

**ANALISIS TOPIK TERHADAP KOMENTAR TIKTOK  
MENGENAI UIN SUNAN AMPEL SURABAYA MENGGUNAKAN  
METODE *K-MEANS CLUSTERING* DAN *PRINCIPAL  
COMPONENT ANALYSIS***

**SKRIPSI**



**UIN SUNAN AMPEL  
S U R A B A Y A**

Disusun Oleh  
**DEWI MURTHASI'ATUL FITRIYAH**  
**09010222005**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL  
SURABAYA**

**2026**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : DEWI MURTHASI' ATUL FITRIYAH  
NIM : 09010222005  
Program Studi : Matematika  
Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "ANALISIS TOPIK TERHADAP KOMENTAR TIKTOK MENGENAI UIN SUNAN AMPEL SURABAYA MENGGUNAKAN METODE *K-MEANS CLUSTERING* DAN *PRINCIPAL COMPONENT ANALYSIS*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 05 Januari 2026

Yang menyatakan,



DEWI MURTHASI' ATUL FITRIYAH  
NIM. 09010222005

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : DEWI MURTHASI' ATUL FITRIYAH  
NIM : 09010222005  
Judul proposal skripsi : ANALISIS TOPIK TERHADAP  
KOMENTAR TIKTOK MENGENAI  
UIN SUNAN AMPEL SURABAYA  
MENGUNAKAN METODE K-  
MEANS CLUSTERING DAN PRINCIPAL  
COMPONENT ANALYSIS (PCA)

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



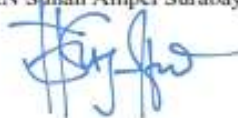
Nurissaidah Ulinnuha, M.Kom  
NIP. 199011022014032004

Pembimbing II



Aris Fanani, M.Kom  
NIP. 198701272014031002

Mengetahui,  
Ketua Program Studi Matematika  
UIN Sunan Ampel Surabaya



Dr. Yuniar Farida, M.T.  
NIP. 197905272014032002

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh:

Nama : DEWI MURTHASI' ATUL FITRIYAH

NIM : 09010222005

Judul : ANALISIS TOPIK TERHADAP KOMENTAR TIKTOK  
MENGENAI UIN SUNAN AMPEL SURABAYA  
MENGUNAKAN METODE *K-MEANS CLUSTERING*  
DAN *PRINCIPAL COMPONENT ANALYSIS*

Telah dipertahankan di depan Tim Penguji  
pada tanggal 05 Januari 2026.

Mengesahkan,  
Tim Penguji

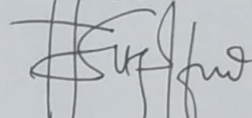
Penguji I



Wika Dianita Utami, M.Sc

NIP. 199206102018012003

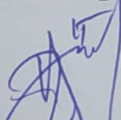
Penguji II



Dr. Yuniar Farida, M.T.

NIP. 197905272014032002

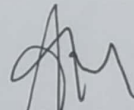
Penguji III



Nurissaidah Ulinnuha, M.Kom

NIP. 199011022014032004

Penguji IV



Aris Fanani, M.Kom

NIP. 198701272014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya



Dr. A. Saepul Hamdani, M.Pd

NIP. 196507312000031002

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : DEWI MURTHASI 'ATUL FITRIYAH  
NIM : 09010222005  
Fakultas/Jurusan : Sains dan Teknologi / Matematika  
E-mail address : dewimurthasiatulfitriyah@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)  
yang berjudul :

ANALISIS TOPIK TERHADAP KOMENTAR TIKTOK MENGENAI  
UIN SUNAN AMPEL SURABAYA MENGGUNAKAN METODE  
K-MEANS CLUSTERING DAN PRINCIPAL COMPONENT ANALYSIS

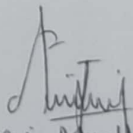
beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 13 Januari 2026

Penulis



(Dewi Murthasi Atul Fitriyah)  
nama terang dan tanda tangan

## ABSTRAK

### ANALISIS TOPIK TERHADAP KOMENTAR TIKTOK MENGENAI UIN SUNAN AMPEL SURABAYA MENGGUNAKAN METODE *K-MEANS CLUSTERING* DAN *PRINCIPAL COMPONENT ANALYSIS*

Universitas Islam Negeri Sunan Ampel (UINSA) merupakan salah satu kampus yang telah melakukan branding di media sosial termasuk pada TikTok. Tingginya interaksi pengguna pada komentar TikTok mencerminkan beragam persepsi, opini, serta harapan masyarakat terhadap kampus UINSA. Namun, komentar yang dihasilkan bersifat tidak terstruktur, berjumlah besar, dan memiliki variasi topik yang beragam, sehingga sulit dianalisis secara manual untuk mengetahui pola topik tersebut. Penelitian ini bertujuan untuk melakukan analisis topik terhadap komentar TikTok mengenai kampus UINSA. Data pada penelitian ini diambil dari komentar video TikTok terkait kampus UINSA. Data komentar tersebut akan dilakukan *pre-processing text* dan pembobotan kata dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil dari pembobotan kata tersebut akan dikelompokkan menggunakan K-Means dan *Principal Component Analysis* (PCA) untuk mereduksi dimensi. Evaluasi hasil dari *K-Means* dan PCA diukur menggunakan *Silhouette Coefficient*. Sementara itu, visualisasinya akan ditampilkan dalam bentuk *word cloud*. Penelitian ini menunjukkan bahwa algoritma *K-Means* dengan reduksi dimensi menggunakan *Principal Component Analysis* memperoleh nilai *silhouette* terbaik pada uji coba  $k = 4$  yaitu sebesar 0,055, dibandingkan dengan metode *K-Means* tanpa reduksi dimensi yaitu sebesar 0,052. Berdasarkan hasil visualisasi kluster topik komentar pada kluster 0 yaitu terkait pengalaman langsung dan opini umum masyarakat terhadap UINSA, topik pada kluster 1 yaitu do'a dan harapan diterima di UINSA, topik kluster 2 yaitu fakultas kedokteran UINSA, kluster 3 yaitu pendaftaran dan seleksi masuk UINSA.

