

Identifikasi Pola Diskusi Publik mengenai Pemindahan Ibu Kota Negara Menggunakan Analisis *TF-IDF* dan *K-Means Clustering*

Salsa Bila Rahmania Sharafi

Sistem Informasi, Universitas Merdeka Malang. 085815708903

e-mail: 0710salsabila@gmail.com

ABSTRAK

Kata Kunci:

Pemindahan Ibu Kota
Sentimen Masyarakat
Twitter
TF-IDF
K-Means Clustering

Penelitian ini bertujuan untuk menganalisis pola diskusi masyarakat tentang pemindahan ibu kota negara Indonesia dari Jakarta ke Kalimantan menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *K-Means Clustering* pada komentar *Twitter*. Pengumpulan data dilakukan melalui web scraping pada tweet yang relevan dari tahun 2022 hingga 2024. Data yang terkumpul diolah melalui tahap preprocessing, termasuk penghapusan spam dan retweet, serta pembersihan teks dari noise. Metode *TF-IDF* digunakan untuk menghitung bobot setiap kata, yang kemudian dianalisis menggunakan algoritma *K-Means Clustering* untuk mengelompokkan komentar ke dalam beberapa kategori berdasarkan kemiripan semantik. Hasil penelitian menunjukkan adanya variasi sentimen yang signifikan di kalangan masyarakat, termasuk sentimen positif, negatif, dan netral. Analisis ini memberikan wawasan penting bagi pemerintah dalam mengevaluasi reaksi masyarakat dan mengambil keputusan yang lebih baik terkait kebijakan pemindahan ibu kota negara, serta dapat menjadi kontribusi pada literatur ilmiah tentang isu-isu sosial dan politik di Indonesia.

ABSTRACT

Keyword:

Relocation of the capital
Public sentiment
Twitter
TF-IDF
K-Means Clustering

This research aims to analyze patterns of public discussion about moving the capital city of Indonesia from Jakarta to Kalimantan using the Term Frequency-Inverse Document Frequency (TF-IDF) and K-Means Clustering methods on Twitter comments. Data collection was carried out through web scraping on relevant tweets from 2022 to 2024. The collected data was processed through a preprocessing stage, including removing spam and retweets, as well as cleaning the text from noise. The TF-IDF method is used to calculate the weight of each word, which is then analyzed using the K-Means Clustering algorithm to group comments into several categories based on semantic similarity. The research results show that there are significant variations in sentiment among the public, including positive, negative and neutral sentiment. This analysis provides important insights for the government in evaluating public reactions and making better decisions regarding the policy of moving the country's capital, and can contribute to scientific literature on social and political issues in Indonesia.

I. PENDAHULUAN

Pemindahan ibu kota negara Indonesia dari Jakarta ke Kalimantan Timur telah menjadi salah satu topik diskusi yang hangat di kalangan masyarakat dan pemerhati kebijakan publik. Keputusan ini diumumkan oleh Presiden Joko Widodo pada tahun 2019 dan bertujuan untuk mengatasi berbagai masalah yang dihadapi Jakarta, seperti kemacetan lalu lintas, polusi, dan penurunan tanah. Selain itu, pemindahan ini juga diharapkan dapat mendorong pemerataan pembangunan di luar Pulau Jawa. Namun, keputusan ini memicu beragam reaksi dari masyarakat, yang terungkap melalui berbagai platform media sosial, terutama Twitter. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan pemindahan ibu kota negara dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan K-Means Clustering.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan data teks dari platform Twitter. Data dikumpulkan melalui teknik web *scraping* menggunakan API Twitter untuk periode 2022 hingga 2024, di mana tweet yang relevan dengan topik pemindahan ibu kota diambil sebagai sampel. Data yang terkumpul kemudian diolah melalui beberapa tahap, termasuk pembersihan data, tokenisasi, dan *stemming*.

Metode TF-IDF digunakan untuk mengukur bobot atau kepentingan kata-kata dalam tweet yang diambil. TF-IDF membantu dalam mengidentifikasi kata-kata kunci yang paling relevan dan signifikan dalam diskusi masyarakat. Selanjutnya, K-Means Clustering diterapkan untuk mengelompokkan tweet ke dalam beberapa kelompok atau cluster berdasarkan kemiripan pola diskusi.

