Analisis Nilai Eigen untuk Mendeteksi Tren dalam Data Media Sosial

Muhammad Aulia Azka 13523137
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13523137@itb.ac.id, auliaazka2506@gmail.com

***Abstract*— Konsep nilai eigen dalam aljabar linear memiliki peran sentral dalam berbagai metode analisis data, salah satunya untuk mendeteksi tren pada data berukuran besar seperti dari media sosial. Makalah ini membahas bagaimana penggunaan nilai eigen dalam konteks media sosial dengan memanfaatkan metode seperti *Principal Component Analysis* (PCA). Melalui penerapan nilai eigen dan vektor eigen, variasi data dalam jumlah besar dapat direduksi menjadi komponen utama yang mewakili tren dominan. Hasil analisis ini diharapkan membantu pengambil keputusan atau peneliti dalam memahami pola perilaku pengguna media sosial secara lebih efektif. Makalah ini mencakup pendahuluan, landasan teori, metodologi, hasil, hingga diskusi pemanfaatan nilai eigen yang komprehensif untuk menyingkap tren di platform media sosial.**

Kata Kunci: Nilai Eigen, PCA, Tren, Vektor Eigen

I. Pendahuluan

Kemajuan teknologi dan pertumbuhan pengguna internet yang pesat telah menjadikan media sosial sebagai salah satu sumber data terbesar saat ini. Platform seperti X, Instagram, dan TikTok menghasilkan jutaan unggahan setiap harinya, mencerminkan opini, minat, dan perilaku masyarakat dalam skala global. Mengingat besarnya volume data yang tersedia, proses pengolahan dan analisis data menjadi tantangan tersendiri.

Salah satu pendekatan yang efektif dalam mengekstraksi informasi dan mendeteksi tren dari data berukuran besar adalah penggunaan konsep aljabar linear, khususnya nilai eigen dan vektor eigen. Ketika sebuah data set diolah menggunakan metode Analisis Komponen Utama (PCA), nilai eigen berperan penting dalam menentukan komponen utama yang mewakili variasi terbesar dalam data. Dengan memfokuskan analisis pada komponen tersebut, kita dapat mengekstraksi tren yang paling menonjol dan menginterpretasi pola perilaku pengguna media sosial dengan lebih mudah..

Penelitian ini bertujuan untuk menganalisis secara komprehensif bagaimana nilai eigen dapat diaplikasikan untuk mendeteksi tren dalam data media sosial. Melalui pemahaman yang baik terhadap konsep dasar aljabar linear, diharapkan teknik-teknik ini dapat memperkaya pendekatan analisis data dan memberikan wawasan baru bagi praktisi maupun peneliti di bidang data science

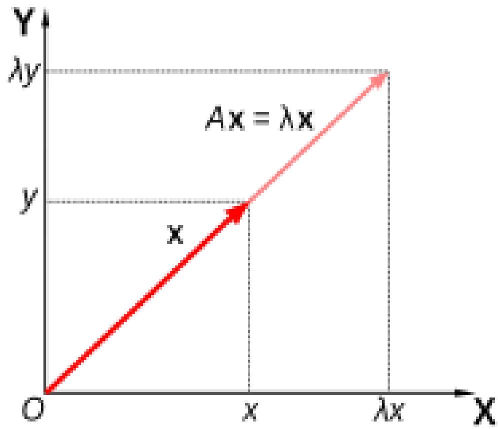
II. Landasan Teori

A. Nilai Eigen

Nilai eigen adalah nilai karakteristik dari suatu matriks yang berukuran n×n. Nilai eigen biasa dilambangkan "λ". Jika A adalah matriks n x n maka vektor tidak-nol x disebut vektor eigen dari A jika Ax sama dengan perkalian suatu skalar λ dengan x, yaitu

$$Ax = \lambda x. \quad (1)$$

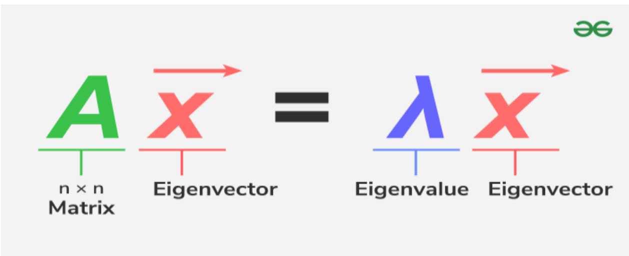
Persamaan (1) menyebabkan vektor x memanjang atau menyusut dengan faktor λ dengan arah yang bergantung pada tanda λ. Jika λ positif, arahnya sama. Jika λ negatif, arahnya berkebalikan.



