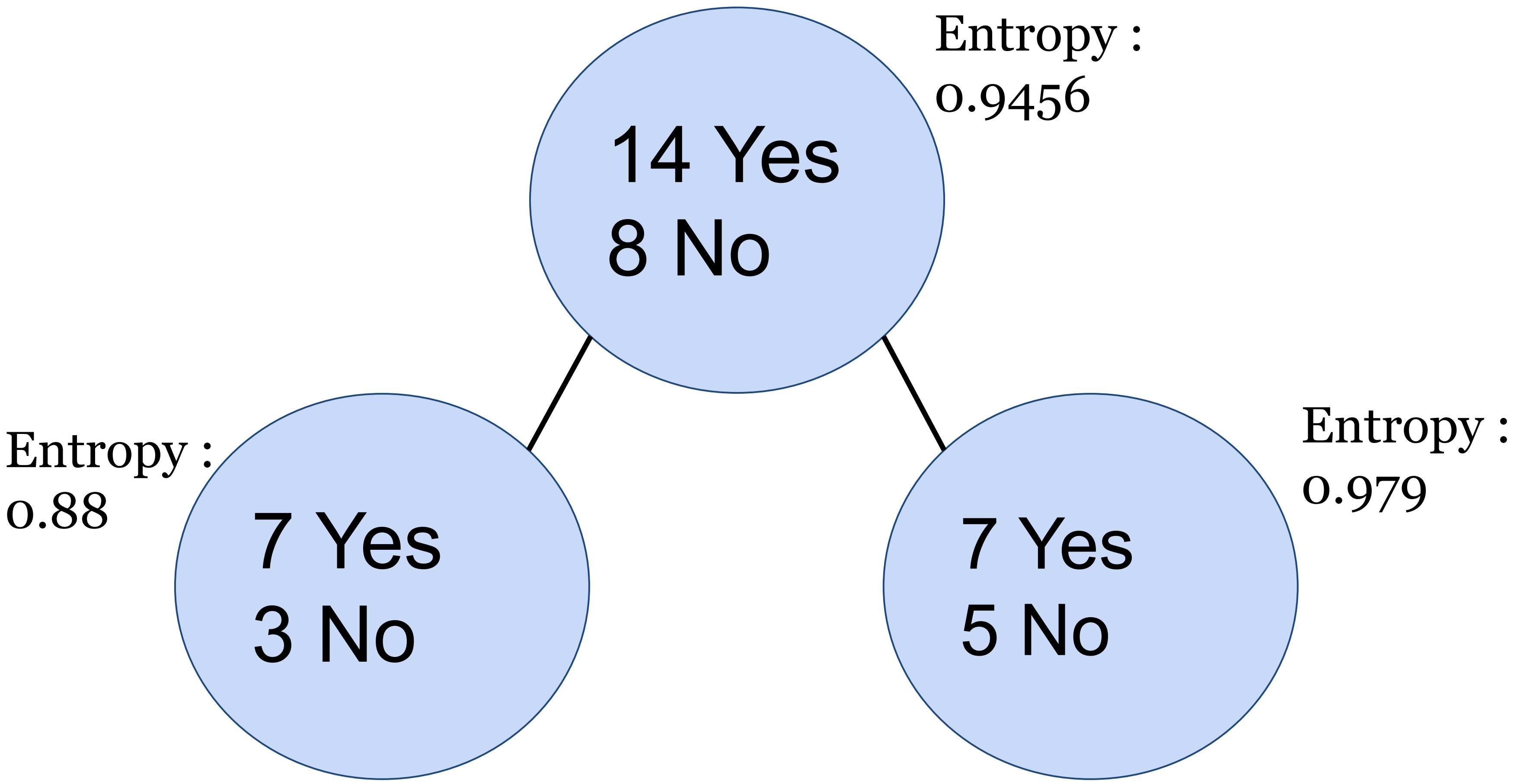


Information Gain Calculation:

Entropy of Parent Node (14 Yes 8 No)
= -(14/22)*log2(14/22) - (8/22)*log2(8/22)
= 0.9456

Decision Tree

How does Decision Tree Work?

