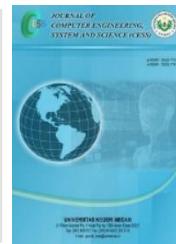


CESS
(Journal of Computer Engineering, System and Science)

Available online: <https://jurnal.unimed.ac.id/2012/index.php/cess>

ISSN: 2502-714x (Print) | ISSN: 2502-7131 (Online)



Pemodelan Topik Menggunakan BERTopic Pada Nama Produk Template Desain di Peterdraw Studio

Topic Modeling Using BERTopic On Product Name Design Template in Peterdraw Studio

Alvero Galen Gaviota^{1*}, Farida Ardiani²

^{1,2}*Informatika, Fakultas Sains dan Teknologi, Universitas Teknologi Yogyakarta*

Jl. Siliwangi (Ringroad Utara), Jombor, Sleman, D.I. Yogyakarta 55285

Email: ¹alverogaviota@gmail.com, ²farida.ardiani@staff.uty.ac.id

**Corresponding Author*

A B S T R A K

Industri kreatif digital mengalami pertumbuhan yang pesat seiring meningkatnya permintaan terhadap produk desain di berbagai *marketplace* daring. Penelitian ini bertujuan untuk mengidentifikasi pola tematik dari kumpulan nama produk *template* desain guna memahami kecenderungan pasar dan tren desain yang diminati. Penelitian ini menggunakan Peterdraw Studio sebagai Studi Kasus, yakni studio desain yang aktif menjual berbagai produk *template* di beberapa *marketplace* digital. Metode yang digunakan adalah BERTopic, yang mengombinasikan *embedding transformer* BERT untuk representasi semantik, reduksi dimensi menggunakan *Uniform Manifold Approximation and Projection* (UMAP), pengelompokan menggunakan *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN), serta ekstraksi kata kunci topik melalui *class-based* TF-IDF dan KeyBERT. Data penelitian diperoleh dari empat *marketplace* digital, yaitu *Creative Market*, *Framer*, *Envato Market*, dan *Envato Elements* dengan total 3.217 entri nama produk. Hasil pemodelan menunjukkan bahwa setiap *marketplace* memiliki topik dominan yang berbeda, mencerminkan variasi kebutuhan pengguna seperti tema *branding*, *presentation*, dan *UI/UX design*. Evaluasi model menghasilkan nilai topic coherence tertinggi sebesar 0.86 dan topic diversity tertinggi sebesar 0.98, menandakan bahwa BERTopic mampu mengidentifikasi topik yang kohesif, beragam, serta relevan untuk menganalisis tren desain pada Peterdraw Studio.

Kata Kunci: *BERTopic; Marketplace Digital; Pemodelan Topik; Pemrosesan Bahasa Alami; Template Desain.*

A B S T R A C T

The digital creative industry has shown rapid growth alongside the increasing demand for design products across various online *marketplaces*. This study aims to identify thematic



patterns from collections of design *template* product names to understand market preferences and emerging design trends. The research uses Peterdraw Studio as a study case, a studio designer who is actively selling *template* products across multiple digital *marketplaces*. The BERTopic method was employed, combining BERT transformer *embeddings* for semantic representation, dimensionality reduction using *Uniform Manifold Approximation and Projection* (UMAP), clustering with Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), and topic keyword extraction through class-based TF-IDF and KeyBERT. The dataset was obtained from four digital *marketplaces* *Creative Market*, *Framer*, *Envato Market*, and *Envato Elements* comprising 3,217 product name entries. The modeling results indicate that each *marketplace* has distinct dominant topics reflecting user preferences, such as branding, presentation, and UI/UX design themes. Model evaluation achieved the highest topic coherence score of 0.86 and topic diversity score of 0.98, demonstrating that BERTopic effectively identifies cohesive, diverse, and contextually relevant topics for analyzing design trends at Peterdraw Studio.

Keywords: *BERTopic*; *Design Template*; *Digital Marketplace*; *Natural Language Processing*; *Topic Modelling*.

1. PENDAHULUAN

Industri kreatif digital di Indonesia telah mengalami pertumbuhan signifikan dalam beberapa tahun terakhir, salah satu sektor yang terdampak adalah sektor bidang desain temsuk di dalamnya *template* desain grafis [1]. Peterdraw studio sebuah studio desain yang berbasis di Daerah Istimewa Yogyakarta, menyediakan *template* desain grafis di berbagai *marketplace* digital. Template desain adalah sebuah kerangka atau pola visual yang telah dibuat sebelumnya untuk membantu dalam pembuatan desain grafis [2]. Marketplace digital seperti *Creative Market*, *Envato Market*, dan *Framer* menyediakan ruang bagi desainer untuk menjual *template* desain dalam berbagai kategori seperti presentasi, website, poster, dan sosial media [3]. Dalam merancang dan menentukan *template* desain yang ingin di jual, Peterdraw Studio menjalankan proses bisnisnya melalui empat tahapan utama yaitu riset, produksi, pengajuan ke *marketplace*, dan pemasaran. Tahapan pertama dilakukan dengan mengamati tren produk di berbagai platform mulai dari *marketplace* desain dan sosial media serta membandingkannya dengan hasil penjualan produk sebelumnya. Tahapan kedua, studio merancang *template* desain yang ditentukan menggunakan software seperti Figma dan Canva. Selanjutnya, di tahap ketiga studio mulai mengajukan *template* desain yang telah dibuat untuk di unggah ke *marketplace* digital *Creative Market*, *Envato Market*, dan *Framer*. Terakhir studio melakukan pemasaran, dimana tim marketing studio mempromosikan produk melalui media sosial dan email marketing untuk meningkatkan penjualan serta menarik lebih banyak pelanggan [4]. Meskipun peterdraw studio telah memiliki proses bisnis yang sistematis dalam pengembangan dan penjualan *template* desain, studio menghadapi kebutuhan dalam memahami kecenderungan pasar secara lebih mendalam terkait tema *template* desain apa yang paling sering terjual.

