

# Analisis Topik Konten Dakwah pada Kanal YouTube Indonesia Menggunakan BERTopic

Zanzibar<sup>1</sup>, Anni Zulfia<sup>2</sup>

<sup>1</sup> Program Studi Komunikasi dan Penyiaran Islam, Fakultas Ushuluddin Adab dan Dakwah, Universitas Islam Negeri Sultanah Nahrasiyah Lhokseumawe.

<sup>2</sup> Program Studi Teknik Informatika, Jurusan Teknik Informatika, Fakultas Teknik, Universitas Malikussaleh  
Email: zanzibar@uinsuna.ac.id, annizulfia@unimal.ac.id

## ABSTRAK

Perkembangan teknologi informasi telah mendisrupsi penyampaian pesan agama, menjadikan YouTube medium utama dakwah digital di Indonesia. Seiring peningkatan eksponensial jumlah konten, pemetaan tema secara manual menjadi tidak efisien. Penelitian ini bertujuan mengidentifikasi, mengelompokkan, dan mendeskripsikan kecenderungan topik konten dakwah pada kanal YouTube Indonesia menggunakan pendekatan *Natural Language Processing* (NLP). Pemodelan topik dilakukan menggunakan BERTopic yang memanfaatkan *transformer-based embeddings*. Dataset mencakup 1.250 dokumen transkrip dari 10 kanal dakwah terkemuka periode 2022–2026. Hasil menunjukkan terdapat 12 topik utama, di mana topik "Akhlak Muslim" paling dominan (17,4%), diikuti "Fikih Ibadah" (14,8%), dan "Pendidikan Anak" (11,8%), sedangkan "Politik dan Umat" terendah (2,2%). Kesimpulannya, konten dakwah digital di Indonesia berorientasi kuat pada pembinaan karakter. BERTopic secara efektif mengekstraksi kluster semantik koheren dari teks berbahasa Indonesia, memberikan wawasan pengembangan strategi Komunikasi dan Penyiaran Islam era digital. Abstrak ini telah disesuaikan agar informatif, mandiri, dan memiliki panjang dalam rentang yang disyaratkan.

Kata kunci: BERTopic, dakwah digital, *natural language processing*, *topic modeling*, YouTube

Penulis koresponden : Zanzibar

Tanggal terbit : 15 Juni 2026

Tautan : <https://jurnal.komputasi.org/index.php/jst/article/view/44>

## 1. PENDAHULUAN

Transformasi digital telah membawa perubahan fundamental dalam penyebaran nilai-nilai agama, dengan berekspansinya ruang dakwah ke platform virtual seperti YouTube [1]. Meningkatnya jumlah konten dakwah secara eksponensial di Indonesia menciptakan fenomena limpahan informasi. Audiens, peneliti, maupun praktisi dakwah kini dihadapkan pada kesulitan untuk memetakan dan mengidentifikasi tema-tema kajian yang sedang berkembang jika dilakukan secara manual [2]. Latar belakang ini menuntut adanya otomatisasi komputasional untuk menganalisis teks dalam skala masif.

Pernyataan masalah utama dalam penelitian ini adalah sulitnya mengidentifikasi secara pasti apa saja topik utama yang dibahas, topik apa yang paling dominan, serta bagaimana distribusi dan relasi antar topik dalam lanskap dakwah YouTube Indonesia. Analisis komprehensif sangat dibutuhkan untuk memahami orientasi dan kecenderungan konsumsi informasi religius masyarakat saat ini [3].

Tinjauan literatur relevan menunjukkan bahwa sebagian besar penelitian terdahulu mengenai *topic modeling* pada media sosial masih bergantung pada algoritma konvensional seperti *Latent Dirichlet Allocation* (LDA) [4]. Dalam konteks dakwah, pendekatan yang sering digunakan masih didominasi oleh metode kualitatif pada sampel yang sangat kecil [5]. LDA seringkali memiliki keterbatasan dalam memahami konteks polisemi dan struktur semantik kalimat yang utuh [6].

Sebagai pendekatan dan solusi yang diusulkan, penelitian ini menerapkan algoritma BERTopic. BERTopic adalah metode *topic modeling* modern yang memanfaatkan model bahasa berbasis *transformer* (*Sentence Transformers*) untuk menghasilkan representasi kluster yang jauh lebih kaya dan kontekstual pada teks berbahasa Indonesia [7]. Pendekatan ini memungkinkan pemrosesan transkrip video lisan (*spoken language*) dengan akurasi yang lebih tinggi dibandingkan metode tradisional.

