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

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 🫡

Bayes Theorem, Total Probability & Independence [50 points]

1. [10 points] OKJek adalah perusahaan ride hailing. Melihat SOPnya, customer diharuskan untuk mengisi rating (**puas** atau **tidak puas**) perjalanan sebelum memulai perjalanan berikutnya. Diketahui 80% dari user mengisi dengan baik dan 20%-nya tidak. Dari yang mengisi dengan baik, 60% nya mengisi dengan rating **puas** dan 40% mengisi dengan rating **tidak puas**.

   Misalkan seorang user memberikan rating **puas** 3x berturut-turut, **berapa peluang kalau mereka adalah user yang tidak baik dalam mengisi rating**?
   Jawaban:
    1. Buat Ruang Sampel
       Dalam konteks ini, kita ingin menyusun sample space dari 3 rating berturut-turut, di mana setiap rating bisa berupa:
       P (puas)
       T (tidak puas)
       Maka, sample space :
       S={PPP,PPT,PTP,PTT,TPP,TPT,TTP,TTT}

Namun, kita hanya tertarik pada kasus di mana ratingnya adalah PPP (puas 3x), jadi kita akan fokus pada peluang terjadinya PPP untuk dua tipe user:
User baik (B)
User tidak baik (NB)

2. Probabilitas tiap user
   P(B) = 80% = 0.8
   P(NB) = 20% = 0.2

   Untuk user baik:
   P(P|B) = 0.6
   P(PPP|B) = 0.6^3 = 0.216

   User tidak baik diasumsikan acak karena dianggap tidak konsisten atau tidak punya pola khusus dalam mengisi rating, tidak ada data pasti yang menjelaskan, dan bisa dianggap seperti gambling antara puas atau tidak puas maka dari itu diasumsikan berikut ini:
   P(P|NB) = 0.5
   P(PPP|NB) = 0.5^3 = 0.125

3. Total probability dari PPP
   P(PPP) = P(PPP|B) x P(B) + P(PPP|NB) x P(NB)
   = 0.216 x 0.8 + 0.125 x 0.2
   = 0.1728 + 0.025
   = 0.1978

4. Hitung Peluang User Tidak baik dalam mengisi rating puas 3x berturut dengan teorema bayes
   P(NB|PPP) = (P(PPP|NB) X P(NB)) / P(PPP)
   = (0.125 x 0.2) / 0.1978
   = 0.1264 atau 12.64%

5. Jadi, peluang user adalah user yang tidak baik dalam mengisi rating jika dia memberikan 3 rating puas berturut turut adalah sebesar 12.64%

6. Python

```
[10] # Probabilitas user baik dan tidak baik
     P_B = 0.8
     P_NB = 0.2

     # Probabilitas rating puas untuk masing-masing user
     P_P_given_B = 0.6
     P_P_given_NB = 0.5  # asumsi acak

     # Probabilitas user memberikan rating puas 3 kali berturut-turut
     P_PPP_given_B = P_P_given_B ** 3
     P_PPP_given_NB = P_P_given_NB ** 3

     # Total probabilitas mendapatkan PPP (puas 3x berturut)
     P_PPP = P_PPP_given_B * P_B + P_PPP_given_NB * P_NB

     # Hitung probabilitas user tidak baik jika diketahui PPP terjadi (Teorema Bayes)
     P_NB_given_PPP = (P_PPP_given_NB * P_NB) / P_PPP

     # Tampilkan hasil dalam persen
     print(f"Peluang user tidak baik diberikan rating puas 3x berturut-turut: {P_NB_given_PPP * 100:.2f}%")

Peluang user tidak baik diberikan rating puas 3x berturut-turut: 12.64%
```

2. [20 points] Anda sedang ada di tim B2B. Tim marketing memiliki beberapa data leads bisnis yang pernah berinteraksi dengan penawaran produk sebagai berikut.

