Modul Analisis Biplot

Dikutip dari :

- *Modul Dosen Statistika ITS (Prof. Dr. Bambang Widjanarko Otok, M.Si)*
- *Buku Sidik Peubah Ganda dengan Menggunakan SAS (Ahmad Ansori Mattjik & I Made Sumertajaya) – Statistika IPB*
- *Analisis Biplot Pengguna Beberapa Jenis Kartu Prabayar*

Analisis Biplot diperkenalkan oleh Gabriel tahun 1971. Analisis ini bertujuan memperagakan suatu matriks dengan menumpang tindihkan vektor-vektor yang merepresentasikan vektorvektor baris dengan vektor-vektor yang merepresentasikan vektorvektor kolom matriks tersebut. Biplot merupakan penggambaran grafis sembarang matriks berpangkat dua atau lebih dengan pendekatan matriks berpangkat dua.

Dengan metode Biplot dapat diperoleh informasi tentang posisi relatif dari variabel asal. Gabungan antara plot variabel asal dengan plot pengamatan melalui *superimpose* akan memberi informasi tentang hubungan antara variabel dengan pengamatan.

Dari matriks data:

$$
{}_n\mathbf{X}_p = \begin{pmatrix} x_{11} & \cdots & x_{1i} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{ki} & \cdots & x_{kp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{ni} & \cdots & x_{np} \end{pmatrix}
$$

akan dibangkitkan matriks $\mathbf{G}$ dan $\mathbf{H}$ sebagai berikut:

$$
\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} \\ \vdots & \vdots \\ g_{k1} & g_{k2} \\ \vdots & \vdots \\ g_{n1} & g_{n2} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_1^T \\ \vdots \\ \mathbf{g}_k^T \\ \vdots \\ \mathbf{g}_n^T \end{pmatrix}
$$

$$
H = \begin{pmatrix} h_{11} & h_{12} \\ \vdots & \vdots \\ h_{i1} & h_{i2} \\ \vdots & \vdots \\ h_{p1} & h_{p2} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_i^T \\ \vdots \\ \mathbf{h}_p^T \end{pmatrix}
$$

dimana diinginkan:

$$
\mathbf{g}_k^T = \begin{pmatrix} g_{k1} & g_{k2} \end{pmatrix} \quad \text{representasi dari} \quad \mathbf{x}_k^T = \begin{pmatrix} x_{k1} & \cdots & x_{ki} & \cdots & x_{kp} \end{pmatrix}
$$

$$
\mathbf{h}_i^T = \begin{pmatrix} h_{i1} & h_{i2} \end{pmatrix} \quad \text{representasi dari} \quad \mathbf{x}_i^T = \begin{pmatrix} x_{1i} & \cdots & x_{ki} & \cdots & x_{ni} \end{pmatrix}
$$

Misalkan matrik ${}_n\mathbf{Y}_p$ merupakan matriks data dan ${}_n\mathbf{X}_p$ merupakan matriks data yang telah terkoreksi terhadap nilai tengahnya, yaitu $\mathbf{X} = \mathbf{Y} - (\mathbf{JY})/n$, dimana $\mathbf{J}$ merupakan matriks berunsur bilangan satu dan berukuran nxn. Dengan dekomposisi nilai singular diperoleh :

$$ {}_n\mathbf{X}_p = {}_n\mathbf{U}_r\ \mathbf{L}_r\ \mathbf{A}_p \qquad (1) $$

dimana :

$\mathbf{U}$ dan $\mathbf{A}$ adalah matriks dengan kolom orthonormal ($\mathbf{U}^T\mathbf{U} = \mathbf{A}^T\mathbf{A} = {}_r\mathbf{I}_r$)

$\mathbf{L}$ merupakan matriks diagonal dengan elemen diagonal berupa eigen value

Persamaan di atas dapat pula ditulis sebagai:

$\mathbf{X} = \mathbf{U}\ \mathbf{L}^{\alpha}\ \mathbf{L}^{1-\alpha}\mathbf{A}'$ (2)

$\mathbf{X} = {}_n\mathbf{G}_r\ {}_r\mathbf{H}_p$

Dengan mendefinisikan $\mathbf{G} = \mathbf{U}\ \mathbf{L}^{\alpha}$ dan $\mathbf{H} = \mathbf{L}^{1\alpha}\mathbf{A}'$

