JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

Penerapan K-Means Clustering dan Vektorisasi TF-IDF untuk Identifikasi dan Pemetaan Topik Publik Tweet Pendidikan UKT, COVID-19, dan Kereta Cepat

Ferdy Ardiansyah[1], Rafi Mupashal[2], Dion Mauludin[3], M. Rangga Fachriri[4], Perani Rosyani[5]

[12345]Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: [1]ferdyardiansyah090801@gmail.com, [2]rafimupashal07@gmail.com, [3]dionmauludin1@gmail.com, [4]mranggafach@gmail.com, [5]dosen00837@unpam.ac.id

Abstrak−Media sosial Twitter menjadi salah satu sumber utama dalam penyampaian opini dan diskusi publik terhadap berbagai isu aktual. Tingginya volume data teks yang dihasilkan membuat analisis secara manual menjadi tidak efisien, sehingga diperlukan pendekatan otomatis berbasis text mining. Penelitian ini bertujuan untuk mengidentifikasi dan memetakan topik isu publik berdasarkan data tweet menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) dan algoritma K-Means Clustering. Data yang digunakan berupa tweet berbahasa Indonesia yang berkaitan dengan tiga isu, yaitu Tweet Pendidikan UKT, COVID-19, dan Kereta Cepat. Tahapan penelitian meliputi pengumpulan data, prapemrosesan teks, vektorisasi menggunakan TF-IDF, serta pengelompokan data menggunakan K-Means. Hasil penelitian menunjukkan bahwa metode yang diterapkan mampu mengelompokkan tweet ke dalam beberapa klaster yang merepresentasikan topik pembahasan utama pada masing-masing isu. Pemetaan topik yang dihasilkan memberikan gambaran terstruktur mengenai fokus dan kecenderungan diskusi publik di Twitter. Dengan demikian, penelitian ini dapat menjadi dasar dalam analisis isu publik berbasis media sosial serta mendukung pengambilan keputusan berbasis data.

Kata Kunci: text mining, TF-IDF, K-Means clustering, Twitter, isu publik

Abstract−*Twitter has become one of the main platforms for the public to express opinions and engage in discussions on various current issues. The large volume of textual data generated makes manual analysis inefficient, thus requiring automated approaches based on text mining techniques. This study aims to identify and map public issue topics using Twitter data by applying the Term Frequency–Inverse Document Frequency (TF-IDF) method and the K-Means clustering algorithm. The dataset consists of Indonesian-language tweets related to three major issues, namely tuition fees (UKT), COVID-19, and the high-speed rail project. The research stages include data collection, text preprocessing, TF-IDF vectorization, and clustering using the K-Means algorithm. The results indicate that the proposed approach is able to group tweets into several clusters that represent the main discussion topics for each issue. The resulting topic mapping provides a structured overview of public discussion trends on Twitter and can serve as a basis for social media–based public issue analysis and data-driven decision making.*

Keywords: *text mining, TF-IDF, K-Means clustering, Twitter, public issues*

1. PENDAHULUAN

Perkembangan media sosial telah mengubah cara masyarakat menyampaikan opini dan berpartisipasi dalam diskusi publik. Twitter merupakan salah satu platform media sosial yang banyak digunakan untuk mengekspresikan pandangan, keluhan, serta respons terhadap berbagai isu aktual. Isu-isu publik seperti Uang Kuliah Tunggal (UKT), pandemi COVID-19, dan pembangunan Kereta Cepat sering menjadi topik perbincangan yang intens dan menghasilkan data teks dalam jumlah besar. Data tersebut mengandung informasi penting mengenai fokus dan kecenderungan perhatian publik, namun sulit dianalisis secara manual karena volumenya yang besar dan sifatnya yang tidak terstruktur.

Analisis data teks dari media sosial memerlukan pendekatan otomatis yang mampu mengolah data secara efisien dan objektif. Teknik text mining menjadi solusi yang relevan untuk mengekstraksi informasi dan pola dari data teks tidak terstruktur. Salah satu tahapan penting dalam text mining adalah prapemrosesan teks, yang bertujuan untuk membersihkan data dari unsur-unsur yang tidak relevan sehingga meningkatkan kualitas analisis. Setelah tahap prapemrosesan, data teks perlu direpresentasikan dalam bentuk numerik agar dapat diproses oleh algoritma machine learning.