**Kata kunci:** UINSA, TikTok, Klasterisasi, K-Means, PCA

## ABSTRACT

### TOPIC ANALYSIS OF COMMENTS TIKTOK REGARDING UIN SUNAN AMPEL SURABAYA USING K-MEANS CLUSTERING AND PRINCIPAL COMPONENT ANALYSIS METHODS

Sunan Ampel State Islamic University (UINSA) is one of the universities that has implemented branding on social media, including TikTok. The high level of user interaction in TikTok comments reflects the diverse perceptions, opinions, and expectations of the community towards UINSA. However, the comments are unstructured, large in number, and have a wide variety of topics, making it difficult to analyze manually to determine the topic patterns. This study aims to conduct a topic analysis of TikTok comments regarding UINSA. The data in this study were taken from TikTok video comments related to UINSA. The comment data will be subjected to text pre-processing and word weighting using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The results of the word weighting will be grouped using K-Means and Principal Component Analysis (PCA) to reduce dimensionality. Evaluation of the results of K-Means and PCA is measured using the Silhouette Coefficient. Meanwhile, the visualization will be displayed in the form of a word cloud. This study shows that the K-Means algorithm with dimension reduction using Principal Component Analysis obtained the best silhouette value in the  $k = 4$  trial, which was 0.055, compared to the K-Means method without dimension reduction, which was 0.052. Based on the visualization results of the comment topic clusters in cluster 0, which are related to direct experience and general public opinion towards UINSA, the topic in cluster 1 is prayer and hope for acceptance at UINSA, the topic in cluster 2 is the UINSA medical faculty, cluster 3 is registration and selection for admission to UINSA.

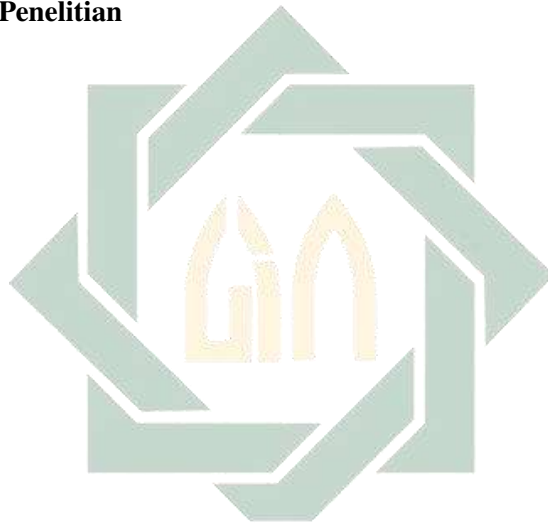
**Keywords:** UINSA, TikTok, Clustering, K-Means, PCA

## DAFTAR ISI

|                                                                          |            |
|--------------------------------------------------------------------------|------------|
| <b>HALAMAN JUDUL</b>                                                     | <b>i</b>   |
| <b>ABSTRAK</b>                                                           | <b>vi</b>  |
| <b>ABSTRACT</b>                                                          | <b>vii</b> |
| <b>DAFTAR ISI</b>                                                        | <b>1</b>   |
| <b>I PENDAHULUAN</b>                                                     | <b>4</b>   |
| 1.1. Latar Belakang Masalah . . . . .                                    | 4          |
| 1.2. Rumusan Masalah . . . . .                                           | 11         |
| 1.3. Tujuan Penelitian . . . . .                                         | 11         |
| 1.4. Manfaat Penelitian . . . . .                                        | 11         |
| 1.5. Batasan Masalah . . . . .                                           | 12         |
| 1.6. Sistematika Penulisan . . . . .                                     | 13         |
| <b>II TINJAUAN PUSTAKA</b>                                               | <b>14</b>  |
| 2.1. Kampus UINSA . . . . .                                              | 14         |
| 2.2. TikTok . . . . .                                                    | 15         |
| 2.3. <i>Text Mining</i> . . . . .                                        | 16         |
| 2.4. <i>Pre-processing Text</i> . . . . .                                | 16         |
| 2.4.1. <i>Case Folding</i> . . . . .                                     | 17         |
| 2.4.2. <i>Cleaning Text</i> . . . . .                                    | 17         |
| 2.4.3. <i>Normalisasi Slangword</i> . . . . .                            | 17         |
| 2.4.4. <i>Tokenisasi</i> . . . . .                                       | 17         |
| 2.4.5. <i>Stopword Removal</i> . . . . .                                 | 17         |
| 2.4.6. <i>Stemming</i> . . . . .                                         | 18         |
| 2.5. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> . . . . . | 18         |