Pemodelan topik adalah sebuah teknik analisis dalam pemrosesan bahasa alami yang bekerja dengan cara mengelompokan kata – kata dalam dokumen untuk menemukan topik

yang tersembunyi di dalamnya [5]. BERTopic merupakan metode pemodelan topik yang menggunakan representasi *embedding* BERT, *clustering*, dan prosedur TF-IDF berbasis kelas (*class-based* TF-IDF) untuk menghasilkan topik yang lebih berhubungan dan interpretatif [6]. BERTopic memiliki tahapan arsitektur, dimulai dengan transformasi teks menjadi representasi vektor menggunakan model BERT (*Bidirectional Transformers for Language Understanding*) yang melalui mekanisme *self-attention* mampu menangkap konteks dalam dua arah secara mendalam [7]. Vektor *embedding* berdimensi tinggi tersebut kemudian direduksi dimensinya dengan UMAP (*Uniform Manifold Approximation and Projection*), sehingga struktur manifold data semantik dapat dipertahankan dalam ruang berdimensi rendah [8]. Setelah itu, algoritma HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) digunakan untuk mengidentifikasi klaster dokumen berdasarkan kepadatan, memungkinkan penentuan topik tanpa perlu menetapkan jumlah klaster sebelumnya [9]. Setelah itu setiap topik klaster dihitung frekuensi kemunculan kata dengan menggunakan *countvectorizer* [10]. Selanjutnya topik dianalisis menggunakan class-based TF-IDF, yakni modifikasi dari TF-IDF tradisional yang menghitung bobot istilah dengan mempertimbangkan distribusi frekuensi dalam masing-masing klaster dibandingkan keseluruhan korpus, sehingga menghasilkan representasi kata kunci topik yang lebih kontras dan interpretatif [6], [11]. Sebagai pelengkap, BERTopic dapat melakukan *fine-tuning* dengan memanfaatkan KeyBERT untuk mengekstraksi kata kunci secara kontekstual dari *embedding*, memperdalam kualitas representasi topik dengan mempertimbangkan hubungan semantik antar kata [12]. Tahapan berantai ini, mewujudkan kemampuan BERTopic dalam mengekstraksi topik yang konsisten dan mudah diinterpretasikan dari kumpulan teks.

Beberapa penelitian mengenai pemodelan topik dengan metode BERTopic telah dilakukan dalam berbagai bidang. Penelitian yang dilakukan oleh Bagas Raditya Nur Listyawan, Nanang Yudi Setiawan, dan Mochamad Chandra Saputra pada tahun 2024 berhasil mengidentifikasi topik utama dengan nilai koherensi topik 0.625 dan keberagaman topik 0.828 pada data aktivitas pengembangan perangkat lunak [13]. Berikutnya penelitian yang dilakukan oleh Aziida Nanyonga, Keith Joiner, dan Ugur Turhan melakukan perbandingan metode BERTopic dengan *Probabilistic Latent Semantic Analysis* (PLSA) dalam menemukan kunci topik utama di laporan keselamatan penerbangan. Di penelitian tersebut ditemukan bahwa BERTopic mampu mengungguli dengan nilai koherensi topik sebesar 0.41 dibandingkan 0.37 pada metode PLSA [14]. Selanjutnya penelitian oleh Dini Aryani, Ivana Lucia Kharisma, Alun Sujjada, dan Kamdan di tahun 2024 mengaplikasikan BERTopic pada pemberitaan pemilu 2024 di detik.com. Penelitian ini berhasil memproses 2,500 artikel berita dan mengidentifikasi 10 topik utama yang merefleksikan dinamika partisipasi generasi muda, distribusi isu kampanye, dan sentimen publik dengan nilai skor rata-rata koherensi topik 0.58 [15].

Meskipun pemodelan topik banyak diterapkan pada jenis data teks yang panjang seperti berita, laporan, dan deskripsi. Studi yang secara khusus menelaah nama produk *template* desain yang berupa teks pendek dalam industri kreatif digital masih terbatas. Penelitian sebelumnya umumnya menelaah data berdimensi panjang sehingga kurang menggambarkan efektivitas teknik pemodelan topik BERTopic bekerja pada teks pendek, dan mengandung istilah khusus desain. Selain itu, variasi penamaan nama produk *template* desain antar *marketplace* serta keterbatasan jumlah kata dapat mempengaruhi kestabilan dan keterbacaan

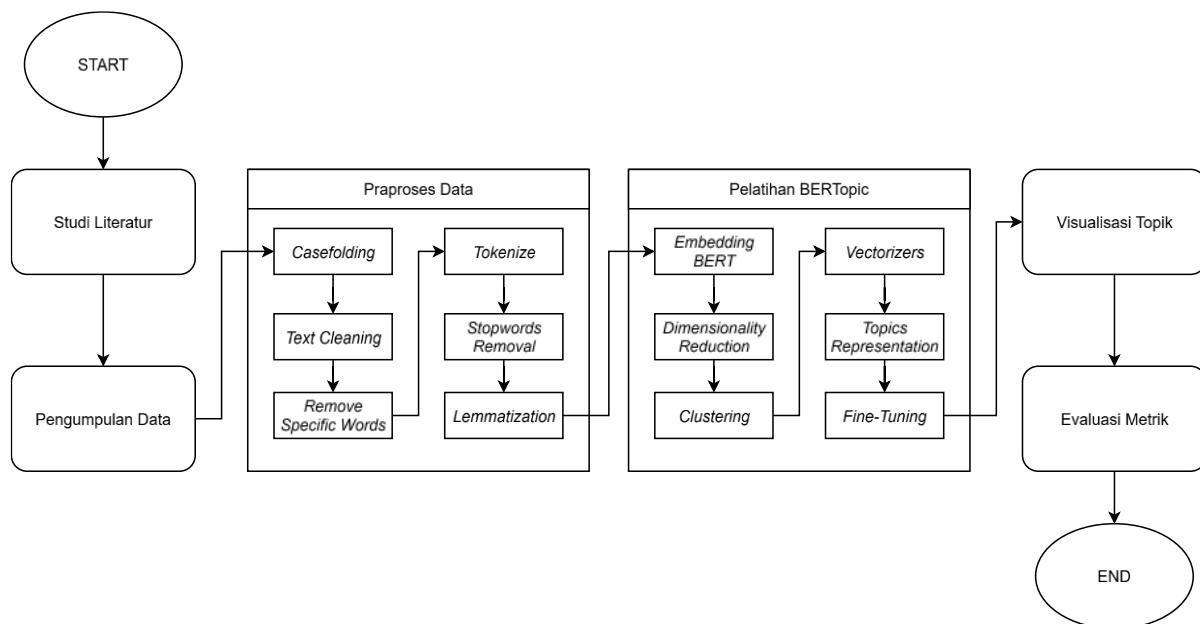
topik. Kebutuhan untuk menyediakan visualisasi yang mudah diinterpretasikan oleh pelaku industri serta evaluasi kuantitatif yang menjamin kualitas dan keberagam topik untuk mendukung keputusan pemasaran dan perancangan produk juga kurang mendapat perhatian. Oleh karena itu, diperlukan penelitian yang secara khusus menguji penerapan BERTopic pada nama produk *template* desain antar *marketplace* dengan pendekatan evaluasi kuantitatif dan visualisasi interpretatif yang mampu menghasilkan wawasan berguna bagi praktik industri digital kreatif.

