

Analisis Aspek dan Sentimen Diskusi *Green Economy* dan Energi Baru Terbarukan (EBT) di Media Sosial X Menggunakan IndoBERT Multi-Task

Amanda Putri Aprilliani (105222001)

Raihan Akira Rahmaputra (105222040)

Anom Wajawening (105222029)

Gema Fitri Ramadani (105222009)

Program Studi Ilmu Komputer, Fakultas Sains dan Ilmu Komputer, Universitas Pertamina 2025

Abstrak— Transisi energi menuju Energi Baru Terbarukan (EBT) dan Green Economy di Indonesia menghadapi tantangan tidak hanya dari sisi teknis dan kebijakan, tetapi juga dalam membentuk persepsi publik. Penelitian ini bertujuan untuk menganalisis aspek dan sentimen publik terhadap isu EBT dan Green Economy melalui data media sosial X (Twitter) menggunakan pendekatan multi-task learning berbasis IndoBERT. Dataset terdiri dari 1.095 tweet yang dikumpulkan selama akhir tahun 2024. Proses penelitian melibatkan tahapan preprocessing teks, pemodelan topik dengan Latent Dirichlet Allocation (LDA), pelabelan data berbantuan LLM dan validasi manusia, serta klasifikasi aspek dan sentimen secara simultan menggunakan arsitektur IndoBERT multi-head. Hasil LDA menghasilkan 11 topik utama yang dikategorikan ke dalam lima aspek tematik, dengan fokus utama pada dampak lingkungan dan kebijakan pemerintah. Sementara itu, analisis sentimen menunjukkan bahwa mayoritas tweet bersifat netral, dengan proporsi kecil yang positif dan negatif. Temuan ini menunjukkan pentingnya strategi komunikasi yang lebih efektif dan berbasis data untuk menjembatani kesenjangan pemahaman publik dalam mendukung kebijakan transisi energi di Indonesia.

Kata kunci— Energi Terbarukan, Analisis Sentimen, IndoBERT, Topic Modeling, Multi-Task Learning.

I. PENDAHULUAN

Sebagai negara dengan populasi terbesar keempat di dunia dan ekonomi yang terus berkembang, Indonesia menghadapi tantangan signifikan dalam memenuhi kebutuhan energinya yang terus meningkat. Ketergantungan terhadap bahan bakar fosil seperti batu bara, minyak bumi, dan gas alam telah menimbulkan berbagai dampak negatif, mulai dari peningkatan emisi gas rumah kaca hingga degradasi kondisi lingkungan. Menurut[1], Indonesia menduduki peringkat ke-tujuh sebagai penghasil emisi karbon terbesar di dunia, di mana sebagian besar emisi tersebut berasal dari sektor energi dan deforestasi hutan. Pembangkit listrik dan industri masih sangat bergantung pada batu bara, sementara permintaan minyak kelapa sawit terus menjadi pendorong utama deforestasi dan emisi sektor penggunaan lahan yang meluas. Oleh karena itu, transisi menuju Energi Baru Terbarukan (EBT) tidak hanya menjadi pilihan untuk Indonesia, melainkan adalah keharusan untuk memastikan keberlanjutan ekonomi, lingkungan, dan kesejahteraan masyarakat.

EBT mencakup berbagai sumber energi seperti tenaga surya, angin, air, geothermal, dan biomassa yang tentunya memiliki potensi besar untuk dikembangkan di Indonesia, mengingat keberagaman dan kekayaan sumber daya yang ada di dalamnya. Pemerintah Indonesia telah menetapkan target untuk meningkatkan sektor EBT dalam bauran energi nasional menjadi 23% pada tahun 2025, sesuai dengan apa yang telah diatur dalam Rencana Umum Energi Nasional (RUEN) [2]. Target ini ditujukan untuk memenuhi Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals/SDGs*). Sayangnya, hingga saat ini capaian EBT masih jauh dari target yang ditetapkan.

Dilansir dari data Kementerian Energi dan Sumber Daya Mineral (ESDM) menunjukkan bahwa implementasi EBT dalam bauran energi baru mencapai sekitar 13% pada tahun 2023 [2].

Transisi energi menuju EBT tidak hanya berkaitan dengan aspek teknis dan lingkungan, tetapi juga merupakan bagian integral dari transformasi menuju Green Economy atau ekonomi hijau. Konsep Green Economy, sebagaimana didefinisikan oleh United Nations Environment Programme (UNEP), merujuk pada sistem ekonomi yang bertujuan meningkatkan kesejahteraan manusia dan kesetaraan sosial sambil mengurangi risiko lingkungan dan kelangkaan ekologis [3]. Dalam konteks Indonesia, implementasi Green Economy melalui pengembangan sektor EBT diharapkan dapat menciptakan lapangan kerja hijau, meningkatkan efisiensi energi, mengurangi ketergantungan impor energi fosil, serta mendorong inovasi teknologi berkelanjutan. Namun, keberhasilan transformasi menuju Green Economy sangat bergantung pada dukungan dan partisipasi aktif masyarakat, yang memerlukan pemahaman dan persepsi positif manfaat ekonomi dan lingkungan dari transisi energi berkelanjutan.

