

Analisis Opini *User* di Twitter pada Proses Pembangunan Ibu Kota Nusantara (IKN) dengan Metode *DBSCAN*

Ni Wayan Gauri Prasanti¹, Ni Luh Gede Pivin Suwirmayanti², Ni Made Dewi Kansa Putri³

¹Teknologi Informasi, ²Sistem Komputer, ³Bisnis Digital

Institut Teknologi dan Bisnis STIKOM Bali

Denpasar, Indonesia

e-mail: 210040104@stikom-bali.ac.id, pivin@stikom-bali.ac.id, kansa@stikom-bali.ac.id

Abstrak

Twitter merupakan salah satu platform *micro-blogging* yang digunakan oleh jutaan pengguna untuk menyampaikan opini dan berdiskusi mengenai berbagai topik, termasuk pembangunan Ibu Kota Nusantara (IKN). Penelitian ini bertujuan untuk menganalisis opini pengguna Twitter terkait proses pembangunan IKN dari tahun 2010 hingga 2024. Metode yang digunakan dalam penelitian ini adalah *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)*, yang memiliki keunggulan dalam mendeteksi outlier atau noise serta mampu mengenali bentuk kluster yang tidak beraturan, yang sulit dikenali oleh beberapa algoritma klusterisasi lainnya. Tahapan awal penelitian mencakup proses *preprocessing* untuk membersihkan data sebelum dilakukan klusterisasi. Evaluasi model dilakukan menggunakan *Silhouette Index* untuk mengukur kualitas kluster yang terbentuk. Hasil penelitian menunjukkan bahwa parameter optimal untuk *DBSCAN* adalah $Eps = 0.1$ dan $MinPts = 9$, yang diterapkan pada 2.011 data yang telah melalui tahap pembersihan. Hasil ini menunjukkan bahwa model yang digunakan mampu membentuk kluster opini dengan kualitas yang cukup baik. Dengan demikian, penelitian ini dapat memberikan wawasan yang lebih mendalam mengenai sentimen publik terhadap pembangunan IKN, yang dapat dimanfaatkan dalam pengambilan kebijakan dan strategi komunikasi pemerintah.

Kata kunci: *Opini User, Twitter, IKN, DBSCAN*

1. Pendahuluan

Saat ini kita hidup di era teknologi informasi dan komunikasi. Kemajuan teknologi telah menyediakan informasi dan sarana komunikasi jauh melampaui apa yang telah dimiliki manusia. Meskipun dalam beberapa dekade terakhir peran informasi kurang diperhatikan, namun kebutuhan akan informasi dan komunikasi tidak sepeenting kebutuhan masyarakat akan sandang dan pangan.

Twitter merupakan salah satu layanan *micro-blogging* yang menyatukan jutaan dari pengguna. Memungkinkan pengguna untuk mempublikasikan dan bertukar pesan singkat, juga dikenal sebagai *tweet*, melalui berbagai macam klien. Pengguna dapat memposting *tweet* mereka dengan mengirim email, pesan teks SMS, langsung dari smartphone mereka dan beragam layanan berbasis Web[1].

Pemindahan ibu kota memang bukan hal yang baru dalam perbincangan. Rencana pemindahan pusat pemerintahan Indonesia dari Jakarta ke daerah lain telah lama ada di bawah Presiden Susilo Bambang Yudhoyono. Akan tetapi, rencana itu tidak pernah terwujud. Pada masa jabatan Presiden Joko Widodo, Kalimantan Timur dipilih sebagai ibu kota baru Indonesia setelah melalui pertimbangan matang, kajian, dan konsultasi publik[2]. Masyarakat sering membahas bidang politik pada media sosial Twitter sehingga menarik untuk dikaji. Bidang politik dalam hal ini mengenai pemindahan Ibu Kota Indonesia, yang saat ini di Jakarta yang teletak di pulau Jawa, akan dipindah ke Ibu Kota Nusantara (IKN) yang terletak di pulau Kalimantan, lebih tepatnya di Kalimantan Timur.

