



## Analisis Rekomendasi Pembuatan Produk Menggunakan RStudio Dan Twitter (Studi Kasus : Git Solution)

Fiyas Mahananing Puri<sup>1\*</sup>, Surya Tri Atmaja Ramadhani<sup>2</sup>, Dewi Anisa Istiqomah<sup>3</sup>, Vikky Aprelia Windarni<sup>4</sup>, Kardilah Rohmat Hidayat<sup>5</sup>

<sup>1,5</sup>Program Studi Sistem Informasi, <sup>2</sup>Program Studi Teknik Informatika, <sup>3</sup>Program Studi Manajemen Informatika, <sup>4</sup>Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, Sleman, Indonesia

Email: [fiyas@amikom.ac.id](mailto:fiyas@amikom.ac.id)<sup>1</sup>, [surya@amikom.ac.id](mailto:surya@amikom.ac.id)<sup>2</sup>, [dewianisaist@amikom.ac.id](mailto:dewianisaist@amikom.ac.id)<sup>3</sup>, [vikkyaprelia@amikom.ac.id](mailto:vikkyaprelia@amikom.ac.id)<sup>4</sup>, [kardilah.rh@amikom.ac.id](mailto:kardilah.rh@amikom.ac.id)<sup>5</sup>

### Abstrak

Dalam lanskap korporat kontemporer, internet telah muncul sebagai alat fundamental untuk meningkatkan penjualan dan layanan pelanggan, terutama melalui penerapan sistem rekomendasi yang banyak digunakan khususnya dalam *e-commerce*. Namun, banyak perusahaan, termasuk PT GIT Solution, masih belum sepenuhnya memanfaatkan potensi big data dalam analisis pasar, dan masih bergantung pada data historis serta metode konvensional seperti *follow-up* atau kunjungan langsung. Hal ini menyebabkan kurang optimalnya pemanfaatan data untuk menghasilkan rekomendasi yang relevan dan personal bagi pelanggan. Penelitian ini menggunakan pendekatan analisis data Twitter dengan algoritma pengelompokan *K-Means* untuk mengelompokkan data dan metode perankingan *Simple Additive Weighting* (SAW) untuk memberikan rekomendasi produk yang disesuaikan dengan preferensi pelanggan. Data yang digunakan diperoleh melalui *crawling* data Twitter, dan dianalisis menggunakan RStudio untuk mengidentifikasi tren serta preferensi konsumen. Hasil penelitian menunjukkan bahwa metode yang digunakan memberikan hasil evaluasi yang memuaskan dengan *Indeks Davies-Bouldin* (DBI) sebesar 0,10%, yang menunjukkan kualitas pengelompokan yang baik. Implikasi dari hasil ini adalah perusahaan dapat memanfaatkan analisis big data dari media sosial untuk meningkatkan kualitas rekomendasi produk, yang pada akhirnya dapat meningkatkan kepuasan pelanggan dan efektivitas pemasaran.

**Kata Kunci:** *K-Means*; *Simple Additive Weighting*; Twitter; RStudio; *Clustering*

### ABSTRACT

*In the contemporary corporate landscape, the internet has emerged as a fundamental tool for enhancing sales and customer service, notably through the implementation of extensively utilized recommendation systems, particularly in e-commerce. PT GIT Solution, a prominent software development company nationally, has acquired numerous clients across diverse sectors, including both governmental and private entities. The objective of the recommendation system is to deliver tailored product suggestions based on the preferences and interests of each consumer. The software development cycle has progressed to an annual or even weekly phase. Customer satisfaction is the primary responsibility of the organization. Twitter, a prevalent micro-blogging site with over 200 million active users, serves as a significant supply of big data for use. The research results on product recommendations utilizing the k-means clustering algorithm involve grouping data according to the proximity of specified criteria. Subsequently, the clustering phase is assessed using the Davies-Bouldin Index (DBI), yielding an evaluation result of 0.70%, which is sufficiently close to 0, indicating satisfactory performance. Ranking was conducted to derive the optimal product data outcomes with the simple additive weighting (SAW) algorithm, resulting in six product ratings suitable for recommendations.*

**Keywords:** *K-Means*; *Simple Additive Weighting*; Twitter; RStudio; *Clustering*

## 1. PENDAHULUAN

PT GIT Solution, sebagai pengembang perangkat lunak nasional, melayani pelanggan pemerintah dan swasta dengan menggunakan sistem rekomendasi produk untuk meningkatkan penjualan dan kepuasan pelanggan. Namun, metode konvensional seperti data historis, follow-up telepon, dan kunjungan langsung memiliki keterbatasan, karena hanya mencerminkan pola perilaku masa lalu dan memakan waktu. Dengan berkembangnya teknologi dan meningkatnya interaksi di media sosial, PT GIT Solution berpeluang memanfaatkan big data dari media sosial seperti Twitter, yang memiliki lebih dari 200 juta pengguna aktif, untuk lebih efektif menangkap preferensi dan tren pelanggan, serta meningkatkan rekomendasi produk (Taherdoost, 2023).

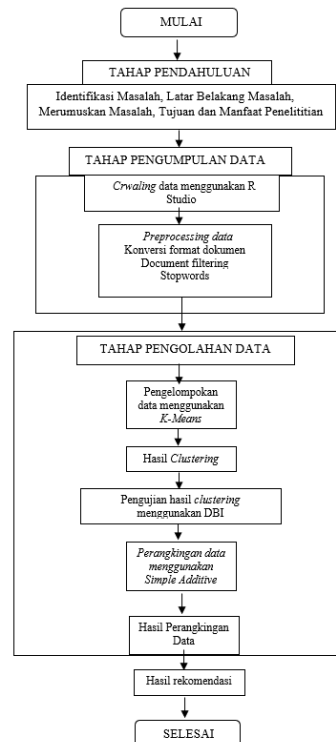
Penelitian ini bertujuan memanfaatkan big data dari Twitter untuk meningkatkan rekomendasi produk dan membantu perusahaan lebih memahami kebutuhan konsumen. Data media sosial memungkinkan perusahaan mengidentifikasi kebutuhan pelanggan, mengikuti tren pasar, dan memberikan solusi yang lebih personal. Pemanfaatan big data dalam pengambilan keputusan strategis menjadi sangat penting, terutama di era industri 4.0. Twitter, dengan banyak pengguna aktif, menjadi sumber data besar, di mana fitur hashtag sering digunakan sebagai indikator tren (Giannakis, 2022).