Indonesia menghadapi tantangan kompleks di level pemahaman dan persepsi publik terhadap energi berkelanjutan dan *green economy*. Kesenjangan antara kebijakan pemerintah dan pemahaman publik menjadi salah satu hambatan utama dalam implementasi transisi energi. Survei yang dilakukan oleh Institute for Essential Services Reform (IESR) terhadap 1.000 responden Indonesia menunjukkan paradoks yang menarik: meskipun 80% responden menganggap perubahan iklim sebagai masalah yang sangat mendesak, hanya 31% yang menilai bahwa pemerintah telah merespons masalah tersebut dengan memadai [4]. Studi yang sama juga mengungkapkan adanya dukungan publik yang kuat terhadap energi terbarukan, dengan 68% responden memilih tenaga surya, 60% memilih tenaga air, dan 39% memilih tenaga angin sebagai sumber energi yang harus diprioritaskan dalam pembangkitan listrik Indonesia [4]. Perbedaan antara tingginya kesadaran publik dan korporat dengan rendahnya persepsi terhadap langkah pemerintah serta lambatnya implementasi konkret EBT menunjukkan adanya *communication gap* yang signifikan dalam strategi transisi energi di Indonesia, yang memerlukan pendekatan analisis lebih mendalam untuk memahami diskusi publik di media sosial.

Media sosial memiliki peran penting sebagai *platform* diskusi publik yang memungkinkan masyarakat untuk berbagi informasi, opini, dan pandangan terkait isu-isu tertentu, termasuk transisi Energi Baru Terbarukan (EBT) dan *green economy*. Peran media sosial adalah menjembatani komunikasi antara masyarakat, akademisi, dan pembuat kebijakan, serta memberikan peluang untuk menganalisis tren transisi energi di kawasan ASEAN, salah satunya negara Indonesia [5]. Platform seperti X (Twitter) menjadi penting dimana masyarakat mengekspresikan pandangan, kekhawatiran, dan harapan mereka terhadap kebijakan EBT. Analisis sentimen aspek diskusi di media sosial dapat memberikan *insight* berharga tentang persepsi publik, identifikasi *communication gap*, dan membantu pemerintah merancang strategi komunikasi yang lebih efektif untuk mendorong terwujudnya target energi terbarukan dalam SDGs.

Beberapa penelitian terdahulu telah mengeksplorasi penggunaan analisis sentimen untuk memahami persepsi publik terhadap isu energi dan lingkungan melalui data media sosial. [6] mengembangkan model berbasis RoBERTa untuk menganalisis sentimen publik terhadap energi surya menggunakan 266,686 *tweets* dan mencapai akurasi 80,2% dengan hasil temuan variasi sentimen yang signifikan antar wilayah berdasarkan karakteristik kebijakan dan kondisi pasar energi. Dalam konteks Indonesia, meskipun telah tersedia model bahasa IndoBERT yang dikembangkan khusus untuk bahasa Indonesia, penelitian yang secara spesifik menganalisis diskusi EBT dan Green Economy di media sosial Indonesia menggunakan pendekatan *multi-task learning* masih sangat terbatas.

Kondisi ini menciptakan *research gap* yang signifikan dalam pemahaman diskusi publik Indonesia tentang transisi berkelanjutan.

Penelitian ini bertujuan untuk menganalisis aspek dan sentimen diskusi Green Economy dan Energi Baru Terbarukan (EBT) di media sosial X dengan menggunakan pendekatan *multi-task learning* IndoBERT. Melalui kombinasi *topic modeling* menggunakan Latent Dirichlet Allocation (LDA) untuk mengidentifikasi topik-topik utama diskusi, *LLM-assisted labeling* dengan validasi manual untuk pembentukan *ground truth*, dan pengembangan model IndoBERT *multi-task* yang dapat mengklasifikasikan sentimen dan aspek secara bersamaan, penelitian ini berupaya mengidentifikasi pola sentimen publik, tema diskusi yang dominan, serta *communication gap* dalam strategi transisi energi Indonesia. Hasil analisis ini diharapkan dapat memberikan *insights data-driven* untuk perbaikan strategi komunikasi kebijakan EBT dan mendukung percepatan implementasi energi berkelanjutan di Indonesia melalui pemahaman yang lebih komprehensif terhadap diskusi publik di media sosial.

II. KAJIAN PUSTAKA

A. Sentimen Analisis untuk Kebijakan Publik

Analisis sentimen atau yang biasa disebut juga sebagai *opinion mining* adalah bidang studi komputasional yang fokus pada ekstraksi, identifikasi, kuantifikasi, dan studi mengenai opini, sentimen, serta emosi yang diekspresikan dalam data teks [7]. Proses ini dasarnya bertujuan untuk mengklasifikasikan polaritas opini yang terkandung dalam sebuah unit teks—baik pada level dokumen, kalimat, maupun aspek—ke dalam kategori positif, negatif, atau netral [8]. Seiring dengan ledakan data tidak terstruktur dari internet, terutama dari media sosial, analisis sentimen telah berkembang dari proses manual yang terbatas menjadi teknik otomatis yang mampu mengolah jutaan opini dalam waktu singkat.