Membalas email pertama	Kategori harga produk yang ditawarkan	Pitching dengan direksi	Membeli
Ya	Mahal	Ya	Tidak
Ya	Medium	Tidak	Ya
Ya	Murah	Ya	Ya
Tidak	Mahal	Ya	Ya
Tidak	Medium	Tidak	Tidak
Tidak	Murah	Ya	Ya
Tidak	Murah	Tidak	Ya
Ya	Mahal	Ya	Tidak

Apabila leads bisnis baru memiliki informasi

- Membalas email pertama
- Produk yang ditawarkan masuk ke kategori medium
- Dan tidak sempat pitching dengan direksi

Prediksi apakah leads tersebut akan membeli atau tidak? *Hint: Gunakan Bayesian Theorem*.

Jawab:

1. Hitung probabilitas prior
   Membeli = 5 leads
   P(Beli) = ⅝
   P(Tidak Beli) = ⅜

2. Hitung likelihood

   Kondisi:

   1. **Membalas Email = Ya**
      Leads Beli = Leads no 2,3,4,6,7
      Leads Beli dan Email Ya = Leads no 2,3
      P(Email = Ya | Beli) = ⅖

      Leads Tidak Beli = Leads no 1,5,8
      Leads Tidak belli dan Email Ya = 1,8
      P(Email = Ya | Tidak Beli) = ⅔

   2. **Harga = Medium**
      Leads Beli = Leads no 2,3,4,6,7
      Leads Beli dan Harga Medium = Leads no 2
      P(Harga = Medium | Beli) = ⅕

      Leads Tidak Beli = Leads no 1,5,8
      Leads Tidak Beli dan Harga Medium = Leads no 5
      P(Harga = Medium | Tidak Beli) = ⅓

   3. **Pitching**
      Leads Beli = Leads no 2,3,4,6,7
      Leads Beli dan Tidak Pitching = Leads no 2, 7
      P(Pitching = Tidak | Beli) = ⅖

      Leads Tidak Beli = Leads no 1,5,8
      Leads Tidak Beli dan Tidak Pitching = Leads no 5
      P(Pitching = Tidak| Tidak Beli) = ⅓

3. Hitung Joint Probability
   Untuk P(Beli dan Kondisi)
   P(Beli) x P(Email = Ya|Beli) x P(Harga=Medium | Beli) x P(Pitching = Tidak | Beli)
   = ⅝ x ⅖ x ⅕ x ⅖
   = 20/1000
   = 0.02
   Untuk P(Tidak Beli dan Kondisi)

P(Tidak Beli) x P(Email = Ya | Tidak Beli) x P(Harga = Medium | Tidak Beli) x P(Pitching = Tidak| Tidak Beli)