Penelitian berjudul "Event Recommendation using Social Media" menyimpulkan bahwa peneliti berhasil memprediksi popularitas acara di masa depan dan merekomendasikan acara berdasarkan popularitas tersebut. Mereka juga mengusulkan metode untuk mengidentifikasi tagar yang relevan dengan menggunakan algoritma pembelajaran untuk peringkat dalam identifikasi konten (Madisetty, 2019). Dalam penelitian yang berjudul "*Food Trend Based on Social Media Big Data Analysis Using K-Mean Clustering Algorithm (A Case Study on Yogyakarta Culinary Industry)*", menunjukkan hasil kesimpulan, penggalan tren makanan dari media sosial Twitter menggunakan *K-Means Clustering* menghasilkan akurasi data 70%-80% dibandingkan dengan penjualan pada beberapa restoran di Yogyakarta (Mihuandayani, 2018).

## 2. METODE

Bagian metode penelitian ini berisi ringkasan metode penelitian, meliputi jenis penelitian, subyek penelitian, teknik pengumpulan data, teknik analisis data serta pengujian keabsahan data ( jika menggunakan metode kualitatif). Untuk penelitian kuantitatif hindari penulisan rumus-rumus matematik dan statistik secara berlebihan. Untuk penelitian kajian

teori uraikan secara ringkas alur jalannya penelitian. Penelitian ini menggunakan jenis penelitian eksperimen untuk menghasilkan rekomendasi pembuatan produk software dengan data dari Twitter. Penelitian ini juga bersifat deskriptif untuk mengetahui hasil rekomendasi menggunakan R Studio, *K-Means*, dan SAW. Pendekatan kuantitatif digunakan untuk menghasilkan informasi rekomendasi produk yang bertujuan meningkatkan pendapatan perusahaan (Albayati & Ortakci, 2022).



**Gambar 1. Diagram Alur Penelitian**

Data untuk analisis rekomendasi produk dikumpulkan melalui *crawling* dari akun Twitter RDataMining menggunakan teknik *text mining* dan *topic modelling* dengan Twitter API dan R Packages. Algoritma *K-Means Clustering* dan *Simple Additive Weighting* (SAW) digunakan. Data diproses melalui *preprocessing*, dikelompokkan dengan *K-Means*, lalu produk diberi peringkat menggunakan SAW untuk menghasilkan rekomendasi (Rahmadian & Fajaryanti, 2023). Dalam penerapan *K-Means Clustering*, jumlah *cluster* ditentukan berdasarkan nilai optimal dari *indeks Davies-Bouldin*, dan metrik jarak euclidean digunakan untuk mengukur kedekatan data. Proses pengelompokan berhenti ketika *centroid* stabil atau iterasi maksimal tercapai. Setelah pengelompokan selesai, metode *Simple Additive Weighting* (SAW) digunakan untuk perangkingan, mengevaluasi setiap cluster berdasarkan bobot kriteria yang ditentukan untuk menghasilkan rekomendasi produk yang sesuai dengan preferensi pelanggan. Tahapan penelitian ini dapat dilihat melalui diagram pada Gambar 1.

## 2.1 K-Means Clustering

*Clustering* digunakan untuk meningkatkan variasi dalam suatu *cluster* dan antar *cluster*, dengan mengelompokkan data yang memiliki atribut sama ke dalam satu *cluster* dan data dengan atribut berbeda ke *cluster* lain. Tujuan *clustering* adalah mengurangi jumlah fungsi objek yang ditetapkan selama proses *clustering* (Faesal, 2020). Rumus euclidean yang digunakan untuk menentukan jarak dapat dilihat pada persamaan 1.

$$\text{dist} = \sqrt{\sum_{k=1}^n (P_k - q_k)^2} \quad (1)$$

Keterangan :

dist = Jarak obyek

n = Jumlah *cluster*

p<sub>k</sub> = Koordinat dari objek p

q<sub>k</sub> = Koordinat dari objek q

k = Urutan dari koordinat

## 2.2 Indeks Validitas Silhouette

Silhouette  $S(i)$  adalah ukuran statistik yang digunakan untuk menentukan jumlah cluster  $K$  yang optimal. Nilai rata-rata maksimum dari silhouette memberikan gambaran seberapa baik setiap objek terletak dalam clusternya, membantu menentukan jumlah cluster optimal (Li, 2018). Indeks validitas silhouette dapat dijabarkan seperti persamaan 2 berikut ini.

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

Keterangan :

$a(i)$  = Jarak rata-rata antara  $i$  dan semua objek lain dalam cluster yang sama

$b(i)$  = Jarak rata-rata minimum antara  $i$  ke semua objek lain dalam cluster apapun

$S(i)$  = Nilai silhouette

Nilai  $S(i)$  dari seluruh objek dalam suatu *cluster* menunjukkan tingkat kemiripan objek dalam suatu cluster, yang juga menunjukkan seberapa tepat objek telah dikelompokkan. Nilai  $S(i)$  yang lebih dekat dengan 1 menunjukkan pengelompokan objek yang lebih baik, sedangkan nilai  $S(i)$  yang lebih rendah menunjukkan pengelompokan objek yang lebih buruk. Seperti pada Tabel 1 yang merupakan nilai setiap dari  $S(i)$  berdasarkan strukturnya.