III. HASIL DAN PEMBAHASAN

Dari analisis yang dilakukan, ditemukan bahwa terdapat tiga kelompok utama dalam diskusi masyarakat mengenai pemindahan ibu kota: kelompok yang mendukung, kelompok yang menolak, dan kelompok yang bersikap netral. Kelompok pendukung cenderung menyoroti manfaat ekonomi dan peningkatan infrastruktur yang diharapkan terjadi dengan adanya pemindahan ini. Mereka melihat pemindahan ibu kota sebagai langkah strategis yang dapat membawa Indonesia menuju kemajuan yang lebih merata.

Di sisi lain, kelompok yang menolak mengkhawatirkan dampak negatif seperti kerusakan lingkungan, biaya yang besar, dan potensi ketidakstabilan sosial. Mereka mempertanyakan urgensi pemindahan ini dan menyarankan agar fokus pemerintah lebih diarahkan pada perbaikan kondisi di Jakarta. Sementara itu, kelompok netral menunjukkan bahwa meskipun mereka tidak secara tegas mendukung atau menolak, mereka menuntut kejelasan lebih lanjut dari pemerintah terkait rencana, anggaran, dan dampak jangka panjang dari kebijakan ini.

Hasil *clustering* menunjukkan bahwa cluster terbesar adalah kelompok yang mendukung kebijakan ini, diikuti oleh kelompok yang netral, dan terakhir kelompok yang menolak. Analisis ini menunjukkan bahwa meskipun ada dukungan yang signifikan, kekhawatiran masyarakat juga tidak bisa diabaikan.

TF (Term Frequency) dan IDF (Inverse Document Frequency)

Melibatkan proses perhitungan nilai bobot TF dan IDF (untuk setiap kata dengan menggunakan rumus TF-IDF yang tertera di bawah ini:

$$tF_{ij} = \frac{f_{ij}}{\sum_k f_{kj}} \quad (1)$$

Keterangan: f_{ij} = jumlah eksistensi (i) term pada dokumen (j) $\sum_k f_{kj}$ = jumlah (i) term yang terdapat dalam dokumen

$$IDF_i = \log\left(\frac{N}{n_i}\right) + \quad (2)$$

Keterangan: N = Jumlah dokumen n_i = Jumlah dokumen (i) term atau kata tersebut.

Perhitungan TF-IDF dapat dilakukan dengan menggunakan persamaan

$$TFIDF_{ij} = tF_{ij} \times IDF_i \quad (3)$$

TF-IDF merupakan suatu teknik statistik yang digunakan untuk memberikan nilai bobot pada kata-kata dalam sebuah dokumen *teks*. Pendekatan ini banyak diterapkan dalam bidang Pemrosesan Bahasa Alami (NLP), pencarian informasi, dan penambangan *teks*. Tujuan utamanya adalah untuk mengidentifikasi tingkat signifikansi suatu kata dalam konteks kumpulan dokumen. Algoritma TF-IDF digunakan untuk meningkatkan hubungan antara kata dan dokumen. TF-IDF adalah metrik statistik yang mengukur bobot kata dalam setiap dokumen atau kumpulan kata. Keberadaan kata yang lebih sering dalam dokumen tersebut menunjukkan pentingnya kata tersebut dalam konteks dokumen tersebut. Sebaliknya, kemunculan kata di banyak dokumen mengidentifikasi kata tersebut umum dan kurang signifikan. Bobot kata akan tinggi jika frekuensinya tinggi dalam suatu dokumen dan lebih rendah jika frekuensinya tinggi di berbagai dokumen.

Penelitian ini menggunakan algoritma TF-IDF karena sifatnya yang efisien, mudah diimplementasikan, dan berkinerja tinggi. Algoritma ini mampu menghasilkan hasil akurat dengan waktu pemrosesan yang cepat, bahkan tanpa memerlukan perangkat keras dan canggih. Selain itu, TF-IDF meratakan pengaruh kata-kata yang sering muncul di semua , sehingga metode ini efektif dalam menentukan kata kunci yang termasuk dalam StopList. StopList adalah kumpulan kata-kata yang dianggap tidak relevan untuk penelitian, biasanya kata-kata umum dalam bahasa Indonesia. Dalam penelitian ini, penulis akan menggunakan

algoritma TF-IDF untuk proses seleksi fitur. Jumlah kata yang muncul dalam suatu dokumen dikenal sebagai “TF” (Term Frequency). Nilai TF dipengaruhi oleh frekuensi kemunculan frasa dalam dokumen; semakin tinggi frekuensi, semakin tinggi nilai TF-nya. Sebaliknya, IDF adalah sejumlah kata yang didistribusikan ke semua dokumen yang tersedia. Jika frekuensi kemunculan sebuah kata semakin sering, IDF akan meningkat, dan nilainya akan sejalan dengan TF.