= $\frac{3}{8} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}$

= 6/216

= 0.0278

Total Probability = 0.02 + 0.0278

= 0.0478

4. Hitung

   P(Beli | Kondisi) = P(Beli dan Kondisi) / Total Probability

   = 0.02 / 0.0478

   = 0.419

   P(Tidak beli | Kondisi) = P(Tidak beli dan Kondisi) / Total Probability

   = 0.0278 / 0.0478

   = 0.581

5. P(Tidak Beli | Kondisi) = 0.581 > P(Beli | Kondisi) = 0.419

6. Maka, Leads baru diprediksi tidak akan membeli
7. Python

```
# Probabilitas prior
p_beli = 5/8
p_tidak_beli = 3/8

# Likelihood
# P(Email=Ya | Beli), P(Harga=Medium | Beli), P(Pitching=Tidak | Beli)
p_email_ya_given_beli = 2/5
p_harga_medium_given_beli = 1/5
p_pitching_tidak_given_beli = 2/5

# P(Email=Ya | Tidak Beli), P(Harga=Medium | Tidak Beli), P(Pitching=Tidak | Tidak Beli)
p_email_ya_given_tidak_beli = 2/3
p_harga_medium_given_tidak_beli = 1/3
p_pitching_tidak_given_tidak_beli = 1/3

# Hitung posterior probability (tanpa normalisasi karena hanya bandingkan relatif)
p_beli_given_kondisi = p_beli * p_email_ya_given_beli * p_harga_medium_given_beli * p_pitching_tidak_given_beli
p_tidak_beli_given_kondisi = p_tidak_beli * p_email_ya_given_tidak_beli * p_harga_medium_given_tidak_beli * p_pitching_tidak_given_tidak_beli
total_prob = p_beli_given_kondisi + p_tidak_beli_given_kondisi

result_beli_kondisi = p_beli_given_kondisi / total_prob
result_tidak_beli_kondisi = p_tidak_beli_given_kondisi / total_prob

# Tampilkan hasil
print("P(Beli | Kondisi):", round(result_beli_kondisi, 3))
print("P(Tidak Beli | Kondisi):", round(result_tidak_beli_kondisi, 3))

# Prediksi
if result_beli_kondisi > result_tidak_beli_kondisi:
    print("Prediksi: Leads akan membeli")
else:
    print("Prediksi: Leads tidak akan membeli")
```

```
P(Beli | Kondisi): 0.419
P(Tidak Beli | Kondisi): 0.581
Prediksi: Leads tidak akan membeli
```

3. [20 points] Anda adalah data scientist di startup teknologi yang baru berdiri 2020. Setelah startup tersebut meluncurkan produknya, hasilnya ternyata tidak sesuai dengan

yang diharapkan. Di kondisi tersebut, Anda diminta oleh CEO untuk menentukan probability startup Anda akan sukses atau tidak. Anda mengumpulkan data beberapa startup yang mirip dengan startup Anda dan waktu hidupnya (tidak bangkrut) dengan rentang studi dari 2000 hingga 2024.

Startup_id	Tahun berdiri	Status	Usia hidup (tahun)
1	2023	Tidak bangkrut	1+ *(simbol + menandakan masih hidup)*
2	2019	Tidak bangkrut	5+
3	2018	Bangkrut	6
4	2018	Bangkrut	6
5	2015	Tidak bangkrut	9+
6	2014	Bangkrut	10
7	2014	Bangkrut	10
8	2014	Tidak bangkrut	10+
9	2012	Bangkrut	12
10	2012	Bangkrut	12
11	2012	Bangkrut	12
12	2012	Bangkrut	12
13	2012	Tidak bangkrut	12+
14	2011	Tidak bangkrut	13+
15	2009	Tidak bangkrut	15+
16	2008	Tidak bangkrut	16+
17	2004	Tidak bangkrut	20+
18	2000	Bangkrut	24
19	2000	Tidak bangkrut	24+

Berapakah peluang startup Anda masih tetap hidup hingga tahun 2030?

Hint: Cari probability tahun ini masih hidup ketika tahun lalu hidup. Cara mencari probability-nya memanfaatkan jumlah startup yang memiliki usia tersebut dan jumlah startup yang bangkrut di usia tersebut. Tag: survival analysis, Kaplan Meier Method.

Jawab:

1. Tahun berdiri start up = 2020
   Target tahun = 2030
   P(Bertahan 10 Tahun)

2. Kaplan-Meier Estimator
   Untuk setiap tahun ke-t, probabilitas survival:

$$S(t) = \prod_{i=1}^{t}\left(1 - \frac{d_i}{n_i}\right)_i$$

   d_i = jumlah start up yang bangkrut di tahun ke-i
   n_i = jumlah start up yang hidup tepat sebelum tahun ke-i

3. Hitung banyak startup yang
   - Masih hidup pada awal tiap tahun
   - Bangkrut pada tahun itu (dilihat dari usia hidup)

   Hitung survival rate = 1 - di/ni dengan rumus kaplan-meier

Tahun ke-	ni	di	Survival rate tahun ke-i
1	19	0	1,00
2	18	1	0,94
3	17	0	1,00
4	17	0	1,00
5	17	0	1,00
6	17	2	0,88
7	15	0	1,00
8	15	0	1,00
9	15	0	1,00
10	15	2	0,87