**Tabel 1. Nilai Silhoutte Kaufman dan Rousseeuw**

| Nilai Silhoutte Coefficient | Struktur          |
|-----------------------------|-------------------|
| $0.7 < SC \leq 1$           | Struktur Kuat     |
| $0.5 < SC \leq 0.7$         | Struktur Sedang   |
| $0.25 < SC \leq 0.5$        | Struktur Lemah    |
| $SC \leq 0.25$              | Tidak Terstruktur |

### 2.3 *Davies-Bouldin Index (DBI)*

Pada metode pengelompokan, indeks Davies-Bouldin mengukur validitas *cluster*, kohesi adalah jumlah dari kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti (Jumadi Dehotman Sitompul, 2019). Indeks Davies-Bouldin digunakan untuk mengukur separasi berdasarkan jarak antar pusat *cluster*. Indeks ini berupaya memaksimalkan jarak antar *cluster* dan meminimalkan jarak antar titik dalam setiap *cluster*. Jika jarak antar *cluster* maksimal, maka perbedaan karakteristik antar *cluster* akan lebih jelas (Konno, 2018).

### 2.4 *Simple Additive Weighting (SAW)*

Metode SAW menggunakan proses normalisasi matriks keputusan (X) ke suatu skala yang dapat diperbandingkan dengan semua rating alternatif yang ada untuk menghitung penjumlahan terbobot dari rating kinerja pada setiap alternatif dari semua atribut (Muhammad Dini et al., 2023).

Diberikan persamaan 3 sebagai berikut :

$$V_i = \sum_{j=1}^n w_j r_{ij} \quad (3)$$

Dengan keterangan  $V_i$  adalah nilai preferensi,  $W_j$  adalah bobot ranking,  $F_{ij}$  adalah rating kinerja ternormalisasi dan nilai  $V_i$  yang lebih besar mengindikasikan bahwa alternatif  $A_i$  lebih terpilih.

### 2.5 **RStudio**

RStudio adalah *frontend* R yang disukai karena mudah digunakan. RStudio memungkinkan pembuatan proyek seperti Shiny, serta penulisan laporan menggunakan RMarkdown atau RNotebook. RStudio juga memudahkan pengaturan direktori kerja dan impor file data tanpa perlu mengetik sintaks Commander, cukup melalui menu yang tersedia (Okazaki, 2020).

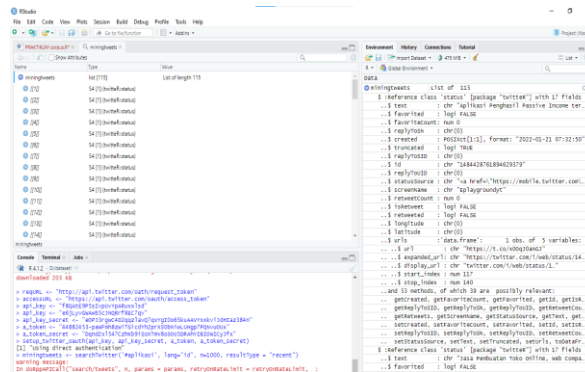
### 2.6 *Text Mining*

*Text mining* adalah proses penambangan pengetahuan dari data teks yang tidak terstruktur, seperti artikel web, chat, atau tweet. Lebih dari 80% data internet tidak terstruktur, termasuk teks, video, audio, dan gambar. Untuk mencapai akurasi tinggi dalam pemodelan,

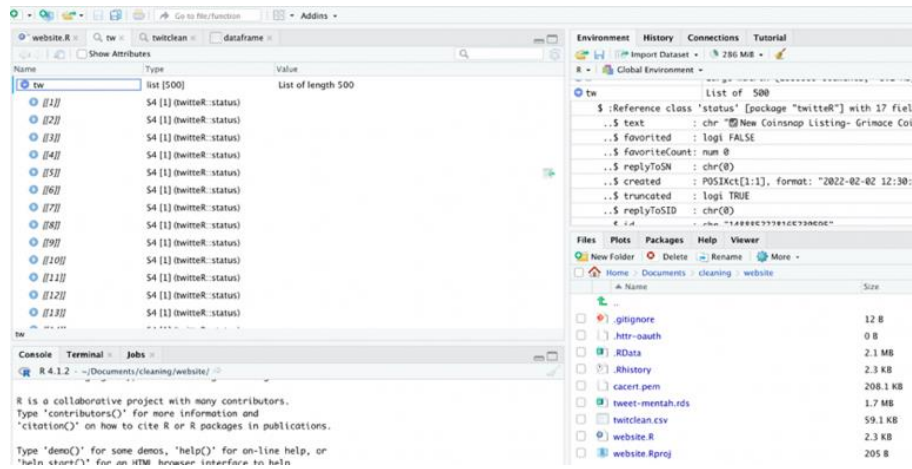
data harus bebas dari *noise* melalui *preprocessing*. *Preprocessing* melibatkan *case folding*, tokenisasi, *stopwords*, dan *stemming*, yang mengubah kata menjadi bentuk dasar, seperti "mengerjakan" menjadi "kerja".