|                                                              |           |
|--------------------------------------------------------------|-----------|
| 2.6. <i>Principal Component Analysis (PCA)</i> . . . . .     | 20        |
| 2.7. <i>Clustering (Klasterisasi)</i> . . . . .              | 21        |
| 2.8. <i>K-Means</i> . . . . .                                | 22        |
| 2.9. <i>Silhouette Coefficient</i> . . . . .                 | 23        |
| 2.10. <i>Word Cloud</i> . . . . .                            | 25        |
| 2.11. Integrasi Keislaman . . . . .                          | 26        |
| <b>III METODE PENELITIAN</b>                                 | <b>30</b> |
| 3.1. Jenis Penelitian . . . . .                              | 30        |
| 3.2. Sumber Data . . . . .                                   | 30        |
| 3.3. Tahapan Penelitian . . . . .                            | 30        |
| <b>IV HASIL DAN PEMBAHASAN</b>                               | <b>34</b> |
| 4.1. <i>Crawling Data TikTok</i> . . . . .                   | 34        |
| 4.2. <i>Pre-processing Text</i> . . . . .                    | 34        |
| 4.2.1. <i>Case Folding</i> . . . . .                         | 35        |
| 4.2.2. <i>Cleaning Text</i> . . . . .                        | 35        |
| 4.2.3. <i>Normalisasi Slangword</i> . . . . .                | 36        |
| 4.2.4. <i>Tokenizing</i> . . . . .                           | 36        |
| 4.2.5. <i>Stopwords</i> . . . . .                            | 37        |
| 4.2.6. <i>Stemming</i> . . . . .                             | 37        |
| 4.3. Pembobotan Kata . . . . .                               | 38        |
| 4.4. Reduksi Dimensi dengan PCA . . . . .                    | 40        |
| 4.5. Klasterisasi dengan <i>K-Means Clustering</i> . . . . . | 46        |
| 4.5.1. Uji Coba $k = 2$ . . . . .                            | 47        |
| 4.5.2. Uji Coba $k = 3$ . . . . .                            | 51        |
| 4.5.3. Uji Coba $k = 4$ . . . . .                            | 55        |
| 4.6. Evaluasi Hasil Klaster . . . . .                        | 61        |
| 4.7. Visualisasi Hasil Klaster . . . . .                     | 65        |

|                                    |           |
|------------------------------------|-----------|
| 4.8. Integrasi Keislaman . . . . . | 73        |
| <b>V PENUTUP</b>                   | <b>77</b> |
| 5.1. Kesimpulan . . . . .          | 77        |
| 5.2. Saran . . . . .               | 78        |
| <b>DAFTAR PUSTAKA</b>              | <b>78</b> |
| <b>A Sumber Data Penelitian</b>    | <b>94</b> |



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR PUSTAKA

- Abror, M. (2022). Kisah Abu Bakar Mendapat Julukan Ash-Shiddiq dan Al-'Atiq.
- Abubakar, H. D., Umar, M., and Bakale, M. A. (2022). Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec. *SLU Journal of Science and Technology*, 4(1&2):27–33.
- Adna, R. and Hayadi, B. H. (2025). Definisi dan Perbedaan antara Klasifikasi dan Klusterisasi. *Jurnal Pendidikan dan Pembelajaran Matematika Indonesia*.
- Afida, D., Udayanti, E. D., and Kartikadharma, E. (2021). Aplikasi Text Mining untuk Klusterisasi Aduan Masyarakat Kota Semarang Menggunakan Algoritma K-means. *Jurnal Transformatika*, 18(2):215–224.
- Agustina, V. and Herliana, A. (2025). Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan Text Mining dan Natural Language Processing. *Jurnal Media Informatika*, 6(3):2182–2194.
- Ahmed, M. H., Tiun, S., Omar, N., and Sani, N. S. (2023). Short Text Clustering Algorithms , Application and Challenges : A Survey. *applied sciences*, 13(1):1–38.
- Aji, B. G., Sondawa, D. C. A., Gifari, M. R., and Wijayanto, S. (2023). Penerapan Algoritma K-Means untuk Clustering Harga Rumah di Bandung. *Jurnal Ilmiah Informatika Global*, 14(2), 14(2):17–23.
- Akbar, M. N., Darmatasia, D., Mustikasari, M., and Syahwal, M. (2021). Analisis Clustering Teks Tanggapan Masyarakat di Twitter Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma K-Means. *Jurnal INSYPRO (Information System and Processing)*, 6(1):1–9.
- Aldalalah, O. M. A. A. (2022). Employment the Word Cloud in Brainstorming Via the Web and its Effectiveness in Developing the

- Design Thinking Skill. *International Journal of Instruction*, 15(1):1045–1064.
- Aldino, A. A., Darwis, D., Prastowo, A. T., and Sujana, C. (2021). Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *Journal of Physics: Conference Series*, 1751(012038).
- Aleqabie, H. J., Sfoq, M. S., Albeer, R. A., and Abd, E. H. (2024). A Review Of TextMining Techniques : Trends , and Applications In Various Domains. *Iraqi Journal for Computer Science and Mathematics*, 5(1).
- Alfawas, T. I., Rahim, A., and Rudiman, R. (2024). Penerapan Fitur Ekstraksi TF-IDF untuk Analisis Sentimen Ulasan Game Bus Simulator Indonesia dengan Algoritma Naive Bayes. *Innovative: Journal Of Social Science Research*, 4(5):3177–3193.
- Alyousf, M. and Alhalabi, M. F. (2025). A Survey of Document Stemming Algorithms in Information Retrieval Systems. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 24(4):1–28.
- Amin, A. S., Kurniadi, D., Nurzaman, M. Z., Nurfadillah, R. S., Khoerunisa, S., Khaerunisa, N., Ajiz, R. N., Jembar, T. H., and Faisal, R. N. (2024). Analisis Sentimen Pengguna Twitter dalam Pemilihan Presiden ( PILPRES ) 2024 dengan Menggunakan Algoritma K-Means. *Jurnal Algoritma*, 21(1):314–321.
- Andayani, B. T. (2022). Branding Menurut Al Quran, Sejak Kapan.
- Angelina, S. J., Negara, A. B. P., and Muhardi, H. (2023). Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia. *JUARA (Jurnal Aplikasi dan Riset Informatika)*, 02(1):165–173.
- Apriliana, T. and Widodo, E. (2023). Analisis Cluster Hierarki untuk Pengelompokan Provinsi di Indonesia berdasarkan Jumlah Base Transceiver Station dan Kekuatan Sinyal. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 3(2):286–296.