Gambar 1. Diambil dari Wikipedia

Nilai eigen dapat dicari dengan cara persamaan (1) dikalikan dengan matriks identitas (I) pada kedua ruas. Lalu didapatkan persamaan sebagai berikut.

$$(\lambda I - A)x = 0 \quad (2)$$



Gambar 2. Diambil dari GeeksforGeeks

Pada persamaan (2), x = 0 adalah solusi trivial. Supaya persamaan (2) memiliki solusi tidak-nol, maka determinan dari (λI – A) haruslah sama dengan 0.

$$\det(\lambda I - A) = 0 \quad (3)$$

Persamaan (3) disebut sebagai persamaan karakteristik. Akar – akar dari persamaan tersebut, yaitu λ, dinamakan akar – akar karakteristik atau nilai – nilai eigen.

B. *Principal Component Analysis* (PCA).

Principal Component Analysis (PCA) adalah metode yang digunakan untuk mereduksi dimensi yang bertujuan untuk menyederhanakan sekumpulan data menjadi beberapa komponen utama dengan tetap mempertahankan sebanyak mungkin informasi. Langkah – langkah PCA meliputi:

1. Standarisasi atau Normalisasi Data

Pada tahap ini, data akan distandardisasi dengan cara menghitung rata – rata tiap kolom dari matriks data dan menghitung data *centering*.

$$X\textit{centered} = X - X\textit{mean} \quad (4)$$

Persamaan (4) adalah persamaan menghitung data centering.

2. Membentuk Matriks Kovarian

Matriks kovarian dapat dihitung dengan matriks yang sudah distandarisasi.

$$c = \frac{1}{m-1} X^T centered\ Xcentered$$

3. Menentukan Nilai Eigen dan Vektor Eigennya

Dekomposisi nilai eigen dilakukan pada C. Nilai eigen λi dan vektor eigen Vi diperoleh dengan persamaan:

$$CV_i = \lambda_i V_i$$

Nilai eigen terbesar λ_1 merefleksikan sumbu (komponen utama) dengan varian terbesar; selanjutnya λ_2 , λ_3, dan seterusnya.

4. Memilih dan Menginterpretasi Komponen Utama

Pilih komponen utama yang mencakup persentase varian kumulatif tinggi (misal 80 – 90%).

Data X lalu diproyeksikan ke sumbu – sumbu baru V1, V2, ..., menghasilkan data ringkas yang berdimensi rendah

$$PCi = \frac{\lambda i}{\sum_{j-1}^{n} \lambda j} \times 100\%$$

III. Perhitungan

Pada bagian ini, dijelaskan secara rinci langkah-langkah perhitungan yang dilakukan untuk menerapkan konsep nilai eigen dalam metode Analisis Komponen Utama (PCA). Pembahasan mencakup pengumpulan data set, membuat matriks berdasarkan data set, menghitung rata – rata tiap kolom, membuat matriks *centering*, menghitung matriks kovarian, hingga diperolehnya nilai eigen dan vektor eigen.

A. Data set

Mula – mula, data set yang nanti akan dihitung dan diuji harus ditentukan parameternya. Saya mengambil 5 postingan dengan mengambil durasi *vidio*, *like*, *comment*, *share,* dan *save* sebagai parameter.

Post	Durasi (detik)	Like	Share	Comment	Save
1	15	150	10	8	20
2	30	400	25	20	60
3	50	800	50	35	90
4	70	1500	80	60	180
5	90	2500	100	85	300

Selanjutnya dari data – data di atas bentuk matriks X dengan ukuran 5 x 5.

$$X = \begin{bmatrix} 15 & 150 & 10 & 8 & 20 \\ 30 & 400 & 25 & 20 & 60 \\ 50 & 800 & 50 & 35 & 90 \\ 70 & 1500 & 80 & 60 & 180 \\ 90 & 2500 & 100 & 85 & 300 \end{bmatrix}$$

Setiap baris mewakili satu postingan, sedangkan setiap kolom merepresentasikan parameter – parameter. Contoh kolom pertama merepresentasikan durasi.