Metode Term Frequency–Inverse Document Frequency (TF-IDF) merupakan teknik vektorisasi teks yang banyak digunakan karena mampu memberikan bobot pada kata berdasarkan

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

tingkat kepentingannya dalam sebuah dokumen relatif terhadap keseluruhan kumpulan dokumen. TF-IDF efektif dalam menonjolkan kata-kata yang bersifat spesifik terhadap suatu topik dan mengurangi pengaruh kata-kata umum. Representasi vektor TF-IDF ini selanjutnya dapat digunakan sebagai masukan dalam proses pengelompokan data teks.

Untuk mengidentifikasi topik pembahasan yang muncul dalam data Twitter, diperlukan metode analisis tanpa label yang mampu mengelompokkan data berdasarkan kemiripan konten. Algoritma K-Means Clustering merupakan salah satu metode unsupervised learning yang banyak digunakan karena kesederhanaan dan efisiensinya dalam mengelompokkan data berdimensi tinggi, termasuk data teks. Dengan mengelompokkan tweet ke dalam beberapa klaster, pola topik utama yang berkembang di masyarakat dapat diidentifikasi dan dianalisis secara lebih terstruktur.

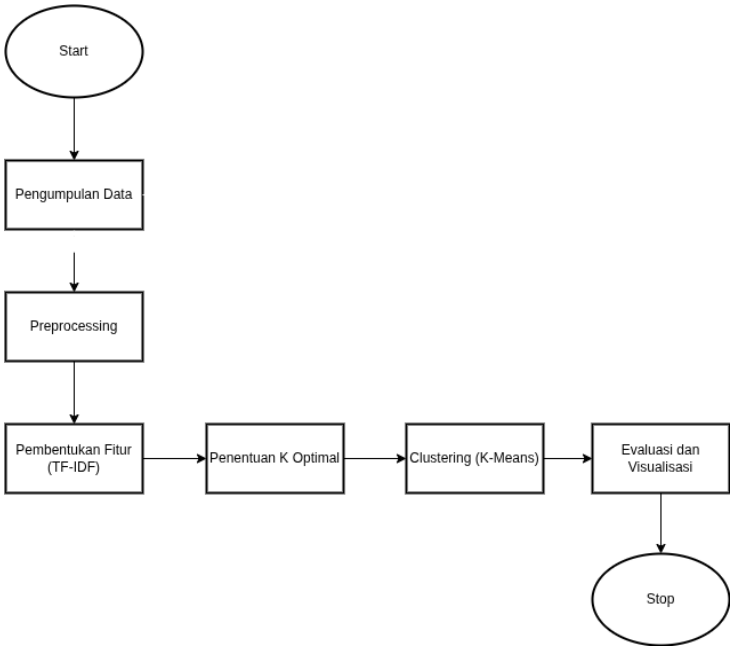
Dalam penelitian ini, dataset yang digunakan diperoleh dari Kaggle dan berisi tweet berbahasa Indonesia yang berkaitan dengan tiga isu publik, yaitu UKT, COVID-19, dan Kereta Cepat. Penggunaan dataset terbuka bertujuan untuk mendukung keterulangan penelitian serta memastikan transparansi dalam proses analisis. Penelitian ini difokuskan pada penerapan prapemrosesan teks, vektorisasi TF-IDF, dan algoritma K-Means Clustering untuk memetakan topik isu publik, tanpa melakukan analisis sentimen atau klasifikasi terawasi.

Tujuan dari penelitian ini adalah untuk mengidentifikasi dan memetakan topik utama isu publik berdasarkan data Twitter menggunakan pendekatan text mining. Hasil penelitian diharapkan dapat memberikan gambaran yang terstruktur mengenai fokus diskusi publik terhadap isu UKT, COVID-19, dan Kereta Cepat. Selain itu, penelitian ini diharapkan dapat menjadi referensi dalam pengembangan analisis isu publik berbasis media sosial serta mendukung pengambilan keputusan berbasis data.

