

Nama : Hilm Fawwaz Sa'ad

NRP : 5025221103

Kelas : Pembelajaran Mesin (F)

1. Dengan menggunakan metode K-Nearest Neighbor, prediksilah class pada data ke-11 jika  $K=3$  dan  $K=5$ . (Gunakan Jarak Euklidian)

SepalLength	SepalWidth	PetalLength	PetalWidth	Class		
4.6	3.2	1.4	0.2	Iris-setosa	SP010	?
5.3	3.7	1.5	0.1	Iris-setosa	SP011	?
5	3.3	1.4	0.2	Iris-setosa	SP012	?
7	3.2	4.7	1.4	Iris-versicolor	SP013	?
6.4	3.2	4.5	1.5	Iris-versicolor	SP014	?
6.9	3.1	4.9	1.5	Iris-versicolor	SP015	?
6.5	3	5.8	2.2	Iris-virginica	SP016	?
7.6	3	6.6	2.1	Iris-virginica	SP017	?
4.9	2.5	4.5	1.7	Iris-virginica	SP018	?
7.3	2.9	6.3	1.8	Iris-virginica	SP019	?
5.7	3	4.2	1.2	?	SP020	?

Jawab :

- Menentukan parameter  $K \rightarrow K=3$  dan  $K=5$
- Menghitung jarak antara data uji dengan semua data latih (Euclidean Distance)

$$d(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2 + (a_4 - b_4)^2}$$

$$1. d(1, 11) = \sqrt{(4.6 - 5.7)^2 + (3.2 - 3)^2 + (1.4 - 4.2)^2 + (0.2 - 1.2)^2} = 3.176$$

$$2. d(2, 11) = \sqrt{(5.3 - 5.7)^2 + (3.7 - 3)^2 + (1.5 - 4.2)^2 + (0.1 - 1.2)^2} = 3.025$$

$$3. d(3, 11) = \sqrt{(5 - 5.7)^2 + (3.3 - 3)^2 + (1.4 - 4.2)^2 + (0.2 - 1.2)^2} = 3.069$$

$$4. d(4, 11) = \sqrt{(7 - 5.7)^2 + (3.2 - 3)^2 + (4.7 - 4.2)^2 + (1.4 - 1.2)^2} = 1.421$$

$$5. d(5, 11) = \sqrt{(6.4 - 5.7)^2 + (3.2 - 3)^2 + (4.5 - 4.2)^2 + (1.5 - 1.2)^2} = 0.843$$

$$6. d(6, 11) = \sqrt{(6.9 - 5.7)^2 + (3.1 - 3)^2 + (4.9 - 4.2)^2 + (1.5 - 1.2)^2} = 1.425$$

$$7. d(7, 11) = \sqrt{(6.5 - 5.7)^2 + (3 - 3)^2 + (5.8 - 4.2)^2 + (2.2 - 1.2)^2} = 2.049$$

$$8. d(8, 11) = \sqrt{(7.6 - 5.7)^2 + (3 - 3)^2 + (6.6 - 4.2)^2 + (2.1 - 1.2)^2} = 3.191$$

$$9. d(9, 11) = \sqrt{(4.9 - 5.7)^2 + (2.5 - 3)^2 + (4.5 - 4.2)^2 + (1.7 - 1.2)^2} = 1.109$$

$$10. d(10, 11) = \sqrt{(7.3 - 5.7)^2 + (2.9 - 3)^2 + (6.3 - 4.2)^2 + (1.8 - 1.2)^2} = 2.709$$

- Mengurutkan data di atas (hasil perhitungan jarak) dari kecil ke besar

$$1. d(5, 11) = 0.843 \rightarrow \text{Iris-versicolor} \quad (\text{data ke-5})$$

$$2. d(9, 11) = 1.109 \rightarrow \text{Iris-virginica} \quad (\text{data ke-9})$$

$$3. d(4, 11) = 1.421 \rightarrow \text{Iris-versicolor} \quad (\text{data ke-4})$$

$$4. d(6, 11) = 1.425 \rightarrow \text{Iris-versicolor} \quad (\text{data ke-6})$$

$$5. d(7, 11) = 2.049 \rightarrow \text{Iris-virginica} \quad (\text{data ke-7})$$

$$6. d(10, 11) = 2.709 \rightarrow \text{Iris-virginica} \quad (\text{data ke-10})$$

$$7. d(2, 11) = 3.025 \rightarrow \text{Iris-setosa} \quad (\text{data ke-2})$$

$$8. d(3, 11) = 3.069 \rightarrow \text{Iris-setosa} \quad (\text{data ke-3})$$

$$9. d(1, 11) = 3.176 \rightarrow \text{Iris-setosa} \quad (\text{data ke-1})$$

$$10. d(8, 11) = 3.191 \rightarrow \text{Iris-virginica} \quad (\text{data ke-8})$$