   P(Bertahan 10 tahun) = 1 x 0,94 x 1 x 1 x 1 x 0,88 x 1 x 1 x 1 x 0,87
   = 0.72

4. Jadi, peluang start up saya tetap hidup hingga tahun 2030 adalah 0.72 atau 72%

5. Python

```python
import numpy as np

n_i = [19, 18, 17, 17, 17, 17, 15, 15, 15, 15]
d_i = [0, 1, 0, 0, 0, 2, 0, 0, 0, 2]

survival_rates = [1 - (d / n) for d, n in zip(d_i, n_i)]

print("Survival rate tahun ke-i:")
for i, rate in enumerate(survival_rates, start=1):
    print(f"Tahun ke-{i}: {rate:.2f}")

S_t = np.cumprod(survival_rates)

print(f"\nPeluang startup tetap hidup hingga tahun ke-10: {S_t[-1]:.2f}")
```

```
Survival rate tahun ke-i:
Tahun ke-1: 1.00
Tahun ke-2: 0.94
Tahun ke-3: 1.00
Tahun ke-4: 1.00
Tahun ke-5: 1.00
Tahun ke-6: 0.88
Tahun ke-7: 1.00
Tahun ke-8: 1.00
Tahun ke-9: 1.00
Tahun ke-10: 0.87

Peluang startup tetap hidup hingga tahun ke-10: 0.72
```

Random Variable dan Expected Value [50 points]

4. [8 points] Kamu adalah marketing analyst di suatu industri fintech. Fokus pekerjaanmu adalah menjaga agar customer yang mencicil dengan baik tetap menggunakan fasilitas cicilannya. Apabila diketahui ada 2 orang customer sebagai berikut:

   Customer A
   - Peluang churn dari produk cicilan = 0.82
   - Peluang gagal bayar cicilan = 0.21
   - Rata-rata revenue yang didapat kantor = Rp 150.000.000

   Customer B
   - Peluang churn dari produk cicilan = 0.42
   - Peluang gagal bayar cicilan = 0.71
   - Rata-rata revenue yang didapat kantor = Rp 150.000.000

   Tentukan
   - **Berapa ekspektasi revenue yang didapatkan oleh masing-masing customer**?

   Jawab:
   Customer A

P(Tidak churn) = 1- P(Churn) = 1 - 0.82 = 0.18
P(Tidak gagal bayar) = 1- P(Gagal bayar) = 1 - 0.21 = 0.79
Avg Revenue = Rp 150.000.000

Ekspektasi Revenue Cust A

E_A = P(Tidak churn) x P(Tidak gagal bayar) x Avg Revenue
= 0.18 x 0.79 x 150.000.000
= Rp 21.330.000

Customer B

P(Tidak churn) = 1- P(Churn) = 1 - 0.42 = 0.58
P(Tidak gagal bayar) = 1- P(Gagal bayar) = 1 - 0.71 = 0.29
Avg Revenue = Rp 150.000.000

Ekspektasi Revenue Customer B

E_A = P(Tidak churn) x P(Tidak gagal bayar) x Avg Revenue
= 0.58 x 0.29 x 150.000.000
= Rp 25.230.000

- **Customer mana yang sebaiknya diprioritaskan untuk dijaga**?