Nilai baru atau inovasi dari penelitian ini terletak pada pemanfaatan *state-of-the-art NLP* secara spesifik pada korpus transkrip dakwah berskala besar. Kajian literatur yang telah dilakukan penulis digunakan untuk menjelaskan perbedaan naskah ini dengan makalah lain. Hasil dari pemetaan data ini diharapkan mampu memberikan kontribusi lintas disiplin, baik bagi pengembangan ilmu Informatika maupun strategi pengembangan konten pada bidang Komunikasi dan Penyiaran Islam [8].

## 2. METODELOGI

Penelitian ini menggunakan pendekatan *Text Mining* kuantitatif yang dijalankan melalui serangkaian tahapan secara kronologis. Pertama, pengumpulan data dilakukan menggunakan YouTube Data API v3 untuk mengunduh metadata dan mengekstraksi transkrip otomatis dari 10 kanal dakwah di Indonesia pada periode 2022–2026. Data yang terkumpul berjumlah 1.250 dokumen transkrip teks [9].

Prosedur pemrosesan dilanjutkan dengan tahapan *preprocessing*. Teks mentah dibersihkan melalui proses *case folding*, penghapusan karakter khusus (*cleaning*), tokenisasi, penghapusan kata hubung (*stopword removal*), dan *stemming* agar optimal saat diproses oleh mesin [10]. Deskripsi jalannya penelitian ini didukung oleh referensi agar dapat diterima secara ilmiah.

Tahap akhir adalah pemodelan menggunakan BERTopic. Kalimat diubah menjadi representasi vektor menggunakan *Sentence Transformers* (paraphrase-multilingual-MiniLM-L12-v2). Reduksi dimensi dilakukan dengan UMAP, dilanjutkan klusterisasi menggunakan HDBSCAN, dan ekstraksi kata kunci representatif dengan c-TF-IDF [11], [12].

## 3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil ekstraksi data dan diskusi komprehensif yang disusun untuk menjawab tiga rumusan masalah penelitian secara berurutan. Pertama, pemodelan BERTopic digunakan untuk mengidentifikasi apa saja topik utama yang dibahas dalam konten dakwah YouTube Indonesia. Kedua, perhitungan frekuensi dokumen dilakukan untuk menentukan topik apa yang paling dominan di antara kluster yang terbentuk. Ketiga, ekstraksi jarak semantik disajikan untuk menjelaskan bagaimana distribusi dan relasi antar topik dalam lanskap dakwah digital tersebut.

### 3.1 Deskripsi Dataset

Penelitian ini berhasil menghimpun jutaan kata dari ribuan dokumen teks transkrip. Penggunaan tabel merujuk dengan gaya tabel berikut.

Tabel 1. Statistik Dataset

| Nomor | Parameter                  | Nilai     |
|-------|----------------------------|-----------|
| 1     | Jumlah kanal               | 10        |
| 2     | Jumlah video / dokumen     | 1.250     |
| 3     | Total kata                 | 5.438.721 |
| 4     | Rata-rata kata per dokumen | 4.351     |

### 3.2 Hasil Topic Modeling

Proses BERTopic mengidentifikasi 12 kluster utama. Label topik diberikan berdasarkan kata kunci dominan (seperti: *sabar, ikhlas, syukur* untuk Topik Akhlak).

Tabel 2. Distribusi Topik Utama

| Nomor | Label Topik       | Jumlah Dokumen |
|-------|-------------------|----------------|
| 1     | Akhlak Muslim     | 218            |
| 2     | Fikih Ibadah      | 185            |
| 3     | Pendidikan Anak   | 147            |
| 4     | Tauhid dan Aqidah | 132            |
| 5     | Keluarga Islami   | 119            |
| 6     | Ekonomi Syariah   | 103            |
| 7     | Tafsir Al-Qur'an  | 96             |

|    |                                |    |
|----|--------------------------------|----|
| 8  | Sejarah Islam                  | 88 |
| 9  | Dakwah Digital                 | 64 |
| 10 | Akhir Zaman                    | 53 |
| 11 | Politik dan Umat               | 28 |
| 12 | Tidak Terklasifikasi (Outlier) | 17 |

### 3.3 Analisis dan Diskusi Topik

Distribusi topik menunjukkan bahwa tema **Akhlak Muslim** merupakan topik paling banyak dibahas dengan proporsi sekitar 17,4%. Hal ini mengindikasikan bahwa audiens digital cenderung mencari konten perbaikan moral dan motivasi spiritual. Topik **Fikih Ibadah** menempati posisi fundamental kedua (14,8%) karena sifatnya yang aplikatif menjawab persoalan ibadah sehari-hari [13].

