

Nama : Shafa Aldiena Meraldin
Email : shafaaldienameraldin@gmail.com
Kelas : SUP MAX November 2024 - Data Science

Soal Mentoring Probability

Job Preparation Program, Data Science, Pacmann AI

Catatan:

- Make a copy docs ini sebelum menjawab.
- Ada dua fokus dalam menjawab soal ini
 - Menemukan jawaban yang benar
 - Memahami cara mendapatkan jawaban yang benar
- Selama mengerjakan soal, Anda **wajib melampirkan** cara & proses menjawab. **BUKAN CUMA** pasang rumus, tapi **berikan elaborasi** kenapa Anda melakukan hal tersebut.
- Silahkan lampirkan proses Anda dengan menuliskan langsung pada docs atau melampirkan foto proses (yang dapat dibaca).
- Kumpulkan docs Anda sesuai dengan link submission yang tersedia.
- Selamat mengerjakan 😊

Set dan Kombinatorika [20 points]

1. **[5 points]** Kimia adalah perusahaan yang bergerak dibidang produksi dan distribusi obat-obatan. Ada 10 apotek berbeda yang tiap bulannya dikirimkan obat-obatan oleh Kimia. **Hitung ada berapa pilihan rute pengiriman yang bisa dilakukan.**

Catatan:

Tulis jawaban disini

Diketahui ada 10 Apotek berbeda. Misal Apotek 1 sampai Apotek 10.

Karena urutan pengiriman penting (misalnya, rute $1 \rightarrow 2 \rightarrow 3$ berbeda dari $2 \rightarrow 3 \rightarrow 1$), maka soal ini dapat diselesaikan dengan permutasi jalur distribusi dari 10 apotek berbeda. Permutasi adalah cara penyusunan dengan urutan pada suatu kumpulan objek.

Permutasi yang digunakan adalah tanpa pengulangan karena agar jalur distribusi efektif dan efisien maka tidak boleh mendatangi apotek yang sama lebih dari 1 kali.

1. Formula permutasi dari n objek adalah:
$$P(n) = n! = n(n-1)(n-2)\dots 1 = n!$$
2. Perhitungan permutasi dari jalur distribusi 10 apotek berbeda :
$$P(10) = 10! = 10(10-1)(10-2)\dots 1$$
$$= 3.628.800 \text{ pilihan rute}$$
3. Pengerjaan di python :

```

from itertools import permutations
from math import factorial

# Jumlah rute
jumlah_rute = factorial(10)
print("Jumlah pilihan rute:", jumlah_rute)

# Menampilkan 10 rute pertama
apotek = [1,2,3,4,5,6,7,8,9,10]
perm = permutations(apotek)

print("\n10 rute pengiriman pertama:")
for i, rute in enumerate(perm):
    if i == 10:
        break
    print(f"Rute {i+1}: {' '.join(map(str, rute))}")

```

Jumlah pilihan rute: 3628800

10 rute pengiriman pertama:
Rute 1: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 8 → 9 → 10
Rute 2: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 8 → 10 → 9
Rute 3: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 9 → 8 → 10
Rute 4: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 9 → 10 → 8
Rute 5: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 10 → 8 → 9
Rute 6: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 10 → 9 → 8
Rute 7: 1 → 2 → 3 → 4 → 5 → 6 → 8 → 7 → 9 → 10
Rute 8: 1 → 2 → 3 → 4 → 5 → 6 → 8 → 7 → 10 → 9
Rute 9: 1 → 2 → 3 → 4 → 5 → 6 → 8 → 9 → 7 → 10
Rute 10: 1 → 2 → 3 → 4 → 5 → 6 → 8 → 9 → 10 → 7

2. [5 points] Tim marketing butuh melakukan analisa performa dari marketing campaign kepada tim Data Analyst. Setelah requirement gathering, ternyata butuh 2 orang data analyst agar bisa menyelesaikan analisa performa tersebut. Tim data analyst sekarang terdiri dari 5 orang. **Hitung banyak pasangan data analyst yang bisa dipilih untuk membantu tim marketing.**

Catatan:

Tulis jawaban disini

1. Karena urutan pemilihan tidak penting (misalnya, memilih Analyst A dan B sama saja dengan memilih B dan A), maka ini adalah soal kombinasi. Kombinasi adalah cara memilih sejumlah objek dari sekumpulan objek tanpa memperhatikan urutan. Kombinasi digunakan saat yang penting hanya siapa yang dipilih, bukan urutannya.
2. Perhitungan kombinasi 2 DA dari 5 orang tim DA

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\begin{aligned}
C(5,2) &= 5! / 2!(5-2)! \\
&= 5 \times 4 \times 3! / 2 \times 1 \times 3! \\
&= 20/2 \\
&= 10 \text{ pasangan}
\end{aligned}$$

3. Jadi, ada 10 pilihan pasangan analyst untuk membantu tim marketing
4. Python :

```
from math import comb

# Jumlah kombinasi memilih 2 dari 5
jumlah_pasangan = comb(5, 2)
print("Jumlah pasangan data analyst:", jumlah_pasangan)
```

Jumlah pasangan data analyst: 10

3. Tim product berencana melakukan user interview agar bisa memperbaiki productnya. Untuk melakukan user interview, tim product mencari user yang memiliki kriteria
- Sudah menggunakan product lebih dari 30 hari, **dan**
 - Memberikan feedback negatif lebih dari 25% dari total transaksi yang dilakukan.

Tim data membantu melakukan query data berdasarkan karakteristik di atas. Karena tim datanya masih junior, informasi yang didapat adalah sebagai berikut

- Ada 500 user yang menggunakan product lebih dari 30 hari
- Ada 350 user yang memberikan feedback negatif lebih dari 25% dari total transaksi yang dilakukan
- Ada total 799 user yang memenuhi minimal satu dari dua kriteria tersebut.

Pertanyaannya

- **[5 points]** Diasumsikan tim product butuh minimal mewawancarai 50 user, **apakah tim product dapat menemukan user tersebut dengan kriteria yang telah dijelaskan?**

Catatan:

Tulis jawaban disini

1. Kita misalkan informasi di atas sebagai set
Set A : user yang menggunakan produk > 30 hari = 500
Set B: user yang memberikan feedback negatif > 25% = 350
 $|A \cup B|$: total user yang memenuhi minimal satu kriteria = 799
2. Cari apakah $A \cap B \geq 50$ user
3. $|A \cap B| = |A| + |B| - |A \cup B|$
 $= 500 + 350 - 799$
 $= 51$

Maka total user yang memenuhi 2 kriteria A dan B adalah 51 user

4. Jadi, tim product dapat menemukan user dengan 2 kriteria tersebut.
- **[5 points]** Ada berapa banyak cara untuk bisa memilih 50 user yang siap diwawancarai tim product?

Catatan:

Tulis jawaban disini

1. Karena urutan pemilihan tidak penting (misalnya, memilih User 1 - 50 sama saja dengan memilih user 2 - 51), maka ini adalah soal kombinasi.

Kombinasi adalah cara memilih sejumlah objek dari sekumpulan objek tanpa memperhatikan urutan. Kombinasi digunakan saat yang penting hanya siapa yang dipilih, bukan urutannya.

2. Karena hanya ingin memilih 50 dari 51 user **tanpa memperhatikan urutan**, maka gunakan **kombinasi**

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\begin{aligned} C(51,50) &= 51! / 50! (51-50)! \\ &= 51! / 50! \times 1! \\ &= 51 \text{ cara memilih user} \end{aligned}$$

3. Jadi, terdapat 51 cara memilih user dari 51 yang memenuhi 2 kriteria.
4. Python :

```
[2] from math import comb

# Jumlah kombinasi memilih 50 dari 51
jumlah_cara = comb(51, 50)
print("Jumlah cara memilih user:", jumlah_cara)
```

Jumlah cara memilih user: 51

Probability [30 points]

4. [7 points] Tim product mengembangkan fitur “Easter Egg” untuk menambah retensi dari user. Feature tersebut ketika digunakan akan memberikan promo belanja secara acak senilai Rp 0, Rp 5.000, Rp 10.000, Rp 20.000, dan Rp 50.000. Rudi menggunakan feature “Easter Egg” selama 4x. **Berapa peluang Rudi mendapatkan total promo lebih besar dari Rp 25.000 dalam 4x penggunaan?**