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Crawling Data



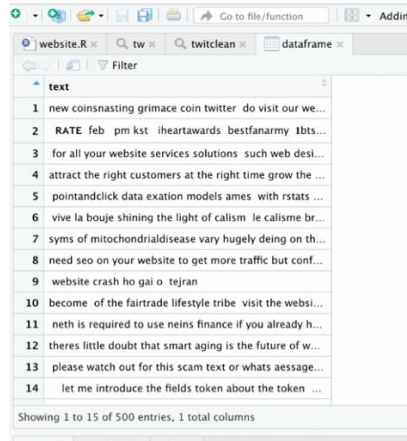
Gambar 2. Visualisasi Proses Crawling Data



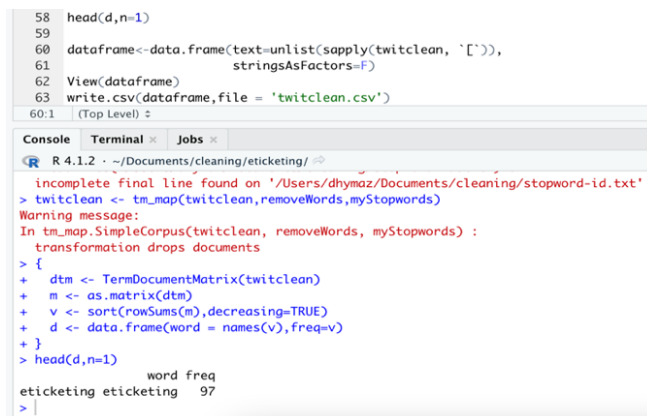
Gambar 3. Contoh Data Sebelum Dibersihkan Format .rds

Pada Gambar 2 menampilkan proses *crawling* data dengan menggunakan RStudio dengan 12 kata kunci yaitu #2dgame, #3dgame, #AugmentedReality, #TrainingIT, #SistemInformasiManajemen, #Website, #Virtual Reality, #E-planning, #E-Budgeting, #E-ticketing dan #gamesimulator. Gambar tersebut menunjukkan bahwa dari nilai 5000 tweet terbaru yang di ingin dikumpulkan, semuanya sukses di kumpulkan dan dapat diolah sebagai dataset. Tahap awal untuk mengolah data yang dilakukan adalah melihat melihat hasil *crawling* data twitter yang dilakukan di R Studio. Hasil tersebut dapat dilihat kedalam 2 bentuk yaitu dengan format .rds, seperti terlampir pada Gambar 3.

Pada Gambar 3 dan 4 memperlihatkan contoh hasil *crawling* data dengan kata kunci #website. Data tersebut diatas merupakan data mentah sebelum dilakukan *cleaning* csv di RStudio. Setelah dilakukan *cleaning* data dapat terlihat seperti pada Gambar 4. Proses ekstraksi dilakukan untuk mendapatkan representasi angka seperti pada Gambar 5.



Gambar 4. Data Setelah Dibersihkan



Gambar 5. Cleaning Data

Dari keseluruhan data yang diperoleh selanjutnya akan dilakukan perhitungan yaitu dengan 2 tahap, pengelompokkan menggunakan *K-Means Clustering* dan perangkan data menggunakan *Simple Additive Weighting* (SAW). Data keseluruhan yang diperoleh seperti pada Tabel 2 berikut.

Tabel 2. Hasil Crawling Data

| No. | Produk                            | User | Tweet |
|-----|-----------------------------------|------|-------|
| 1   | <i>2d game</i>                    | 254  | 381   |
| 2   | <i>3d game</i>                    | 122  | 183   |
| 3   | <i>Augmented Reality</i>          | 315  | 473   |
| 4   | <i>E-budgeting</i>                | 48   | 72    |
| 5   | <i>E-planning</i>                 | 36   | 54    |
| 6   | <i>Website</i>                    | 433  | 650   |
| 7   | <i>Sistem Informasi Manajemen</i> | 153  | 230   |
| 8   | <i>Training IT</i>                | 252  | 378   |

| No. | Produk          | User | Tweet |
|-----|-----------------|------|-------|
| 9   | Photo VR        | 112  | 168   |
| 10  | E-ticketing     | 145  | 218   |
| 11  | Game simulator  | 156  | 234   |
| 12  | Virtual Reality | 298  | 447   |

### 3.2 Hasil Clustering

Dalam pengelompokan data digunakan algoritma *K-Means Clustering* dengan menghitung jarak pusat *cluster* menggunakan rumus Euclidean hingga iterasi 3 dengan hasil seperti pada Tabel 3 berikut.

- a. Hasil *Cluster* Baru Iterasi 4

**Tabel 3. Hasil Cluster Baru Iterasi 4**

|                        |         |         |
|------------------------|---------|---------|
| Cluster baru yang ke-1 | 433     | 650     |
| Cluster baru yang ke-2 | 279.75  | 419.75  |
| Cluster baru yang ke-3 | 110.285 | 165.571 |

- b. Melakukan Perhitungan Jarak Pusat *Cluster* Baru Iterasi 4

**Tabel 4. Melakukan perhitungan Jarak Pusat Cluster Baru Iterasi 4**

| No. | Produk                   | User | Tweet | C1          | C2          | C3          | Jarak Terpendek |
|-----|--------------------------|------|-------|-------------|-------------|-------------|-----------------|
| 1   | 2d game                  | 254  | 381   | 323.1129833 | 46.52553063 | 258.9657609 | 46.52553063     |
| 2   | 3d game                  | 122  | 183   | 561.079317  | 284.4918716 | 20.99951409 | 20.99951409     |
| 3   | Augmented Reality        | 315  | 473   | 212.7275253 | 63.86019887 | 369.3511409 | 63.86019887     |
| 4   | E-budgeting              | 48   | 72    | 694.4847011 | 417.8972661 | 112.4060606 | 112.4060606     |
| 5   | E-planning               | 36   | 54    | 716.118007  | 439.5305735 | 134.0393637 | 134.0393637     |
| 6   | Website Sistem Informasi | 433  | 650   | 0           | 276.5874636 | 582.0786468 | 0               |
| 7   | Manajemen                | 153  | 230   | 504.7771786 | 228.189888  | 77.30168834 | 77.30168834     |
| 8   | Training IT              | 252  | 378   | 326.7185333 | 50.13107819 | 255.3602098 | 50.13107819     |
| 9   | Photo VR                 | 112  | 168   | 579.1070713 | 302.5196275 | 2.972664578 | 2.972664578     |
| 10  | E-ticketing              | 145  | 218   | 519.1993837 | 242.6120875 | 62.87954146 | 62.87954146     |
| 11  | Game simulator           | 156  | 234   | 499.7849538 | 223.1975022 | 82.29377441 | 82.29377441     |
| 12  | Virtual Reality          | 298  | 447   | 243.790894  | 32.7967224  | 338.2878861 | 32.7967224      |