(KIKY)

- Mengambil data untuk  $K=3$  dan  $K=5$

$K=3$

Data	Distance	Class	
5	0.893	Iris-versicolor	2 Iris-versicolor, 1 Iris-virginica
9	1.109	Iris-virginica	
9	1.421	Iris-versicolor	

$K=5$

Data	Distance	Class	
5	0.893	Iris-versicolor	3 Iris-versicolor, 2 Iris-virginica
9	1.109	Iris-virginica	
4	1.421	Iris-versicolor	
6	1.425	Iris-versicolor	
7	2.049	Iris-virginica	

#### • Hasil prediksi

Untuk  $K=3$  menghasilkan prediksi class Iris-versicolor.

Untuk  $K=5$  menghasilkan prediksi class Iris-versicolor.

2. Selesaikan permasalahan berikut menggunakan Decision Tree berdasarkan:

a. Nilai Gini Index. Gambarkan Treenya.

b. Nilai Entropy. Gambarkan Treenya.

Hari ke-	Cuaca	Suhu	Kelambapan	Angin	Bermain Tenis
1	Cerah	Panas	Tinggi	Lemah	Tidak
2	Cerah	Panas	Tinggi	Kuat	Tidak
3	Mendung	Panas	Tinggi	Lemah	Ta
4	Hujan	Sedang	Tinggi	Lemah	Ta
5	Hujan	Dingin	Normal	Lemah	Ta
6	Hujan	Dingin	Normal	Kuat	Tidak
7	Mendung	Dingin	Normal	Kuat	Ta
8	Cerah	Sedang	Tinggi	Lemah	Tidak
9	Cerah	Dingin	Normal	Lemah	Ta
10	Hujan	Sedang	Normal	Lemah	Ta
11	Cerah	Sedang	Normal	Kuat	Ta
12	Mendung	Sedang	Tinggi	Kuat	Ta
13	Mendung	Panas	Normal	Lemah	Ta
14	Hujan	Sedang	Tinggi	Kuat	Tidak
15	Hujan	Sedang	Tinggi	Lemah	?

Jawab:

a. Gini Index

$$\text{Rumus : } \text{GINI}(t) = 1 - \sum_j [p(j|t)]^2 \quad \text{GINI}_{\text{split}} = \sum_{i=1}^n \frac{n_i}{n} \text{GINI}(i)$$

(KIKY)

- Cek atribut data satu per satu untuk menghitung gini

Cuaca	Ya	Tidak	
Cerah	2	3	Gini ( $C = \text{Cerah}$ ) = $1 - (2/5)^2 - (3/5)^2 = 0.48$
Mendung	4	0	Gini ( $C = \text{Mendung}$ ) = $1 - (4/5)^2 - (1/5)^2 = 0$
Hujan	3	2	Gini ( $C = \text{Hujan}$ ) = $1 - (3/5)^2 - (2/5)^2 = 0.48$

$$\text{Gini}_{\text{split}}(\text{Cuaca}) = 0.48 \times \frac{5}{9} + 0 \times \frac{4}{9} + 0.48 \times \frac{5}{9} = 0.342857143$$

Suhu	Ya	Tidak	
Panas	2	2	Gini ( $S = \text{Panas}$ ) = $1 - (2/4)^2 - (2/4)^2 = 0.5$
Sedang	4	2	Gini ( $S = \text{Sedang}$ ) = $1 - (4/6)^2 - (2/6)^2 = 0.444$
Dingin	3	1	Gini ( $S = \text{Dingin}$ ) = $1 - (3/4)^2 - (1/4)^2 = 0.375$

$$\text{Gini}_{\text{split}}(\text{Suhu}) = \frac{2}{9} \times 0.5 + \frac{6}{9} \times 0.444 + \frac{1}{9} \times 0.375 = 0.4402857143$$