penulis melanjutkan dengan metode K – Means Clustering menggunakan aplikasi Orange Data Mining. K-Means Clustering merupakan algoritma pembelajaran yang didasarkan pada (Unsupervised Learning) yang banyak digunakan dalam bidang Machine Learning. Algoritma ini bekerja dengan mengelompokkan data menjadi sejumlah cluster yang ditentukan. Salah satu parameter penting dalam K-Means Clustering adalah nilai “k” yang merepresentasikan jumlah cluster yang akan dibentuk. Sebelum memulai proses clustering, peneliti perlu menentukan nilai k terlebih dahulu. Nilai k ini pada dasarnya mendefinisikan jumlah centroid, yaitu titik pusat dari setiap cluster, yang dibutuhkan untuk mengelompokkan data dalam data set. (Nursyafitri, 2022).

Tabel 1. Hasil Cluster dan Silhouette

| Kamus Kata | Cluster | Silhouette |
|-------------|---------|------------|
| Kerja | C1 | 0.508783 |
| Mengurangi | C1 | 0.560206 |
| Tekanan | C1 | 0.576824 |
| Lingkungan | C1 | 0.591208 |
| Jawa | C1 | 0.550973 |
| Bersih | C1 | 0.506304 |
| Sebagai | C2 | 0.467869 |
| Mendorong | C2 | 0.496471 |
| Pertumbuhan | C2 | 0.465388 |
| ekonomi | C2 | 0.478434 |

Hail tersebut dapat di dapatkan dengan melewati beberapa algoritma di bawah ini

Implementasi Rumus =LOWER

| | |
|---|---|
| 1 | Pemindahan Ibu Kota negara bukanlah bentuk egoisme pemerintah yg selama ini dilontarkan oleh mereka yang anti terhadap pemerintah =LOWER(B3) |
|---|---|

Gambar 1. rumus =LOWER

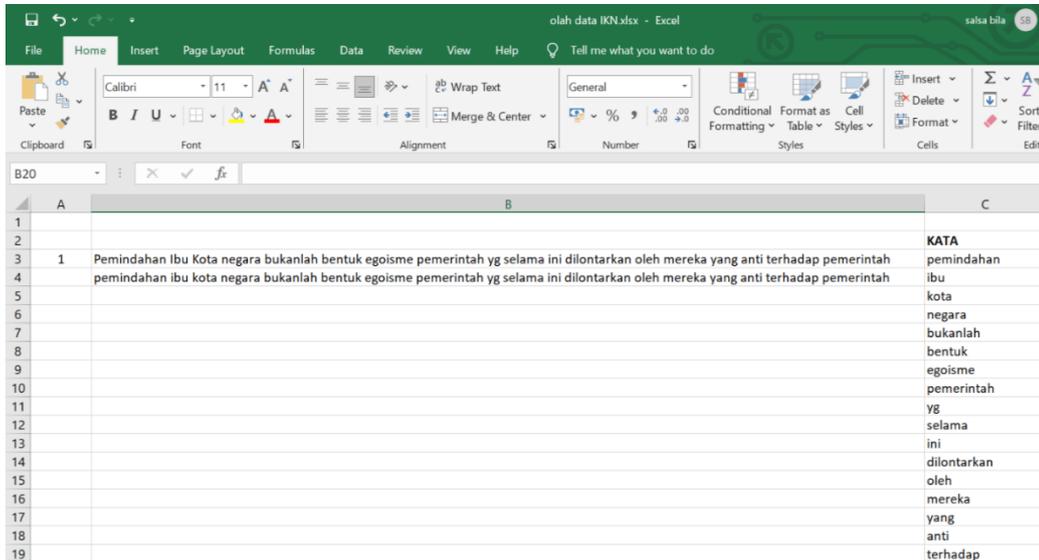
Rumus yang di gunakan adalah =LOWER berfungsi untuk mengonversi semua huruf besar dalam string teks ke huruf kecil.