Sebaliknya, topik **Politik dan Umat** hanya mencapai 2,2%. Temuan ini merupakan anomali dari asumsi umum bahwa media sosial penuh dengan polarisasi politik [14]. Para pendakwah di YouTube cenderung menghindari materi sensitif untuk menjaga retensi penonton dan jangkauan algoritma [15]. Pengelompokan jarak antar topik (*intertopic distance*) membuktikan bahwa BERTopic mampu memisahkan kluster "Ekonomi Syariah" secara jauh dari kluster ibadah personal, membuktikan keandalan *transformer* dalam memahami konteks leksikal [16].

### 3.4 Analisis dan Diskusi Topik

Berdasarkan temuan data, analisis dan diskusi pada bagian ini difokuskan pada tiga aspek utama:

#### a. Identifikasi Topik Utama Konten Dakwah

Berdasarkan hasil pemodelan komputasional, algoritma BERTopic terbukti mampu mengekstraksi 12 kluster topik utama dari total dataset. Topik-topik tersebut merepresentasikan variasi tema kajian di ruang digital, yang mencakup dimensi vertikal (seperti Tauhid dan Aqidah, Fikih Ibadah) hingga dimensi horizontal (seperti Akhlak Muslim, Pendidikan Anak, dan Ekonomi Syariah). Kemampuan model Sentence Transformers dalam memahami konteks leksikal memungkinkan mesin untuk mengidentifikasi ragam tema ini secara otomatis dari bahasa lisan (*spoken language*) masyarakat Indonesia tanpa perlunya pelabelan manual sejak awal.

#### b. Dominasi Topik Dakwah Digital

Analisis terhadap proporsi dokumen menunjukkan bahwa tema Akhlak Muslim merupakan topik yang paling dominan dengan proporsi tertinggi mencapai 17,4% dari keseluruhan dataset. Hal ini mengindikasikan bahwa audiens digital saat ini sangat meminati konten-konten self-improvement religius, perbaikan moral, dan motivasi spiritual. Narasi ceramah pada topik ini kental dengan anjuran aplikatif terkait cara mengendalikan emosi dan bersyukur. Topik Fikih Ibadah menempati posisi fundamental kedua (14,8%), yang menegaskan bahwa YouTube telah bertransformasi menjadi medium rujukan syariat harian umat. Sebaliknya, topik Politik dan Umat berada pada frekuensi terendah (2,2%). Temuan ini menjadi anomali dari asumsi umum bahwa media sosial penuh dengan polarisasi. Para pembuat konten dakwah tampaknya secara sadar menghindari materi sensitif atau diskursus politik guna meramu pesan yang lebih inklusif, mempertahankan harmonisasi ruang digital, serta menjaga stabilitas retensi penonton dari pembatasan algoritma (*shadowbanning*).

#### c. Distribusi dan Relasi Semantik Antar Topik

Lebih lanjut, pola sebaran topik menunjukkan adanya pemusatan (*skewed*), di mana tiga topik teratas (Akhlak, Fikih, Pendidikan Anak) mendominasi hampir 44% dari total porsi dakwah digital. Dari sisi relasi spasial, pengelompokan jarak antar topik (*intertopic distance*) membuktikan bahwa BERTopic mampu mengenali perbedaan kosakata secara tegas. Kluster "Ekonomi Syariah", misalnya, divisualisasikan terpisah jauh dari kluster ibadah personal karena penggunaan terminologi teknis bisnis yang berbeda. Di sisi lain, kluster "Tauhid dan Aqidah" dengan "Akhlak Muslim" memiliki kedekatan jarak ruang vektor (*embedding space*). Hal ini secara logis merepresentasikan gaya ceramah di Indonesia, di mana pemaparan materi fundamental tentang keimanan (aqidah) hampir selalu dikontekstualisasikan langsung dengan pesan-pesan moral aplikatif (akhlak) oleh para pendakwah.

#### 4. KESIMPULAN

Penelitian ini mengonfirmasi apa yang diharapkan di bagian pendahuluan, yaitu bahwa pemodelan komputasional secara otomatis mampu memetakan lanskap informasi agama. Algoritma BERTopic terbukti akurat dalam mengekstraksi 12 topik utama dari 1.250 dokumen transkrip, di mana "Akhlaq Muslim" dan "Fikih Ibadah" mendominasi ekosistem dakwah YouTube Indonesia. Hal ini menunjukkan tren dakwah digital yang inklusif dan berorientasi pada pengembangan karakter individu. Prospek aplikasi studi lanjutan dapat diarahkan pada penggunaan *Dynamic Topic Modeling* berbasis *time-series* atau analisis lintas platform seperti TikTok dan Instagram untuk menangkap evolusi dakwah secara lebih komprehensif.