Penelitian ini bertujuan untuk mengimplementasikan pemodelan topik dengan metode BERTopic pada nama produk *template* desain yang terjual oleh peterdraw studio. Kesenjangan ini penting karena karakteristik data nama produk yang singkat dan beragam dari sumber *marketplace* menuntut kepastian apakah pendekatan BERTopic dapat memberikan wawasan yang bermakna bagi pelaku industri desain. Dengan penelitian ini, Peterdraw studio dapat memahami kecenderungan pelanggan terhadap tema atau topik *template* desain apa yang paling diminati. Dengan mengidentifikasi pola topik dalam nama produk *template* desain, studio mendapatkan wawasan dari data *template* desain untuk meningkatkan strategi pemasaran, peningkatan laba, serta pengambilan keputusan yang berbasis data untuk merancang *template* desain yang lebih optimal [16]. Penelitian menggunakan data penjualan *template* desain studio dari *marketplace* digital seperti *Creative Market*, *Framer*, dan *Envato*. Dalam data penjualan, terdapat nama produk *template* desain berbentuk teks pendek yang akan dianalisis. Metode BERTopic telah dibuktikan berguna untuk menganalisis teks yang memiliki jumlah volume yang pendek, sehingga metode tersebut sesuai dengan data yang digunakan di penelitian ini [17]. Selain itu, penelitian ini menyajikan visualisasi hasil pemodelan dengan meliputi representasi distribusi topik dan kata kunci dominan sebagai alat bantu analisis yang memperjelas struktur tematik dalam data penjualan. Kualitas topik diuji secara kuantitatif menggunakan metrik *topic coherence* dan *topic diversity* untuk menghitung seberapa jelas topik yang saling terkait serta keberagaman topik yang dihasilkan, sehingga validitas interpretasi topik dapat diinterpretasikan secara statistik [18]. Dengan kombinasi domain data yang spesifik, pendekatan BERTopic yang adaptif, serta evaluasi metrik yang terukur, studi ini memberikan kontribusi baru pada literatur pemodelan topik di bidang desain digital.

2. METODE PENELITIAN

Bab ini menguraikan metodologi yang diterapkan dalam pemodelan topik pada nama produk *template* desain di Peterdraw Studio. Penelitian ini disusun melalui enam tahapan yang disusun secara terstruktur untuk menjamin validitas dan reproduktibilitas hasil. Tahap pertama studi literatur, difokuskan pada tinjauan pustaka terkait teknik pemodelan topik, evaluasi metrik kualitas topik, serta pemilihan BERTopic sebagai metode utama karena kemampuannya memadukan representasi *embedding*, *clustering*, dan *class-based TF-IDF* dalam satu *pipeline*. Tahap kedua, pengumpulan data dilakukan dengan mengajukan data penjualan *template* desain dari Peterdraw Studio yang terdapat pada *marketplace Envato Market*, *Envato Elements*, *Creative Market*, dan *Framer*, kemudian menyusunnya dalam format yang sesuai untuk di analisis. Selanjutnya tahap ketiga data praproses, data yang diperoleh akan diolah melalui enam langkah: *casefolding* untuk menyamakan huruf pada teks,

text cleaning yang mencakup penghapusan tanda baca, angka, karakter non-alfabet, dan *whitespace*, lalu penghilangan kata-kata spesifik yang terdapat pada teks seperti merek atau kata lain yang tidak merepresentasikan makna dan tema dari *template* desain, tokenisasi untuk memecah kata dalam kalimat menjadi token, *stopwords removal* untuk menghapus kata – kata umum pada bahasa, *lemmatization* untuk mengambil kata dasar. Enam langkah tersebut perlu dilakukan agar bentuk data siap untuk diproses pada tahap pelatihan model. Di tahap keempat, data yang telah diproses akan melalui pelatihan BERTopic. Pelatihan model BERTopic mencakup 6 subproses. transformasi teks menjadi vektor *embedding* menggunakan *embedding* model transformer BERT dalam bentuk pra-latih, reduksi dimensi *embedding* menggunakan UMAP guna mempertahankan struktur *manifold* semantik, *clustering* dokumen melalui HDBSCAN untuk membentuk klaster topik, *vectorizers* menggunakan *CountVectorizer* untuk menyusun matriks dokumen-kata berdasarkan frekuensi istilah, mengekstraksi kata kunci tiap klaster melalui *class-based TF-IDF* untuk menonjolkan istilah paling representatif, *fine-tuning* representasi topik menggunakan KeyBERT untuk memperkaya pilihan kata kunci berdasarkan kesamaan kontekstual. Tahap kelima, hasil dari pelatihan model disajikan dalam bentuk visualisasi untuk merepresentasikan tema atau topik – topik yang dibentuk. Terakhir pada tahap keenam, mengevaluasi kinerja model dengan melibatkan perhitungan *metrik topic coherence* untuk menilai koherensi semantik antar istilah dalam setiap topik dan *topic diversity* untuk mengukur keberagaman topik yang dihasilkan. Dari keseluruhan enam tahap ini, penelitian dapat dilaksanakan secara beraturan. Alur tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Alur Tahapan Penelitian

2.1. Data Penelitian

Data yang digunakan untuk penelitian ini bersumber dari catatan penjualan internal *template* desain milik Peterdraw Studio yang diperoleh dari tiga *marketplace* yang digunakan, yaitu *Envato*, *Creative Market*, dan *Framer*. Penelitian akan dikembangkan berdasarkan ketiga

marketplace tersebut sebagai dasar dalam proses perhitungan dan analisis. Pengumpulan data dilakukan melalui wawancara mendalam dengan *Chief Executive Officer* (CEO) dari Peterdraw Studio. Wawancara dilakukan melalui komunikasi email, di mana pihak Peterdraw Studio mengirimkan kumpulan data dalam bentuk folder yang berisi file data penjualan dari tiga *marketplace* yang tersedia. File data tersebut disampaikan dalam format *Comma-Separated Values* (CSV) untuk memudahkan proses pengolahan dan analisis lebih lanjut. Data penjualan mencakup periode dari tahun 2023 hingga 2025. Khusus pada *marketplace Framer*, data yang tersedia hanya mencakup periode 2024 hingga 2025. Untuk Envato, data terbagi ke dalam dua layanan, *Envato Market* yang menjual *template* per satuan dan *Envato Elements* yang berbasis langganan. Kedua sumber tetap mencatat nama produk secara tekstual sebagai basis analisis.