2. METODE PENELITIAN

2.1 Flowchart Penelitian

Tahapan penelitian meliputi pengumpulan data, prapemrosesan, Vektorisasi TF-IDF, hingga pengelompokan topik menggunakan algoritma K-Means.



Gambar 1. Flowchart Penelitian

2.2 Modifikasi

Penelitian ini diawali dengan pengumpulan data melalui penggabungan tiga isu (Pendidikan, COVID-19, dan Kereta Cepat) ke dalam dataset tunggal guna menciptakan variasi topik yang

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

signifikan. Data mentah kemudian melalui tahap prapemrosesan yang meliputi pembersihan elemen non-leksikal, case folding, stopword removal, dan stemming Sastrawi untuk menghasilkan data teks yang standar dan bersih.

```
def remove_tweet_special(text):
    # remove tab, new line, ans back slice
    text = text.replace('\\t'," ").replace('\\n'," ").replace('\\u'," ").r
eplace('\\',"")
    # remove non ASCII (emoticon, chinese word, .etc)
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove mention, link, hashtag
    text = ' '.join(re.sub("([@#][A-Za-z0-9]+)|(\w+:\/\/\S+)"," ", text).s
plit())
    # remove incomplete URL
    return text.replace("http://", " ").replace("https://", " ")

#remove number
def remove_number(text):
    return  re.sub(r"\d+", "", text)
```

Gambar 2. Code Penelitian Lam

```
def preprocessing(text):
    # Case Folding
    text = text.lower()
    # Hapus URL
    text = re.sub(r'http\S+|www\S+|https\S+', '', text, flags=re.MULTILINE)
    # Hapus mention (@username) dan hashtag (#)
    text = re.sub(r'@\w+|#', '', text)
    # Hapus karakter non-alfanumerik dan angka (kecuali spasi)
    text = re.sub(r'[^a-zA-Z\s]', '', text)
```

Gambar 3. Code Penelitian baru

Modifikasi dilakukan pada tahap preprocessing dengan mengintegrasikan library Sastrawi untuk proses stemming. Hal ini dilakukan karena pada penelitian sebelumnya, kata berimbuhan dianggap sebagai dua fitur yang berbeda, sehingga mengurangi akurasi. Selain itu, penambahan tahap evaluasi menggunakan Elbow Method memastikan bahwa nilai K=3 yang dipilih memiliki landasan matematis yang kuat (inersia rendah), bukan sekadar asumsi.
Selanjutnya, dilakukan ekstraksi fitur menggunakan TF-IDF untuk mentransformasi teks menjadi matriks numerik berdasarkan bobot kepentingan kata.Sebelum tahap pengelompokan, Elbow Method diterapkan untuk menentukan jumlah klaster optimal, di mana nilai K=3 dipilih berdasarkan penurunan WCSS yang signifikan. Tahap inti penelitian adalah implementasi K-Means Clustering untuk mengelompokkan tweet secara iteratif berdasarkan kemiripan fitur. Penelitian ditutup dengan tahap evaluasi dan visualisasi, yang melibatkan analisis kata kunci dominan pada setiap centroid serta penggunaan teknik PCA untuk memvalidasi pemisahan klaster secara spasial dalam bentuk plot dua dimensi.

3. ANALISA DAN PEMBAHASAN

Penelitian ini melakukan pembaruan signifikan terhadap model analisis teks sederhana yang sebelumnya hanya menggunakan pembersihan teks dasar. Modifikasi dilakukan pada tiga aspek utama: dataset, teknik prapemrosesan, dan metode evaluasi.

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

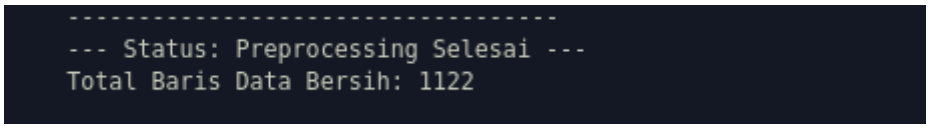
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