Penerapan analisis sentimen menjadi sangat relevan dalam konteks perumusan kebijakan publik, terutama di sektor energi yang kompleks. Kemampuan untuk memahami persepsi publik secara *real-time* memberikan masukan yang tak ternilai bagi pemerintah [8]. Sebagai contoh konkret, sebuah studi menggunakan analisis sentimen dan korelasi pada data media sosial untuk menyelidiki pandangan publik terhadap hubungan antara energi terbarukan dan energi nuklir. Hasilnya menunjukkan sebuah dualisme yang menarik, yakni publik cenderung melihat keduanya sebagai sumber energi yang saling melengkapi (*complementary*) saat mengekspresikan sentimen positif, namun melihatnya sebagai sumber energi yang saling menggantikan (*substitute*) ketika mengekspresikan sentimen negatif [9]. Temuan seperti ini tergolong sangat penting karena dapat menginformasikan penyusunan rencana dan kebijakan energi nasional oleh pemegang kuasa atau pembuat keputusan yang lebih selaras dengan dinamika opini masyarakat.

B. Aspect-Based Analysis

Analisis Sentimen Berbasis Aspek atau *Aspect Based Sentiment Analysis* (ABSA) adalah salah satu sub-bidang penting dalam *Natural Language Processing* (NLP) yang berfokus pada analisis opini yang lebih mendalam. Meskipun analisis sentimen pada level dokumen (teks langsung) memberikan gambaran umum, seringkali masih terlalu kasar untuk menghasilkan wawasan yang dapat ditindaklanjuti. Di sinilah ABSA berperan, dengan tujuan untuk mengidentifikasi opini terhadap entitas dan aspek-aspek spesifik dari entitas tersebut [10]. Sebagai contoh, dalam sebuah ulasan "Kebijakan EBT bagus, tetapi implementasinya lambat," ABSA dapat mengidentifikasi

sentimen positif terhadap aspek "Kebijakan EBT" dan sentimen negatif terhadap aspek "Implementasi". Tugas atau proses ini secara formal distandardisasi melalui *benchmark* internasional seperti *SemEval-2016 Task 5*, yang mendefinisikan sub-tugas inti ABSA, termasuk mengidentifikasi entitas target, aspek kategorinya, serta polaritas sentimen untuk setiap aspek [10].

Untuk dapat melakukan ABSA, model komputasional harus mampu menangani interaksi kompleks antara sebuah aspek dan kata-kata lain dalam konteks kalimatnya. Sementara itu, pendekatan modern seringkali memanfaatkan model *deep learning* yang canggih. Sebagai contoh, sebuah penelitian terbaru mengusulkan metode ABSA yang mengombinasikan mekanisme atensi multi-kanal (*multi-channel attention*) dengan strategi pembelajaran multi-tugas (*multi-task learning*) [11]. Pendekatan ini secara eksplisit dirancang untuk meningkatkan akurasi identifikasi istilah aspek dan klasifikasi sentimennya secara simultan. Kemampuan untuk membedah opini publik hingga ke level aspek dengan akurasi tinggi inilah yang membuat ABSA menjadi alat analisis kebijakan yang sangat kuat.

C. Multi-Task Learning dalam NLP

Secara tradisional, tugas-tugas kompleks dalam NLP seperti ABSA sering dipecah menjadi beberapa sub-tugas yang diselesaikan secara terpisah. Namun, pendekatan ini mengabaikan fakta bahwa banyak tugas NLP secara inheren saling terkait. Paradigma *Multi-Task Learning* (MTL) hadir sebagai solusi dengan cara melatih sebuah model tunggal untuk menyelesaikan beberapa tugas secara bersamaan [12]. Menurut tinjauan fundamental oleh Ruder [13], MTL meningkatkan generalisasi model dengan memanfaatkan domain informasi dari sinyal pelatihan tugas-tugas terkait. Ini berfungsi sebagai bentuk regularisasi implisit yang mengurangi risiko *overfitting* serta sebagai *inductive bias* yang mengarahkan model untuk mempelajari representasi yang lebih kuat dan universal.

Secara arsitektural, implementasi MTL yang paling umum dalam *deep learning* ialah menggunakan pola yang dikenal sebagai *HydraNet* [13]. Arsitektur ini terdiri dari satu "badan" (*body*) bersama yang berfungsi sebagai *encoder* atau *backbone* untuk mengekstraksi fitur, dan beberapa "kepala" (*heads*) yang merupakan lapisan-lapisan *output* terpisah untuk setiap tugas spesifik. Dalam konteks proyek ini, model IndoBERT berfungsi sebagai "badan" yang menghasilkan representasi kontekstual dari teks input, sementara dua kepala klasifikasi terpisah—satu untuk sentimen dan satu untuk aspek—berfungsi sebagai "kepala" dari *HydraNet* tersebut. Pola ini sangat efisien karena sebagian besar parameter komputasi (di dalam badan IndoBERT) digunakan bersama oleh kedua tugas.

Secara teknis, proses pelatihan model MTL melibatkan optimalisasi fungsi kerugian (*loss function*) gabungan. Total kerugian dihitung sebagai jumlah tertimbang (*weighted sum*) dari kerugian masing-masing tugas [13]. Dengan meminimalkan kerugian gabungan ini, model secara simultan belajar untuk menjadi ahli dalam setiap tugas. Pendekatan ini sangat cocok untuk ABSA, di mana tugas-tugas seperti identifikasi aspek dan klasifikasi sentimennya dapat dilatih secara bersamaan untuk meningkatkan akurasi. Untuk memastikan keberhasilan arsitektur canggih ini pada data berbahasa Indonesia, penggunaan IndoBERT sebagai fondasi menjadi sangat penting. Kemampuan IndoBERT untuk memahami kekhasan linguistik pada teks media sosial Indonesia memberikan dasar yang kuat bagi kepala-kepala klasifikasi untuk bekerja secara efektif [14].