- Arunadevi, J. and Sowjanya, A. M. (2022). Recognizing Fake Headlines Using Clustering Algorithms. *Mathematical Statistician and Engineering Applications*, 71(2):111–121.
- Ashari, I. F., Banjarnahor, R., Farida, D. R., Aisyah, S. P., Dewi, A. P., and Humaya, N. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*, 6(1):07–15.
- Asher, J. and Rachmawati, E. P. (2025). Analisis Sentimen Ulasan Bintang Lima Aplikasi Instagram di Google Play Store Menggunakan IndoBERT. *DINAMIK*, 30(2):318–330.
- Bharadiya, J. P. (2023). A Tutorial on Principal Component Analysis for Dimensionality Reduction in Machine Learning. *International Journal of Innovative Research in Science Engineering and Technology*, 8(5):2028–2032.
- Bur, R., Ayuningtyas, F., and Muqsith, M. A. (2023). Pemanfaatan TikTok Sebagai Media Informasi Baru Generasi Z. *Jurnal Komunikasi Nusantara*, 5(2):189–198.
- Butar, R. P. B. (2023). Analisis Hierarchical dan Non-Hierarchical Clustering Untuk Pengelompokan Potensi Ekonomi Kelautan Indonesia 2021. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 11(3):543–553.
- Chairudin, M. and Widodo, L. (2024). Transformasi dan Inovasi Perguruan Tinggi Islam Menjadi Universitas Kelas Dunia. *DAARUS TSAQOFAH Jurnal Pendidikan Pascasarjana Universitas Qomaruddin*, 1(2):146–155.
- Faizin, M. (2023). 6 Etika Berbicara dalam Ayat-ayat Al-Qur'an.
- Fani, S. M., Santoso, R., and Suparti, S. (2021). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Akun Bibli pada Media Sosial Twitter Menggunakan K-Means Clustering. *Jurnal Gaussian*, 10(4):583–593.
- Farida, Y., Khariri, A. F., Yuliati, D., and Khaulasari, H. (2022). Clustering Couples of Childbearing Age to Get Family Planning Counseling using

- K-Means Method. *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, 22(1):189–200.
- Guntara, M. and Lutfi, N. (2023). Cacah Klaster pada Klasterisasi dengan Algoritma KMeans Menggunakan Silhouette Coefficient dan Elbow Method. *JuTI "Jurnal Teknologi Informasi"*, 2(1):43–52.
- Gyamerah, S., Soori, G. T., Korda, D. R., Tawiah, J. K., Akolgo, E. A., and Dapaah, E. O. (2023). Comparative Analysis of Feature Extraction of High Dimensional Data Reduction Using Machine Learning Techniques. *American Journal of Electrical and Computer Engineering*, 7(2):27–39.
- Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing pada Data Mining Dengan Menggunakan Machine Learning. *JBASE - Journal of Business and Audit Information Systems*, 4(2):16–22.
- Haq, M. A., Purnomo, W., and Setiawan, N. Y. (2023). Analisis Clustering Topik Survey menggunakan Algoritme K-Means (Studi Kasus: Kudata). *Teknologi Informasi dan Ilmu ...*, 7(7):3498–3506.
- Haryanto, M., Sidauruk, A. C., Hendy, Y. B., Sabailaket, J. A., Purba, D. R., and Handoyo, E. R. (2023). Pengaruh Penggunaan Aplikasi TikTok terhadap Produktivitas Belajar Mahasiswa di Yogyakarta. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 3(2):330–341.
- Hasan, B. M. S. and Abdulazeez, A. M. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, 2(1):20–30.
- Hasanah, N. N. and Purnomo, A. S. (2022). Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus : Perpustakaan Politeknik LPP Yogyakarta). *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 4(2):300–311.
- Hasanah, S. N. (2024). Perbandingan Metode Clustering pada Kelompok Dokumen Teks untuk Menentukan Nilai Ambang Batas Similaritas. *Nusantara Journal of Multidisciplinary Science*, 1(8):611–621.

Hatami, M. H. (2025). Pendidikan Qur'ani: Kajian Tafsir QS. Al-Qashash Ayat 77 terhadap Nilai dan Prinsip. *Al-Fahmu: Jurnal Ilmu Al-Qur'an dan Tafsir*, 4(1):1–13.

Hayati, N. (2025). UINSA Resmi Miliki Fakultas Kedokteran.

Hediyati, D. and Suartana, I. M. (2021). Application of Principal Component Analysis (PCA) for Dimension Reduction in the Clustering Process of Agricultural Production Data in Kab Bojonegoro. *JIEET (Journal of Information Engineering and Educational Technology)*, 5(2):49–54.