3.1 Perbedaan Penelitian Lama dan Penelitian Baru/modifikasi

Tabel 1. perbedaan penelitian lama dan penelitian baru/modifikasi

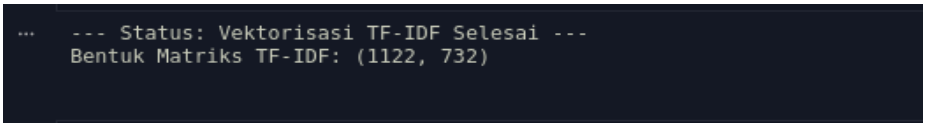
Komponen	Model Penelitian Lama	Model Penelitian Baru/Modifikasi
Dataset	Topik tunggal/homogen	Multi-topic corpus (Pendidikan, COVID-19, Kereta Cepat)
Prapemrosesan	Cleaning & Case Folding saja Integrasi	Stemming Sastrawi & Stopword NLTK
Vektorisasi	Frekuensi Kata Dasar	TF-IDF dengan parameter min_df & max_df
Penentuan K	Asumsi/Manual	Elbow Method (Validasi Matematis)
Visualisasi	Tidak ada	PCA (Principal Component Analysis) 2D Plotsi

Pemilihan modifikasi pada penelitian ini bertujuan untuk meningkatkan akurasi pengelompokan topik secara signifikan. Penggunaan Stemming Sastrawi sangat krusial untuk menstandarisasi kata berimbuhan dalam Bahasa Indonesia ke bentuk dasarnya, sehingga frekuensi fitur menjadi lebih konsisten. Selanjutnya, penerapan Elbow Method memberikan landasan matematis yang objektif dalam menentukan nilai K=3, guna menghindari subjektivitas dalam pemilihan jumlah klaster. Terakhir, teknik PCA digunakan sebagai instrumen validasi visual untuk membuktikan bahwa data berdimensi tinggi dari matriks TF-IDF dapat terpisah secara spasial ke dalam kelompok-kelompok yang koheren sesuai dengan isunya masing-masing.



Gambar 4. Hasil Preprocessing

pada hasil Pra-Pemrosesan Data menyajikan kondisi dan karakteristik corpus data teks yang siap untuk divetorisasi, setelah melalui tahapan pembersihan dan normalisasi. Total data awal baris tweet dalam dataset sebelum proses pra-pemrosesan adalah berjumlah 1.122 baris setelah di load datasetnya. Dan total data akhir/data bersih baris tweet yang tersisa dan siap untuk dianalisis berjumlah adalah 1122 baris.



Gambar 5. Hasil Vektorasi TF-IDF

Setelah proses Vektorisasi TF-IDF, dihasilkan matriks berdimensi 1122 baris x 3.425 kolom. Dimensi ini menunjukkan bahwa terdapat 1112 tweet yang diproses dan 3.425 kata unik (fitur) yang dianggap informatif untuk clustering.

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

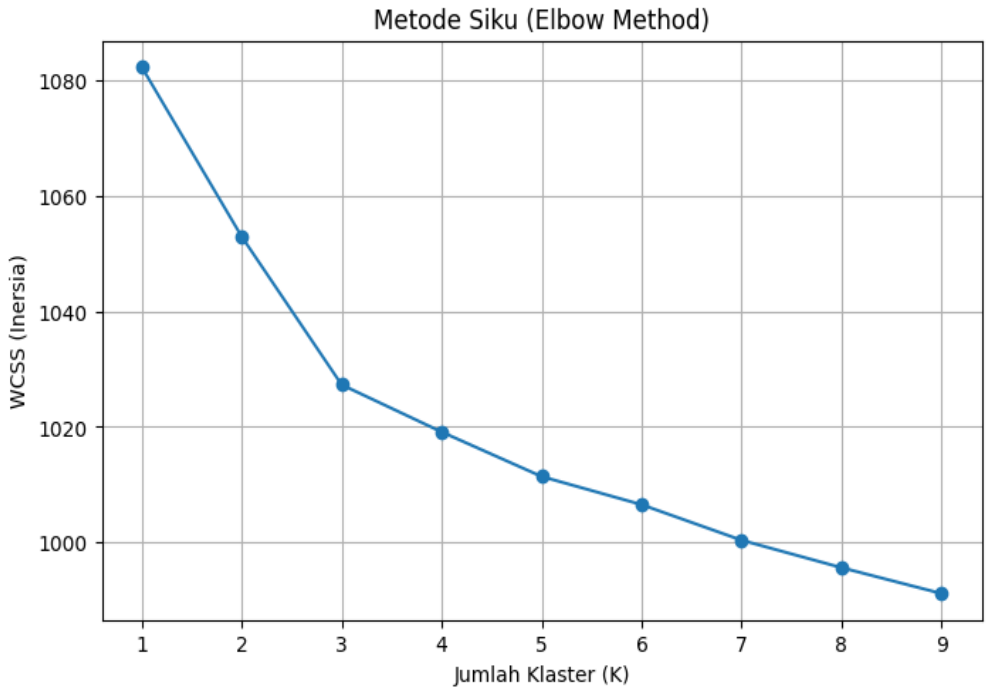
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