Masing – masing file CSV memiliki struktur kolom yang berbeda, namun atribut utama yang dijadikan fokus pemodelan topik adalah kolom nama produk *template* desain. Kolom ini memuat teks pendek yang mendeskripsikan setiap *template* di setiap *marketplace*, dan menjadi variabel utama yang digunakan untuk menemukan topik – topik yang terdapat pada data. Untuk penggambaran yang lebih jelas, berikut adalah sampel data penjualan *template* desain dari *marketplace* framer pada Tabel 1.

Tabel 1. Sample Data Marketplace *Framer*

No	Product_Name
1	Xplore - Travel Agency Website <i>Framer</i> Template
2	LeafLife - Garden Landscape Design Service <i>Framer</i> Template
3	Drivoxe - Car Rental Website <i>Framer</i> Template
4	LuxDin - Luxury Restaurant Website <i>Framer</i> Template
5	Axels - Jewelry Ecommerce <i>Framer</i> Template (Reg)

2.2. Praproses Data

Setelah data dikumpulkan, data akan memasuki tahapan praproses melalui enam langkah. *Casefolding* untuk menyamakan huruf pada teks, *text cleaning* yang mencakup penghapusan tanda baca, angka, karakter non-alfabet, dan *whitespace*, lalu penghilangan kata-kata spesifik yang terdapat pada teks seperti merek atau kata lain yang tidak merepresentasikan makna dan tema dari *template* desain, tokenisasi untuk memecah kata dalam kalimat menjadi token, *stopwords removal* untuk menghapus kata – kata umum pada bahasa, lemmatization untuk mengambil kata dasar. Tahapan praproses ini penting untuk memastikan format data teks dari nama produk *template* desain mampu dilatih dengan metode BERTopic.

2.2.1 Casefolding

Langkah pertama dalam tahapan praproses adalah mengonversikan seluruh huruf dalam teks nama produk menjadi di setiap *marketplace* menjadi huruf kecil. Dengan *casefolding*, seluruh bentuk teks dapat dipastikan dalam format yang konsisten dan berstruktur.

2.2.2 Text Cleaning

Langkah kedua adalah menghapus elemen pada nama teks nama produk *template* desain seperti simbol tanda baca (@, #, ?, !, -, _, dan lain - lain), angka (1, 2, 3, ...), karakter non-alfabet, dan *whitespace*. Karakter elemen ini perlu dihilangkan setiap nama produk *template* desain menggambarkan deskripsi yang jelas.

2.2.3 Remove Specific Words

Selanjutnya langkah ketiga adalah menghapus kata – kata spesifik yang terdapat di nama produk *template* desain. Kata – kata ini biasanya terdapat di awal kalimat sebagai bentuk merek, sehingga setiap *template* menggunakan kata yang sangat spesifik dan tidak merepresentasikan sebuah tema atau topik untuk di analisis.

2.2.4 Tokenization

Berikutnya, Langkah keempat dalam tahapan praproses adalah membagi setiap kata pada teks menjadi unit token. Dengan melakukan *tokenization*, BERTopic lebih mudah untuk menganalisis dan memahami kalimat dan bahasa secara kontekstual.

2.2.5 Stopwords Removal

Lalu, langkah kelima adalah penghapusan kata – kata umum yang tidak memberikan makna signifikan dalam sebuah topik. Seluruh data nama produk *template* desain menggunakan bahasa inggris dan penghapusan menggunakan kamus bahasa inggris dari pustaka spacy. Kata – kata umum yang dihapus seperti (the, a, is, and). langkah ini penting agar pengambilan kata terpusat pada kata yang lebih penting dalam teks nama produk *template* desain.

2.2.6 Lemmatization

Langkah terakhir adalah menghilangkan imbuhan pada sebuah kata dengan mengubah kata menjadi bentuk dasarnya. seperti “*walking*” menjadi “*walk*”. Langkah ini memastikan bahwa setiap kata dengan makna yang sama di perlakukan dengan serupa, sehingga pengelompokan kata menjadi lebih akurat, dan menghasilkan topik yang lebih tematik.

2.3. Pelatihan BERTopic

Tahapan selanjutnya, data yang telah dipraproses akan latih dengan metode BERTopic. BERTopic memiliki struktur langkah pelatihan yang terdiri dari enam langkah. Hal ini dilakukan agar pemodelan topik dapat menangkap struktur teks tematik yang terdapat pada data nama produk *template* desain.

2.3.1 Embedding Transformer

Langkah pertama dalam pelatihan dengan model BERTopic adalah melatih data yang telah dipraproses dengan model transformer pra-latih dari BERT. Model pra-latih yang digunakan bernama “all-MiniLM-L12-v2”, model ini merupakan model pra-latih yang telah dilatih menggunakan 1 miliar dataset pasangan pelatihan, dan digunakan dalam pengelompokan serta mencari kata secara semantik.

2.3.2 Dimensionality Reduction

Langkah kedua, mengurangi dimensi dari data yang telah dilatih menggunakan UMAP (*Uniform Manifold Approximation and Projection*) dengan tetap mempertahankan informasi

yang penting. Langkah ini berguna untuk mensimplifikasi data sehingga model BERTopic menjadi mudah untuk dianalisis dan meningkatkan efisiensi pelatihan. Beberapa parameter yang digunakan pada langkah ini, *n_neighbors*=10 untuk mengatur struktur data local dengan global, *n_components*=5, *min_dist*=0.0 untuk mengatur jarak minimum antar poin dalam dimensi rendah. dan *random_state*=42 untuk pembelajaran metode dengan hasil yang konsisten.