Catatan:

Tulis jawaban disini

Diketahui :

Promo : {0, 5000, 10000, 20000, 50000}

Peluang total promo > 25000 dalam 4x penggunaan (percobaan)

1. Hitung total kemungkinan hasil
Ada 5 kemungkinan dalam 4 percobaan
Maka total kemungkinan hasil -> $n(S) = 5^4 = 625$
2. Cari $n(\text{Promo} > 25000)$ dengan mencari $n(\text{Promo} \leq 25000)$ atau banyak kombinasi yang jumlahnya ≤ 25000 karena lebih mudah dan lebih sedikit kemungkinannya karena $P(\text{Promo} > 25000) = 1 - P(\text{Promo} \leq 25000)$

n(0) -> Kombinasi total = 0

$$0 + 0 + 0 + 0 = 0$$

(1 cara)

n(5000) ->Kombinasi total 5000

Case 1 : (5000, 0, 0, 0) namun urutan dihitung kejadian berbeda

Permutasi multiset $4! = 4$ cara

n(10000) ->Kombinasi total 10000

Case 1 : (10000, 0, 0, 0)

Permutasi multiset $4! / (1! \times 3!) = 4$ cara

Case 2 : (5000, 5000, 0, 0)

Permutasi multiset $4! / (2! \times 2!) = 6$ cara

Total = 10 cara

n(15000) ->Kombinasi total 15000

Case 1 : (5000, 5000, 5000, 0)

Permutasi multiset $= 4! / (3! \times 1!) = 4$ cara

Case 2 : (10000, 5000, 0, 0)

Permutasi multiset $= 4! / (1! \times 1! \times 2!) = 12$ cara

Total = 16 cara

n(20000) ->Kombinasi total 20000

Case 1 : (20000, 0, 0, 0)

Permutasi multiset $= 4! / (1! \times 3!) = 4$ cara

Case 2 : (10000, 5000, 5000, 0)

Permutasi multiset $= 4! / (1! \times 2! \times 1!) = 12$ cara

Case 3 : (10000, 10000, 0, 0)

Permutasi multiset $= 4! / (2! \times 2!) = 6$ cara

Case 4 : (5000, 5000, 5000, 5000)

Permutasi multiset $= 4! / (4!) = 1$ cara

Total = 23 cara

n(25000) -> Kombinasi total 25000

Case 1 : (20000, 5000, 0, 0)

Permutasi multiset $= 4! / (1! \times 1! \times 2!) = 12$ cara

Case 2 : (10000, 5000, 5000, 5000)

Permutasi multiset $= 4! / (1! \times 3!) = 4$ cara

Case 3 : (10000, 10000, 5000,0)

Permutasi multiset $= 4! / (2! \times 1! \times 1!) = 12$ cara

Total = 28 cara

3. Jumlahkan semua kejadian promo kurang dari 25000

$$\begin{aligned} n(\text{Promo} \leq 25000) &= n(0) + n(5000) + n(10000) + n(20000) + n(25000) \\ &= 1 + 4 + 10 + 16 + 23 + 28 \end{aligned}$$

= 82 cara

4. Hitung jumlah kemungkinan kejadian promo lebih dari 25000 dengan mengurangi sample space dengan total kejadian promo < 25000

Maka $n(\text{Promo} > 25000) = n(S) - n(\text{Promo} \leq 25000)$

= $625 - 82$

= 543 cara

5. Hitung probabilitas Rudi mendapat promo > 25000 dengan probability formula

$P(\text{Promo} > 25000) = n(\text{Promo} > 25000) / n(S)$

= $543 / 625 = 0,868$

Jadi probabilitas Rudi mendapat total promo lebih besar dari 25000 adalah 0,868 atau 86,88%