Kelembaran	Ya	Tidak	
Tinggi	3	4	Gini ( $K = \text{Tinggi}$ ) = $1 - (3/7)^2 - (4/7)^2 = 0.490$
Normal	6	1	Gini ( $K = \text{Normal}$ ) = $1 - (6/14)^2 - (1/14)^2 = 0.245$

$$\text{Gini}_{\text{split}}(\text{Kembaran}) = \frac{7}{14} \times 0.490 + \frac{7}{14} \times 0.245 = 0.3675$$

Angin	Ya	Tidak	
Kuat	3	3	Gini ( $A = \text{Kuat}$ ) = $1 - (3/6)^2 - (3/6)^2 = 0.5$
Lemah	6	2	Gini ( $A = \text{Lemah}$ ) = $1 - (6/8)^2 - (2/8)^2 = 0.375$

$$\text{Gini}_{\text{split}}(\text{Angin}) = \frac{6}{14} \times 0.5 + \frac{8}{14} \times 0.375 = 0.428571429$$

$\approx 0.429$

#### • Informasi Gini

$$\text{Gini}(\text{Cuaca}) = 0.342857143 \quad ① \rightarrow \text{Gini indeks berbentuk sebagai root}$$

$$\text{Gini}(\text{Suhu}) = 0.440 \quad ④$$

$$\text{Gini}(\text{Kembaran}) = 0.3675 \quad ②$$

$$\text{Gini}(\text{Angin}) = 0.429 \quad ③$$

#### • Pembentukan Tree

1) Gini indeks berbentik ada di cuaca, maka cuaca menjadi root

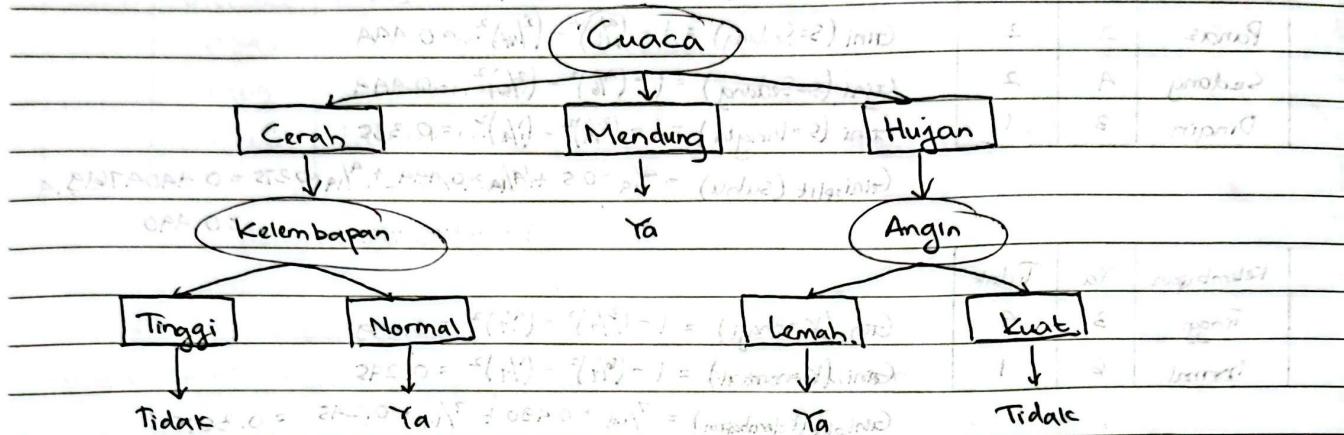
2) Membuat cabang pada cuaca dan mengeceknya apakah data tiap cabang homogen/campuran. Mendung merupakan contoh data homogen. Suhu dan Angin campuran, sehingga perlu gini dan split seperti langkah sebelumnya (per subdata).