Perhitungan *clustering* berhenti jika telah didapatkan hasil dari perhitungan pengelompokan antar iterasi yang sama. Dari hasil diatas perhitungan berhenti di iterasi ke-4 yang mempunyai hasil sama dengan pengelompokan *cluster* pada iterasi ke-3. Hasil perhitungan pengelompokan produk menggunakan algoritma *K-Means* terlihat pada Tabel 5.

**Tabel 5. Pengelompokan Item dalam Cluster**

| No. | Produk  | C1 | C2 | C3 |
|-----|---------|----|----|----|
| 1   | 2d game |    | 1  |    |
| 2   | 3d game |    |    | 1  |



| No. | Produk                            | C1 | C2 | C3 |
|-----|-----------------------------------|----|----|----|
| 3   | <i>Augmented Reality</i>          |    | 1  |    |
| 4   | <i>E-budgeting</i>                |    |    | 1  |
| 5   | <i>E-planning</i>                 |    |    | 1  |
| 6   | <b>Website</b>                    | 1  |    |    |
| 7   | <b>Sistem Informasi Manajemen</b> |    |    | 1  |
| 8   | <i>Training IT</i>                |    | 1  |    |
| 9   | <i>Photo VR</i>                   |    |    | 1  |
| 10  | <i>E-ticketing</i>                |    |    | 1  |
| 11  | <i>Game simulator</i>             |    |    | 1  |
| 12  | <i>Virtual Reality</i>            |    | 1  |    |

Didapatkan hasil bahwa produk Website termasuk kedalam pengelompokan tertinggi yang menjadi perbincangan *user Twitter*. Sehingga pada penelitian ini akan dilakukan perankingan dengan menggunakan semua *cluster 2* menggunakan metode *Simple Additive Weighting (SAW)* .

### 3.3 Pengujian Data Metode *Davies- Bouldin Index (DBI)*

- a. Menghitung *Sum of Square Within Cluster*

**Tabel 6. Hasil Centroid**

| Cluster   | User (x) | Tweet (y) |
|-----------|----------|-----------|
| <b>C1</b> | 433      | 650       |
| <b>C2</b> | 279.75   | 419.75    |
| <b>C3</b> | 110.286  | 165.571   |

- b. Rincian Hasil SSW

**Tabel 7. Rincian Hasil SSW**

| No. | Aplikasi                          | Cluster | User | Tweet | SSW1 | SSW2   | SSW3    |
|-----|-----------------------------------|---------|------|-------|------|--------|---------|
| 1   | <i>2d game</i>                    | 2       | 254  | 381   |      | 46.526 |         |
| 2   | <i>3d game</i>                    | 3       | 122  | 183   |      |        | 21.000  |
| 3   | <i>Augmented Reality</i>          | 2       | 315  | 473   |      | 63.860 |         |
| 4   | <i>E-budgeting</i>                | 3       | 48   | 72    |      |        | 112.406 |
| 5   | <i>E-planning</i>                 | 3       | 36   | 54    |      |        | 134.039 |
| 6   | <b>Website</b>                    | 1       | 433  | 650   | 0    |        |         |
| 7   | <b>Sistem Informasi Manajemen</b> | 3       | 153  | 230   |      |        | 77.302  |
| 8   | <i>Training IT</i>                | 2       | 252  | 378   |      | 50.131 |         |
| 9   | <i>Photo VR</i>                   | 3       | 112  | 168   |      |        | 2.973   |
| 10  | <i>E-ticketing</i>                | 3       | 145  | 218   |      |        | 62.880  |
| 11  | <i>Game simulator</i>             | 3       | 156  | 234   |      |        | 82.294  |
| 12  | <i>Virtual Reality</i>            | 2       | 298  | 447   |      | 32.797 |         |

Dengan total rata-rata Setiap SSW yaitu SSW1 adalah 0, SSW2 senilai 48.33 dan SSW3 yaitu 70.41.

- c. Menghitung *Sum of Square Between Cluster* (SSB)

$$SSB_{i,j} = d(c_i, c_j)$$

$$SSB_{1,2} = \sqrt{(433 - 279.75)^2 + (650 - 419.75)^2} = 276.59$$

$$SSB_{1,3} = \sqrt{(433 - 110.286)^2 + (650 - 165.571)^2} = 582.08$$

$$SSB_{2,3} = \sqrt{(279.75 - 110.286)^2 + (419.75 - 165.571)^2} = 305.49$$

Tabel 8 menunjukkan hasil matrik SSB.

**Tabel 8. Hasil Matrik SSB**

| SSB | 1      | 2      | 3      |
|-----|--------|--------|--------|
| 1   | 0      | 276.59 | 582.08 |
| 2   | 276.59 | 0      | 305.49 |
| 3   | 582.08 | 305.49 | 0      |

- d. Menghitung Nilai Rasio

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

$$R_{1,2} = \frac{0 + 48.33}{276.59} = 0.175, \quad R_{1,3} = \frac{0 + 70.41}{582.08} = 0.121, \quad R_{2,3} = \frac{48.33 + 70.41}{305.49}$$

Tabel 9 menunjukkan hasil perhitungan rasio.