Kasus 1: α = 0, maka G = U, dan H = AL

Fakta yang dapat diperoleh dari kasus ini adalah :

1. $h_i'h_j = (n1)\ s_{ij.}$
   dimana $s_{ij} = (x_{ik} - x_i)\ (x_{jk} - x_j)/\ (n\text{-}1)$
   Artinya perkalian titik antara vektor h_i dan h_j akan memberikan gambaran kovarian antara variabel ke-i dan ke-j.
2. $\|h_i\| = (n\text{-}1)\ s_i$
   Artinya panjang vektor tersebut akan memberikan gambaran keragaman variabel ke-i. Makin panjang vektor h_i makin besar pula keragaman variabel ke-i.
3. cos Θ = r_{ij} dimana Θ adalah sudut antara vektor h_i dengan vektor h_j. Artinya cos sudut antara vektor h_i dengan vektor h_j merupakan korelasi antara variabel kei dengan variabel kej. Bila sudut antara kedua vektor tersebut mendekati nol maka makin besar korelasi positif antara kedua variabel tersebut. Bila sudut tersebut mendekati Π, maka makin besar korelasi negatif antara kedua variabel tersebut. Korelasi sama dengan satu, jika $\Theta = 0$. Jika Θ mendekati Π/2 maka makin kecil korelasi antara kedua variabel dan korelasi sama dengan nol jika Θ = Π/2.
4. Bila rank x = p , untuk p < n maka
   $\delta^2(x_i,x_j) = d^2(x_i,x_j)$ dimana :
   $\delta^2(x_i,x_j) = (x_i,x_j)^T S\ (x_i,x_j)$ = jarak Mahalonobis
   $d^2(x_i,x_j) = (x_i,x_j)^T\ (x_i,x_j)$ = jarak euclidean

Kasus 2 : α = 1, maka G = U L , dan H = A.

Fakta yang dapat diperoleh dari kasus ini adalah:

1. Koordinat h_j' merupakan koefisien variabel ke-j dalam dua komponen utama pertama.
2. $d^2(x_i,x_j) = d^2(g_i,g_j)$, artinya jarak Euclidean antara x_i dan x_j akan sama dengan jarak Euclidean antara g_i dan g_j.
3. Posisi g_i dalam plot akan sama dengan posisi obyek kei dengan menggunakan dua skor dari dua komponen utama pertama.

Pendekatan langsung untuk mendapatkan biplot dimulai dari SVD, dimana sebelumnya kita membuat matrik **Y** yang merupakan matrik X berukuran n x p yang sudah dikoreksi dengan mean,

$$ Y_{nxp} = U_{nxp}\Lambda_{pxp}V'_{pxp} \qquad (3) $$

dimana Λ = diag (λ_1, λ_2, ..., λ_p) dan **V** merupakan matrik orthogonal yang kolomnya adalah eigenvektor dari **Y'Y** yang ekivalen dengan (n-1)S, sehingga

$$V = \hat{E} = [\hat{e}_1, \hat{e}_2, ..., \hat{e}_p] \quad (4)$$

dengan mengalikan persamaan (4) dengan $\hat{E}$, kita mendapatkan

$$Y\hat{E} = U\Lambda \quad (5)$$

membuat baris ke-j sisi kiri persamaan (5) menjadi

$$[(x_j - \bar{x})'e_1, (x_j - \bar{x})'e_2, ..., (x_j - \bar{x})e_p] = [\hat{y}_{j1}, \hat{y}_{j2}, .. \hat{y}_{jp}] \quad (6)$$

yang merupakan nilai komponen utama ke-j. Dari sini bisa diketahui bahwa **UΛ** terdiri dari nilai-nilai komponen utama sedangkan **V** mengandung koefisien-koefisien yang membentuk komponen utama.