Hari Ke -	Cuaca	Suhu	Kelembaran	Angin	Bermain tenis
1	Cerah	Panas	Tinggi	Lemah	Tidak
2	Cerah	Panas	Tinggi	Kuat	Tidak
8	Cerah	Sedang	Tinggi	Lemah	Tidak
9	Cerah	Angin	Normal	Lemah	Ya
11	Cerah	Sedang	Normal	Kuat	Ya

(KARY)

Hari Ke-	Cuaca	Suhu	Kelentapan	Angin	Bermain tenis
1	Hujan	Sedang	Tinggi	Lemah	Tidak
5	Hujan	Dingin	Normal	(Lemah)	Ya
6	Hujan	Dingin	Normal	Kuat	Tidak
10	Hujan	Sedang	Normal	Lemah	Ya
14	Hujan	Sedang	Tinggi	Kuat	Tidak

3) Lanjutkan sampai semua leaf homogen atau tidak ada atribut lagi



### b) Entropy

$$\text{Rumus : Entropy } (t) = - \sum_j p(j|t) \log_2 p(j|t)$$

$$\text{GAIN}_{\text{split}} = \text{Entropy}(p) - \left( \sum_{i=1}^n \frac{n_i}{n} \text{Entropy}(i) \right)$$

• Entropy terapis untuk keperluan split

Bermain tenis

$$\text{Ya} \quad \text{Entropy(Bermain tenis)} = - \frac{3}{14} \log_2 \left( \frac{3}{14} \right) - \frac{5}{14} \log_2 \left( \frac{5}{14} \right) = 0.94028596$$

$$\text{Tidak} \quad = 0.94028596$$

$$\text{Total} \quad = 0.94$$

• Cek atribut data satu per satu

\* Data sumbu seperti bagian Gini Index, disini hanya perhitungan.

$$\text{Cuaca} \rightarrow \text{Entropy}(\text{Cerah}) = - \left( \frac{2}{5} \right) \log_2 \left( \frac{2}{5} \right) - \left( \frac{3}{5} \right) \log_2 \left( \frac{3}{5} \right) = 0.971$$

$$\text{Entropy}(\text{Mendung}) = - \left( \frac{4}{14} \right) \log_2 \left( \frac{4}{14} \right) - \left( \frac{10}{14} \right) \log_2 \left( \frac{10}{14} \right) = 0.918$$

$$\text{Entropy}(\text{Hujan}) = - \left( \frac{3}{14} \right) \log_2 \left( \frac{3}{14} \right) - \left( \frac{2}{14} \right) \log_2 \left( \frac{2}{14} \right) = 0.971$$

$$\text{Gain}_{\text{split}}(\text{Cuaca}) = 0.94 - \left( \frac{5}{14} \times 0.971 \right) - \left( \frac{4}{14} \times 0.918 \right) - \left( \frac{3}{14} \times 0.971 \right) = 0.24692857 \approx 0.246$$

$$\text{Suhu} \rightarrow \text{Entropy}(\text{Panas}) = - \left( \frac{2}{4} \right) \log_2 \left( \frac{2}{4} \right) - \left( \frac{2}{4} \right) \log_2 \left( \frac{2}{4} \right) = 1$$

$$\text{Entropy}(\text{Sedang}) = - \left( \frac{4}{14} \right) \log_2 \left( \frac{4}{14} \right) - \left( \frac{2}{14} \right) \log_2 \left( \frac{2}{14} \right) = 0.918$$

$$\text{Entropy}(\text{Dingin}) = - \left( \frac{3}{14} \right) \log_2 \left( \frac{3}{14} \right) - \left( \frac{1}{14} \right) \log_2 \left( \frac{1}{14} \right) = 0.811$$

$$\text{Gain}_{\text{split}}(\text{Suhu}) = 0.94 - \left( \frac{4}{14} \times 1 \right) - \left( \frac{4}{14} \times 0.918 \right) - \left( \frac{1}{14} \times 0.811 \right) = 0.029142857 \approx 0.029$$

(KIKY)

Nama : Hilmi Fawwaz Sa'ad

NRP : 5025221103

Kelas : Pembelajaran Mesin (F)