D. Model Berbahasa Indonesia (IndoBERT)

Meskipun bahasa Indonesia digunakan hampir 200 juta orang dan merupakan bahasa ke-10 yang paling banyak digunakan di dunia, bahasa Indonesia masih kurang terwakili dalam penelitian *Natural Language Processing* (NLP) [15]. Dalam konteks model bahasa berbasis *deep learning*, sebagian besar model seperti BERT [16] hanya berfokus pada bahasa Inggris, dengan sedikit perhatian pada bahasa lain seperti bahasa Indonesia [17]. Oleh karena itu, diadakan pengembangan model monolingual BERT untuk bahasa Indonesia (IndoBERT) untuk konteks NLP berbasis bahasa Indonesia. IndoBERT adalah model berbasis transformer dalam gaya BERT, namun dilatih murni sebagai *masked language* model menggunakan framework Huggingface, mengikuti konfigurasi default BERT-base (*uncased*) [15]. Model ini memiliki spesifikasi teknis sebagai berikut [15]:

- Jumlah layer: 12 hidden layers dengan dimensi 768
- Attention heads: 12 multi-head attention
- Feed forward layers: 3072 dimensi
- Vocabulary size: 31.923 Indonesian WordPiece vocabulary
- Konfigurasi training: 512 token per batch

IndoBERT dilatih menggunakan total 220 juta kata yang dikumpulkan dari tiga sumber utama, Wikipedia Indonesia (74 juta kata), artikel berita dari Kompas, Tempo, dan Liputan6 (55 juta kata total), dan Indonesian Web Corpus (90 Juta kata) [15]. IndoBERT telah dievaluasi menggunakan IndoLEM, sebuah *benchmark* Indonesia yang terdiri dari tujuh tugas bahasa Indonesia, meliputi morfologi-sintaksis, semantik, dan wacana [18]. Eksperimen menunjukkan bahwa IndoBERT mencapai performa *state-of-the-art* (yang terbaik saat ini) pada sebagian besar tugas dalam IndoLEM [15].

Dalam konteks analisis sentimen, IndoBERT menunjukkan keunggulan signifikan dibanding model multilingual. Studi perbandingan kinerja delapan model BERT *pre-trained* untuk analisis sentimen di sepuluh bahasa daerah Indonesia menunjukkan bahwa model yang dilatih pada data Indonesia, khususnya IndoBERT, mengungguli model multilingual BERT [15]. Model IndoBERT tersedia melalui HuggingFace dengan *identifier* “indolem/indobert-base-uncased” [18], memudahkan implementasi dalam proyek penelitian. Kombinasi antara performa *state-of-the-art* dan kemudahan akses menjadikan IndoBERT pilihan optimal untuk analisis teks berbahasa Indonesia, khususnya dalam domain energi terbarukan dan Green Economy.

E. Topic Modeling untuk Sosial Media dengan LDA

Topic Modeling adalah teknik *machine learning unsupervised* yang digunakan untuk mengidentifikasi struktur tematik tersembunyi dalam koleksi dokumen besar [19]. Dalam konteks media sosial, *topic* modeling memungkinkan analisis otomatis terhadap tema-tema yang dibicarakan pengguna, memberikan wawasan berharga untuk berbagai bidang seperti pemasaran, keamanan, pendidikan, dan manajemen [20]. *Topic Modeling* dengan *Latent Dirichlet Allocation* (LDA) adalah teknis analisis konten komputasional yang dapat digunakan untuk menyelidiki struktur tematik ‘tersembunyi’ dan koleksi teks tertentu [21]. Metode ini mengasumsikan bahwa setiap dokumen merupakan campuran dari beberapa topik, dan setiap topik didefinisikan sebagai distribusi probabilitas atas kosakata tetap [22].

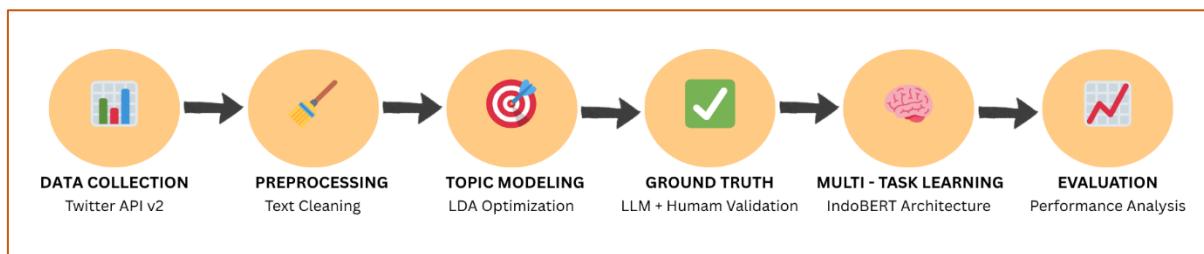
LDA merupakan model probabilistik generatif yang menganalisis kata-kata dalam setiap dokumen dan menghitung distibusi probabilitas bersama antara yang teramat (kata-kata dalam dokumen) dan yang tidak

teramat (topik yang tersembunyi) [22]. Model ini menggunakan pendekatan ‘Bag of Words’ di mana semantik dan makna kalimat tidak dievaluasi, melainkan metode ini mengevaluasi frekuensi kata dengan asumsi bahwa kata-kata paling sering dalam suatu topik akan mewakili esensi topik tersebut [22]. Dataset media sosial dengan 90.257 record menunjukkan bahwa LDA telah terbukti efektif dalam mendeteksi aspek topik dan mengekstrak dinamikanya dari waktu ke waktu, di mana lima topik utama berhasil ditemukan secara akurat menurut komentar para ahli domain [20].