| | |
|---|--|
| 1 | Pemindahan Ibu Kota negara bukanlah bentuk egoisme pemerintah yg selama ini dilontarkan oleh mereka yang anti terhadap pemerintah pemindahan ibu kota negara bukanlah bentuk egoisme pemerintah yg selama ini dilontarkan oleh mereka yang anti terhadap pemerintah |
|---|--|

Gambar 2. Hasil dari rumus = LOWER

Susunan per Kata

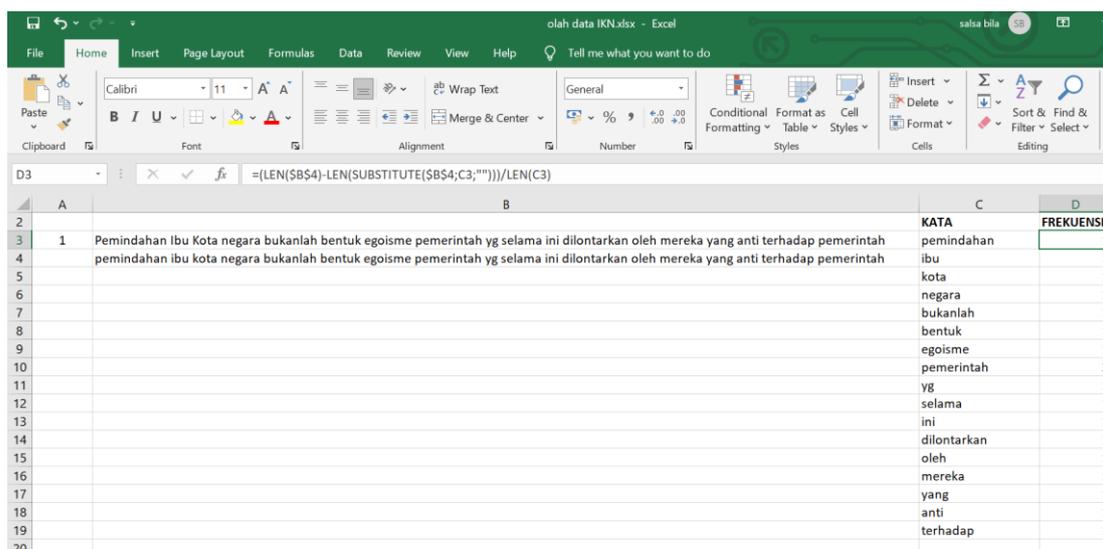
Setelah selesai dengan rumus =LOWER , kalimat yang utuh di susun menjadi per kata, jika di dalam satu kalimat ada kata ganda maka hanya satu saja yang di tulis proses ini untuk mempermudah perhitungan frekuensi



Gambar.3 Susunan Kalimat Per Kata

Frekuensi

Perhitungan data $LEN(\$B\$4) - LEN(SUBSTITUTE(\$B\$4;C3;"")) / LEN(C3)$ perintah =LEN berfungsi menghasilkan jumlah karakter dalam suatu teks sedangkan perintah SUBTITUTE berfungsi mengganti satu atau semua kemunculan string dalam teks untuk ditelusuri. SUBSTITUTE tidak dapat digunakan untuk mengganti beberapa, kecuali semua kemunculan dalam satu pemanggilan. Fungsi ini menampilkan teks sebagai output.