Herman, H., Siahaan, A. M., Andy, A., and Aryanto, D. (2024). Implementasi Algoritma Rabin Karp dan Optimasi dengan Algoritma Stemmer Sastrawi dalam Deteksi Plagiat pada Jurnal Skripsi Mahasiswa. *Jurnal Armada Informatika*, 8(1):23–31.

Hilmi, N. F. and Irwiensyah, F. (2024). Analisis Sentimen Terhadap Aplikasi Tiktok dari Ulasan pada Google Playstore Menggunakan Metode Naïve Bayes. *SMATIKA Jurnal : STIKI Informatika Jurnal*, 14(1):146–156.

Indah, Y. M. and Octaviana, S. (2025). Analisis Perbandingan Metode Klaster Hierarki dan Non-Hierarki Terhadap Tingkat Pengangguran di Pulau Jawa Tahun 2023. *BIAStatistics: Journal Of Statistics Theory and Applications*, 19(1):92–104.

Indriyani, F. A., Fauzi, A., and Faisal, S. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, 10(2):176–184.

Jennifer, D. P., Ningrum, N. C., Pinasty, S., Edma, S. M. N., Andini, W. N., and Widodo, E. (2024). Analisis Sentimen pada Pengguna Tiktok Menggunakan Metode Random Forest ( Studi Kasus : Jessica-Mirna ). *Innovative: Journal Of Social Science Research*, 4(3):14477–14489.

Kamil, M. I. and Wibowo, A. P. (2024). Implementation of the Naive Bayes

- Classifier Algorithm for Classifying Toddler Nutritional Status. *Journal of Applied Informatics and Computing*, 8(2):421–427.
- Kammoun, A., Ravier, P., and Buttelli, O. (2024). Impact of PCA Pre-Normalization Methods on Ground Reaction Force Estimation Accuracy. *Sensors*, 24(4):1–17.
- Karimah, A. and Fitriana, I. N. L. (2025). Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Pembangunan Literasi Masyarakat 2023 menggunakan Metode Klaster Hierarki dan non Hierarki. *Prosiding Seminar Nasional Sains dan Teknologi "SainTek"*, 2(1):394–407.
- Karo, I. M. K., Dewi, S., Ramadhani, F., and Harliana, P. (2023). K-means and K-medoids Algorithm Comparison for Clustering Forest Fire Location in Indonesia. *Jurnal Ecotipe (Electronic Control Telecommunication Information and Power Engineering)*, 10(1):86–94.
- Ke, Z. T., Ji, P., Jin, J., and Li, W. (2024). Recent Advances in Text Analysis. *Annual Review of Statistics and Its Application*, 11(1):347–372.
- Kementerian Komunikasi dan Digital (2024). Infrastruktur Digital – Fondasi Akselerasi Inovasi dan Pertumbuhan Ekonomi.
- Kemp, S. (2025a). Digital 2026: 2 in 3 People on Earth Now Use Social Media.
- Kemp, S. (2025b). TikTok Users, Stats, Data & Trends for 2025.
- Khairuna, R., Nurdin, N., and Razi, A. (2025). Analisis Perbandingan Algoritma K-Means dan K-Medoids untuk Klusterisasi Teks Ulasan pada Aplikasi Cookpad. *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, 10(2):445–458.
- Kholidah, I. A. and Indriyanti, A. D. (2025). Perbandingan Penerapan Text Mining pada Aplikasi Wargaku Surabaya untuk Aduan Masyarakat Terhadap Satpol Pp Kota Surabaya dengan Algoritma K-Means Clustering dan Fuzzy C-Means Clustering. *Journal of Informatics and Computer Science (JINACS)*, 06(04):1180–1189.
- Kurniawan, A. (2020). Tafsir Surat Al-Baqarah Ayat 31.

- Kusumandaru, A. D. and Rahmawati, F. P. (2022). Implementasi Media Sosial Aplikasi Tik Tok sebagai Media Meningkatkan Literasi Sastra dalam Pembelajaran Tematik di Sekolah Dasar. *Jurnal basicedu*, 6(3):4876–4886.
- Laila, K. N. and Mashuri, M. M. (2024). Konsep Keseimbangan Hidup Sebagai Solusi Burnout dalam Al-Qur'an (Analisis Deskriptif Qs. Al-Qashash Ayat 77). *Jurnal Mafhum*, Vol. 9(No. 2):8–15.
- Laksana, J. P., Shela, S., Irsyad, H., and Rahman, A. (2025). Analisis Topik Dominan dalam Paper Ilmu Komputer menggunakan TF-IDF dan K-Means. *Buletin Ilmiah dan Informatika Teknologi*, 3(3):78–84.
- Lusiana, N. L., Fatimah, Z. F., Muna, S., and Rahmawati, A. (2024). Keseimbangan Hidup Dalam Al-Qur'an: Telaah Tafsir Tarbawiy. *MUSHAF JOURNAL : Jurnal Ilmu Al Quran dan Hadis*, 4(3):443–449.
- Mahendra, M. H., Murdiansyah, D. T., and Lhaksana, K. M. (2023). Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer. *DIKE : Jurnal Ilmu Multidisiplin*, 1(2):37–43.
- Mana, L. H. A. (2021). Respon Siswa Terhadap Aplikasi Tiktok Sebagai Media Pembelajaran Bahasa Indonesia. *Jurnal Inovasi dan Riset Akademik*, 2(4):418–429.
- Manuaba, I. B. N. W., Dantes, G. R., and Indrawan, G. (2022). Analisis Sentimen Data Provider Layanan Internet Pada Twitter Menggunakan Support Vector Machine dengan Penambahan Algoritma Levenshtein Distance. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 5(2):9–17.
- Margaretha, A. P., Ulinnuha, N., and Intan, P. K. (2024). Clustering Data Kecelakaan Lalu Lintas melalui Algoritma K-Means dengan Seleksi Fitur Chi-Square. *INTEGER: Journal of Information Technology*, 10(2):215–224.
- Masduki, M., Prihartini, E., and Abdullah, D. (2023). Analisis Faktor-Faktor yang Mempengaruhi Keputusan Kuliah di Perguruan Tinggi