Taksiran terbaik rank 2 untuk matrik **Y** diperoleh dengan mengganti Λ menjadi Λ^* = diag (λ_1, λ_2, 0, ...,0) menggunakan teorema Eckart-Young. Sehingga matrik **Y** menjadi,

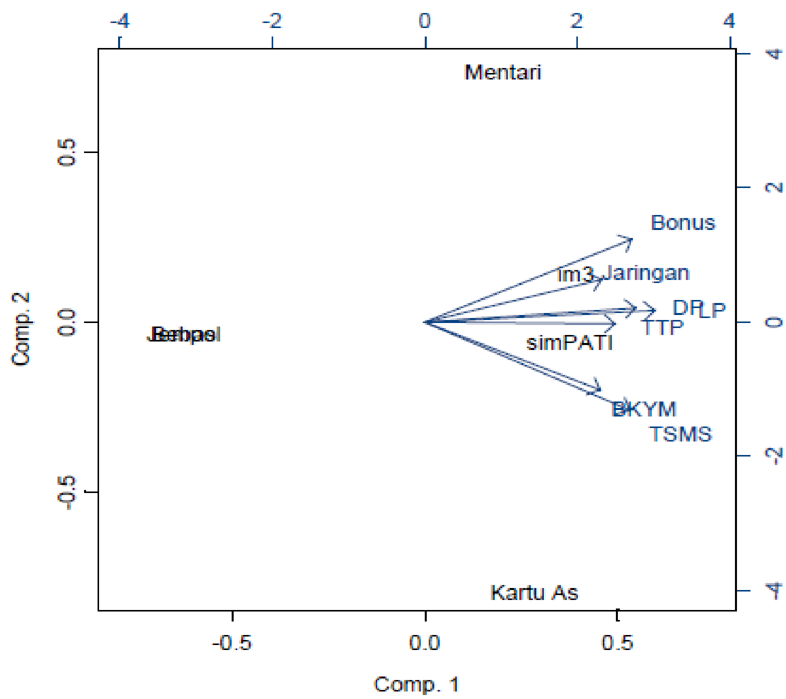
$$Y = U\Lambda^* V' = [\hat{y}_1, \hat{y}_2] \begin{bmatrix} \hat{e}_1' \\ \hat{e}_2' \end{bmatrix} \quad (7)$$

dimana $\hat{y}_1$ merupakan vektor berukuran n x 1 dari komponen utama pertama dan $\hat{y}_2$ merupakan vektor berukuran n x1 dari komponen utama kedua.

Pada Biplot masing-masing baris dari matrik data atau item ditunjukkan oleh titik dalam pasangan nilai komponen utama. Sedangkan kolom ke-i dari matrik data atau variabel ditunjukkan tanda panah.

Interpretasi analisis biplot :

1. Panjang vektor peubah sebanding dengan keragaman peubah tersebut. Semakin panjang vektor suatu peubah maka keragaman peubah tersebut semakin tinggi.
2. Nilai cosinus sudut antara dua vektor peubah menggambarkan korelasi kedua peubah. Semakin sempit sudut yang dibuat antara dua peubah maka semakin positif tinggi korelasinya. Jika sudut yang dibuat tegak lurus maka korelasi keduanya rendah. Sedangkan jika sudutnya tumpul (berlawanan arah) maka korelasinya negatif.
3. Posisi objek yang searah dengan suatu vektor peubah diinterpretasikan sebagai besarnya nilai peubah untuk objek yang searah dengannya. Semakin dekat letak objek dengan arah yang ditunjuk oleh suatu peubah maka semakin tinggi peubah tersebut untuk objek itu. Sedangkan jika arahnya berlawanan, maka nilainya rendah.
4. Kedekatan letak/posisi dua buah objek diinterpretasikan sebagai kemiripan sifat dua objek. Semakin dekat letak dua buah objek maka sifat yang ditunjukkan oleh nilai-nilai peubahnya semakin mirip.



Gambar 2. Biplot Pengguna Kartu Parbayar

Bila daerah dalam Biplot tersebut dibagi menjadi empat kuadran, maka Mentari dan IM3 menempati kuadran pertama, Jempol dan Bebas menempati kuadran ketiga dan Kartu As dan simPATI menempati kuadran keempat. Pengelompokan ini menunjukkan bahwa setiap kelompok jenis kartu prabayar yang berada dalam satu kuadran memiliki persamaan yang cukup dekat dibandingkan dengan kartu prabayar yang berada pada kuadran lain.