Penerapan *topic modeling* pada data media sosial menghadapi tantangan unik dibandingkan dengan dokumen teks panjang tradisional. Karakter pendek, berat teks, dan sifat tidak terstruktur dari konten media sosial sering menyebabkan tantangan metodologis dalam pengumpulan dan analisis data [23]. Dalam konteks analisis diskusi Energi Baru Terbarukan (EBT) di Indonesia, *topic modeling* dengan LDA memberikan pendekatan yang baik untuk identifikasi tema dominan, *temporal analysis*, memberikan masukan untuk strategi komunikasi kebijakan yang lebih efektif, dan pemahaman persepsi dan kekhawatiran masyarakat terhadap transisi energi.

III. METODE PENELITIAN

A. Kerangka Penelitian



Gambar 1 Pipeline Flow Penelitian

Penelitian ini menggabungkan teknik *unsupervised learning* untuk eksplorasi topik dan supervised learning untuk klasifikasi multi-task. Kerangka penelitian dirancang untuk menjawab tiga rumusan masalah melalui enam tahapan sistematis yang saling terhubung. Tiga rumusan masalah tersebut adalah:

1. Bagaimana pola sentimen publik terhadap diskusi EBT di media sosial X?
2. Aspek apa saja yang mendominasi Green Economy dan EBT di Indonesia?
3. Bagaimana efektivitas IndoBERT multi-task learning untuk analisis aspek dan sentimen secara simultan?

Alur penelitian dimulai dari pengumpulan data diskusi EBT di media sosial X, dilanjutkan dengan preprocessing untuk membersihkan dan menormalisasi teks mentah. Tahap ketiga adalah penggunaan topic modeling dengan LDA untuk mengidentifikasi tema-tema dominan secara otomatis. Output dari LDA kemudian digunakan untuk membentuk kategori aspek yang akan menjadi dasar untuk ground truth generation menggunakan bantuan Large Language Model (LLM) dengan validasi manusia. Tahap kelima ialah mengimplementasikan multi-task learning dengan IndoBERT untuk klasifikasi sentimen dan aspek secara bersamaan. Terakhir, evaluasi dilakukan untuk menganalisis performa model dan menganalisis hasil model.

B. Pengumpulan Data dan Preprocessing

Pengumpulan data dalam penelitian ini menggunakan Twitter API V2 untuk mengakses data tweet terkait Energi Baru Terbarukan (EBT) dan Green Economy di Indonesia. Proses pengumpulan data dilakukan dengan memanfaatkan library Python Tweepy sebagai antarmuka untuk mengakses Twitter API.

Untuk mendapatkan data yang relevan, penelitian ini menggunakan kombinasi kata kunci berikut dalam query pencarian:

- “renewable energy” OR “energi terbarukan” OR “EBT” OR “NRE”
- "energi baru" OR "solar panel" OR "panel surya" OR "wind power"
- "geothermal" OR "biomass"
- "ESDM" OR "Bahlil Lahadalia" OR "energi hidro"
- "PLTP" OR "energi bersih" OR "climate change" OR "deforestasi"
- "transisi energi" OR "energi hijau" OR "keberlanjutan energi"

Query pencarian juga dilengkapi dengan filter bahasa Indonesia (lang:id) dan mengecualikan retweet serta reply untuk menghindari duplikasi data. Beberapa *hashtag* seperti #SwasembadaEnergi, #PresidenPrabowo, #KetahananPangan, #AstaCita, #TransisiEnergi, dan #SobatNKRI juga dikeluarkan dari pencarian untuk menghindari bias politik. Data tweet dikumpulkan dalam rentang waktu 28 Desember 2024 hingga 2 Januari 2025. Pemilihan periode ini memungkinkan untuk menangkap dinamika diskusi terkini tentang EBT di Indonesia.

Melalui proses *crawling* data menggunakan Twitter API, diperoleh total 1095 tweet yang memenuhi kriteria pencarian. Setiap tweet yang dikumpulkan memiliki atribut UserID, Date Created, Tweet Text, dan Tweet URL. Data yang terkumpul disimpan dalam format CSV untuk memudahkan proses pengolahan data lebih lanjut. Langkah selanjutnya adalah melakukan *preprocessing text* untuk mempersiapkan data bagi analisis topik menggunakan *Latent Dirichlet Allocation* (LDA).