(Lanjutan)

$$\text{Kelembapan} \rightarrow \text{Entropy (Tinggi)} = -\left(\frac{2}{7}\right) \log_2 \left(\frac{2}{7}\right) - \left(\frac{4}{7}\right) \log_2 \left(\frac{4}{7}\right) = 0.985$$

$$\text{Entropy (Normal)} = -\left(\frac{4}{7}\right) \log_2 \left(\frac{4}{7}\right) - \left(\frac{1}{7}\right) \log_2 \left(\frac{1}{7}\right) = 0.592$$

$$\text{Gain}_{\text{split}}(\text{Kelembapan}) = 0.99 - \left(\frac{7}{14} \times 0.985 + \frac{7}{14} \times 0.592\right) = 0.1515$$

$$\approx 0.152$$

$$\text{Angin} \rightarrow \text{Entropy (Kuat)} = -\left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) \log_2 \left(\frac{3}{6}\right) = 1$$

$$\text{Entropy (Lemah)} = -\left(\frac{6}{8}\right) \log_2 \left(\frac{6}{8}\right) - \left(\frac{2}{8}\right) \log_2 \left(\frac{2}{8}\right) = 0.811$$

$$\text{Gain}_{\text{split}}(\text{Angin}) = 0.99 - \left(\frac{6}{14} \times 1 + \frac{6}{14} \times 0.811\right) = 0.098$$

• Informasi Gain

$$\text{Gain}(\text{Cuaca}) = 0.296 \quad \textcircled{1} \rightarrow \text{Gain terbaik sebagai root}$$

$$\text{Gain}(\text{Suhu}) = 0.029 \quad \textcircled{2}$$

$$\text{Gain}(\text{Kelembapan}) = 0.152 \quad \textcircled{3}$$

$$\text{Gain}(\text{Angin}) = 0.098 \quad \textcircled{4}$$

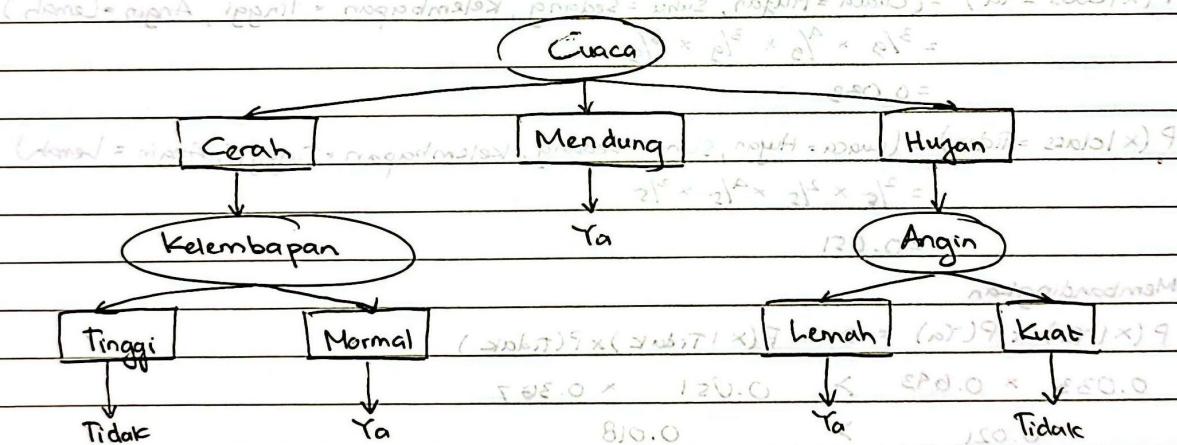
• Pembentukan Tree

1) Gain terbaik ada di Cuaca, maka cuaca menjadi root

2) Membuat cabang pada cuaca. Data homogen hanya yang mendung? Selain mendung datanya berupa campuran, jadi harus mencari entropy dan gain split lagi.

3) Lanjutkan sampai semua leaf homogen atau tidak ada atribut lagi

$$(\text{Cerah} = \text{normal} = \text{angin lemah}) \quad (\text{Mendung} = \text{angin normal}) \quad (\text{Hujan} = \text{angin kuat})$$



3. Berdasarkan data pada nomor 2, prediksilah data ke-15 menggunakan Naive Bayes!

Jawab:

• Mencari peluang dari Class

$$P(Ya) = 9/14 = 0.643$$

$$P(Tidak) = 5/14 = 0.357$$

• Mencari peluang dari masing-masing atribut

Cuaca	Ya	Tidak		
Cerah	2	3	$P(\text{Cuaca} = \text{Cerah}   Ya) = 2/5$	$P(\text{Cuaca} = \text{Cerah}   \text{Tidak}) = 3/5$
Mendung	4	0	$P(\text{Cuaca} = \text{Mendung}   Ya) = 4/9$	$P(\text{Cuaca} = \text{Mendung}   \text{Tidak}) = 0/5$
Hujan	3	2	$P(\text{Cuaca} = \text{Hujan}   Ya) = 3/9$	$P(\text{Cuaca} = \text{Hujan}   \text{Tidak}) = 2/5$
Total	9	5		