Dari Biplot ini terlihat bahwa variabel jaringan yang jarang mengalami gangguan **(Jaringan),** variabel tarif telpon yang murah **(TTP**) dan variabel banyak kenalan yang menggunakan kartu prabayar yang digunakan responden **(BKYM**) mempunyai nilai keragaman yang paling kecil dari pada nilai keragaman dari variabel lain. Hal ini dikarenakan ketiga variabel tersebut mempunyai vektor variabel yang paling pendek. Ini berarti persentase penggunaan kartu prabayar karena jaringan yang jarang mengalami gangguan **(Jaringan**), tarif telpon yang murah **(TTP)** dan banyak kenalan yang menggunakan kartu prabayar yang digunakan responden **(BKYM)** untuk masing -masing kartu prabayar hampir sama besar.

Variabel-variabel yang diteliti dalam penelitian ini mempunyai korelasi yang positif. Korelasi ini positif karena sudut yang dibentuk oleh dua garis berarah dari variabel mempunyai sudut yang sempit (lancip). Ini berarti, semakin baik pelayanan yang diberikan oleh operator kartu prabayar maka akan semakin meningkat penggunaan suatu jenis kartu prabayar. Salah satu contoh adalah variabel Bonus dan Tarif SMS yang murah. Garis berarah dari kedua variabel tersebut membentuk sudut yang kecil (lancip). Oleh karena itu, korelasi antar variabel Bonus dan variabel Tarif SMS yang murah bernilai positif. Ini berarti, semakin banyak Bonus yang ditawarkan oleh operator kartu pra bayar akan mempunyai kecenderungan bahwa operator kartu prabayar tersebut akan menawarkan Tarif SMS yang murah **(TSMS).**

Penghitungan Manual Analisis Biplot

Tahapan analisis biplot adalah

1. Siapkan data dalam bentuk matriks X*
2. Matriks X adalah X* yang sudah dikoreksi terhadap rataannya
3. Hitung X'X
4. Cari nilai eigen dari X'X(λ_i) →
5. Lakukan SVD untuk matriks X'X sehingga diperoleh matriks U, L, A
6. Hitung nilai G = U L^{α} serta H' = $L^{1-\alpha}$ A, nilai α terletak diantara 0 dan 1. kalau ingin menggambarkan objeknya saja, alfa = 0, sementara jika ingin menggambarkan peubahnya saja, alfa = 1, tetapi jika keduanya, alfa = 0.5
7. Plot G dan H
8. G = Skor Objek, H = Skor Peubah
9. Buat interpretasinya

Lakukan analisis biplot dari data berikut, gunakan α = 0,5

Observasi	X_1 (Nilai Fisika)	X_2 (Nilai Matematika)
1	70	65
2	75	70
3	80	75
4	85	80
5	90	85
6	75	80
7	80	90
8	70	100
9	65	85
10	60	80

1. Data dalam bentuk matriks X*

$$\mathbf{X^*} = \begin{bmatrix} 70 & 65 \\ 75 & 70 \\ 80 & 75 \\ 85 & 80 \\ 90 & 85 \\ 75 & 80 \\ 80 & 90 \\ 70 & 100 \\ 65 & 85 \\ 60 & 80 \end{bmatrix} \qquad \bar{\mathbf{X}} = \begin{bmatrix} 75 \\ 81 \end{bmatrix}$$

2. Matriks X adalah X* yang sudah dikoreksi terhadap rataannya

$$\mathbf{X}=\begin{bmatrix}-5 & -16\\ 0 & -11\\ 5 & -6\\ 10 & -1\\ 15 & 4\\ 0 & -1\\ 5 & 9\\ -5 & 19\\ -10 & 4\\ -15 & -1\end{bmatrix}\quad \mathbf{X^T}=\begin{bmatrix}-5 & 0 & 5 & 10 & 15 & 0 & 5 & -5 & -10 & -15\\ -16 & -11 & -6 & -1 & 4 & -1 & 9 & 19 & 4 & -1\end{bmatrix}$$