6. Python :

```
[12] from itertools import product

# Nilai promo yang mungkin
promos = [0, 5000, 10000, 20000, 50000]

# Semua kemungkinan 4 kali pemakaian
kombinasi = list(product(promos, repeat=4))

# Hitung jumlah kombinasi dengan total <= 25.000
count_kurang_sama_25rb = sum(1 for k in kombinasi if sum(k) <= 25000)

# Total kombinasi
total_kombinasi = len(kombinasi)

# Hitung kombinasi dengan total > 25.000
count_lebih_25rb = total_kombinasi - count_kurang_sama_25rb

# Peluang
peluang = count_lebih_25rb / total_kombinasi

print(f"Jumlah kombinasi total <= 25.000: {count_kurang_sama_25rb}")
print(f"Jumlah kombinasi total > 25.000: {count_lebih_25rb}")
print(f"Total kombinasi: {total_kombinasi}")
print(f"Peluang Rudi dapat total > 25.000: {peluang:.4f} atau {peluang*100:.2f}%")
```

Jumlah kombinasi total <= 25.000: 82
Jumlah kombinasi total > 25.000: 543
Total kombinasi: 625
Peluang Rudi dapat total > 25.000: 0.8688 atau 86.88%

5. [3 points] **Riiseet!** adalah perusahaan yang bergerak di bidang media penelitian ilmiah. Untuk bisa menjaga subscribarnya, **Riiseet!** memberikan 1 rekomendasi bacaan dari bacaan yang terakhir user baca. Asumsikan munculnya rekomendasi yang bagus & buruk adalah seimbang. **Jika Anda mendapatkan 5x rekomendasi, berapa peluang Anda mendapatkan 3 rekomendasi bacaan yang sesuai?**

Catatan:

Tulis jawaban disini

Soal ini dapat dikerjakan dengan permutasi multiset karena kita menyusun urutan dari elemen-elemen yang identik (berulang) dan tetap memperhatikan urutan karena tiap rekomendasi bergantung dengan apa yang user baca sebelumnya.

1. Hitung ruang sampel
 Karena tiap rekomendasi bisa bagus/tidak (50:50)
 Maka $n(s) = 2^5 = 32$
2. Hitung kemungkinan kejadian susunan 3 baik dan 2 buruk
 $n(3Ba2Bu) = 5! / (3! \times 2!) = 10$ susunan
3. Hitung peluang $P(3Ba3Bu)$ dengan probability formula

$$P(3Ba2Bu) = n(3Ba2Bu) / n(S)$$

$$= 10/32$$

$$= 0.3125$$

Jadi probabilitas mendapatkan 3 rekomendasi bacaan yang baik adalah 0.3125 atau 31.25%

6. Sebagai tim marketing yang *data-driven*, Anda membuat promo marketing berikut "Belanja 10x dengan Pay Later dapat voucher 10 juta".

Pertanyaannya

- **[5 points] Berapa peluang seseorang melakukan belanja 10x berturut-turut dengan Pay Later?**

Asumsikan seseorang pasti belanja dan hanya memiliki opsi bayar dengan Pay Later atau tidak. Asumsikan juga media bayar antar transaksi tidak saling berpengaruh.

Catatan:

Tulis jawaban disini

1. Peluang user Pay Later atau tidak Pay = 50:50
 Karena harus belanja berulang 10 kali
2. Hitung kemungkinan probabilitas

$$P(\text{PayLater } 10x) = P(\text{PayLater}) * P(\text{PayLater } 2) * \dots * P(\text{PayLater } 10)$$

$$= 0,5 ^ 10$$

$$= 1/(2^{10})$$

$$= 1/1024$$

$$= 0.0009766$$
3. Jadi, peluang user melakukan belanja 10x berturut-turut dengan metode Pay Later adalah sebesar 0.0009766 atau 0,097% atau sangat rendah sekali
4. Python:

```
[13] p = 0.5 # peluang bayar Pay Later per transaksi
      n = 10 # jumlah transaksi berturut-turut

      peluang = p ** n

      print(f"Peluang bayar Pay Later 10x berturut-turut adalah {peluang:.6f} atau {peluang*100:.4f}%")
```

Peluang bayar Pay Later 10x berturut-turut adalah 0.000977 atau 0.0977%

- **[3 points] Berapa budget marketing yang harus disiapkan untuk promo ini?**
Asumsikan ada 10.000 user yang dapat mengakses promo ini.