- Swasta. *Entrepreneur: Jurnal Bisnis Manajemen dan Kewirausahaan*, 4(1):205–219.
- Melati, R. and Islami, M. R. R. (2024). Analisis Sentimen Data Twitter Menggunakan Metode K-Means Clustering pada Studi Kasus Pemindahan Ibu Kota Nusantara (IKN). *Technology Acceptance Model) Jurnal TAM*, 15(1):66–73.
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., and Foiadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, 11(November):119596–119633.
- Momahhed, S. S., Sefiddashti, S. E., Minaei, B., and Shahali, Z. (2023). K-Means Clustering of Outpatient Prescription Claims for Health Insureds in Iran. *BMC Public Health*, 23(788):1–15.
- Moseley, B. and Wang, J. R. (2023). Approximation Bounds for Hierarchical Clustering : Average Linkage , Bisecting K-means , and Local Search. *Journal of Machine Learning Research*, 24(1):1–36.
- Mostafa, G., Ahamed, M. S., and Morol, M. K. (2024). Improve the Sentiment of Bengali Language Texts with Stopword Removal. *International Conference on Computing Advancements (ICCA)*, 17(18):414–422.
- MR, N. (2023). UINSA Surabaya, Filosofi Twin-Towers dan Integrasi Nalar Islami dengan Keilmuan Masa Kini.
- Munthe, S. R., Suryadi, S., and Laksono, F. (2024). Analisis Klasifikasi Teks pada Kata Slang di Media Sosial Menggunakan Pengolahan Bahasa Alami untuk Trending Topik. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 5(1):230–241.
- Mustofa, B. A. and Saptomo, W. L. Y. (2025). Use of Natural Language Processing in Social Media Text Analysis. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2):1235–1238.
- Naeem, S., Ali, A., Anam, S., and Ahmed, M. M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*, 13(1):911–921.

- Nafi', A., Harjanta, A. T., Herlambang, B. A., and Fahmi, S. (2024). Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna. *J-Intech*, 12(02):330–339.
- Nahdhudin, M., Wibowo, N. C. H., Handayani, M. R., and Umam, K. (2025). Klasifikasi Sentimen Masyarakat Terhadap Aplikasi Tiktok Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 1(3):107–112.
- Novanto, A., Indra, D., and Astuti, W. (2024). Analisis Pre-processing Sentimen Terhadap Komentar Layanan Indihome Pada Twitter. *Buletin Sistem Informasi dan Teknologi Islam*, 5(1):30–36.
- NU Online (2025a). Surat Al-Hujurat Ayat 6.
- NU Online (2025b). Surat Al-Qashash Ayat 77: Arab, Latin, Terjemah dan Tafsir.
- NU Online (2025c). Surat An-Nahl Ayat 125: Arab, Latin, Terjemah dan Tafsir.
- NU Online (2025d). Surat Ar-Ra'd Ayat 11: Arab, Latin, Terjemah dan Tafsir Lengkap.
- Nugraha, C. A., Kesuma, M. A., Cahyani, O. I., Wati, M., and Haviluddin, H. (2025). Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means. *JUKI : Jurnal Komputer dan Informatika*, 7(1):80–88.
- Nur (2023). Tiga Pesan Rektor Jadi Kunci Pengelolaan Media Uinsa.
- Nurhalizah, R. S., Ardianto, R., and Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning : Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1):61–72.
- Nurputra, W. R., Witanti, W., and Komarudin, A. (2025). Principal Component Analysis (PCA) untuk Meningkatkan Hasil Klasterisasi

- Penjualan Video Game Menggunakan Algoritma K-Means. *JURNAL LOCUS: Penelitian dan Pengabdian*, 4(8):4683–4703.
- Oktaviyanti, A., Firmansyah, F., and Kadafi, A. R. (2025). Analisis Sentimen Ulasan Pelecehan di Media Sosial Tiktok menggunakan SVM dan Pengelompokan Pola Kejadian dengan K-Means Clustering. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1):1938–1945.
- Olabanjo, O., Wusu, A., Afisi, O., Asokere, M., Padonu, R., Olabanjo, O., Ojo, O., Folorunso, O., Aribisala, B., and Mazzara, M. (2023). From Twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election. *Heliyon*, 9(5):e16085.
- Pardede, J. and Darmawan, D. (2025). Perbandingan Algoritma Stemming Porter, Sastrawi, Idris, Dan Arifin & Setiono Pada Dokumen Teks Bahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 12(1):69–76.
- Pardianti, M. S. and Valiant, V. (2022). Pengelolaan Konten Tiktok Sebagai Media Informasi. *Ikon –Jurnal Ilmiah Ilmu Komunikasi*, 27(2):187–210.
- Permana, E., Nury, N., Oktavia, S., and Firdaus, R. Z. (2025). Pemanfaatan TikTok sebagai Media Pemasaran Brand Fashion Arkline untuk Menarik Konsumen Gen Z. *MENAWAN : Jurnal Riset Dan Publikasi Ilmu Ekonomi*, 3(4):21–35.
- Prasetyo, T., Waskita, A. A., and Taryo, T. (2025). Analisis Sentimen Pengguna Seputar Kendaraan Listrik di Twitter dengan Penerapan Algoritma Naïve Bayes, KNN, dan Decision Tree untuk Klasifikasi. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, VIII(2):108–117.
- Prastyo, P. A., Berlilana, B., and Tahyudin, I. (2025). Sentiment Analysis on Slang Enriched Texts Using Machine Learning Approaches. *Journal of Applied Data Sciences*, 6(2):1076–1087.
- Prayoga, Y. (2025). Khutbah Jumat: Maulid Nabi dan Kejujuran Rasulullah dalam Berdagang.