Customer A memang lebih rendah potensi gagal bayarnya dibanding customer B, namun tingkat churn customer A (tingkat hilangnya pelanggan A) sangat tinggi dibanding potensi churn customer B. Setelah dihitung, ekspektasi revenue customer B lebih tinggi dibanding customer A. Hal ini menunjukkan customer B lebih profitable untuk diprioritaskan oleh perusahaan. **Maka, customer B lah yang sebaiknya dijaga oleh perusahaan agar tetap mencicil.**

```python
# Data Customer A
churn_A = 0.82
fail_A = 0.21
revenue_A = 150_000_000

# Data Customer B
churn_B = 0.42
fail_B = 0.71
revenue_B = 150_000_000

# Hitung ekspektasi revenue Customer A
not_churn_A = 1 - churn_A
not_fail_A = 1 - fail_A
expected_A = not_churn_A * not_fail_A * revenue_A

# Hitung ekspektasi revenue Customer B
not_churn_B = 1 - churn_B
not_fail_B = 1 - fail_B
expected_B = not_churn_B * not_fail_B * revenue_B

# Print hasil
print(f"Ekspektasi revenue Customer A: Rp {expected_A:,.0f}")
print(f"Ekspektasi revenue Customer B: Rp {expected_B:,.0f}")

# Keputusan prioritas
if expected_A > expected_B:
    print("Customer A lebih layak diprioritaskan.")
else:
    print("Customer B lebih layak diprioritaskan.")
```

```
Ekspektasi revenue Customer A: Rp 21,330,000
Ekspektasi revenue Customer B: Rp 25,230,000
Customer B lebih layak diprioritaskan.
```

5. [8 points] Anda sedang **war** tiket konser Taylor Swift. Penyelenggara sudah memberitahu bahwa jumlah seluruh tiket yang ada adalah 1000 tiket. Penyelenggara membuat aturan akan mengeluarkan 10 tiket per menit selama war berlangsung. **Berapa rata-rata waktu tunggu (menit) untuk mendapatkan tiket konser tersebut**?

   Jawab:

   1. Hitung total waktu war
      Total waktu war = jumlah total tiket / jumlah tiket per menit
      = 1000 / 10
      = 100 menit
   2. Rata-rata waktu tunggu
      Asumsikan semua orang yang mendapatkan tiket berada dalam antrian dan memiliki peluang yang sama. Maka, waktu tunggu setiap orang tersebar merata dari menit ke-1 hingga menit ke-100. Waktu tunggu diasumsikan beridstribusi uniform T~U(a,b). Dengan:

a = 1 (menit pertama tiket tersedia)
b = 100 (menit terakhir tiket tersedia)
Rumus Distribusi Uniform untuk ekspektasi rata-rata waktu tunggu:

$$\mathbb{E}[T] = \frac{a+b}{2}$$

Sehingga,
Rata-rata waktu tunggu dari distribusi uniform antara 1 hingga 100 menit:
Rata-rata waktu tunggu = (1+100) / 2 = 50,5 menit

```
[4] # Parameter distribusi uniform kontinu
    a = 1     # waktu paling awal (menit)
    b = 100   # waktu paling akhir (menit)

    # Menghitung ekspektasi (rata-rata) dari distribusi uniform kontinu
    expected_wait_time = (a + b) / 2

    print(f"Rata-rata waktu tunggu: {expected_wait_time} menit")

    Rata-rata waktu tunggu: 50.5 menit
```

6. [8 points] Anda bekerja di **KebahagianTiadaBerakhir**, suatu sosial media yang sharing pengalaman-pengalaman usernya selama menikmati kuliner di kota Bandung. Untuk mendapatkan revenue, perusahaan mengeluarkan iklan di laman utamanya. Saat ini ada dua opsi penayangan iklan.
   Opsi 1: Setiap 25 post user di laman utama, akan ada 1 iklan ditampilkan.
   Opsi 2: Setiap post di laman utama memiliki peluang 4% menjadi suatu iklan.

   Jika Anda menggunakan opsi 2, berapa peluang user mendapatkan 1 iklan dalam 100 post di laman utama?

   Jawab:
   1. Diketahui :
      - Jumlah percobaan (n) tetap yaitu 100 post
      - Ada 2 kemungkinan hasil (iklan atau bukan iklan)
      - Probabilitas sukses (p) tetap di setiap percobaan (n) yaitu sebesar 4% atau 0.04
      - Percobaan independen karena post satu tidak memengaruhi post lainnya

      Dengan catatan di atas, peluang terjadinya sejumlah keberhasilan munculnya iklan dapat dihitung dengan distribusi binomial yang mensyaratkan jumlah percobaan tetap, ada 2 kemungkinan hasil, probabilitas sukses konstan, dan percobaan diskrit dan independen.