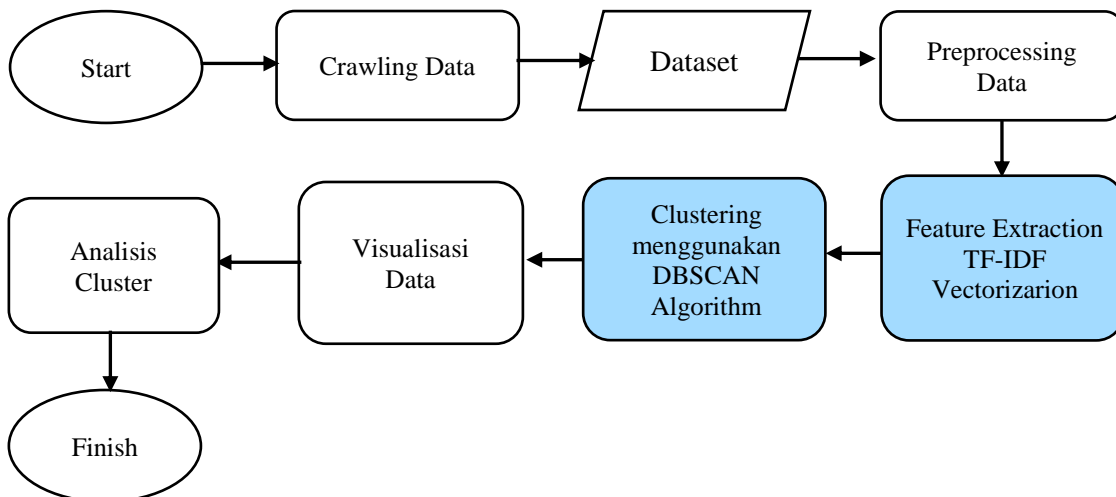
Clustering merupakan teknik yang penting dalam *Data Mining*[3]. *DBSCAN* merupakan algoritma pengelompokan berbasis kepadatan yang paling populer dan mampu memulihkan kluster non-bulat. Kelemahan utama dari *DBSCAN* adalah memperlakukan semua fitur secara setara. *DBSCAN* adalah algoritma pengelompokan spasial fundamental, tetapi algoritma ini memiliki hambatan dalam kasus terburuk, kompleksitas waktu proses adalah $O(n^2)$. Algoritma bertahap bergantung pada penerapan *DBSCAN* dengan koreksi tepi yang berulang pada data. Analisis tersebut menunjukkan parameterisasi yang ditentukan untuk memungkinkan *DBSCAN* mengidentifikasi kluster seluler yang wajar untuk kondisi yang berbeda[4]. Penambangan data dalam *Data Mining* merupakan sebuah istilah untuk menjelaskan cara mencari informasi dalam sebuah *database*[5].

Informasi adalah data yang telah diolah sehingga berguna bagi orang yang membutuhkan. Informasi dapat dibagi menjadi dua, yaitu fakta dan opini. Fakta merupakan pernyataan yang objektif disertai dengan bukti-bukti yang jelas. Sedangkan opini bersifat subjektif, pernyataan ini memuat tanggapan seseorang mengenai sesuatu sesuai persepsi dan asumsi mereka masing-masing[6]. Opini *user* yang dimaksud ini merupakan persepsi dan asumsi *user* mengenai sesuatu hal yang dibahas pada penelitian ini.

Silhouette Coefficient atau *Silhouette Index* (SI) adalah cara menafsirkan gugus validasi objek. Teknik ini memberikan representasi grafis ringkas tentang seberapa tepat setiap objek dilokalisasi dalam suatu kluster. Koefisien Siluet pertama kali dikembangkan oleh Rousseeuw pada tahun 1986. Koefisien Silhouette merupakan indeks rasio yang didasarkan pada kekompakan dan keterpisahan kluster. Tujuan teknik validasi kluster ini adalah untuk mengevaluasi hasil pengelompokan. Hasil evaluasi ini dapat digunakan untuk menentukan jumlah kluster dalam dataset[7].