2.3.3 Clustering

Langkah ketiga merupakan pengelompokan data menjadi beberapa klaster menggunakan HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*). Data dikelompokkan berdasarkan kemiripan antar kata sehingga membentuk kelompok antar topik. Parameter yang digunakan dengan HDBSCAN adalah *min_cluster_size*=10 untuk mengatur minimum jumlah kelompok yang dibentuk, *metric*=“euclidean” jarak metrik yang digunakan untuk komputasi, *cluster_selection_method*=“eom” untuk memilih klaster dalam dalam hirarki pohon

2.3.4 Vectorizers

Selanjutnya langkah keempat, teks dari masing – masing klaster dianalisis menggunakan *CountVectorizer* untuk membentuk kata kunci topik. Paramater yang digunakan adalah *ngram_range*(1, 3) untuk menentukan jumlah kata yang terbentuk dalam topik. Dan *stopwords*=“english” untuk menghapus kata – kata umum yang terdapat pada nama produk *template* desain di kamus bahasa inggris.

2.3.5 Topic Representation

Berikutnya langkah kelima, melakukan perhitungan dengan mengekstraksi kata kunci tiap klaster menggunakan c-TF-IDF sehingga menghasilkan representasi dari pemodelan topik.

2.3.6 Fine Tuning

Langkah terakhir melakukan *fine-tuning* dengan meningkatkan kualitas dengan memperhalus representasi topik menggunakan KeyBERT. Langkah ini bertujuan untuk menampilkan kata – kata yang paling representatif dari setiap topik sehingga hasil yang ditampilkan lebih bermakna dan mudah diinterpretasikan.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data penelitian ini berasal dari *marketplace* tempat Peterdraw Studio menjual produk *template* desain, yaitu *Creative Market*, *Framer*, *Envato Market*, dan Envato Elements. Setiap *marketplace* memiliki jumlah entri data yang berbeda yang ditunjukkan pada Tabel 2.

Tabel 2. Jumlah Data Per Marketplace

Marketplace	Jumlah Data
<i>Creative Market</i>	390
<i>Framer</i>	202
<i>Envato Market</i>	1,129
<i>Envato Elements</i>	1,496
Total	3,217

Perbedaan jumlah data di masing-masing *marketplace* menunjukkan variasi model bisnis dan perilaku pengguna. *Envato Elements* memiliki volume data tertinggi karena sistem berlangganan memungkinkan pengguna mengunduh banyak produk sekaligus, sementara *Framer* dan *Creative Market* memiliki pola transaksi yang lebih individual. Variasi ini secara langsung memengaruhi keragaman topik yang terbentuk pada proses pemodelan selanjutnya.

3.2. Hasil Praproses Data

Setiap dataset dari *marketplace Creative Market, Framer, Envato Market, dan Envato Elements* dilakukan serangkaian tahapan praproses teks yang meliputi *casefolding, text cleaning, remove specific words, tokenization, stopwords removal, and lemmatization*. Proses ini bertujuan untuk menyeragamkan bentuk teks, menghapus elemen yang tidak relevan, serta meminimalkan noise yang dapat memengaruhi kualitas representasi semantik pada tahap pemodelan BERTopic. Seluruh tahapan dilakukan secara konsisten pada tiap *marketplace* menggunakan pipeline praproses yang sama, sehingga data akhir memiliki struktur dan format seragam. Tahapan *casefolding* menurunkan semua huruf menjadi bentuk kecil, diikuti *text cleaning* untuk menghapus tanda baca, angka, dan karakter non-alfabet. Kemudian dilakukan penghapusan kata spesifik (misalnya “*template*”, “*design*”, atau nama platform) untuk mengurangi dominasi kata umum. Selanjutnya teks dipecah menjadi token, kata berhenti dihapus, dan bentuk kata dikembalikan ke leksikal dasarnya melalui *lemmatization*. Hasil dari tahapan ini menghasilkan teks yang lebih bersih, ringkas, dan informatif untuk dianalisis lebih lanjut. Sebagai contoh, hasil praproses data dari *marketplace Framer* dapat dilihat pada Tabel 3, yang menampilkan perbandingan antara teks mentah dan hasil praproses akhir.

Tabel 3. Sample Perbandingan Praproses Data *Framer*

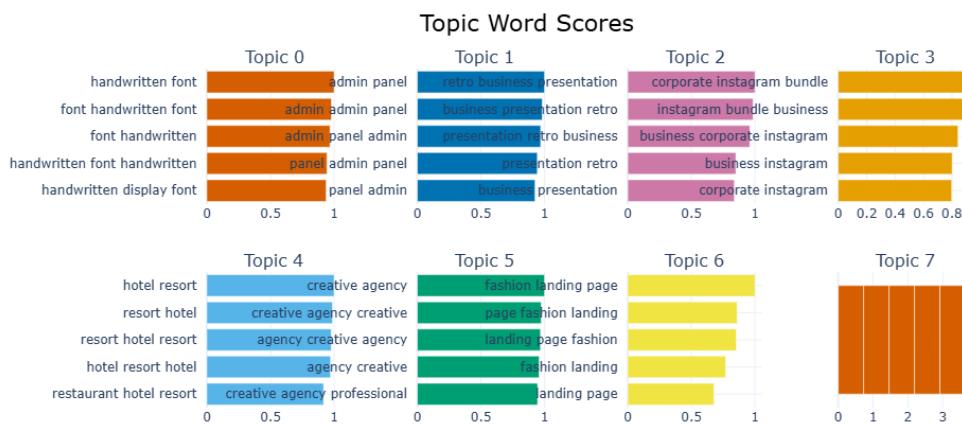
Sebelum	Sesudah
<i>Xplore - Travel Agency Website Framer Template</i>	<i>travel agency</i>
<i>LeafLife - Garden Landscape Design Service Framer Template</i>	<i>garden design service</i>
<i>Drivoxe - Car Rental Website Framer Template</i>	<i>car rental</i>
<i>LuxDin - Luxury Restaurant Website Framer Template</i>	<i>luxury restaurant</i>
<i>Axels - Jewelry Ecommerce Framer Template (Reg)</i>	<i>jewelry ecommerce</i>

Hasil pada Tabel 3 memperlihatkan bahwa praproses data berhasil menghapus elemen non-esensial dan menyisakan kata-kata bermakna utama yang menggambarkan konteks desain *template*. Dengan demikian, hasil praproses ini memperkuat pondasi semantik bagi tahap *embedding* dan pemodelan topik menggunakan BERTopic.