Tahap preprocessing dilakukan untuk mengubah data tidak terstruktur menjadi data yang lebih terstruktur dan berkualitas tinggi, meliputi:

- Penghapusan duplikat untuk memastikan keunikan data
- URL cleaning dengan replacement token standar
- User mention normalization [USER] token
- Hastag cleaning
- Slang normalization menggunakan custom Indonesian dictionary
- Whitespace normalization dan punctuation cleaning

| Tahap | Hasil |
|-----------------------|---|
| Input | "Nangis bet 🥺🥺🥺 Buat tugas kuliah aja gue sampe pusing eneg nyari akibat dari #deforestasi sama penanaman #sawit, ini buat negara ga research sama sekali??? @jokowi @kementerianesdm https://t.co/G8imeKJ3YV " |
| URL Cleaning | "Nangis bet 🥺🥺🥺 Buat tugas kuliah aja gue sampe pusing eneg nyari akibat dari #deforestasi sama penanaman #sawit, ini buat negara ga research sama sekali??? @jokowi @kementerianesdm" |
| User Mention Cleaning | "Nangis bet 🥺🥺🥺 Buat tugas kuliah aja gue sampe pusing eneg nyari akibat dari #deforestasi sama penanaman #sawit, ini buat negara ga research sama sekali??? [USER] [USER]" |
| Hastag cleaning | " Nangis bet 🥺🥺🥺 Buat tugas kuliah aja gue sampe pusing eneg nyari akibat dari sama penanaman , ini buat negara ga research sama sekali??? [USER] [USER]" |

| | |
|---------------------------------------|--|
| Excessive punctuation removal | " Nangis bet 🥺🥺🥺 Buat tugas kuliah aja gue sampe pusing eneg nyari akibat dari sama penanaman , ini buat negara ga research sama sekali? [USER] [USER]" |
| Whitespace Normalization | "Nangis bet 🥺🥺🥺 Buat tugas kuliah aja gue sampe pusing eneg nyari akibat dari sama penanaman , ini buat negara ga research sama sekali? [USER] [USER]" |
| Indonesian Slang Normalization | Nangis bet 🥺🥺🥺 Buat tugas kuliah aja saya sampe pusing eneg nyari akibat dari deforestasi sama penanaman sawit, ini buat negara tidak research sama sekali? [USER] [USER]" |
| Final output | "nangis bet 🥺🥺🥺 buat tugas kuliah aja saya sampe pusing eneg nyari akibat dari deforestasi sama penanaman sawit, ini buat negara tidak research sama sekali?" |

Dalam proses *preprocessing* ini, emoji sengaja dipertahankan karena memiliki nilai penting untuk analisis sentimen, di mana emoji dapat memberikan informasi tambahan tentang emosi dan sikap pengguna terhadap topik EBT. Kata bahasa Inggris seperti "research" juga dipertahankan karena sudah umum digunakan dalam konteks bahasa Indonesia dan merupakan bagian dari covabulary yang dipahami oleh IndoBERT. Slang dictionary yang digunakan dalam tahap ini berfokus pada normalisasi kata-kata informal Indonesia seperti "gue" menjadi "saya" dan "ga" menjadi tidak. Hasil akhir dari *preprocessing* ini menghasilkan teks yang telah dibersihkan dan dinormalisasi yang siap untuk tahap selanjutnya yaitu pemodelan topik dengan LDA.

C. Topic Modeling dengan LDA

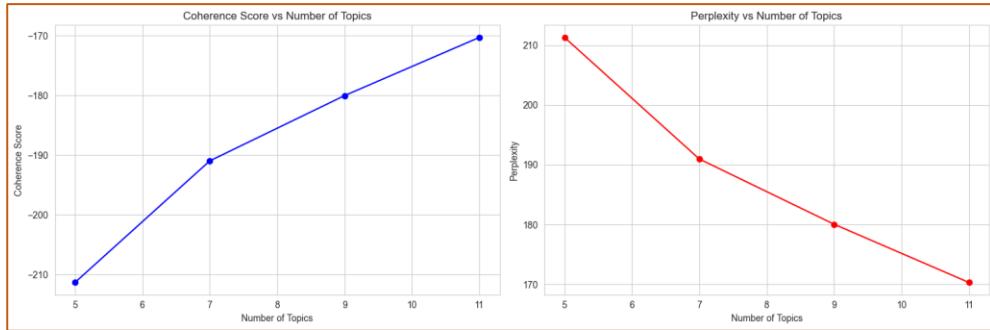
Penelitian ini menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik-topik dalam tweet terkait EBT dan Green Economy. Proses pemodelan topik dilakukan dalam beberapa tahap :

1. Persiapan data untuk LDA
 - a. Konstruksi Document-Term Matrix
 1. Mengonversi tweet hasil *preprocessing* menjadi format numerik menggunakan CountVectorizer dari scikit-learn
 2. Mengonfigurasi parameter vectorizer dengan max_features=800 untuk membatasi ukuran vocabulary
 3. Menerapkan ngram_range=(1,2) untuk menginklusikan unigrams dan bigrams
 4. Mengatur min_df=3 dan max_df=0.8 untuk filtering term frequency
 - b. Preprocessing untuk Topic Modeling
 1. Mengonversi teks ke representasi *Bag of Words* (BoW) format
 2. Menyimpan informasi frekuensi kemunculan setiap kata dalam dokumen
 3. Memvalidasi ukuran vocabulary (13,633 tokens) dan distribusi term frequency
2. Penentuan jumlah topik optimal

Penentuan jumlah topik merupakan langkah krusial dalam pemodelan LDA karena akan memengaruhi kualitas dan hasil analisis. Untuk menemukan jumlah topik yang optimal, penelitian ini melakukan pengujian secara bertahap dengan beberapa nilai K (jumlah topik) yang berbeda dan mengevaluasi model menggunakan *coherence score* dan perplexity.

1. Membuat model dengan range jumlah topik (5-12)
2. Interval pengujian: step 2 yaitu 5, 7, 9, 11
3. Melakukan evaluasi model dengan cara menghitung *coherence score* untuk setiap model dengan metrik C_v.