Catatan:

Tulis jawaban disini

1. $P(\text{PayLater } 10x) = 0.0009766$
2. Hitung ekspektasi banyak user yang menang promo

$$\begin{aligned}\text{Budget} &= P(\text{PayLater } 10x) \times \text{Jumlah User} \times \text{Harga Voucher} \\ &= 0.0009766 \times 10_000 \times 10_000_000 \\ &= 9,7 \times 100_000_000 \\ &\approx 10 \text{ Users} \times 10_000_000 \\ &= \text{Rp. } 100_000_000\end{aligned}$$
3. Jadi, jumlah total budget yang perlu disiapkan adalah Rp 100.000.000 jika asumsi user yang mengakses adalah 10.000 users.
4. Python:

```
import math
jumlah_user = 10000
peluang_sukses = (1/2) ** 10 # peluang 10x berturut-turut pakai Pay Later
nilai_voucher = 10_000_000 # Rp 10 juta

# Ekspektasi jumlah user yang berhasil
ekspektasi_user_berhasil = math.ceil(jumlah_user * peluang_sukses)

# Total budget marketing
total_budget = ekspektasi_user_berhasil * nilai_voucher

# Cetak hasil
print(f"Ekspektasi jumlah user yang berhasil: {ekspektasi_user_berhasil:.2f}")
print(f"Total budget marketing: Rp {total_budget:,.0f}")
```

Ekspektasi jumlah user yang berhasil: 10.00
 Total budget marketing: Rp 100,000,000

- **[12 points] Apabila rata-rata revenue yang didapat perusahaan dari user yang bertransaksi dengan Pay Later adalah Rp 2.000/transaksi, apakah Anda menghasilkan untung dengan menjalankan promo marketing tersebut?**

Catatan:

Tulis jawaban disini

Untuk mengetahui apakah ada profit atau loss maka dilakukan perhitungan dengan menghitung selisih antara budget yang dikeluarkan dan revenue yang masuk. Perlu juga memperhatikan behavior transaksi dari 10.000 user yang ada untuk mengetahui berapa user melakukan berapa transaksi.

1. Diketahui bahwa ada asumsi media bayar antar transaksi tidak saling berpengaruh
 Peluang bayar Pay Later = 50:50 atau 0.5 menunjukkan 2 kemungkinan yang seimbang (pakai pay later atau tidak)
 Percobaan berbentuk 10 kali transaksi identik
 Maka, kita asumsikan peluang penggunaan n pay later dan estimasi jumlah user untuk tiap kategori jumlah transaksi mengikuti **distribusi binomial**

2. Hitung total peluang user melakukan transaksi Pay Later sebanyak n kali ($n = 1-10$)
Distribusi binomial $B(n=10, p=0.5)$
 $= 0.5^n \times 0.5$
 $= 0.5^{n+1}$ untuk $n = 1, 2, 3, \dots, 9$
Untuk $n = 10$ maka $P(10) = 0.5^{10}$ karena batas maksimal 10 kali transaksi
Contoh perhitungan :
 $P(9) = 0.5^9 \times 0.5 = 0.0009766$
3. Hitung total estimasi jumlah user per transaksi dengan distribusi binomial juga
Distribusi binomial $B(n=10, p=0.5)$
 $= 0.5^n \times 0.5$
 $= 0.5^{n+1}$ untuk $n = 1, 2, 3, \dots, 9$
4. Untuk mempermudah, perhitungan dilakukan dengan python. Berikut perhitungan peluang dan distribusi n kali transaksi paylater, estimasi jumlah user tiap kategori kemungkinan transaksi pay later, dan revenue dari tiap kategorinya:

```

from math import comb, floor

# Parameter
n = 10 # jumlah transaksi
p = 0.5 # peluang PayLater per transaksi
total_users = 10000
revenue_per_transaction = 2000 # dalam Rupiah

# Hitung probabilitas dan estimasi user dengan floor
probabilities = [comb(n, k) * (p ** k) * ((1 - p) ** (n - k)) for k in range(n + 1)]
raw_users = [prob * total_users for prob in probabilities]
floored_users = [floor(u) for u in raw_users]

# Hitung sisa pembulatan dan alokasikan agar total = 10000
remainder = total_users - sum(floored_users)
remainders = [(raw - floor(raw), idx) for idx, raw in enumerate(raw_users)]
remainders.sort(reverse=True) # urutkan berdasarkan sisa terbanyak

# Tambahkan sisa ke kategori terbanyak
for i in range(remainder):
    floored_users[remainders[i][1]] += 1

# Langkah 3: Hitung total revenue
total_revenue = 0
print(f"{'PayLater x Times':<18}{'Estimated Users':<20}{'Revenue (Rp)':<15}")

for k in range(n + 1):
    estimated_users = floored_users[k]
    revenue = estimated_users * k * revenue_per_transaction
    total_revenue += revenue
    print(f"{k:<18}{estimated_users:<20}{revenue:<15}")

print(f"Total Revenue dari semua user: Rp {total_revenue:,}")