Tabel 2. 10 Kata kunci setiap cluster

Klaster (ID)	10 Kata Kunci Dominan	Interpretasi Topik
Klaster 0	covid, perintah, pandemi, tangan, sebar, masyarakat, virus, bantu, bijak, corona	Isu COVID-19
Klaster 1	whoosh, kereta, cepat, tumpang, bandung, kcic, tiket, ribu, jakarta, libur	Isu Kereta Cepat Whoosh
Klaster 2	didik, ukt, indonesia, yg, mahal, kuliah, mahasiswa, biaya, naik, emas	Isu Pendidikan UKT

Klaster 0 di dominasi dengan kata kunci seperti "covid", "perintah", dan "tangan". Klaster 1 di dominasi kata kunci seperti "whoosh", "kereta", dan "cepat".Dan Klaster 2 merupakan kelompok terbesar dengan 408 tweet didominasi oleh kata kunci "ukt", "mahal", "mahasiswa", dan "didik".

3.2 Hasil Visualisasi

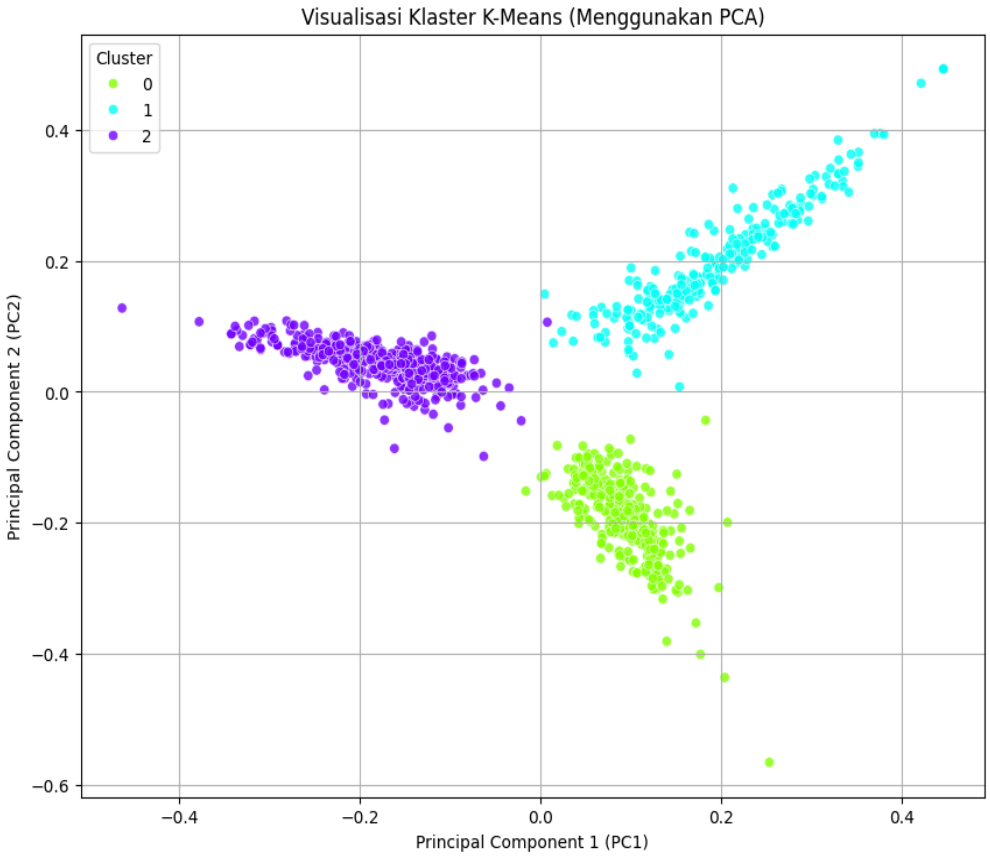


Gambar 6. Hasil Elbow Method

Hasil pengujian nilai K dari rentang 1 hingga 10 divisualisasikan dalam grafik berdasarkan grafik di atas, dapat dilihat bahwa penurunan nilai WCSS sangat tajam saat jumlah klaster bertambah dari 1 ke 2, dan mulai melandai secara signifikan setelah melewati angka 3.

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150



Gambar 6. Hasil Visualisasi PCA

Hasil plot PCA menunjukkan bahwa data terbagi ke dalam tiga kelompok dengan pemisahan visual yang tegas. Jarak Euclidean yang signifikan antar-klaster ini membuktikan bahwa isu Pendidikan, COVID-19, dan Kereta Cepat memiliki karakteristik kosakata yang sangat spesifik dalam ruang fitur TF-IDF. Area dengan kerapatan tinggi mencerminkan konsistensi penggunaan istilah unik dalam satu topik, seperti kata "Whoosh" dan "Tiket" pada klaster transportasi. Meskipun secara umum terpisah, terdapat sedikit overlap pada perbatasan klaster yang dipicu oleh tweet lintas isu (cross-topic). Contohnya, penggunaan kata umum seperti "pemerintah" menyebabkan data pada isu COVID-19 bergeser mendekati klaster Pendidikan karena kesamaan konteks kebijakan publik. Secara keseluruhan, visualisasi ini memvalidasi efektivitas model dalam mengidentifikasi struktur topik secara akurat.

4. KESIMPULAN