Gambar4. Hasil Perhitungan Frekuensi

Kamus Kata

Membuat kamus kata, kamus kata adalah kumpulan kata yang tidak boleh ada kata ganda di dalamnya dari semua kalimat 100 data set opini Masyarakat tentang pemindahan ibu kota negara, proses ini peneliti lakukan secara manual.

| KATA | FREKUENSI | Kamus Kata |
|-------------|-----------|-------------|
| pemindahan | 1 | pemindahan |
| ibu | 1 | ibu |
| kota | 1 | kota |
| negara | 1 | negara |
| bukanlah | 1 | bukanlah |
| bentuk | 1 | bentuk |
| egoisme | 1 | egoisme |
| pemerintah | 2 | pemerintah |
| yg | 1 | yg |
| selama | 1 | selama |
| ini | 1 | ini |
| dilontarkan | 1 | dilontarkan |
| oleh | 1 | oleh |
| mereka | 1 | mereka |
| yang | 1 | yang |
| anti | 1 | anti |

Gambar 5. Kamus Kata

Perhitungan TF-IDF

Tahap berikutnya dalam seleksi fitur adalah melakukan mengurutkan fitur berdasarkan bobot totalnya. Urutan ini dilakukan dari yang bobot tertinggi hingga terendah. Bobot total dari setiap fitur diperoleh melalui penjumlah bobot dari setiap kata (term) yang terdapat pada seluruh dokumen yang tersedia. Berikut cara perhitungannya

| | F | G | H | I | J | K | L | M | N | O | P |
|-------------|----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|---|
| | Term Frequency | | | | | | | | | | |
| Kamus Kata | Dokumen 1 | Dokumen 2 | Dokumen 3 | Dokumen 4 | Dokumen 5 | Dokumen 6 | Dokumen 7 | Dokumen 8 | Dokumen 9 | Dokumen 10 | |
| pemindahan | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ibu | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| kota | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| negara | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| bukanlah | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| bentuk | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| egoisme | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| pemerintah | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| yg | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| selama | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ini | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| dilontarkan | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| oleh | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| mereka | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| yang | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| anti | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| terhadap | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| dukung | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| nusantara | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| sebagai | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |

Gambar 6. Hasil TF

Dokumen tersebut kemudian diolah dengan menghitung nilai TF/IDF seperti di bawah ini.

$$\text{Inverse Document Frequency Dokumen 1 Dokumen 2 Dokumen} \\ =\text{IFERROR}(\text{LN}(\text{COUNTA}(\$G\$2:\$DB\$2)/\text{COUNTIF}(G3:DB3,"<>0")),0)$$

Gambar 7. Perhitungan TF/IDF

Berikut adalah hasil dari perhitungan TF/IDF dengan menggunakan rumus di atas.

Tabel 1. Hasil Hitung Nilai TF-IDF

| No | Kata | TF | IDF |
|----|-------------|----|------|
| 1 | Pemindahan | 1 | 0,15 |
| 2 | Ibu | 1 | 0,12 |
| 3 | Kota | 1 | 0,12 |
| 4 | Negara | 1 | 1,83 |
| 5 | Bukanlah | 1 | 4,61 |
| 6 | Bentuk | 1 | 4,61 |
| 7 | Egoism | 1 | 4,61 |
| 8 | Pemerintah | 2 | 3,00 |
| 9 | Yg | 1 | 3,00 |
| 10 | Selama | 1 | 4,61 |
| 11 | Ini | 1 | 3,22 |
| 12 | Dilontarkan | 1 | 4,61 |
| 13 | Oleh | 1 | 3,91 |
| 14 | Mereka | 1 | 3,91 |
| 15 | Yang | 1 | 1,56 |
| 16 | Anti | 1 | 4,61 |
| 17 | Terhadap | 1 | 4,61 |
| 18 | Dukung | 0 | 3,51 |
| 19 | Nusantara | 0 | 1,14 |
| 20 | Sebagai | 0 | 2,21 |