**Tabel 9. Hasil Perhitungan Rasio**

| R | 1     | 2     | 3     | RMax  |
|---|-------|-------|-------|-------|
| 1 | 0     | 0.175 | 0.121 | 0.175 |
| 2 | 0.175 | 0     | 0.389 | 0.389 |
| 3 | 0.121 | 0.389 | 0     | 0.389 |

- e. Perhitungan *Davies-Bouldin Index* (DBI)

$$DBI = 1/k \sum_{i=1}^k i \neq j^{Max}(R_{i,j})$$

$$DBI = \frac{0.175 + 0.389 + 0.389}{3} = 0.10$$

Hasil penelitian menunjukkan bahwa algoritma *K-Means Clustering* yang digunakan dalam mengelompokkan data tweet menghasilkan tiga *cluster* yang cukup baik, dengan nilai Indeks Davies-Bouldin (DBI) sebesar 0,10%, yang mendekati angka 0 dan menunjukkan kualitas pengelompokan yang memuaskan. Selain itu, metode *Simple Additive Weighting* (SAW) digunakan untuk memberikan peringkat terhadap produk berdasarkan preferensi yang

diperoleh dari data Twitter, menghasilkan rekomendasi enam produk yang relevan untuk pelanggan.

### 3.4 Perhitungan Metode *Simple Additive Weighting* (SAW)

Data yang digunakan dalam perhitungan *Simple Additive Weighting* adalah 6 data produk dari *Cluster 2* dengan ditambahkan 1 anggota pada *Cluster 1* yang telah dilakukan perhitungan sebelumnya menggunakan *K-Means Clustering*.

**Tabel 10. Data Cluster II**

| Kode | Produk                   | User | Tweet |
|------|--------------------------|------|-------|
| 1    | <i>2d game</i>           | 254  | 381   |
| 2    | <i>Augmented Reality</i> | 315  | 473   |
| 3    | <i>Training IT</i>       | 252  | 378   |
| 4    | <i>Virtual Reality</i>   | 298  | 447   |
| 5    | Website                  | 433  | 650   |

**Tabel 11. Daftar Bobot Atribut**

| Kode | Keterangan | Bobot |
|------|------------|-------|
| K1   | User       | 0.40  |
| K2   | Tweet      | 0.60  |

**Tabel 12. Daftar Kriteria Tweet**

| Nilai | Keterangan    | Range   |
|-------|---------------|---------|
| 1     | Sangat Rendah | <= 200  |
| 2     | Rendah        | 201-400 |
| 3     | Cukup         | 401-600 |
| 4     | Tinggi        | 601-800 |
| 5     | Sangat Tinggi | > 800   |

**Tabel 13. Daftar Kriteria User**

| Nilai | Keterangan    | Range   |
|-------|---------------|---------|
| 1     | Sangat Rendah | <= 50   |
| 2     | Rendah        | 51-149  |
| 3     | Cukup         | 150-299 |
| 4     | Tinggi        | 300-449 |
| 5     | Sangat Tinggi | > 450   |

Proses normalisasi dilakukan dengan pengkalian kriteria dengan bobot yang sudah ditentukan. Dengan terlebih dahulu menghitung vektor atau nilai alternatif. Tabel 14 menunjukkan hasil proses normalisasi. Dari hasil normalisasi terbobot didapatkan perangkingan berdasarkan kriteria yang telah ditentukan, yang ditampilkan pada Tabel 15.

**Tabel 14. Normalisasi**

| Alternatif-1 (A1) | Alternatif-2 (A2) | Alternatif-3 (A3) | Alternatif-4 (A4) | Alternatif-5 (A5) |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| $r_{1,1} = 0.429$ | $r_{2,1} = 0.44$  | $r_{3,1} = 0.429$ | $r_{4,1} = 0.375$ | $r_{5,1} = 0.444$ |
| $r_{1,2} = 0.571$ | $r_{2,2} = 0.556$ | $r_{3,2} = 0.571$ | $r_{4,2} = 0.625$ | $r_{5,2} = 0.556$ |

**Tabel 15. Hasil Perangkingan SAW**

| Kode | Produk                   | Bobot | Ranking |
|------|--------------------------|-------|---------|
| A4   | <i>Virtual Reality</i>   | 0.525 | 1       |
| A3   | <i>Training IT</i>       | 0.514 | 2       |
| A1   | <i>2D Game</i>           | 0.514 |         |
| A2   | <i>Augmented Reality</i> | 0.511 | 3       |
| A5   | Website                  | 0.511 |         |

Dengan menggunakan algoritma perhitungan yang sama, dilakukan juga penghitungan kemungkinan lain dalam pengolahan data, yaitu dari proses pengelompokan sampai perangkingan *Simple Additive Weighting* dengan 2 *clustering*. Perhitungan *clustering* dilakukan dengan perhitungan jarak terpendek dengan menggunakan 2 *centroid* awal. Tabel 16 menunjukkan hasil pengelompokan data.