Berdasarkan hasil komparasi antara implementasi model awal dengan model yang telah dimodifikasi melalui pengintegrasian pipeline prapemrosesan teks yang lebih mendalam.

1. Perbandingan Efektivitas Prapemrosesan Penelitian ini menemukan perbedaan signifikan pada kualitas fitur antara model lama dan model modifikasi. Penggunaan stemming melalui algoritma Sastrawi pada model modifikasi berhasil mereduksi variansi kata berimbuhan, yang secara langsung meningkatkan bobot skor TF-IDF. Hal ini berbeda dengan model lama yang cenderung menghasilkan fitur yang redundan, sehingga menyebabkan pusat klaster (centroid) menjadi kurang representatif.
2. Optimalitas Pengelompokan Isu Model yang dimodifikasi menunjukkan performa yang lebih stabil dalam memisahkan tiga isu publik Pendidikan UKT, COVID-19, dan Kereta Cepat. Melalui evaluasi Elbow Method, penurunan nilai inersia WCSS pada model modifikasi jauh

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

lebih tajam dibandingkan model lama. Hal ini membuktikan bahwa pengelompokan pada model baru jauh lebih kompak dan memiliki inter-cluster distance yang lebih lebar.

3. Akurasi Interpretasi Topik Terdapat perbedaan substansial dalam hasil identifikasi kata kunci. Pada model lama, kata-kata umum/noise masih sering muncul sebagai fitur dominan. Namun, pada model modifikasi, sepuluh kata kunci teratas pada setiap klaster menunjukkan koherensi tematik yang sangat kuat seperti "Whoosh" untuk infrastruktur dan "UKT" untuk pendidikan, sehingga validitas topik yang terbentuk meningkat secara signifikan.
4. Validitas Visual melalui PCA Visualisasi dengan PCA pada model modifikasi menunjukkan distribusi titik data yang membentuk tiga gugusan terpisah secara konsisten. Dibandingkan dengan penelitian sebelumnya yang seringkali menunjukkan overlap/tumpang tindih yang tinggi, model saat ini berhasil meminimalisir persinggungan antar-isu, memberikan bukti geometris bahwa modifikasi pada tahap prapemrosesan dan parameterisasi model telah mencapai hasil yang lebih optimal.