- Priantiwi, T. N. and Abdurrahman, M. (2023). Analisis Konten Pembelajaran Bahasa Arab pada Media Tiktok. *Jurnal Ilmiah Profesi Pendidikan*, 8(3):1365–1371.
- Pulungan, Z. and Dalimunthe, S. S. (2023). Sejarah Berdirinya Perguruan Tinggi Islam di Indonesia. *AL-USWAH: Jurnal Riset dan Kajian Pendidikan Agama Islam*, 6(1):57–67.
- Punhani, A., Faujdar, N., Mishra, K. K., and Subramanian, M. (2022). Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means. *IEEE Access*, 10(September):115025–115032.
- Purnajaya, A. R., Lieputra, V., Tayanto, V., and Salim, J. G. (2022). Implementasi Text Mining untuk Mengetahui Opini Masyarakat Tentang Climate Change. *Journal of Information System and Technology*, 3(3):36–44.
- Purwo, R. H. S. and Puspasari, D. (2020). Peran Humas dalam Meningkatkan Citra Positif pada Badan Pengembangan Sumber Daya Manusia (BPSDM) Provinsi Jawa Timur. *Jurnal Pendidikan Administrasi Perkantoran (JPAP)*, 8(3):458–467.
- Putra, B. A., Mukhtar, H., Bangun, E. T. B., Gusnanda, A., Maisyarah, A., Kurniawan, M. I., Pradipa, R., and Ali, Z. M. (2025). Optimisasi Algoritma K-Means Dengan Metode Reduksi Dimensi Untuk Pengelompokan Big Data Dalam Arsitektur Cloud Computing. *Jurnal Software Engineering and Information System (SEIS)*, 5(1):1–8.
- Rahardaya, A. K. and Irwansyah, I. (2021). Studi Literatur Penggunaan Media Sosial Tiktok Sebagai Sarana Literasi Digital pada Masa Pandemi Covid-19. *Jurnal Teknologi dan Informasi Bisnis*, 3(2):308–319.
- Rahayu, S. C. A., DwiYansaputra, R., and Husodo, A. Y. (2025). Klustering Topik pada Kolom Komentar Instagram Tentang Kabinet Merah Putih Menggunakan Metode K-Means. *Jurnal Teknologi Informasi, Komputer dan Aplikasinya (JTika)*, 7(1):90–97.

- Rahmadani, A. and Nursyahira, N. (2025). Implementation of the K-Means Algorithm for Inventory Data Clustering. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 5(1):1–11.
- Rahmayani, M., Ramdhani, M., and Lubis, F. O. (2021). Pengaruh Penggunaan Aplikasi Tiktok Terhadap Perilaku Kecanduan Mahasiswa. *Syntax Literate: Jurnal Ilmiah Indonesia*, 6(7):3327–3343.
- Ramadhan, R. F., Wijoyo, S. H., and Saputra, M. C. (2023). Penerapan Metode K-Means Clustering pada Ulasan Perumahan PT XYZ di Google Maps untuk Formulasi Strategi Bisnis dengan Analisis SWOT. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7(6):2879–2888.
- Ramdani, N. S., Hadiapurwa, A., and Nugraha, H. (2021). Potensi Pemanfaatan Media Sosial TikTok sebagai Media Pembelajaran dalam Pembelajaran Daring. *Akademika*, 10(2):425–435.
- Ramadhan, D., Dwilestari, G., Dana, R. D., Ajiz, A., and Kaslani, K. (2022). Clustering Data Persediaan Barang dengan Menggunakan Metode. *MEANS (Media Informasi Analisa dan Sistem)*, 7(1):1–9.
- Ramli, R. and Rama, B. (2024). Perkembangan Perguruan Tinggi Islam di Indonesia. *JUPEIS : Jurnal Pendidikan dan Ilmu Sosial*, 3(1):42–54.
- Rifaldi, D., Fadlil, A., and Herman, H. (2023). Teknik Preprocessing pada Text Mining Menggunakan Data Tweet “Mental Health”. *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2):161–171.
- Rifki, M. and Imelda, I. (2022). Analisis Sentimen Wacana Kenaikan Harga Tiket Candi Borobudur Menggunakan Multinomial Naïve Bayes. *Jiko (Jurnal Informatika dan Komputer)*, 5(2):156–163.
- Rinandyaswara, R., Sari, Y. A., and Furqon, M. T. (2022). Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(4):717–724.