Selanjutnya dilakukan perhitungan TF-IDF kata pada seluruh dokumen

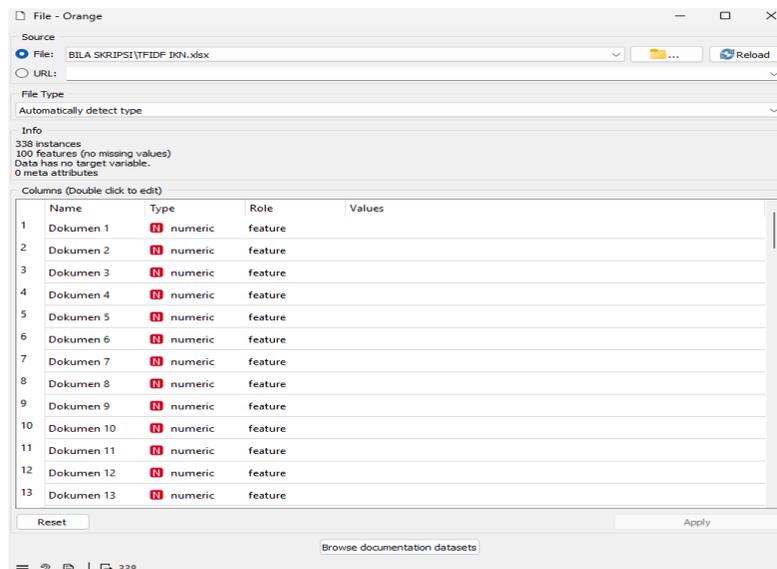
Tabel 2. Hasil perhitungan TF-IDF

| No | Kata | D1 | D2 | D3 | D4 | D5 |
|----|------------|--------|--------|--------|--------|--------|
| 1 | Pemindahan | 0,1508 | 0,1508 | 0,1508 | 0,1508 | 0,1508 |
| 2 | Ibu | 0,1165 | 0,1165 | 0,1165 | 0,1165 | 0,1165 |
| 3 | Kota | 0,1165 | 0,1165 | 0,1165 | 0,1165 | 0,1165 |
| 4 | Negara | 1,8325 | 1,8325 | 1,8235 | 1,8325 | 1,8325 |
| 5 | Bukanlah | 4,6051 | 0 | 0 | 0 | 0 |
| 6 | Bentuk | 4,6051 | 0 | 0 | 0 | 0 |
| 7 | Egoism | 4,6051 | 0 | 0 | 0 | 0 |
| 8 | Pemerintah | 5,9914 | 0 | 0 | 0 | 0 |
| 9 | Yg | 2,9957 | 0 | 0 | 0 | 0 |
| 10 | Selama | 4,6051 | 0 | 0 | 0 | 0 |
| 11 | Ini | 3,2118 | 0 | 0 | 0 | 0 |

| | | | | | | |
|----|-------------|--------|--------|--------|--------|---|
| 12 | Dilontarkan | 4,6051 | 0 | 0 | 0 | 0 |
| 13 | Oleh | 3,9120 | 0 | 0 | 0 | 0 |
| 14 | Mereka | 3,9120 | 0 | 0 | 0 | 0 |
| 15 | Yang | 1,5606 | 0 | 0 | 1,5605 | 0 |
| 16 | Anti | 4,6051 | 0 | 0 | 0 | 0 |
| 17 | Terhadap | 4,6051 | 0 | 0 | 0 | 0 |
| 18 | Dukung | 0 | 3,5065 | 0 | 0 | 0 |
| 19 | Nusantara | 0 | 1,1394 | 1,1394 | 0 | 0 |
| 20 | Sebagai | 0 | 2,2072 | 0 | 2,2072 | 0 |

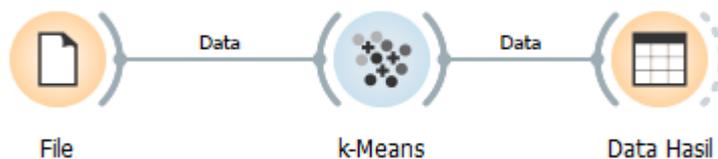
K – Means Clustering

Pertama peneliti memasukan perhitungan TF-IDF ke dokumen excel yang baru, lalu di input ke orange data mining



Gambar 8. Input Data di Orange Data Mining

Setelah terinput selanjutnya peneliti memakai algoritma k-mean clustering seperti yang tersusun di bawah



Gambar 9. Susunan Algoritma K-Means

Di bawah ini adalah hasil dari K-Means clustering menggunakan Orange Data Mining Metode Silhouette yang merupakan alat evaluasi yang digunakan untuk menilai kualitas pemisah cluster dalam algoritma clustering. Metode ini mengukur seberapa baik setiap titik data point berada dalam clusternya sendiri dibandingkan dengan cluster tetangga terdekat. Nilai Silhouette dapat bervariasi dari -1 hingga 1. Nilai positif: menunjukkan bahwa data point lebih mirip dengan anggota clusternya sendiri dibandingkan dengan cluster tetangga terdekat. Ini mengindikasikan bahwa data point ditempatkan dengan tepat di dalam cluster yang sesuai. Nilai negative: menunjukkan bahwa data point lebih mirip dengan

anggota *cluster* tetangga terdekat dibandingkan dengan anggota clusternya sendiri. Hal ini menunjukkan bahwa data point mungkin di tempatkan dalam cluster yang salah. Nilai mendekati 0 menunjukkan bahwa data point berada di dekat batas antara dua cluster. Semakin tinggi nilai silhouette positif, semakin baik kualitas pemisahan *cluster*.