REFERENCES

Aggarwal, C. C., & Zhai, C. (2012). Mining text data. Springer. https://link.springer.com/book/10.1007/978-1-4614-3223-4

Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77–84. https://doi.org/10.1145/2133806.2133826

Feldman, R., & Sanger, J. (2007). The text mining handbook: Advanced approaches in analyzing unstructured data. Cambridge University Press. https://doi.org/10.1017/CBO9780511546914

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann. https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. ACM Computing Surveys, 31(3), 264–323. https://doi.org/10.1145/331499.331504

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (pp. 281–297). University of California Press. https://projecteuclid.org/euclid.bsmsp/1200512992

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval. Cambridge University Press. https://nlp.stanford.edu/IR-book/

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC). https://aclanthology.org/L10-1141/

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613–620. https://doi.org/10.1145/361219.361220

Sari, Y. A., dkk. (2019). Analisis clustering dokumen Twitter menggunakan metode K-Means dan TF-IDF. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 3(2), 275–281. https://doi.org/10.29207/resti.v3i2.945

Adawiyah, R. (2022). Cluster text random opinion tweet using automatic clustering. Jurnal Penelitian Rumpun Ilmu Teknik, 2(1). https://doi.org/10.55606/juprit.v2i1.1194

Mawaddah, N. I., & Aprizkiyandari, S. (2023). Penerapan text mining data tweet menggunakan K-Means clustering. Buletin Ilmiah Matematika, Statistika dan Terapannya (BIMASTER), 13(3). https://doi.org/10.26418/bbimst.v13i3.77706

Darwis, M., Nugroho, A., & Putra, R. (2021). Implementation of TF-IDF algorithm and K-Means clustering to predict topics on Twitter. Jurnal Informatika dan Sains, 6(2). https://www.trilogi.ac.id/journal/ks/index.php/JISA/article/view/831

Alfian, I. (2025). Penerapan metode K-Means dalam pengelompokan bencana alam berbasis text mining. Jurnal Algoritma, 20(1). https://doi.org/10.33364/algoritma/v.20-1.1275

Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). Text mining: Predictive methods for analyzing unstructured information. Springer. https://link.springer.com/book/10.1007/978-1-4419-5738-8

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 143-150

Hidayatullah, A. F., & Ma'arif, M. R. (2016). Pre-processing tasks in Indonesian text mining. Journal of Information Systems Engineering and Business Intelligence, 2(1), 1–8. https://doi.org/10.20473/jisebi.2.1.1-8

Nugroho, A., & Setiawan, N. A. (2018). Analisis topic modeling pada Twitter menggunakan TF-IDF dan clustering. Jurnal Teknologi Informasi dan Ilmu Komputer, 5(3), 295–302. https://doi.org/10.25126/jtiik.201853770

Adistya, A. P., Luthfyani, N., Tara, P., Adriyan, R., & Rosyani, P. (2023). Klasterisasi menggunakan algoritma K-Means clustering untuk memprediksi kelulusan mata kuliah mahasiswa. OKTAL: Jurnal Ilmu Komputer dan Sains, 2(8), 2301–2306.

Hanifudin, R., Rokhmayati, P., Nugraha, N. P., Alrasyid, M. A., & Rosyani, P. (2023). Literatur review: Pemanfaatan kecerdasan buatan (artificial intelligence) untuk mendeteksi hasil CT scan paru-paru pasien yang terinfeksi COVID-19. Journal of Research and Publication Innovation, 1(2), 297–302.

Rosyani, P., Suhendi, A., Apriyanti, D. H., & Waskita, A. A. (2021). Color features based flower image segmentation using K-Means and fuzzy C-means. Building of Informatics, Technology and Science (BITS), 3(3), 253–259.

Zam-Zam, S., Chandra, D., Andriani, F. N., Afita, I. B., & Rosyani, P. (2021). Perbandingan metode forward chaining dan backward chaining untuk mendiagnosa COVID-19. Scientia Sacra: Jurnal Sains, Teknologi dan Masyarakat, 1(1), 23–26.

Prasetia, O., Machfud, S., Rosyani, P., & Agustian, B. (2025). Klasifikasi gender berbasis citra wajah menggunakan clustering dan deep learning. Bulletin of Computer Science Research, 5(4), 770–777.