2. Metode Penelitian

Dalam penelitian ini, dilakukan tujuh tahapan untuk menyelesaikan penelitian ini. *Crawling data* untuk memperoleh data *tweet*, dan pemrosesnya untuk menghasilkan data yang bersih. Kemudian, perhitungan centrality bertujuan untuk menyaring data dalam beberapa bagian dalam perhitungan yang ada dalam *centrality*. DBSCAN *Clustering* digunakan untuk mendapatkan *cluster* dan *Silhouette Score*. *Cluster Analysis* berfungsi untuk mengetahui tipe opini dari *user*. Untuk lebih jelas, dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Metode Penelitian

2.1 Crawling Data

Crawling data adalah sebuah aktivitas dimana kita mengumpulkan data dari suatu web atau database[8]. Ada dua sistem pencarian untuk merayapi data di twitter, berdasarkan pengguna dan kata kunci. Pencarian kata kunci didasarkan pada fragmen kata atau tagar dan dapat mencari hingga 100 *tweet* per unduhan. Pengguna sekarang dapat mencari berdasarkan nama akun twitter dan dapat mengunduh hingga 200 *tweet* per unduhan. Kami mengekstrak fitur data pengguna dari indeks pengguna twitter dalam bentuk jumlah total *tweet*, jumlah total *like*, sumber, jumlah total pengikut, *id* pengguna, jumlah total pengikut, akun, lokasi, situs web, dan nama. Saat ini, ekstraksi fitur yang diperoleh dari indeks data *tweet* twitter tersedia dalam bentuk jumlah suka, URL, *mention*, tagar, dan *retweet*[9]. Dataset diperoleh dengan cara crawling data pada platform twitter menggunakan Twitter API. Perolehan API Key dilakukan pada website Twitter Developers. Setelah key didapatkan, crawling data baru bisa dilakukan dengan bahasa pemrograman python. Pencarian dataset menggunakan keyword “#IKN” dengan format “.csv”.

2.2 Preprocessing Data

Preprocessing mencakup semua proses yang terjadi dalam mempersiapkan seluruh data yang akan diproses dalam operasi penemuan pengetahuan dari *text mining*. *Preprocessing data* adalah proses analisis data yang melibatkan transformasi dan pembersihan data mentah agar dapat digunakan secara efektif dalam model analisis atau *machine learning*. Data mentah sering kali mengandung *noise*, *missing value*, atau format yang tidak konsisten, sehingga perlu dilakukan tahapan *preprocessing* terlebih dahulu. Dataset yang

telah di-*crawl* akan dilanjutkan dalam *preprocessing* untuk mendapatkan hasil yang lebih optimal, karena data yang sudah di-*preprocessing* akan menjadi data yang bersih dan tidak mengandung *noise*[10].

2.3 Fitur Ekstraksi

Metode yang digunakan untuk menentukan signifikansi kata dalam dokumen disebut TF-IDF. Frekuensi munculnya sebuah kata dalam dokumen menentukan signifikansi atau tidaknya kata tersebut. Sebuah metode ekstraksi fitur yang disebut *Term Frequency – Inverse Document Frequency* (TF-IDF) digunakan untuk menentukan kata yang penting dalam sekelompok dokumen. Karena, data dalam penelitian ini yang akan diolah dengan teknik clustering adalah data teks, maka tahapan ini mutlak dilakukan. Vektorisasi teks, atau mengubah teks menjadi angka, akan dilakukan pada saat ini. Formulasi TF-IDF adalah sebagai berikut[11]:

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

Mengukur seberapa sering kata t muncul dalam dokumen d .

Vektor TF-IDF akan diperoleh dan kemudian dinormalisasi menggunakan Euclidean Norm:

$$V_{norm} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (2)$$

v adalah vektor TF-IDF, v_1, \dots, v_n adalah komponen vektor TF-IDF untuk setiap kata dalam dokumen.