4. Menganalisis perplexity sebagai metrik tambahan untuk mengikuti kemampuan prediksi model
5. Hasil optimal diperoleh 11 topik dengan coherence score 0.6576



Gambar 2 Evaluasi Topik Optimal LDA

3. Training dan pembuatan model LDA

Berdasarkan hasil penentuan jumlah optimal (11 topik), model LDA dibangun menggunakan LatentDirichletAllocation dari scikit-learn dengan konfigurasi random_state=42, max_iter=100, learning_method='batch', dan n_jobs=-1 untuk paralelisasi. Model kemudian divalidasi melalui perhitungan final *coherence score* C_v (0.6576) dan UMass (-3.3650), analisis *perplexity* (170.3105), dan evaluasi kualitas topik melalui interpretasi kata-kata dominan dalam setiap topik. Berikut adalah 11 topik yang didapatkan dari LDA:

| Topic ID | Top Words | Word Weights | Core Theme | Interpretasi |
|----------|--|---|----------------------|--|
| Topic 0 | kita, dari, untuk, dengan, sebuah, baru, ini, perubahan | kita: 92.09, dari: 87.69, untuk: 85.80, dengan: 84.58, sebuah: 82.09 | OTHER | Topik umum dengan kata-kata generic dan stop words |
| Topic 1 | baru, tahun, energi, dan, tahun baru, energi baru, dengan, yang | baru: 236.64, tahun: 124.05, energi: 112.40, dan: 99.86, tahun baru: 86.13 | ENERGY_TECHNOLOGY | Diskusi tentang teknologi energi baru dan perkembangan tahunan |
| Topic 2 | kelapa, kelapa sawit, dan, energi, baru, terbarukan, energi baru, usah | kelapa: 87.34, kelapa sawit: 83.65, dan: 74.86, energi: 69.20, baru: 37.46 | ENVIRONMENTAL_IMPACT | Diskusi kelapa sawit dalam konteks energi terbarukan |
| Topic 3 | rakyat, bumi, gas, untuk rakyat, untuk, gas bumi, energi, yang | rakyat: 76.23, bumi: 73.57, gas: 68.28, untuk rakyat: 65.18, untuk: 63.29 | ENERGY_ACCESS | Akses gas bumi untuk kepentingan rakyat |
| Topic 4 | erupsi, di, user, via, via user, hari, dan, pada | erupsi: 166.09, di: 121.71, user: 103.19, via: 88.09, via user: 88.09 | OTHER | Noise dan artifak dari preprocessing (via user, erupsi) |
| Topic 5 | untuk, rakyat, gas, untuk rakyat, gas bumi, bumi, ppn, no | untuk: 92.51, rakyat: 91.95, gas: 80.90, untuk rakyat: 79.00, gas bumi: 79.10 | ENERGY_ACCESS | Kebijakan gas bumi untuk rakyat dan isu PPN |
| Topic 6 | energi, di, indonesia, pln, terbarukan, transisi, transisi energi, dan | energi: 146.14, di: 85.25, indonesia: 53.84, pln: 46.43, terbarukan: 44.67 | ENERGY_TECHNOLOGY | Transisi energi terbarukan di Indonesia dan peran PLN |

| | | | | |
|-----------------|--|---|----------------------|--|
| Topic 7 | bahlil, lahadalia, bahlil lahadalia, menteri, esdm, menteri esdm, golkar, partai | bahlil: 78.09, lahadalia: 63.09, bahlil lahadalia: 63.09, menteri: 59.09, esdm: 44.37 | GOVERNMENT_POLICY | Diskusi Menteri ESDM Bahlil Lahadalia dan politik |
| Topic 8 | deforestasi, sawit, tidak, yang, hutan, ini, itu, di | deforestasi: 419.36, sawit: 242.81, tidak: 228.82, yang: 222.90, hutan: 145.09 | ENVIRONMENTAL_IMPACT | Isu deforestasi, sawit, dan dampak terhadap hutan |
| Topic 9 | esdm, dan, di, dengan, kementerian, listrik, daya, untuk | esdm: 67.10, dan: 59.29, di: 57.85, dengan: 41.90, kementerian: 36.19 | GOVERNMENT_POLICY | Kementerian ESDM, listrik, dan kebijakan energi |
| Topic 10 | prabowo, sawit, deforestasi, dan, tanpa, lahan, penghargaan, lahan sawit | prabowo: 73.34, sawit: 63.37, deforestasi: 39.82, dan: 37.28, tanpa: 33.09 | ENVIRONMENTAL_IMPACT | Prabowo, sawit, deforestasi, dan penghargaan lahan |

4. Kategorisasi dan Mapping Topik

Hasil 11 topik optimal kemudian dikategorisasi menjadi 5 core themes berdasarkan interpretasi semantik kata-kata dominan yang dilakukan secara manual oleh peneliti:

1. ENERGY_TECHNOLOGY (Topics 1, 6): Fokus pada teknologi energi terbarukan dan infrastruktur
2. ENVIRONMENTAL_IMPACT (Topics 2, 8, 10): Diskusi dampak lingkungan, deforestasi, dan kelapa sawit
3. GOVERNMENT_POLICY (Topics 7, 9): Kebijakan pemerintah, regulasi, dan aspek administratif
4. ENERGY_ACCESS (Topics 3, 5): Akses energi untuk masyarakat dan pemerataan energi
5. OTHER (Topics 0, 4): Topik umum dan noise yang tidak masuk kategori spesifik