```

PayLater x Times	Estimated Users	Revenue (Rp)
0	10	0
1	97	194,000
2	439	1,756,000
3	1172	7,032,000
4	2051	16,408,000
5	2461	24,610,000
6	2051	24,612,000
7	1172	16,408,000
8	439	7,024,000
9	98	1,764,000
10	10	200,000
Total Revenue dari semua user:		Rp 100,008,000

5. Hitung Total Revenue

Revenue 10 user = $2000 \times 10 \times 10$
= 200_000

Revenue 9990 user = $\text{sum}(n \times 2000 \times \text{revenue})$
= 99_808_000

Total 10_000 user = 100_008_000

6. Hitung profit/loss

Profit/loss = Revenue - Budget
= 100_008_000 - 100_000_000
= Rp 8.000

7. Jadi, penerapan promo ini hanya menghasilkan untung sebesar Rp 8.000. Untuk reward yang sangat besar, profit yang dihasilkan sangatlah kecil. Promo ini tidak direkomendasikan karena skema promo ini tidak

efisien secara finansial. Perlu dilakukan observasi alternatif lain misalnya mengurangi jumlah reward secara proporsional.

Conditional Probability [50 points]

7. [12 points] Anda adalah seorang senior underwriter (risk analyst) di perusahaan leasing. Seseorang yang ingin mengajukan pinjaman (misal untuk kredit mobil) harus di review lebih dahulu. Anda memiliki tim yang terdiri atas beberapa underwriter. Setelah 1 tahun bekerja, Anda memahami bahwa
- 35% dari underwriter Anda buruk karena selalu menyetujui (approve) pinjaman tanpa mengevaluasi dengan baik.
 - Sisa underwriter adalah underwriter yang bagus dimana 65% pinjaman disetujui dan 35% ditolak.

Pertanyaannya

- **Berapa peluang pinjaman yang diajukan akan disetujui?**

Catatan:

Tulis jawaban disini

Untuk menghitung peluang pinjaman disetujui kita harus menggunakan teorema probabilitas total.

Kita akan mempertimbangkan 2 jenis underwriter

- Buruk -> menyetujui 100% pinjaman
- Bagus -> menyetujui 65% pinjaman

1. Tentukan probabilitas

$$P(\text{Buruk}) = 35\% = 0.35$$

$$P(\text{Bagus}) = 65\% = 0.65$$

$$P(\text{Disetujui} \mid \text{Buruk}) = 1$$

$$P(\text{Disetujui} \mid \text{Bagus}) = 0.65$$

2. Teorema total probabilitas

$$P(\text{Disetujui}) = P(\text{Disetujui} \mid \text{Buruk}) + P(\text{Buruk}) + P(\text{Disetujui} \mid \text{Bagus}) + P(\text{Bagus})$$

$$= (1 \times 0.35) + (0.65 \times 0.65)$$

$$= 0.7725$$

3. Jadi, probabilitas pinjaman akan disetujui adalah sebesar 0.7725 atau 77.25%

4. Python:

```

[3] # Probabilitas
p_buruk = 0.35
p_bagus = 0.65

# Probabilitas pinjaman disetujui oleh masing-masing tipe underwriter
p_disetujui_buruk = 1.0
p_disetujui_bagus = 0.65

# Total probabilitas pinjaman disetujui
p_disetujui = (p_disetujui_buruk * p_buruk) + (p_disetujui_bagus * p_bagus)

# Output
print(f"Peluang pinjaman disetujui: {p_disetujui:.4f} atau {p_disetujui * 100:.2f}%")

```

→ Peluang pinjaman disetujui: 0.7725 atau 77.25%

8. Tim data pemilu di negara Huba-Huba membuat model machine learning untuk membantu proses rekapitulasi hasil pemilu. Ada 2 model yang dibuat, model 1 untuk memprediksi apakah foto rekapitulasi suara yang diinput adalah foto valid / tidak valid, dan model 2 untuk mengekstrak hasil suara menggunakan OCR.