3. Hitung X'X

$$\mathbf{X'X}=\begin{bmatrix}750 & 25\\ 25 & 890\end{bmatrix}$$

4. Cari nilai eigen dari X'X(λ_i) (**lihat kembali modul analisis faktor untuk penghitungan eigen manual)**

$$\boldsymbol{\lambda}=\begin{bmatrix}894.3303\\ 745.6697\end{bmatrix}\quad \mathbf{e_1}=\begin{bmatrix}0.1706723\\ 0.9853278\end{bmatrix}\ \mathbf{e_2}=\begin{bmatrix}-0.9853278\\ 0.1706723\end{bmatrix}$$

5. Lakukan SVD untuk matriks X'X sehingga diperoleh matriks U, L, A

$$\mathbf{A}=\begin{bmatrix}0.1706723 & -0.9853278\\ 0.9853278 & 0.1706723\end{bmatrix}$$

$$\mathbf{L}=\begin{bmatrix}\sqrt{894.3303} & 0\\ 0 & \sqrt{745.6697}\end{bmatrix}=\begin{bmatrix}29.905356 & 0\\ 0 & 27.3069533\end{bmatrix}$$

$$\mathbf{U_1}=\frac{1}{29.9053}\begin{bmatrix}-5 & -16\\ 0 & -11\\ 5 & -6\\ 10 & -1\\ 15 & 4\\ 0 & -1\\ 5 & 9\\ -5 & 19\\ -10 & 4\\ -15 & -1\end{bmatrix}\begin{bmatrix}0.170672\\ 0.985328\end{bmatrix}=\begin{bmatrix}-16.6186 & -0.55571\\ -10.8386 & -0.36243\\ -5.05861 & -0.16915\\ 0.721395 & 0.024123\\ 6.501396 & 0.217399\\ -0.98533 & -0.03295\\ 9.721312 & 0.325069\\ 17.86787 & 0.59748\\ 2.234588 & 0.074722\\ -3.54541 & -0.11855\end{bmatrix}$$

$$\mathbf{U_2} = \frac{1}{27.3069}\begin{bmatrix} -5 & -16 \\ 0 & -11 \\ 5 & -6 \\ 10 & -1 \\ 15 & 4 \\ 0 & -1 \\ 5 & 9 \\ -5 & 19 \\ -10 & 4 \\ -15 & -1 \end{bmatrix}\begin{bmatrix} -0.98533 \\ 0.170672 \end{bmatrix} = \begin{bmatrix} 2.195882 & 0.080415 \\ -1.8774 & -0.06875 \\ -5.95067 & -0.21792 \\ -10.024 & -0.36708 \\ -14.0972 & -0.51625 \\ -0.17067 & -0.00625 \\ -3.39059 & -0.12417 \\ 8.169413 & 0.29917 \\ 10.53597 & 0.385835 \\ 14.60924 & 0.535001 \end{bmatrix}$$

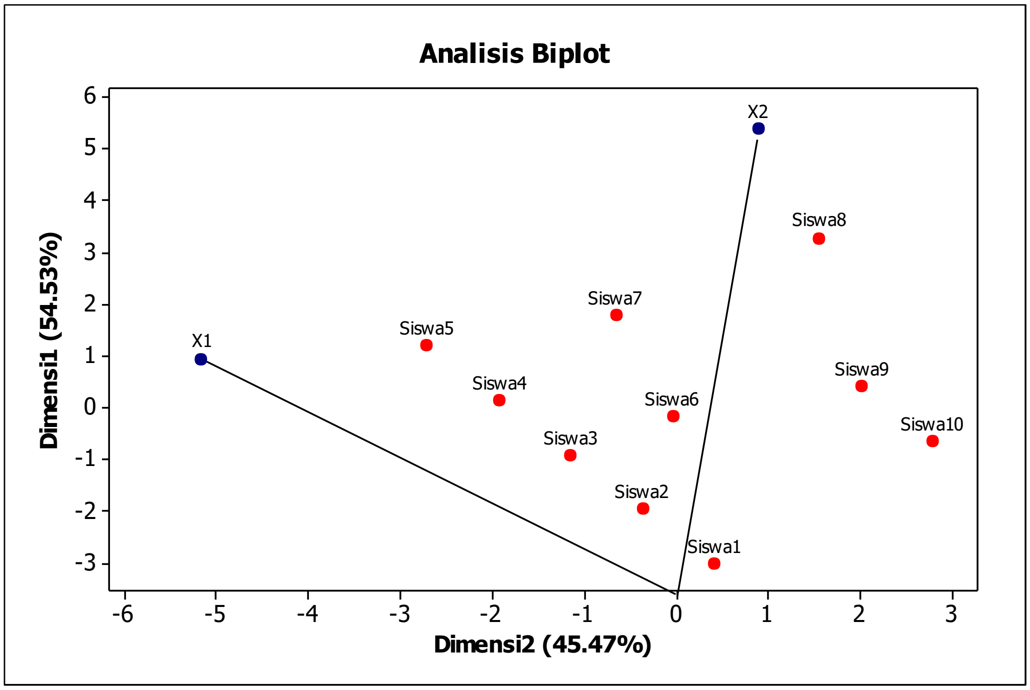
6. Hitung nilai G = U L^{α} serta H' = $L^{1-\alpha}$ A