B. Rata - Rata

Cari Xjmean (Rata – rata kolom matriks X ke j) untuk setiap kolom (j = 1, 2, 3, 4, 5):

$$Xjmean = \frac{1}{m} \times \sum_{i=1}^{m} xij$$

1. Durasi

$$X1mean = \frac{15 + 30 + 50 + 70 + 90}{5} = \frac{255}{5} = 51$$

2. Like

$$X2mean = \frac{150 + 400 + 800 + 1500 + 2500}{5} = 1070$$

3. Share

$$X3mean = \frac{10 + 25 + 50 + 80 + 100}{5} = \frac{265}{5} = 53$$

4. Comment

$$X4mean = \frac{8 + 20 + 35 + 60 + 85}{5} = \frac{208}{5} = 41.6$$

5. Save

$$X5mean = \frac{20 + 60 + 90 + 180 + 300}{5} = \frac{650}{5} = 130$$

C. Menyusun Matriks *Centered*

Kurangi elemen X dengan rata – rata kolomnya masing - Masing. Hasilnya berupa matriks X*Centered* berukuran 5 x 5

$$Xcentered = X - Xmean$$

Xmean adalah matriks 5 x 5 yang kolomnya diisi dengan Xjmean. Setiap elemen Xij akan diganti dengan Xij – Xjmean

i = 1, 2, 3, 4, 5

j = 1, 2, 3, 4, 5

Baris 1:

(15−51,150−1070,10−53,8−41.6,20−130) = (−36,−920,−43,−33.6,−110)

Baris 2:

30−51,400−1070,25−53,20−41.6,60−130) = (−21,−670,−28,−21.6,−70)

Baris 3:

(50−51,800−1070,50−53,35−41.6,90−130) = (−1,−270,−3,−6.6,−40)

Baris 4:

(70−51,1500−1070,80−53,60−41.6,180−130) = (19,430,27,18.4,50)

$$XCentered = \begin{bmatrix} -39 & -920 & -43 & -33.6 & -110 \\ -21 & -670 & -28 & -21.6 & -70 \\ -1 & -270 & -3 & -6.6 & -40 \\ 19 & 430 & 27 & 18.4 & 50 \\ 39 & 1430 & 47 & 43.4 & 170 \end{bmatrix}$$

D. Membuat Matriks Kovarian C

Langkah berikutnya adalah membuat matriks kovarian C dari data yang sudah dinormalisasi.

$$C = \frac{1}{m-1} \times (X^T centered \times Xcentered)$$

Dengan m = 5.

Berikut adalah matriks kovarian yang sudah dihitung

$$C \approx \begin{bmatrix} 552 & 20180 & 624 & 489 & 1680 \\ 20180 & 844700 & 26630 & 20910 & 72200 \\ 624 & 26630 & 870 & 680 & 2330 \\ 489 & 20910 & 680 & 543 & 1810 \\ 1680 & 72200 & 2330 & 1810 & 6200 \end{bmatrix}$$

Elemen diagonal Cjj menggambarkan variansi dari fitur ke-j. Elemen *off*-diagonal Cjk (j $\neq$ k) menggambarkan kovariansi antara fitur ke-j dan ke-k. Jika kovariannya tinggi dan positif, artinya kedua fitur tersebut naik dan turun bersama.

Sebagai contoh sederhana, jika kolom Durasi memiliki nilai – nilai *centering* [-10, -5, 0, 10, 5], maka variansi (elemen diagonal di matriks kovarian) dapat dihitung dengan menjumlahkan kuadratnya dan membaginya dengan (m – 1). Begitu juga dengan elemen *off*-diagonal, di mana kita menjumlahkan hasil perkalian lintas fitur