   2. Rumus distribusi binomial

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

n = jumlah post = 100
k = jumlah iklan expected = 1
p = peluang satu post jadi iklan = 0.04

P(X=1) = 100 x 0.04 x (0.96)^99
= 0.07029 atau 7.03%

3. Jadi, jika menggunakan opsi 2 maka peluang user hanya melihat 1 iklan dalam 100 post adalah 7.03%

```
from scipy.stats import binom

# Parameter
n = 100      # total post
p = 0.04     # peluang 1 post jadi iklan
k = 1        # ingin tahu peluang tepat 1 iklan

# Hitung probabilitas menggunakan PMF (probability mass function)
prob_1_iklan = binom.pmf(k, n, p)

print(f"Peluang tepat 1 iklan dalam 100 post adalah {prob_1_iklan:.4f} atau {prob_1_iklan*100:.2f}%")
```

```
Peluang tepat 1 iklan dalam 100 post adalah 0.0703 atau 7.03%
```

7. [8 points] Perusahaanmu bergerak dibidang telekomunikasi yang menyediakan jasa internet. Ada dua buah paket internet **pasca bayar** berbeda, A dan B. Sehari-hari, kamu melakukan analisa terhadap user yang melakukan deaktivasi service internet dari masing-masing paket tersebut. Setiap selesai melakukan pembayaran paket, customer akan membuat keputusan akan melanjutkan paket atau tidak. Apabila peluang deaktivasi paket A adalah 0.1 dan peluang deaktivasi paket B adalah 0.2, **dalam 30 transaksi kedepan, apakah peluang deaktivasi customer setelah *t*-transaksi dari paket A selalu lebih kecil dibanding peluang deaktivasi customer setelah *t*-transaksi dari paket B?**

Jawab:
Setiap transaksi adalah percobaan bernoulli (pelanggan bisa melanjutkan atau deaktivasi) namun karena transaksi lebih dari 1, tidak hanya fokus pada satu waktu namun waktu (jumlah transaksi) sampai kegagalan pertama maka digunakan distribusi geometri.

$$P(X = k) = (1 - p)^{k-1} \cdot p$$

1. Misal:

Peluang deaktivasi per transaksi

- Paket A = 0.1
- Paket B = 0.2

Peluang tidak deaktivasi per transaksi

- Paket A = 1 - 0.1 = 0.9
- Paket B = 1 - 0.2 = 0.8

Peluang tetap aktif selamat t transaksi

- Paket A = PA(t) = 0.9^t
- Paket B= PB(t) = 0.8^t

2. Maka peluang terjadi deaktivasi setidaknya sekali hingga transaksi ke-t

$$P(X \leq t) = 1 - (1 - p)^t$$

- Paket A
  PdeactA(t) = 1 - 0.9^t
- Paket B
  PdeactB(t) = 1 - 0.8^t

3. Kesimpulan
   Karena 0.9^t > 0.8^t untuk semua t>=1
   maka 1 - 0.9^t < 1 - 0.8^t
   Selalu PdeactA(t) < PdeactB(t)
   Atau untuk setiap transaksi t dari 1-30, peluang deaktivasi customer dari paket A selalu lebih kecil dibanding paket B

   Paket A memiliki **peluang deaktivasi yang lebih kecil per transaksi** dibanding Paket B (10% vs 20%). Artinya, pada setiap siklus pembayaran, pelanggan paket A lebih mungkin **melanjutkan langganan**.

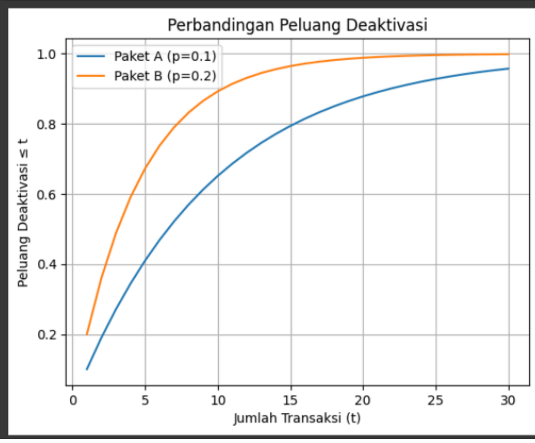
   Karena peluang ini berlaku **berulang tiap transaksi**, maka secara kumulatif, pelanggan Paket A akan **lebih tahan lama** dan **lebih jarang deaktivasi** dibanding Paket B seiring waktu.