$$\mathbf{G} = \begin{bmatrix} -0.55571 & 0.080415 \\ -0.36243 & -0.06875 \\ -0.16915 & -0.21792 \\ 0.024123 & -0.36708 \\ 0.217399 & -0.51625 \\ -0.03295 & -0.00625 \\ 0.325069 & -0.12417 \\ 0.59748 & 0.29917 \\ 0.074722 & 0.385835 \\ -0.11855 & 0.535001 \end{bmatrix}\begin{bmatrix} 5.468579 & 0 \\ 0 & 5.225606 \end{bmatrix} = \begin{bmatrix} -3.03893 & 0.420216 \\ -1.98198 & -0.35927 \\ -0.92503 & -1.13875 \\ 0.131916 & -1.91824 \\ 1.188864 & -2.69772 \\ -0.18018 & -0.03266 \\ 1.777667 & -0.64884 \\ 3.267369 & 1.563343 \\ 0.408623 & 2.016219 \\ -0.64832 & 2.795704 \end{bmatrix}$$

$$\mathbf{H^T} = \begin{bmatrix} 5.468579 & 0 \\ 0 & 5.225606 \end{bmatrix}\begin{bmatrix} 0.170672 & 0.985328 \\ -0.98533 & 0.170672 \end{bmatrix} = \begin{bmatrix} 0.933335 & 5.388343 \\ -5.14893 & 0.891866 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 0.933335 & -5.14893 \\ 5.388343 & 0.891866 \end{bmatrix}$$

7. Plot G dan H



8. Interpretasi Analisis Biplot

Dimensi 1 (54.53%)

Dimensi 2 (45.47%)

Kedua dimensi tersebut menjelaskan **100% variasi data (54.53% + 45.47%)**, artinya seluruh informasi data sudah direpresentasikan cukup baik di bidang dua dimensi ini.

- **Vektor X2 (Matematika)** lebih panjang dan mengarah ke atas.
  N*ilai Matematika* memiliki kontribusi besar dalam membentuk struktur data siswa. Semakin ke arah vektor X2, semakin tinggi nilai Matematika siswa tersebut.
- **Vektor X1 (Fisika)** mengarah ke kiri bawah.
  Nilai Fisika tinggi berada di arah kiri bawah grafik. Semakin menjauh ke arah vektor X1, semakin tinggi nilai Fisika siswa.
- **Sudut antara X1 dan X2 besar (mendekati 90°)**
  Nilai Fisika dan Matematika tidak berkorelasi kuat. Dengan kata lain, siswa yang unggul di Matematika belum tentu unggul di Fisika.
- **Siswa8, Siswa9, dan Siswa10** berada di arah positif vektor **X2 (Matematika)**.
  Mereka **unggul di Matematika**, tapi tidak menonjol di Fisika.

- **Siswa4 dan Siswa5** berada ke arah **X1 (Fisika)**.
  Mereka memiliki **nilai Fisika tinggi**, tetapi nilai Matematikanya cenderung lebih rendah.
- **Siswa1 dan Siswa2** berada di sisi bawah dekat titik asal dan berlawanan arah dengan X1. Mereka **rendah di kedua mata pelajaran**, terutama Fisika.
- **Siswa6 dan Siswa7** berada di posisi agak tengah.
  Mereka memiliki **nilai sedang** di kedua pelajaran, tanpa dominasi kuat di salah satu.