2.4 Clustering menggunakan DBSCAN Algorithm

Algoritma DBSCAN (*Density-based Spatial Clustering of Application with Noise*) akan digunakan dalam proses *clustering* atau pengelompokan data dengan algoritma tersebut, kita dapat mendeteksi *outlier/noise*. Hal tersebut disebabkan konsep *density-based* yang digunakan, yaitu objek yang tidak memiliki kedekatan jarak dengan objek lainnya akan dikenali sebagai *outlier*[12]. DBSCAN adalah metode pengelompokan yang membangun area berdasarkan kepadatan yang terhubung. Setiap objek dalam radius area (*cluster*) harus berisi data minimal. Semua objek yang tidak termasuk dalam cluster dianggap sebagai noise. DBSCAN pada umumnya menggunakan perhitungan jarak Euclidean untuk mencari jarak setiap data[13]. DBSCAN dipilih karena mampu mengenali bentuk cluster yang tidak beraturan dan sulit dikenali oleh beberapa algoritma clustering lainnya. Clustering DBSCAN memiliki 2 parameter yaitu MinPts dan Epsilons. MinPts berfungsi untuk menentukan data minimal yang digunakan sebagai data awal untuk menentukan Batasan suatu cluster dan Epsilons digunakan untuk menentukan jarak antar data dalam suatu cluster. DBSCAN menggunakan fungsi Euclidean Distance untuk menentukan jarak antar item. Rumus Euclidean Distance dapat dilihat pada persamaan (3)[14]:

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2}, i = 1, \dots, n; j = 1, \dots, n, \quad (3)$$

Dimana X_{ia} merupakan variable ke- a dari objek i ($i = 1, \dots, n; a = 1, \dots, p$) dan d_{ij} merupakan nilai *Euclidean Distance*.

2.5 Silhouette Index

Kinerja algoritma dinilai menggunakan nilai *silhouette index* algoritma. Dengan mengukur kohesi intra-*cluster* dan pemisah antar *cluster*, *silhouette* membantu dalam menentukan apakah mengalokasikan titik data ke satu *cluster* daripada yang lain adalah tindakan terbaik. Tujuan dari teknik validasi *cluster* adalah untuk mengevaluasi hasil *cluster*, hasil evaluasi ini dapat digunakan untuk menentukan jumlah *cluster* dan dataset. Teknik ini memberikan representasi grafis ringkas tentang seberapa baik setiap objek diposisikan dalam klasternya. Rumus *Silhouette Index* dapat dilihat pada persamaan (4)[7]:

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

Dimana $S(i)$ merupakan rentang nilai antara -1 hingga 1, $a(i)$ adalah rata-rata jarak titik data i dengan semua titik lainnya dalam *cluster* yang sama (kohesi), $b(i)$ adalah rata-rata jarak titik data i dengan semua titik lainnya yang berbeda (pemisah).

2.6 Visualisasi Data

Visualisasi data adalah data yang direpresentasikan secara visual sehingga membuat audiens lebih mudah memahami informasi[15]. Visualisasi data digunakan untuk menampilkan data dengan berbagai

Analisis Opini User di Twitter pada Proses Pembangunan Ibu Kota Nusantara (IKN) dengan Metode DBSCAN (Ni Wayan Gauri Prasanti)

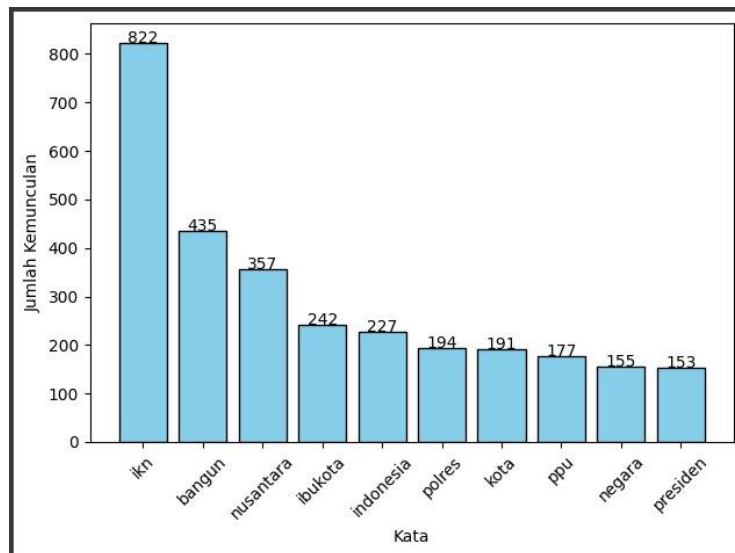
metode, salah satunya adalah dalam bentuk grafik atau bagan. Salah satu cara untuk menampilkan data visualisasi adalah dengan cara menggunakan library Matplotlib untuk menampilkan gambar tersebut agar mudah untuk dilihat.