Pada model 1, didapatkan hasil evaluasi prediksi sebagai berikut

		Hasil Prediksi	
		Valid	Tidak valid
Label aktual	Valid	8500	500
	Tidak valid	990	10

Angka-angka di atas berisi tentang jumlah dari temuan saat melakukan evaluasi model. Misal angka 8.500 artinya ada 8.500 gambar valid yang diprediksi sebagai gambar valid, angka 500 artinya ada 500 gambar valid yang diprediksi sebagai tidak valid (**salah prediksi**).

Pertanyaannya,

- **[8 points]** Berapa peluang model 1 memprediksi gambar dengan akurat?

Catatan:

Tulis jawaban disini

1. Definisikan matrix

Label Valid - Prediksi Valid = True Positive = 8500

Label Tidak Valid - Prediksi Tidak Valid = True Negative = 10

Label Valid - Prediksi Tidak Valid = False Negative = 500

Label Tidak Valid - Prediksi Valid = False Positive = 990

2. Definisikan akurasi dan cara menghitungnya

Akurasi = Jumlah prediksi benar / Total prediksi

Jumlah prediksi benar = True Positive + True Negative

3. Menghitung peluang akurasi model 1

$$\begin{aligned}
 P(\text{Akurasi}) &= (\text{True Positive} + \text{True Negative}) / \text{Total Prediksi} \\
 &= (8500 + 10) / (8500 + 990 + 500 + 10) \\
 &= 8510 / 10000 \\
 &= 0.851
 \end{aligned}$$

4. Jadi, peluang model 1 memprediksi gambar dengan akurat adalah sebesar 0.851 atau 85.1 % (akurasi model cukup tinggi).
5. Python:

```

# Jumlah masing-masing kategori dalam confusion matrix
true_positive = 8500 # Valid diprediksi valid
false_negative = 500 # Valid diprediksi tidak valid
false_positive = 990 # Tidak valid diprediksi valid
true_negative = 10  # Tidak valid diprediksi tidak valid

# Total prediksi benar
prediksi_benar = true_positive + true_negative

# Total semua prediksi
total_prediksi = true_positive + false_negative + false_positive + true_negative

# Hitung akurasi
akurasi = prediksi_benar / total_prediksi

# Tampilkan hasil
print(f"Akurasi model: {akurasi:.3f} atau {akurasi*100:.1f}%")

```

Akurasi model: 0.851 atau 85.1%

- **[8 points] Berapa peluang model 1 dapat menemukan foto yang tidak valid?**

Catatan:

Tulis jawaban disini

1. Definisikan matrix
 - Label Valid - Prediksi Valid = True Positive = 8500
 - Label Tidak Valid - Prediksi Tidak Valid = True Negative = 10
 - Label Valid - Prediksi Tidak Valid = False Negative = 500
 - Label Tidak Valid - Prediksi Valid = False Positive = 990
2. Definisikan ruang sampel Tidak Valid
 - Tidak Valid - Tidak Valid : True Negative
 - Tidak Valid - Valid : False Positive
$$\begin{aligned}
 \text{Ruang sampel Tidak valid} &= n(\text{Tidak Valid}) \\
 &= n(\text{True Negative}) + n(\text{False Positive}) \\
 &= 10 + 990 \\
 &= 1000
 \end{aligned}$$
3. Definisikan jumlah kejadian menemukan foto tidak valid : True Negative

Maka, $n(\text{True Negative}) = 10$
4. Hitung peluang

$$\begin{aligned}
 P(\text{True Negative}) &= n(\text{True Negative}) / n(\text{Tidak Valid}) \\
 &= 10 / 1000 \\
 &= 0.01
 \end{aligned}$$