**Tabel 16. Pengelompokan Data**

| No. | Aplikasi                   | C1 | C2 |
|-----|----------------------------|----|----|
| 1   | <i>2d game</i>             | 1  |    |
| 2   | <i>3d game</i>             |    | 1  |
| 3   | <i>Augmented Reality</i>   | 1  |    |
| 4   | <i>E-budgeting</i>         |    | 1  |
| 5   | <i>E-planning</i>          |    | 1  |
| 6   | Website                    | 1  |    |
| 7   | Sistem Informasi Manajemen |    | 1  |
| 8   | <i>Training IT</i>         | 1  |    |
| 9   | <i>Photo VR</i>            |    | 1  |
| 10  | <i>E-ticketing</i>         |    | 1  |
| 11  | <i>Game simulator</i>      |    | 1  |
| 12  | <i>Virtual Reality</i>     | 1  |    |

Dari hasil *clustering* tersebut dilakukan perangkingan menggunakan *Simple Additive Weighting* dengan mengambil data pada produk dalam *cluster* 1, nilai bobot dan kriteria perangkingan menggunakan nilai yang sama dengan perhitungan yang sebelumnya dilakukan. Tabel 17 menunjukkan hasil normalisasi dengan algoritma *Simple Additive Weighting*. Tabel

18 menunjukkan hasil perangkingan menggunakan metode SAW. Hasil tersebut kemudian diuji menggunakan data para pakar, dimana data pada penelitian ini diperoleh dengan metode wawancara kepada *stakeholder* PT GIT Solution yang berwenang sebagai Direktur Divisi Pengembangan *Software*. Hasil wawancara tersebut memberikan data perangkingan produk berdasarkan skala prioritas produk yang akan dikembangkan secara intens, seperti pada Tabel 19.

**Tabel 17. Hasil Normalisasi dengan Algoritma SAW**

| Kode | Nama Aplikasi            | K1   | K2   |
|------|--------------------------|------|------|
| A1   | <i>2d game</i>           | 0,60 | 0,40 |
| A2   | <i>Augmented Reality</i> | 0,57 | 0,43 |
| A3   | <i>Training IT</i>       | 0,60 | 0,40 |
| A4   | <i>Virtual Reality</i>   | 0,50 | 0,50 |
| A5   | Website                  | 0,50 | 0,50 |

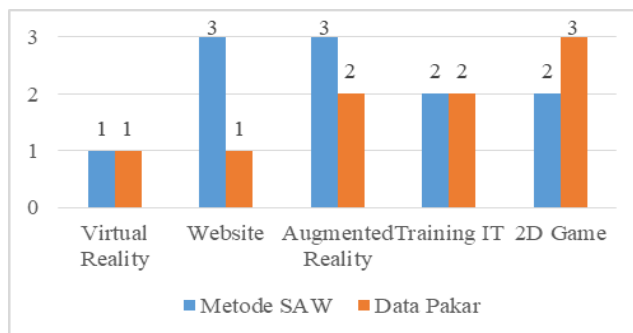
**Tabel 18. Hasil Perangkingan dengan SAW**

| Kode | Produk                   | Bobot | Ranking |
|------|--------------------------|-------|---------|
| A4   | <i>Virtual Reality</i>   | 0.500 | 1       |
| A3   | <i>Training IT</i>       | 0.480 |         |
| A1   | <i>2D Game</i>           | 0.480 | 2       |
| A2   | <i>Augmented Reality</i> | 0.480 |         |
| A5   | Website                  | 0.300 | 3       |

**Tabel 19. Data Pakar Pengembangan Produk**

| No | Kode | Nama Produk              | Skala (1-5) | Prosentase |
|----|------|--------------------------|-------------|------------|
| 1  | A1   | <i>2D Game</i>           | 2           | 40%        |
| 2  | A2   | <i>Augmented Reality</i> | 3           | 60%        |
| 3  | A3   | <i>Training IT</i>       | 3           | 60%        |
| 4  | A4   | <i>Virtual Reality</i>   | 4           | 80%        |
| 5  | A5   | Website                  | 4           | 80%        |

Dari data Tabel 19 dapat disimpulkan perangkingan menurut pakar, produk yang menduduki peringkat pertama yaitu *Virtual Reality* dan Website, kedua *Augmented Reality* dan *Training IT* kemudian peringkat ketiga ialah *2D Game*. Hasil dari data pakar kemudian digunakan untuk menguji akurasi metode *Simple Additive Weighting* yang sudah didapatkan sebelumnya, dengan nilai kecocokan dan perbedaan sebesar 33,33%.



Gambar 6. Grafik Hasil Perangkingan Metode SAW dan Data Pakar

Temuan ini menunjukkan bahwa penerapan *K-Means Clustering* efektif dalam mengelompokkan data tweet berdasarkan kesamaan atribut yang relevan dengan preferensi pelanggan. Nilai DBI yang mendekati 0 menandakan kualitas pengelompokan yang baik, dengan kemiripan tinggi dalam *cluster* dan perbedaan yang jelas antar *cluster*. Hal ini menunjukkan bahwa model dapat menangkap variasi preferensi pelanggan dengan baik dan membaginya ke dalam kelompok yang bermakna. Penerapan metode SAW menunjukkan bahwa produk dengan interaksi tinggi di Twitter, seperti *Virtual Reality* dan *Training IT*, mendapat peringkat lebih baik, menandakan keterlibatan pengguna di media sosial merupakan indikator penting untuk rekomendasi produk. Hasil ini mendukung penelitian sebelumnya bahwa media sosial, terutama Twitter, dapat menjadi sumber data signifikan dalam mengidentifikasi tren pasar dan preferensi konsumen, serta meningkatkan efektivitas sistem rekomendasi melalui analisis sentiment (Giannakis, 2022).

Penelitian ini menunjukkan bahwa data dari Twitter membantu memberikan rekomendasi produk yang lebih relevan dengan preferensi konsumen. *K-Means Clustering* terbukti efektif untuk mengelompokkan data pelanggan, sesuai dengan literatur sebelumnya (Faesal, 2020). Temuan ini mendukung bahwa *K-Means* adalah algoritma yang tepat untuk data media sosial, namun hasil pengelompokan ini sedikit berbeda dari penelitian lain yang mencatat akurasi 70% hingga 80% dalam menganalisis tren media social (Mihuandayani 2018). Dalam penelitian ini, hasil evaluasi menggunakan DBI hanya mencapai 0,10%, yang masih menunjukkan kualitas pengelompokan yang baik, tetapi tidak mencapai tingkat akurasi yang ditunjukkan oleh studi mereka. Hal ini bisa disebabkan oleh perbedaan dalam konteks dan sumber data, serta variasi dalam parameter yang digunakan untuk model *K-Means*. Metode SAW yang digunakan dalam penelitian ini terbukti efektif dalam memberikan bobot kriteria yang berbeda untuk perangkingan produk. Dalam konteks penelitian ini, metode SAW berhasil memberikan peringkat yang mencerminkan minat pengguna terhadap produk tertentu berdasarkan analisis data media sosial (Muhammad Dini 2023).