2.7 Analisis Cluster

Analisis *cluster* adalah teknik multivariat dengan tujuan utama yaitu untuk mengelompokkan objek-objek berdasarkan karakteristiknya[16]. Analisis *cluster* dilakukan untuk menganalisis hasil pengelompokan data dengan frekuensi kata terbanyak yang muncul dari hasil *cluster*. Tahapan ini adalah tahapan terakhir untuk mengetahui kata yang paling banyak muncul.

3. Hasil dan Pembahasan

Dataset yang telah dikumpulkan dengan proses *data crawling* menggunakan *hashtag* yang berkaitan dengan IKN. Dan telah berhasil mendapatkan data sejumlah 2.011 data *tweet* berbahasa Indonesia. Dataset akan dimasukkan ke dalam algoritma *clustering* dapat menghasilkan *cluster* yang lebih optimal. Pada Gambar 2 adalah hasil dari proses *pre-processing* yang divisualisasikan dalam bentuk histogram menggunakan library Matplotlib dan data yang diambil yaitu 10 data dengan kemunculan kata terbanyak.



Gambar 2. Grafik Visualisasi Data

Setelah melakukan *preprocessing*, dilakukan proses ekstraksi fitur menggunakan TF-IDF untuk mengubah kata menjadi *vector* agar dapat mempermudah program dalam melakukan proses *clustering*. Setelah itu, dataset akan dimasukkan ke dalam program algoritma DBSCAN *Clustering*.

3.1 Analisis Hasil Pengujian

Proses *Clustering* menggunakan algoritma DBSCAN *Clustering*, akan dilakukan proses pengklasteran dengan menggunakan parameter yang berbeda untuk mencari hasil yang terbaik. Pada penelitian ini akan dilakukan percobaan untuk mencari parameter dengan menggunakan nilai Epsilons 0.1 dan MinPts dengan rentang dari 2 sampai 9. Untuk mengukur nilai evaluasi pada percobaan, maka dilakukan perhitungan dengan *silhouette index* pada setiap percobaan.

Tabel 1. *Silhouette Index*

Eps	MinPts	Silhouette	Jumlah Cluster
0.1	3	0.9	6
0.2	6	0.82	4
0.3	9	0.78	4
0.4	9	0.49	6
0.5	9	0.24	6

0.6	9	0.15	7
0.7	9	0.03	2
0.8	6	0.02	2
0.9	3	0.02	2

Hasil dari Tabel 1 menunjukkan bahwa *index* terbaik *silhouette* adalah 0.9 karena paling mendekati 1, dengan parameter Eps 0.1 dan MinPts 3, jumlah *cluster* yang dibentuk sebanyak 6 *cluster*.

Tabel 2. Frekuensi Kata Terbanyak

Label	Kata	Jumlah
-1	IKN	787
	Bangun	408
	Nusantara	347
	Ibu Kota	232
	Indonesia	214
0	Bangun	12
	IKN	11
	Target	6
	Perintah	6
	Prabowo	6
1	IKN	3
	IKN	9
2	Presiden	6
	Kantor	6
	Nusantara	6
	RI	3
3	Ibu Kota	7
	Nusantara	4
	Indonesia	4
4	IKN	2
	Budaya	10
	Fokus	5
	Sosial	5
	Proses	5
5	Akulturas	5
	Sosial	29
	Budaya	20
	Selenggara	10
	Dialog	10
	Nasional	10

Dapat dilihat pada Tabel 2, semua kata yang masuk dalam *cluster* adalah kata-kata positif. Karena kata-kata yang negatif jarang muncul sehingga tidak bisa dimasukkan ke dalam 1 kelompok.