5. Jadi, peluang model 1 dapat menemukan foto yang tidak valid adalah sebesar 0.01 atau 1%.
6. Python

```
[7] # Data dari confusion matrix
    true_negative = 10
    false_positive = 990

    # Total gambar yang sebenarnya tidak valid
    total_actual_invalid = true_negative + false_positive

    # Hitung peluang menemukan gambar tidak valid (True Negative Rate)
    prob_find_invalid = true_negative / total_actual_invalid
    percentage = prob_find_invalid * 100

    # Tampilkan hasil dengan kalimat
    print(f"Peluang model 1 dapat menemukan foto yang tidak valid adalah sebesar {prob_find_invalid:.2f} atau {percentage:.1f}%.")
```

Peluang model 1 dapat menemukan foto yang tidak valid adalah sebesar 0.01 atau 1.0%.

- [10 points] Ketika model 1 memprediksi foto sebagai tidak valid, **berapa peluang bahwa foto tersebut memang tidak valid?**

Catatan:

Tulis jawaban disini

1. Definisikan matrix
 Label Valid - Prediksi Valid = True Positive = 8500
 Label Tidak Valid - Prediksi Tidak Valid = True Negative = 10
 Label Valid - Prediksi Tidak Valid = False Negative = 500
 Label Tidak Valid - Prediksi Valid = False Positive = 990
2. Definisikan kejadian foto tersebut memang tidak valid ketika prediksi tidak valid : True Negative
 Maka, $n(\text{Label Tidak Valid} \mid \text{Prediksi Tidak Valid}) = n(\text{True Negative})$
3. Definisikan ruang sampel
 $n(\text{Tidak Valid}) = n(\text{True Negative}) + n(\text{False Negative})$
4. Hitung peluang $P(\text{Label Tidak Valid} \mid \text{Prediksi Tidak Valid})$
 $= n(\text{True Negative}) / (n(\text{True Negative}) + n(\text{False Negative}))$
 $= 10 / (10 + 500)$
 $= 10 / 510$
 ≈ 0.0196
5. Maka, peluang bahwa foto tersebut memang tidak valid ketika prediksi foto tidak valid adalah 0.0196 atau 1.96%
6. Python:

```
# Data dari confusion matrix
true_positive = 8500
true_negative = 10
false_negative = 500
false_positive = 990

# total prediksi yang mengatakan foto tidak valid
total_predicted_invalid = true_negative + false_negative

# Hitung peluang bahwa foto memang tidak valid ketika diprediksi tidak valid
prob_actual_invalid_given_predicted_invalid = true_negative / total_predicted_invalid
percentage = prob_actual_invalid_given_predicted_invalid * 100

# Tampilkan hasil dengan kalimat
print(f"Peluang bahwa foto tersebut memang tidak valid ketika prediksi menyatakan tidak valid adalah sebesar {prob_actual_invalid_given_predicted_invalid:.4f} atau {percentage:.2f}%.")
```

Peluang bahwa foto tersebut memang tidak valid ketika prediksi menyatakan tidak valid adalah sebesar 0.0196 atau 1.96%.

- **[12 points]** Apakah Anda akan merekomendasikan menggunakan model ini untuk memvalidasi foto rekapitulasi suara yang diinput?

Catatan:

Tulis jawaban disini

Tidak. Karena walaupun akurasi model tinggi (85.1%) namun hal ini tidak cukup menjadi dasar bahwa model ini baik untuk digunakan. Alasan model tidak cukup baik:

1. Karena dataset menunjukkan hal yang tidakimbang. Akurasi model tinggi karena data yang ada mayoritas adalah data valid dan model cenderung selalu menunjukkan "valid".
2. Kemampuan model dalam mendeteksi foto tidak valid sangat buruk. Peluang model menemukan mayoritas foto yang tidak valid sangat rendah hanya 1%.
3. Prediksi "tidak valid" hampir selalu salah dilihat dari peluang presisi model dalam kelas tidak valid hanya 1.96%.

Jadi, model ini tidak layak digunakan untuk memvalidasi foto.