(contoh)

Suhu	Ya	Tidak
Panas	2	2
Sedang	4	2
Dingin	3	1
Total	9	5

$$\begin{aligned} P(\text{suhu} = \text{Panas} | \text{Ya}) &= \frac{2}{9} & P(\text{suhu} = \text{Panas} | \text{Tidak}) &= \frac{2}{5} \\ P(\text{suhu} = \text{Sedang} | \text{Ya}) &= \frac{4}{9} & P(\text{suhu} = \text{Sedang} | \text{Tidak}) &= \frac{2}{5} \\ P(\text{suhu} = \text{Dingin} | \text{Ya}) &= \frac{3}{9} & P(\text{suhu} = \text{Hujan} | \text{Tidak}) &= \frac{1}{5} \end{aligned}$$

Kelembapan	Ya	Tidak
Tinggi	3	4
Normal	6	1
Total	9	5

$$\begin{aligned} P(\text{kelembapan} = \text{Tinggi} | \text{Ya}) &= \frac{3}{9} & P(\text{kelembapan} = \text{Tinggi} | \text{Tidak}) &= \frac{4}{5} \\ P(\text{kelembapan} = \text{Normal} | \text{Ya}) &= \frac{6}{9} & P(\text{kelembapan} = \text{Normal} | \text{Tidak}) &= \frac{1}{5} \end{aligned}$$

Angin	Ya	Tidak
Lemah	6	2
Kuat	3	3
Total	9	5

$$\begin{aligned} P(\text{Angin} = \text{Lemah} | \text{Ya}) &= \frac{6}{9} & P(\text{Angin} = \text{Lemah} | \text{Tidak}) &= \frac{2}{5} \\ P(\text{Angin} = \text{Kuat} | \text{Ya}) &= \frac{3}{9} & P(\text{Angin} = \text{Kuat} | \text{Tidak}) &= \frac{3}{5} \end{aligned}$$

- Menghitung peluang data ke-15 untuk setiap faktor cuaca, suhu, kelembapan, angin.

$$\begin{aligned} P(x|class = \text{Ya}) &= (\text{cuaca} = \text{Hujan}, \text{suhu} = \text{Sedang}, \text{kelembapan} = \text{Tinggi}, \text{Angin} = \text{Lemah}) \\ &= \frac{3}{9} \times \frac{4}{9} \times \frac{3}{9} \times \frac{6}{9} \\ &= 0.033 \end{aligned}$$

$$\begin{aligned} P(x|class = \text{Tidak}) &= (\text{cuaca} = \text{Hujan}, \text{suhu} = \text{Sedang}, \text{kelembapan} = \text{Tinggi}, \text{Angin} = \text{Lemah}) \\ &= \frac{2}{9} \times \frac{2}{5} \times \frac{4}{5} \times \frac{2}{5} \\ &= 0.051 \end{aligned}$$

#### • Membandingkan

$$P(x|\text{Ya}) \times P(\text{Ya}) > P(x|\text{Tidak}) \times P(\text{Tidak})$$

$$0.033 \times 0.643 > 0.051 \times 0.357$$

$$0.021 > 0.018$$

Hasil perbandingannya adalah sisi kiri lebih besar ditandingkan sisi kanan, sehingga tanda baca menjadi >.

#### • Hasil

Kesimpulan untuk data ke-15 kolom "Bermain tenis" adalah Ya.

$$P(A) = P(X) = 0.021$$

$$P(B) = P(Y) = 0.018$$

diketahui probabilitas bahwa probabilitas.

$$P(A) = (m_A / m_{total})^2$$

$$P(A) = (n_A / m_{total})^2$$

$$P(A) = (w_A / m_{total})^2$$

$$P(B) = (m_B / m_{total})^2$$

$$P(B) = (n_B / m_{total})^2$$

$$P(B) = (w_B / m_{total})^2$$

$$P(A) = (m_A / m_{total})^2$$

$$P(A) = (n_A / m_{total})^2$$

$$P(A) = (w_A / m_{total})^2$$

(KIRI)