E. Mencari Nilai Eigen dan Vektor Eigen

Setelah terbentuk matriks kovarian C, dibutuhkan nilai Eigen (λ) dan vektor eigen (V) melalui persamaan:

$$\det(C - \lambda I) = 0$$

Persamaan di atas disebut persamaa karakteristik, dan penyelesaiannya menghasilkan sekumpulan λ1, λ2, ..., λn. Setiap λi akan memiliki vektor eigen Vi yang memenuhi

$$CVi = \lambda iVi$$

Umumnya, nilai Eigen diurutkan dari terbesar hingga terkecil contoh

$$\lambda 1 \geq \lambda 2 \geq \lambda 3 \geq \lambda 4 \geq \lambda 5$$

Nilai Eigen terbesar (λ1) menandakan bahwa komponen utama yang menjelaskan varian paling dominan. Setiap komponen utama (PC) memiliki persentasi varian yang dihitung.

$$PCi = \frac{\lambda i}{\sum_{j=1}^{n} \lambda j} \times 100\%$$

Jika λ1 sangat besar, berarti satu dimensi utama (PC1) sudah mewakili persentase besar dari total variasi data.

$$\lambda 1 \approx 92800$$

Nilai dari λ1

$$\lambda 2 \approx 78000$$

Nilai dari λ2

$$\lambda 3 \approx 12000$$

Nilai dari λ3

$$\lambda 4 \approx 2300$$

Nilai dari λ4

$$\lambda 5 \approx 750$$

Nilai dari λ4

Dengan λtotal ≈ 1020050.

F. Persentase Varian (Komponen Utama)

Mengecek persentase masing – masing varian (komponen Utama)

$$\lambda 1 \approx 92800$$

$$\frac{928000}{1020050} \approx 91.0\%$$

Pada PC1 (λ1 ≈ 92800)menghasilkan persentase 91.0%

$$\lambda 2 \approx 78000 \tag{1}$$

$$\frac{78000}{1020050} \approx 7.6\%$$

Pada PC1 (λ1 ≈ 78000)menghasilkan persentase 7.6%

Untuk sisanya (λ3, λ4, λ5) mencakup ≈ 1.4%. Dari sini, hanya 2 komponen utama (PC1 dan PC2) sudah menampung > 98% keseluruhan varian. Bahkan PC1 saja melampaui 90%. PC1 merefleksikan > 90% varian data. Artinya satu “dimensi” (komponen utama) sudah cukup menjelaskan mayoritas pergerakan kelima variabel

IV. KESIMPULAN

Melalui pembahasan ini, terlihat bahwa nilai Eigen dan vektor Eigen dapat dimanfaatkan untuk mengenali tren di dalam data skala besar. Proses reduksi dimensi melalui PCA memperlihatkan bagaimana komponen utama yang diturunkan dari nilai eigen terbesar dapat mewakili sebagian besar variasi data. Dengan mengukur besar nilai eigen yang dihasilkan oleh matriks kovarian, penulis dapat mengekstraksi komponen utama yang menyerap sebagian besar variansi data. Pendekatan ini mempermudah proses identifikasi pola atau tren karena hanya beberapa komponen utama sudah cukup untuk mewakili ratusan atau ribuan kolom fitur. Pada konteks media sosial, di mana data seperti durasi konten, jumlah *like*, *share*, *comment*, dan beragam interaksi lain berpotensi saling berkaitan

V. UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas izin dan kuasanya, penulis dapat menyelesaikan makalah ini dengan kodisi sebaik mungkin. Penulis berterima kasih kepada orang – orang yang telah mendukung penulis antara lain orang tua dan teman – teman. Ucapan terima kasih juga penulis beri kepada Bapak Arrival Dwi Sentosa, S.Kom., Dr. Ir. Rinaldi Munir, dan Dr. Judhi Santoso, M.Sc. selaku dosen mata kuliah Aljabar Linier dan Geometri yang telah mengajarkan banyak ilmu yang bermanfaat sehingga ilmu tersebut dapat penulis terapkan dalam pembuatan makalah ini. Kata terakhir, penulis mohon maaf apabila terdapat salah kata dalam penulisan.

REFERENCES

[1] https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2023-2024/Algeo-19-Nilai-Eigen-dan-Vektor-Eigen-Bagian1-2023.pdf. Diakses 27 Desember 2024

[2] https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2023-2024/Algeo-20-Nilai-Eigen-dan-Vektor-Eigen-Bagian2-2023.pdf

[3] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065). Diakses 29 Januari

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 27 Desember 2024

Muhammad Aulia Azka 13523137