3.3. Hasil Pemodelan BERTopic

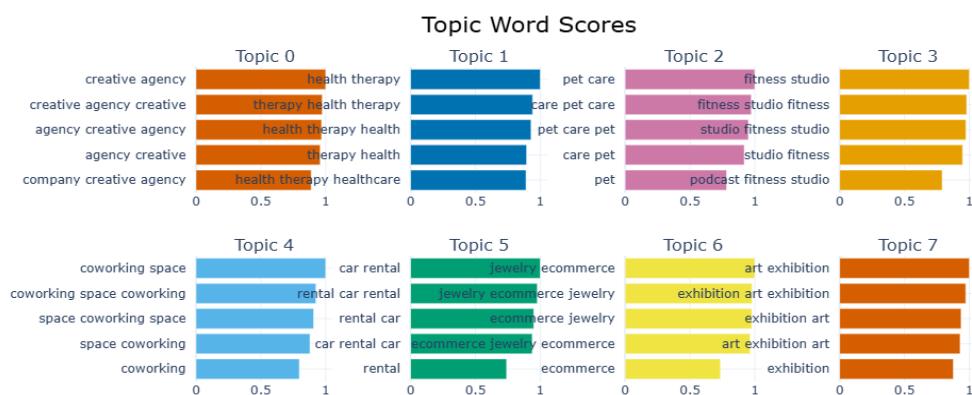
Pemodelan topik dilakukan menggunakan algoritma BERTopic, yang mengintegrasikan enam tahapan utama untuk membentuk struktur tematik dari kumpulan nama produk *template* desain. Proses dimulai dengan representasi semantik melalui *embedding* BERT pralatih yang menangkap konteks antar kata secara mendalam. Representasi tersebut kemudian direduksi dimensinya menggunakan *Uniform Manifold Approximation and Projection* (UMAP) untuk mempertahankan struktur semantik dalam ruang berdimensi rendah. Hasil reduksi selanjutnya dikelompokkan menggunakan algoritma *Hierarchical Density-Based Spatial*

Clustering of Applications with Noise (HDBSCAN) guna memisahkan teks ke dalam kelompok topik yang saling berdekatan secara semantik. Setiap klaster hasil pengelompokan kemudian diubah menjadi matriks frekuensi kata melalui *CountVectorizer*, yang digunakan dalam perhitungan *class-based TF-IDF* (c-TF-IDF) untuk menentukan bobot kata paling representatif di setiap klaster. Tahap akhir dilakukan fine-tuning menggunakan KeyBERT, yang bertujuan memperkuat relevansi semantik kata kunci sehingga topik yang terbentuk lebih interpretatif dan menggambarkan makna kontekstual sebenarnya. Hasil pemodelan divisualisasikan dalam bentuk diagram batang untuk setiap *marketplace* secara terpisah, masing-masing menampilkan distribusi frekuensi topik yang dihasilkan oleh BERTopic. Keempat hasil visualisasi tersebut disajikan pada Gambar 2a hingga Gambar 2d.



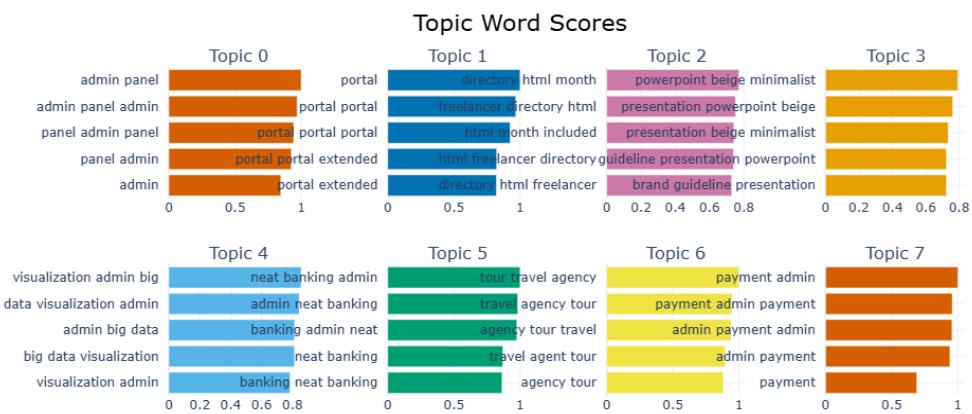
Gambar 2(a). Pemodelan Topik *Creative Market*

Pada *Creative Market*, topik yang dihasilkan menggambarkan orientasi desain yang luas mulai dari *handwritten font* hingga *fashion landing page*. Kemunculan topik seperti *creative agency* dan *retro business presentation* menunjukkan bahwa pasar di platform ini lebih berfokus pada karya-karya desain dengan karakter visual dan gaya artistik yang kuat. Tema seperti *hotel resort* dan *admin panel* juga menandakan keberagaman kebutuhan pengguna terhadap desain *template*, baik untuk keperluan estetika maupun bisnis digital.



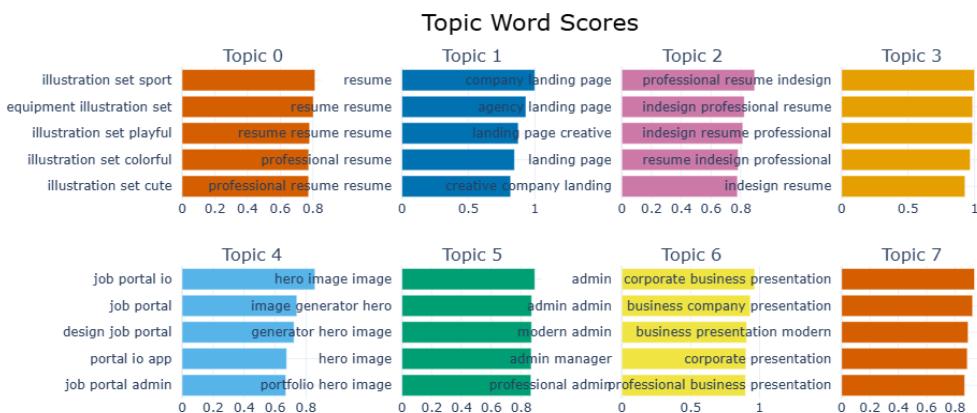
Gambar 3(b). Pemodelan Topik *Framer*

Sementara itu, data *marketplace Framer* memperlihatkan spektrum topik yang lebih terkonsentrasi pada konteks desain berbasis web interaktif. Topik-topik seperti *creative agency*, *coworking space*, *fitness studio*, dan *art exhibition* memperkuat posisi *Framer* sebagai *platform* yang banyak digunakan untuk pengembangan *landing page* atau *prototype website* dengan fokus pada pengalaman pengguna (UX). Kemunculan tema seperti *health therapy* dan *pet care* juga menunjukkan bahwa desain di *Framer* mencakup berbagai sektor jasa kreatif, tidak hanya terbatas pada kebutuhan korporasi.