- Ripa'i, A., Santoso, F., and Lazim, F. (2024). Deteksi Berita Hoax dengan Perbandingan Website Menggunakan Pendekatan Deep Learning Algoritma BERT. *G-Tech : Jurnal Teknologi Terapan*, 8(3):1749–1758.
- Rivaldi, R. C. and Wismarini, T. (2024). Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) (Studi Kasus Zalika Store 88 Shopee). *ELKOM:Jurnal Elektronika dan Komputer*, 17(1):120–128.
- Sahrizal (2025). Perguruan Tinggi Islam dan Masa Depan Indonesia. *Mauriduna: Journal of Islamic Studies*, 6(1):614–624.
- Salsabila, S., Sholeh, M. H., Khaulasari, H., Novitasari, D. C. R., and D.N, Y. (2024). Implementasi K-Means Clustering Dalam Pemetaan Wilayah Rawan Penyakit Leptospirosis Di Jawa Timur. *Jurnal Riset dan Aplikasi Matematika (JRAM)*, 08(02):189–200.
- Santosa, S. and Oktavianti, R. (2023). Berpikir Kreatif Humas di Media Sosial Instagram dalam Membangun Citra Positif Perguruan Tinggi (Studi Kasus pada Akun @Untarjakarta). *Prologia*, 7(1):71–80.
- Santoso, D. P. and Wibowo, W. (2022). Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store. *Jurnal Sains dan Seni ITS*, 11(2):190–196.
- Setiana, E., Marwondo, M., Danestiara, V. R., and Wiyanudin, W. (2023). Analisis Sentimen Pelaksanaan Kuliah Online Menggunakan Algoritma Support Vector Machine. *Nuansa Informatika*, 17(2):66–70.
- Shalih, F. A., Ramadhan, R. A., and Syalaisa, N. (2025). Comprehensive Overview of Principal Component Analysis Applications and Developments. *Jurnal EurekaMatika*, 13(1):25–34.
- Siddiq, D. M., Kencana, R. A. L., Haidar, A., and Sholehah, R. (2025). Pemanfaatan Media Sosial Tiktok Sebagai Media Komunikasi Humas Perguruan Tinggi Uninus. *Journal of Digital Communication and Public Relations*, 1(1):1–12.

- Siregar, M. Z., Elhanafi, A. M., and Irwan, D. (2025). Analisis Sentimen Terhadap Ulasan Aplikasi Media Sosial di Google Play Menggunakan Algoritma Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2):3373–3381.
- Suhaeni, C., Kamila, S. A., Fahira, F., Yusran, M., and Alfa Dito, G. (2025). Exploring a Large Language Model on the ChatGPT Platform for Indonesian Text Preprocessing Tasks. *Indonesian Journal of Statistics and Its Applications*, 9(1):100–116.
- Sukriadi, S., Ismail, I., and Andzar, A. M. (2023). Penerapan Text Mining dalam Klasifikasi Judul Skripsi yang Diusulkan Mahasiswa Menggunakan Metode Naïve Bayes. *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI)*, 6(2):184–196.
- Sunnatullah, S. (2023). Meneladani Para Sahabat Nabi: Tetap Semangat Belajar di Usia Tua.
- Susilowati, S., Armelsa, D., and Duneti, D. (2022). Strategi Hubungan Masyarakat Pascasarjana Sekolah Tinggi Ilmu Administrasi Mandala Indonesia (STIAMI) dalam Membentuk Citra Positif. *Jurnal Public Relations-JPR*, 3(1):44–49.
- Syaibani, I. A. and Zainiyati, H. S. (2022). Penggunaan aplikasi Tik Tok sebagai Media Pembelajaran SKI pada Masa Pandemi Covid-19 di MI Miftahussudur 01 Dagangan. *JURNAL LENTERA : Kajian Keagamaan, Keilmuan Dan Teknologi*, 21(1):79–85.
- Thenata, A. P. (2021). Text Mining Literature Review on Indonesian Social Media. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 7(2):226–232.
- UIN Sunan Ampel Surabaya (2022). *Pedoman Akademik UIN Sunan Ampel Surabaya*. UIN Sunan Ampel Surabaya, 1 edition.
- UIN Sunan Ampel Surabaya (2025). *Sejarah UINSA*.
- Ulummuddin, I., Sari, A. P., and Swari, M. H. P. (2024). Pemanfaatan Data Ulasan Pengguna untuk Membangun Sistem Klasterisasi berdasarkan

- Pain Points menggunakan Algoritma K-Means. *Jurnal Teknologi Terpadu*, 10(1):70–76.
- Wani, A. A. (2024). Comprehensive Analysis of Clustering Algorithms : Exploring Limitations and Innovative Solutions. *PeerJ Computer Science*, 10(e2286):1–45.
- Wijaya, H. W. and Hayati, N. (2025). Natural Language Processing (NLP) for Sentiment Analysis of Seblak Bandung Pedas Kudus Reviews. *JBASE - Journal of Business and Audit Information Systems*, 8(1).
- Wikansari, D. and Setyanto, Y. (2023). Humas dalam Meningkatkan Citra Perguruan Tinggi Melalui Media Sosial Instagram. *Kiwari*, 2(2):243–250.
- Wisesty, U. N. and Mengko, T. R. (2021). Comparison of Dimensionality Reduction and Clustering Methods for SARS-CoV-2 Genome. *Bulletin of Electrical Engineering and Informatics*, 10(4):2170–2180.
- Yanto, R., Alfianini, A., Kesuma, H. D., Afrudi, S., Juminovario, J., and Ahmadi, A. (2025). Implementasi Text Mining Dalam Mengidentifikasi Similarity Judul Penelitian. *Djtechno : Jurnal Teknologi Informasi*, 6(1):237–251.
- Zhou, H. (2022). Research of Text Classification Based on TF-IDF and CNN-LSTM. *Journal of Physics: Conference Series*, 2171:1–8.