Dengan melakukan perbandingan ini, dapat disimpulkan bahwa hasil penelitian ini mendukung literatur sebelumnya tentang penggunaan media sosial untuk analisis pasar dan memberikan wawasan tambahan tentang efektivitas metode *K-Means* dan SAW dalam rekomendasi produk berbasis data Twitter. Penggunaan *big data* dari media sosial, seperti Twitter, memiliki potensi besar dalam analisis pasar dan pengembangan produk. PT GIT Solution dapat fokus pada produk dengan keterlibatan tinggi di media sosial, meningkatkan efektivitas pemasaran, dan menargetkan segmen pasar tertentu. Pendekatan berbasis *big data* lebih tepat dan efisien daripada metode konvensional, memungkinkan perusahaan menyesuaikan produk dengan tren dan kebutuhan konsumen secara proaktif.

#### 4. PENUTUP

##### Simpulan dan Saran

Penelitian ini bertujuan memanfaatkan big data dari Twitter untuk memberikan rekomendasi produk yang lebih relevan bagi pelanggan PT GIT Solution. Algoritma *K-Means Clustering* digunakan untuk mengelompokkan data tweet menjadi tiga *cluster*, dengan hasil evaluasi menggunakan *Indeks Davies-Bouldin* (DBI) sebesar 0,10%, menunjukkan kualitas pengelompokan yang baik metode *Simple Additive Weighting* (SAW) digunakan untuk memberi peringkat produk dalam *cluster*, dan hasilnya menunjukkan bahwa produk dengan interaksi tinggi di media sosial, seperti *Virtual Reality* dan *Training IT*, mendapat peringkat lebih baik. Penelitian ini berhasil memberikan rekomendasi produk berdasarkan preferensi pelanggan dengan menggunakan analisis *big data* dari Twitter, menunjukkan bahwa pemanfaatan data media sosial dapat meningkatkan efektivitas sistem rekomendasi dan membantu perusahaan memahami kebutuhan konsumen secara dinamis dan efisien. Untuk penelitian selanjutnya, disarankan memperluas sumber data dari media sosial lain seperti Instagram dan Facebook agar rekomendasi lebih komprehensif dan mencakup lebih banyak segmen pengguna. Selain itu, pengembangan model yang menggabungkan lebih banyak algoritma *machine learning*, seperti kombinasi *K-Means* dan *deep learning*, dapat dipertimbangkan untuk meningkatkan akurasi dan kualitas rekomendasi produk.

##### DAFTAR PUSTAKA

- Albayati, A. N. K., & Ortakci, Y. (2022). Recommendation Systems on Twitter Data for Marketing Purposes using Content-Based Filtering. *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–5. <https://doi.org/10.1109/HORA55278.2022.9799989>

- Faesal, A., Muslim, A., Ruger, A. H., & Kusriani, K. (2020). Sentimen Analisis Terhadap Komentar Konsumen Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(2), 207–213. <https://doi.org/10.30812/matrik.v19i2.640>
- Giannakis, M., Dubey, R., Yan, S., Spanaki, K., & Papadopoulos, T. (2022). Social media and sensemaking patterns in new product development: demystifying the customer sentiment. *Annals of Operations Research*, 308(1–2), 145–175. <https://doi.org/10.1007/s10479-020-03775-6>
- Jumadi Dehotman Sitompul, B., Salim Sitompul, O., & Sihombing, P. (2019). Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm. *Journal of Physics: Conference Series*, 1235(1). <https://doi.org/10.1088/1742-6596/1235/1/012015>
- Konno, T., Huang, R., Ban, T., & Huang, C. (2018). Goods recommendation based on retail knowledge in a Neo4j graph database combined with an inference mechanism implemented in jess. *2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation, SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI 2017* - , 1–8. <https://doi.org/10.1109/UIC-ATC.2017.8397433>
- Madisetty, S. (2019). Event recommendation using social media. *Proceedings - International Conference on Data Engineering, 2019-April*, 2106–2110. <https://doi.org/10.1109/ICDE.2019.00249>
- Mihuandayani, Ramandita, H. D., Setyanto, A., & Sumafta, I. B. (2018). Food trend based on social media for big data analysis using K-mean clustering and SAW: A case study on yogyakarta culinary industry. *2018 International Conference on Information and Communications Technology, ICOIACT 2018, 2018-January*, 549–554. <https://doi.org/10.1109/ICOIACT.2018.8350805>
- Okazaki, S., Plangger, K., West, D., & Menéndez, H. D. (2020). Exploring digital corporate social responsibility communications on Twitter. *Journal of Business Research*, 117, 675–682. <https://doi.org/https://doi.org/10.1016/j.jbusres.2019.09.006>
- S.M.Shamimul Hasan, D. S. R. K. N. I., Computing and Computational Sciences Directorate, & IEEE Computer Society (2019). (2019). A Scable Graph Analytics Framework for Programming With Big Data in R (pbdR). *IEEE Computer Society*.
- Taherdoost, H. (2023). Analysis of Simple Additive Weighting Method (SAW) as a MultiAttribute Decision-Making Technique: A Step-by-Step Guide. *Journal of Management Science & Engineering Research*, 6(1), 21–24. <https://doi.org/10.30564/jmsr.v6i1.5400>