4. Python

```
[9] import matplotlib.pyplot as plt
    import numpy as np

    t = np.arange(1, 31)
    deact_A = 1 - 0.9**t
    deact_B = 1 - 0.8**t

    plt.plot(t, deact_A, label='Paket A (p=0.1)')
    plt.plot(t, deact_B, label='Paket B (p=0.2)')
    plt.xlabel("Jumlah Transaksi (t)")
    plt.ylabel("Peluang Deaktivasi ≤ t")
    plt.title("Perbandingan Peluang Deaktivasi")
    plt.legend()
    plt.grid(True)
    plt.show()
```



8. [18 points] Diketahui X ~ Uniform(a, b) (Diskrit) . **Buktikan bahwa**

$$\text{var}[X] = \frac{(b-a)(b-a+2)}{12}$$

Jawab:
X ~ Uniform(a, b) (Diskrit)
X memiliki nilai–nilaii integer dari a sampai b
Banyak nilai n = b - a + 1
Setiap nilai punya peluang yang sama:

$$P(X = x) = \frac{1}{b-a+1}, \quad x = a, a+1, ..., b$$

Bukti $Var[X] = \frac{(b-a)(b-a+2)}{12}$

Diketahui :

Expected value uniform $\rightarrow E[X] = \frac{a+b}{2}$

untuk buktikan, turunkan :

$$Var(X) = E(X^2] - (E[X])^2$$

$$E[X^2] = \sum_{x=a}^{b} x^2 \cdot P(X=x) = \frac{1}{n}\sum_{x=a}^{b} x^2$$

$x = a + k$, $k = 0, 1, 2, \dots, n-1$

$x^2 = (a+k)^2$

$$\sum_{k=0}^{n-1} k = \frac{(n-1)n}{2}$$

$$\sum_{k=0}^{n-1} k^2 = \frac{(n-1)n(2n-1)}{6}$$

$$\sum_{x=a}^{b} x^2 = na^2 + 2a \cdot \frac{(n-1)n}{2} + \frac{(n-1)n(2n-1)}{6}$$

$$= na^2 + a(n-1)n + \frac{(n-1)n(2n-1)}{6}$$

substitusi ke $E[X^2]$

$$E[X^2] = \frac{1}{n}\left[na^2 + a(n-1)n + \frac{(n-1)n(2n-1)}{6}\right]$$

$$= a^2 + a(n-1) + \frac{(n-1)(2n-1)}{6}$$

maka :

$$Var(X) = E(X^2] - (E[X])^2$$

$$\Rightarrow a^2 + a(n-1) + \frac{(n-1)(2n-1)}{6} - \left(\frac{a+b}{2}\right)^2$$

substitusikan $n = b - a + 1$

$$\Rightarrow \left(a^2 + a(n-1) + \frac{(n-1)(2n-1)}{6}\right) - \left(\frac{2a+n-1}{2}\right)^2$$

$$\Rightarrow \left(a^2 + a(n-1) + \frac{(n-1)(2n-1)}{6}\right) - \frac{(2a+n-1)^2}{4}$$

$$\Rightarrow \frac{12a^2}{12} + \frac{12a(n-1)}{12} + \frac{2(n-1)(2n-1)}{12} - \frac{3(4a^2+4a(n-1)+(n-1)^2)}{12}$$

$$\Rightarrow \frac{n^2-1}{12}$$

$$= \frac{(b-a+1)^2 - 1}{12}$$

$$= \frac{(b-a)(b-a+2)}{12}$$ Terbukti ✓✓