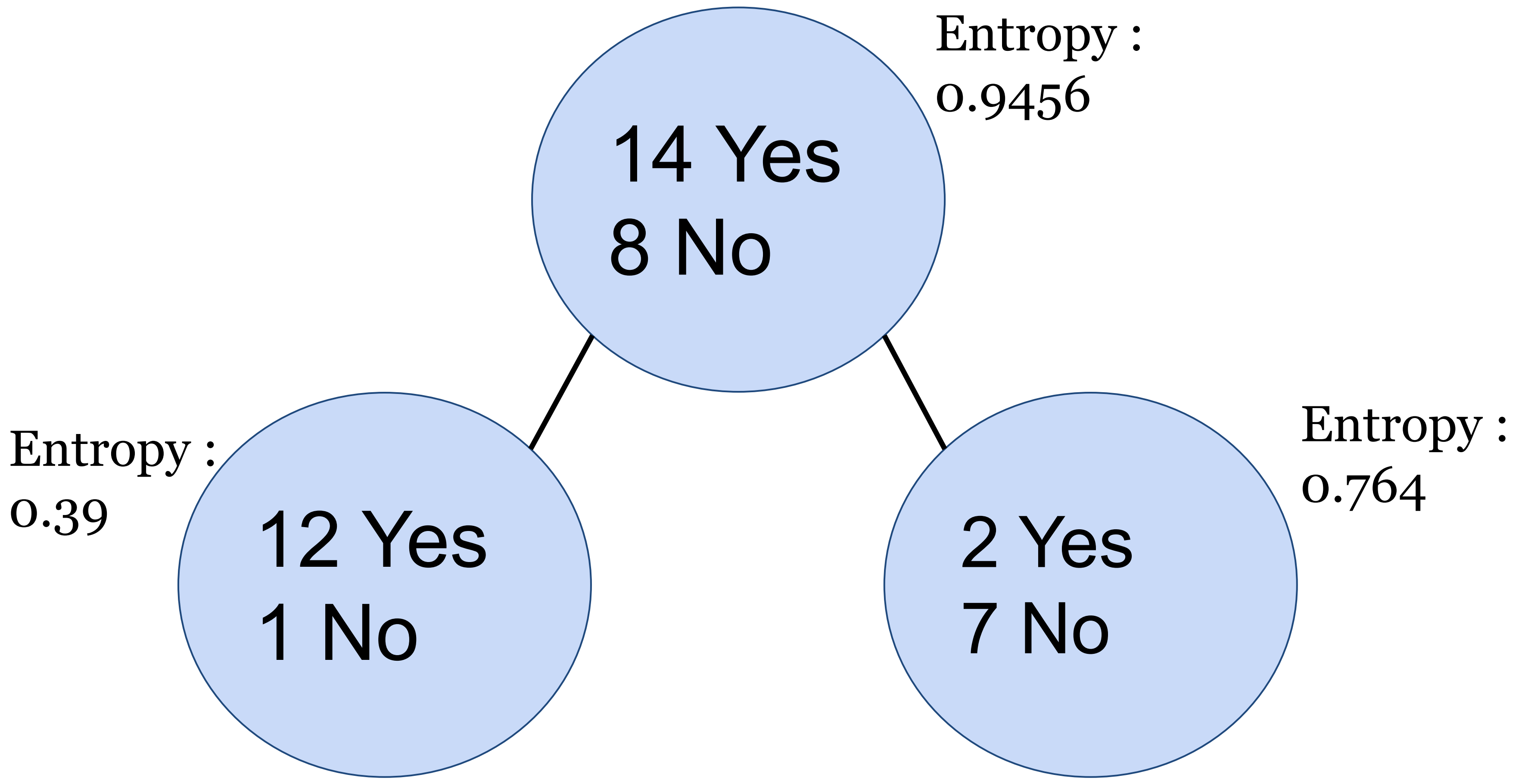


Information Gain Calculation:

0.9456 - (10/22) * 0.88 - (12/22) * 0.979
= 0.0116

Decision Tree

How does Decision Tree Work?



Information Gain Calculation:

0.9456 - (13/22) * 0.39 - (9/22) * 0.764
= 0.4026

Since the Information Gain is greater, it means we reduce more entropy, and this decision tree is better.

Decision Tree

Hands-on Decision Tree in R

Decision Tree

Summary

1. Logistic Regression
   a. Easy to implement
   b. Better performance if data is more correlated with each other

2. Naive Bayes
   a. Assumes independent in all input variables (very rare in real life case, but good enough as a ‘baseline’ model)

3. Decision Tree
   a. If input variables have different magnitude, Decision Tree is less impacted by that problem.
   b. No assumption of relationships between input variables
   c. However, very likely to overfit
   d. How to improve? Create random forest / Gradient Boosted Decision Trees

General Machine Learning Tips

1. For most cases, start out with simple models first.

2. There are 2 types of Machine Learning:
   a. Supervised
      i. Regression: predicting value
      ii. Classification: predicting class
   b. Unsupervised

3. Data preparation and understanding is really important. Don't go and directly put your raw data into your model.

4. It's easier to raise a model's accuracy from 80% to 90% than from 90% to 95%. Prioritize and allocate your time/effort wisely.

Thank you