#### 5. DAFTAR PUSTAKA

- [1] Z. Muttaqin and A. Rahman, "Pemanfaatan YouTube sebagai Media Komunikasi dan Penyiaran Islam di Indonesia," *Jurnal Kajian Media*, vol. 5, no. 2, 2021.
- [2] B. H. Harahap and H. N. Putra, "Strategi Dakwah Digital di Era Disrupsi Informasi," *Jurnal Komunikasi Islam*, vol. 12, no. 2, pp. 110-125, 2022.
- [3] S. M. N. Al-Attas, "Understanding the Paradigm of Islamic Content on YouTube: A Thematic Analysis," *Journal of Islamic Thought and Civilization*, vol. 11, no. 1, 2021.
- [4] T. Hidayat and S. Mulyati, "Analisis Algoritma Topic Modeling LDA dan BERTopic pada Ulasan E-Commerce Indonesia," *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 9, no. 1, pp. 88-95, 2023.
- [5] M. R. K. Siregar, "Metamorfosis Dakwah di YouTube: Studi Netnografi pada Akun Pendakwah Milenial," *Jurnal Ilmu Dakwah*, vol. 42, no. 1, pp. 45-67, 2022.
- [6] I. W. S. Wicaksana, "Perbandingan LDA, NMF, dan BERTopic untuk Ekstraksi Topik Berita Bahasa Indonesia," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 11, no. 4, 2022.
- [7] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv preprint arXiv:2203.05794*, 2022.
- [8] M. H. F. Al-Haj, A. Al-Mousa, and Y. Al-Radaideh, "Natural Language Processing applied to Islamic Digital Content: A Review," *IEEE Access*, vol. 9, pp. 10214-10228, 2021.
- [9] A. F. Hidayatullah, M. R. Ma'arif, and S. C. Wibawa, "Indonesian Text Topic Modeling with BERTopic," *Jurnal Ilmu Komputer dan Informasi*, vol. 15, no. 1, pp. 45-52, 2022.
- [10] D. S. Rini, "Algoritma Pemfilteran Teks dalam Sistem Preprocessing Data Bahasa Indonesia berbasis Python," *Jurnal Ilmiah Komputasi*, vol. 20, no. 3, 2021.
- [11] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv preprint arXiv:1802.03426*, 2018.
- [12] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, Springer, 2013, pp. 160-172.
- [13] R. Rahmawati, "Spiritualitas Digital: Resepsi Milenial terhadap Konten Dakwah di Media Sosial," *Jurnal Studi Agama dan Masyarakat*, vol. 17, no. 2, 2021.
- [14] E. P. W. Purba, "Transformasi Digital dan Implikasinya pada Komunikasi Agama," *Jurnal Ilmu Komunikasi*, vol. 19, no. 2, 2022.
- [15] H. A. R. Ali, M. A. Hasan, and F. A. Rahman, "Trend and Topic Analysis of Da'wah on Social Media Platforms Using NLP," *Int. J. Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.
- [16] A. A. Yahya, "Leveraging Transformer Models for Topic Modeling of Religious Texts," *Information Processing & Management*, vol. 59, no. 4, 2022.
- [17] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing*, 2019, pp. 3982-3992.
- [18] S. Abuzayed and E. Al-Khalifa, "BERT for Arabic Topic Modeling: An Experimental Study on BERTopic," in *2021 Int. Conf. of Women in Data Science*, 2021, pp. 1-5.
- [19] D. B. Setyohadi, "Penerapan Text Mining pada Analisis Sentimen dan Topic Modeling di Media Sosial," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 3, 2021.
- [20] Y. S. Lin, "Evaluating BERTopic on Short Text Social Media Data," *Data Science Journal*, vol. 20, no. 1, 2023.
- [21] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proc. of the 28th Int. Conf. on Computational Linguistics*, 2020.
- [22] A. Kurniawan and D. Suhartono, "Topic Modeling of Indonesian Academic Texts Using BERTopic," *Procedia Computer Science*, vol. 216, pp. 312-319, 2023.
- [23] H. Sujaini, "Analisis Kinerja Stopword Removal pada Data Twitter Bahasa Indonesia," *Jurnal Sistem dan Teknologi Informasi*, vol. 9, no. 1, 2021.

- [24] Z. Abidin, "Pemetaan Jaringan Sosial dan Topik Dakwah di Twitter Menggunakan Social Network Analysis," *Jurnal Komunikasi Penyiaran Islam*, vol. 8, no. 1, 2024.
- [25] A. Vaswani et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.