| | Cluster | Silhouette | Dokumen 1 | Dokumen 2 | Dokumen 3 | Dokumen 4 | Dokumen 5 |
|----|---------|------------|-----------|-----------|-----------|-----------|-----------|
| 1 | C2 | 0.554209 | 0.150823 | 0.150823 | 0.150823 | 0.150823 | 0.150823 |
| 2 | C2 | 0.556626 | 0.116534 | 0.116534 | 0.116534 | 0.116534 | 0.116534 |
| 3 | C2 | 0.556594 | 0.116534 | 0.116534 | 0.116534 | 0.116534 | 0.116534 |
| 4 | C1 | 0.50753 | 1.83258 | 1.83258 | 0 | 1.83258 | 1.83258 |
| 5 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 6 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 7 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 8 | C2 | 0.528328 | 5.99146 | 0 | 0 | 0 | 0 |
| 9 | C2 | 0.534681 | 2.99573 | 0 | 0 | 0 | 0 |
| 10 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 11 | C2 | 0.539302 | 3.21888 | 0 | 0 | 0 | 0 |
| 12 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 13 | C2 | 0.550307 | 3.91202 | 0 | 0 | 0 | 0 |
| 14 | C2 | 0.534675 | 3.91202 | 0 | 0 | 0 | 0 |
| 15 | C2 | 0.525556 | 1.56065 | 0 | 0 | 1.56065 | 0 |
| 16 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 17 | C2 | 0.55144 | 4.60517 | 0 | 0 | 0 | 0 |
| 18 | C1 | 0.505541 | 0 | 3.50656 | 0 | 0 | 0 |
| 19 | C3 | 0.490128 | 0 | 1.13943 | 1.13943 | 0 | 0 |
| 20 | C1 | 0.524717 | 0 | 2.20727 | 0 | 2.20727 | 0 |

Gambar 10. Hasil Silhouette

IV. SIMPULAN

Penelitian ini berhasil mengidentifikasi dan mengelompokkan sentimen masyarakat terhadap pemindahan ibu kota negara menggunakan metode TF-IDF dan K-Means Clustering. Mayoritas masyarakat menunjukkan dukungan terhadap kebijakan ini, meskipun terdapat sejumlah kekhawatiran yang signifikan, terutama terkait dampak lingkungan dan sosial. Hasil penelitian ini diharapkan dapat menjadi masukan bagi pemerintah dalam mempertimbangkan opini publik dan mengambil keputusan yang lebih bijak dalam implementasi kebijakan pemindahan ibu kota.

DAFTAR RUJUKAN

- [1] Fachreza, M. R. (2023). klasifikasi sentimen masyarakat terhadap preoses pemindahan ibu kota negara (IKN) Indonesia pada media sosial Twitter menggunakan Metode Naive Bayes.
- [2] Nursyafitri, G. D. (2022). K-Means Clustering salah satu contoh teknik analisis data populer.
- [3] Permana, D. (2023). Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neightboars. *Jurnal Media Informatika Budidarma*, 7(3) 1306-1314.
- [4] Rachman, D. A. (2021). Implementasi Text Minning Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering. *EKSPONENSIAL*, 11(2) 2085-7829.
- [5] Simha, A. (2021). Memahami TF-IDF untuk Pembelajaran Mesin.

- [6] Syarif, M. (2022). Peranan Komisi Pemilihan Umum (KPU) Dalam Meningkatkan Partisipasi Masyarakat Pada Pemilihan Kepala Daerah (PILKADA) di Desa Sungai Terab Kabupaten Indragiri Hilir Profinsi Riau Tahun 2018-2023. *Diss Universitas Islam Sultan Syarif Kasim Riau*.
- [7] Widaningrum, i. (2022). Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen. *prosiding SISFOTEK*, 6(1), 145-149.