Gambar 4(c). Pemodelan Topik *Envato Market*

Pada *Envato Market*, hasil pemodelan menghasilkan topik yang cenderung teknis dan profesional. Topik seperti admin panel, banking admin, payment admin, serta *visualization dashboard* menunjukkan dominasi desain yang berhubungan dengan antarmuka pengguna (*user interface*) dan sistem bisnis digital. Selain itu, kemunculan tema *tour travel agency* dan *corporate presentation* mengindikasikan permintaan pasar terhadap *template* dengan fungsi bisnis yang spesifik dan siap pakai. Hal ini mencerminkan karakter *Envato Market* sebagai *marketplace* dengan orientasi kuat pada kebutuhan desain komersial.



Gambar 5(d). Pemodelan Topik *Envato Elements*

Adapun *Envato Elements* menampilkan kumpulan topik yang lebih beragam dengan perpaduan antara desain visual dan profesional. Tema seperti *resume*, *corporate presentation*, dan *company landing page* menggambarkan orientasi bisnis dan personal

branding yang kuat, sedangkan *illustration* set dan *hero image* memperlihatkan sisi kreatif visual yang lebih ekspresif. Dominasi topik ini menunjukkan bahwa pengguna *Envato Elements* memanfaatkan platform berbasis langganan tersebut untuk memperoleh aset desain yang serbaguna, baik untuk kebutuhan profesional maupun estetika.

3.4. Hasil Evaluasi Model

Evaluasi performa model BERTopic dilakukan menggunakan dua metrik utama, yaitu *Topic Coherence* dan *Topic Diversity*, untuk menilai kualitas topik yang terbentuk dari nama produk *template* desain di setiap *marketplace*. Metrik *topic coherence* mengukur tingkat keseragaman semantik antar kata dalam satu topik, sedangkan *topic diversity* menilai sejauh mana kata kunci yang dihasilkan oleh model bersifat unik dan tidak berulang antar topik. Hasil evaluasi keseluruhan ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan Evaluasi Model

Marketplace	Topic Coherence	Topic Diversity
<i>Creative Market</i>	0.80	0.93
<i>Framer</i>	0.86	0.83
<i>Envato Market</i>	0.82	0.84
<i>Envato Elements</i>	0.82	0.98

Berdasarkan hasil pada Tabel 4, terlihat bahwa seluruh model menunjukkan performa yang baik dengan nilai *topic coherence* di atas 0.80, menandakan bahwa kata-kata dalam setiap topik memiliki keterkaitan semantik yang kuat dan membentuk makna tematik yang jelas. Nilai tertinggi dicapai oleh *Framer* (0.86), yang menunjukkan bahwa kumpulan nama produk pada *marketplace* ini memiliki struktur semantik yang paling konsisten. Hal ini dapat dikaitkan dengan karakter data *Framer* yang lebih terfokus pada desain web interaktif, sehingga istilah-istilah yang digunakan relatif seragam.

Sementara itu, metrik *topic diversity* memperlihatkan tingkat variasi kata kunci antar topik yang tinggi di seluruh *marketplace*. Nilai tertinggi dicapai oleh *Envato Elements* (0.98), menandakan bahwa model berhasil menghasilkan topik-topik yang sangat berbeda satu sama lain dengan kata kunci yang jarang tumpang tindih. Kondisi ini sesuai dengan karakter platform *Envato Elements* yang menyediakan aset desain dalam berbagai kategori, mulai dari presentasi hingga ilustrasi visual. Sebaliknya, *Framer* memiliki nilai *diversity* yang sedikit lebih rendah (0.83), yang mengindikasikan bahwa beberapa topik memiliki tumpang tindih istilah akibat kesamaan konteks desain, seperti *landing page* dan *UI kit*.

4. KESIMPULAN

Penelitian ini telah berhasil menerapkan metode BERTopic untuk melakukan pemodelan topik pada kumpulan nama produk *template* desain milik Peterdraw Studio yang berasal dari empat *marketplace* digital, yaitu *Creative Market*, *Framer*, *Envato Market*, dan *Envato Elements*. Hasil pemodelan menunjukkan bahwa setiap *marketplace* memiliki karakteristik tematik yang berbeda, di mana *Creative Market* menonjolkan desain visual dan *branding*, *Framer* berfokus pada desain web interaktif, sedangkan *Envato Market* dan *Envato Elements* menampilkan tema profesional seperti *business presentation* dan *resume design*. Evaluasi

model menunjukkan bahwa nilai *topic coherence* tertinggi diperoleh pada *Framer* sebesar 0.86, diikuti oleh *Envato Market* dan *Envato Elements* masing-masing 0.82, serta *Creative Market* sebesar 0.80. Sementara itu, nilai *topic diversity* tertinggi dicapai oleh *Envato Elements* sebesar 0.98, diikuti *Creative Market* 0.93, *Envato Market* 0.84, dan *Framer* 0.83. Nilai-nilai tersebut mengindikasikan bahwa topik yang dihasilkan memiliki keseragaman semantik yang baik serta tingkat keragaman yang tinggi antar topik. Meski demikian, penelitian ini memiliki keterbatasan karena data hanya berasal dari satu studio dan analisis difokuskan pada judul nama produk tanpa verifikasi oleh pakar. Untuk memperkuat temuan, penelitian selanjutnya sebaiknya memperluas sumber data lintas *marketplace*, menambahkan konteks yang lebih deskriptif, melakukan validasi kualitatif dan perbandingan dengan metode pemodelan topik lain. Secara keseluruhan, hasil penelitian ini membuktikan bahwa metode BERTopic efektif dalam mengidentifikasi pola tematik dari teks pendek dan memberikan wawasan strategis bagi Peterdraw Studio dalam memahami tren pasar serta mengarahkan pengembangan produk desain secara lebih tepat sasaran.