4. Kesimpulan

Penerapan metode DBSCAN *Clustering* dalam menganalisis opini pengguna terhadap pembangunan Ibu Kota Nusantara (IKN) terbukti optimal dengan menghasilkan kelompok kata yang paling dominan. Evaluasi menggunakan *Silhouette Index* menunjukkan nilai sebesar 0.9 dengan parameter optimal Eps = 0.1 dan MinPts = 9, yang menghasilkan enam klaster, yaitu *Cluster -1*, *Cluster 0*, *Cluster 1*, *Cluster 2*, *Cluster 3*, *Cluster 4*, dan *Cluster 5*. Analisis lebih lanjut terhadap setiap klaster dilakukan dengan mengidentifikasi kata-kata yang muncul dalam masing-masing kelompok. Hasil penelitian ini menunjukkan bahwa metode DBSCAN efektif dalam mengelompokkan kata dalam dataset yang tidak beraturan, serta mampu mendeteksi *outlier* atau *noise*. Untuk penelitian selanjutnya, metode ini dapat diterapkan dalam analisis sentimen guna membandingkan kata-kata dengan konotasi positif dan negatif

Analisis Opini User di Twitter pada Proses Pembangunan Ibu Kota Nusantara (IKN) dengan Metode DBSCAN (Ni Wayan Gauri Prasanti)

dalam opini publik, sehingga memberikan wawasan lebih mendalam terhadap persepsi masyarakat terkait pembangunan IKN.

Daftar Pustaka

- [1] M. M. Wikantari, Y. Sibaroni, and A. F. Ihsan, "Clustering Content Types and User Motivation Using DBSCAN on Twitter," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 4, pp. 741–748, Aug. 2023, doi: 10.47065/josyc.v4i4.3750.
 - [2] H. Dhery, A. Assyam, and F. N. Hasan, "Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining," *Media Online*, vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.
 - [3] S. Informasi and S. S. Komputer, "Penerapan Algoritma Jaya Untuk Optimasi Pusat Klaster K-Means Batara FS Simangunsong 1) , Ni Luh Gede Pivin Suwirmayanti 2) , I Made Ari Santosa 3)," 2024.
 - [4] L. Küchenhoff *et al.*, "Extended methods for spatial cell classification with DBSCAN-CellX," *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-45190-4.
 - [5] I. Nyoman, Y. Setiawan, N. Luh, P. Suwirmayanti, N. Wayan, and A. Ulandari, "Penerapan Teknik Klasifikasi Untuk Prediksi Status Mahasiswa Menggunakan Algoritma K-Nearest Neighbour," 2023.
 - [6] K. Yogi Setiawan, H. Hidayati, A. Akbar Gozali, J. Telekomunikasi No, and D. Kolot, "Analisis User Opinion Twitter Pada Level Fine-grained Sentiment Analysis Terhadap Tokoh Publik Twitter User Opinion Analysis At Fine-grained Sentiment Analysis Level Toward Public Figure."
 - [7] S. Saitta, B. Raphael, and I. F. C. Smith, "A Bounded Index for Cluster Validity."
 - [8] G. Valkanas, A. Saravanou, and D. Gunopulos, "A Faceted Crawler for the Twitter Service."
 - [9] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," School of Computing, Telkom University, Sep. 2016, pp. 11–16. doi: 10.21108/indosc.2016.111.
 - [10] Ronen. Feldman and James. Sanger, *The text mining handbook : advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.
 - [11] R. Novia, S. S. Prasetyowati, and Y. Sibaroni, "Identify User Behavior Based on The Type of Tweet on Twitter Platform Using Gaussian Mixture Model Clustering," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 502–506, Sep. 2022, doi: 10.47065/josyc.v3i4.2208.
 - [12] M. T. Furqon and L. Muflikhah, "CLUSTERING THE POTENTIAL RISK OF TSUNAMI USING DENSITY-BASED SPATIAL CLUSTERING OF APPLICATION WITH NOISE (DBSCAN)," *Journal of Environmental Engineering & Sustainable Technology JEEEST*, vol. 03, no. 01, pp. 1–8, 2016, [Online]. Available: <http://jeest.ub.ac.id>
 - [13] Mustakim *et al.*, "DBSCAN algorithm: Twitter text clustering of trend topic pilkada pekanbaru," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1363/1/012001.
 - [14] D. Fitriah, W. Gunawan, and R. Algian Kurniাপutra, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot."
 - [15] T. Penulis *et al.*, *ANALISIS DAN VISUALISASI DATA*. [Online]. Available: www.freepik.com
 - [16] D. N. P. Sari and Y. L. Sukestiyarno, "Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 602–610, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
-