DAFTAR PUSTAKA

- [1] R. Rofaida, A. Nur Aryanti, Y. Perdana, and F. Pendidikan Ekonomi dan Bisnis, "Strategi Inovasi pada Industri Kreatif Digital: Upaya Memperoleh Keunggulan Bersaing pada Era Revolusi Industri 4.0," *Jurnal Manajemen Dan Keuangan*, vol. 8, no. 3, pp. 402–414, Dec. 2019, Accessed: Dec. 15, 2025. [Online]. Available: <https://ejurnalunsam.id/index.php/jmk/article/view/1909>
- [2] A. Munthe, K. K. Abstrak, and K. Penulis, "Implementasi Algoritma Dynamic Markov Compression (DMC) Untuk Kompresi Templates Desain Grafis," *JIKTEKS : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 2, no. 01, pp. 23–34, Dec. 2023, Accessed: Dec. 15, 2025. [Online]. Available: <https://jurnal.faaatuatua.com/index.php/JIKTEKS/article/view/55>
- [3] P. Purwono, K. Nisa, J. Indriyanto, and D. Febri Kuncoro, "Jurnal Pengabdian dan Pemberdayaan Masyarakat Indonesia Digital Entrepreneurship Education and Mentoring for PGRI Gumelar High School Students to Enhance Entrepreneurial Skills in International Markets Using Digital Media," vol. 3, no. 6, pp. 285–294, Oct. 2023, [Online]. Available: <https://jppmi.journalptti.com/index.php/jppmi>
- [4] S. Jatmika, T. Aprilianto, B. Poernomo, and T. Prasetyo, "E-Marketing Dengan Media Jejaring Sosial Untuk Peningkatan Omset Penjualan," *Seminar Nasional Sistem Informasi*, vol. 1, pp. 646–654, Sep. 2017, Accessed: Dec. 15, 2025. [Online]. Available: <https://jurnalfti.unmer.ac.id/index.php/senasis/article/view/67>
- [5] Z. N. Muna, B. Darma Setiawan, and R. S. Perdana, "Fakultas Ilmu Komputer Penerapan Pemodelan Topik Komentar Melalui Media Sosial Twitter Menggunakan Latent Dirichlet Allocation (Studi Kasus: Pemerintah Kota Malang)," *JPTIIK : Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 01, no. 1, pp. 2548–964, Jan. 2017, [Online]. Available: <http://j-ptik.ub.ac.id>
- [6] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *ArXiv*, vol. 2203, no. 05794, Mar. 2022, [Online]. Available: <http://arxiv.org/abs/2203.05794>

- [7] J. Devlin, M.-W. Chang, K. Lee, and K. T. Google, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *ArXiv*, vol. 1810, no. 04805, May 2019, [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [8] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," vol. 1802, no. 03426, Sep. 2020, [Online]. Available: <http://arxiv.org/abs/1802.03426>
- [9] L. McInnes and J. Healy, "Accelerated Hierarchical Density Clustering," *ArXiv*, vol. 1705, no. 07321, May 2017, doi: 10.1109/ICDMW.2017.12.
- [10] K. M. Suryaningrum, "Comparison of the TF-IDF Method with the Count Vectorizer to Classify Hate Speech," *Engineering, MAthematics and Computer Science (EMACS) Journal*, vol. 5, no. 2, pp. 79–83, May 2023, doi: 10.21512/emacsjournal.v5i2.9978.
- [11] S. Ghosh and M. S. Desarkar, "Class Specific TF-IDF Boosting for Short-text Classification: Application to Short-texts Generated during Disasters," in *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018*, Lyon: Association for Computing Machinery, Inc, Apr. 2018, pp. 1629–1637. doi: 10.1145/3184558.3191621.
- [12] Q. Huang, "Research on Keywords Extraction of Film Reviews Based on the KeyBERT Model," *Transactions on Computer Science and Intelligent Systems Research*, vol. 5, pp. 2960–2238, Aug. 2024.
- [13] N. R. B. Listyawan, N. Y. Setiawan, and M. C. Saputra, "Fakultas Ilmu Komputer Topic Modelling Pada Aktivitas Pengembangan Perangkat Lunak Menggunakan BERTopic," *JPTIIK : Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, pp. 2548–964, Jan. 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] A. Nanyonga, K. Joiner, G. Wild, and U. Turhan, "Is BERTopic Better than PLSA for Extracting Key Topics in Aviation Safety Reports?," *ArXiv*, vol. 2506, no. 06328, May 2025, Accessed: Dec. 15, 2025. [Online]. Available: <https://arxiv.org/abs/2506.06328>
- [15] D. Aryani, I. Lucia Kharisma, A. Sujjada, and K. Kamdan, "Topic Modeling of the 2024 Election Using the BERTopic Method on Detik.com News Articles," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 171–180, Aug. 2024, doi: 10.25139/inform.v9i2.8429.
- [16] U. Sri Sulistyawati, "Jurnal Manajemen dan Teknologi Decoding Big Data: Mengubah Data Menjadi Keunggulan Kompetitif dalam Pengambilan Keputusan Bisnis," *Jurnal Manajemen dan Teknologi (JMT)*, vol. 1, no. 2, pp. 58–71, Nov. 2024, doi: 10.35870/jmt.vxix.1114.
- [17] Daphney, "Topic Modeling with LDA, NMF, BERTopic, and Top2Vec: Model Comparison, Part 2," Nov. 2023, *Medium*. Accessed: Jul. 19, 2025. [Online]. Available: <https://medium.com/%40daphycarol/topic-modeling-with-lda-nmf-bertopic-and-top2vec-model-comparison-part-2-f82787f4404c>
- [18] A. Riaz, O. Abdulkader, M. J. Ikram, and S. Jan, "Exploring topic modelling: a comparative analysis of traditional and transformer-based approaches with emphasis on coherence and diversity," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 2, p. 1933, Apr. 2025, doi: 10.11591/ijece.v15i2.pp1933-1948.