D. Proses Pembuatan Data Ground Truth

Penelitian ini mengadopsi pendekatan *hybrid* yang menggabungkan *LLM-assisted labeling* dengan *human validation* untuk menghasilkan *ground truth* yang berkualitas tinggi. Pendekatan ini dipilih karena keterbatasan sumber daya untuk manual *annotation* pada *dataset* berskala besar, sekaligus mempertahankan kualitas dan reliabilitas label melalui validasi manusia. Proses dimulai dengan *automatic labeling* menggunakan Large Language Model (LLM) berjenis Claude Sonnet 4, dilanjutkan dengan validasi oleh *human annotators* untuk memastikan akurasi dan konsistensi label.

Berikut adalah prompt yang digunakan saat melakukan labeling:

Saya membutuhkan bantuan untuk melakukan aspect-sentiment analysis pada data tweets tentang Green Economy dan Energi Baru Terbarukan (EBT) di Indonesia. Data telah dipreprocessing dan disimpan dalam bentuk csv (ready_labelling.csv) dengan detail data:
<class 'pandas.core.frame.DataFrame'>

RangeIndex: 949 entries, 0 to 948

Data columns (total 4 columns):

| # Column | Non-Null Count | Dtype |
|------------------------|----------------|--------|
| 0 User ID | 949 non-null | int64 |
| 1 Date Created | 949 non-null | object |
| 2 Tweet URL | 949 non-null | object |
| 3 Tweet Text_processed | 949 non-null | object |

Tweet yang harus anda labelkan adalah kolom Tweet Text_processed Pemodelan topik LDA yang sudah dilakukan menghasilkan:

Definisi Core Themes:

- ENERGY_TECHNOLOGY (Topics 1, 6): Fokus pada aspek teknologi energi terbarukan dan transisi energi Kata kunci: energi, energi baru, terbarukan, transisi, transisi energi, pln, di indonesia, tahun baru, pembangkit, listrik, daya
- ENVIRONMENTAL_IMPACT (Topics 2, 8, 10): Fokus pada dampak lingkungan, terutama deforestasi dan kelapa sawit Kata kunci: deforestasi, sawit, kelapa sawit, hutan, tanpa, lahan, lahan sawit, prabowo, kelapa, keberlanjutan, lingkungan, itu, yang, tidak
- GOVERNMENT_POLICY (Topics 7, 9): Fokus pada kebijakan pemerintah dan tokoh politik Kata kunci: bahlil, lahadalia, menteri, esdm, menteri esdm, kementerian, golkar, partai, listrik, dengan, di, dan, untuk, daya
- ENERGY_ACCESS (Topics 3, 5): Fokus pada akses energi untuk masyarakat Kata kunci: rakyat, bumi, gas, untuk rakyat, gas bumi, ppn, no, untuk, yang, energi
- OTHER (Topics 0, 4): Topik umum atau noise Kata kunci: kita, dari, untuk, dengan, sebuah, baru, ini, perubahan, erupsi, di, user, via, via user, hari, pada

Konteks Penelitian:

Saya sedang melakukan riset tentang “Pemodelan Topik LDA dan Analisis Sentimen untuk Optimalisasi Kebijakan Green Economy dan Energi Baru Terbarukan (EBT) di Indonesia: Studi Media Sosial X”. Saya telah melakukan pemodelan topik dengan LDA dan mendapatkan 11 topik yang dikelompokkan ke dalam 5 aspek yang telah dijelaskan di atas.

Tugas Anda:

Untuk setiap tweet yang saya berikan:

1. Identifikasi ASPEK TOPIK/TEMA yang paling relevan (pilih dari daftar di atas)
2. Tentukan SENTIMEN untuk core theme tersebut (Positif, Negatif, atau Netral)
3. Berikan ALASAN singkat untuk penilaian sentimen tersebut
4. Tentukan RELEVANSI tweet dengan core theme (Tinggi, Sedang, Rendah)

Contoh Format Output yang Diinginkan:

Tweet: “Gas bumi adalah solusi energi bersih yang terjangkau untuk rakyat indonesia. dengan kebijakan bebas ppn sesuai pp no. 49 tahun 2022, pgn memastikan tidak ada tambahan beban pajak bagi pelanggan”

ASPEK: ENERGY_ACCESS

SENTIMEN: Positif ALASAN: Tweet menyoroti keterjangkauan gas bumi dan kebijakan bebas PPN yang mengurangi beban finansial konsumen

RELEVANSI: Tinggi (sangat relevan karena membahas gas bumi untuk rakyat dan kebijakan terkait PPNNya)

Petunjuk Penting:

1. Fokus pada konteks EBT dan green economy di Indonesia
2. Pertimbangkan nuansa bahasa Indonesia, termasuk slang dan ungkapan lokal
3. Berikan analisis yang objektif berdasarkan teks, bukan opini pribadi
4. Jika tweet tidak jelas, ambigu, atau tidak mengandung informasi yang cukup, Anda dapat menandai sentimen sebagai “Netral”
5. Pertimbangkan konteks khusus Indonesia terkait isu sawit, deforestasi, dan kebijakan energi saat ini

Output (hasil labeling) buatkan dalam bentuk dokumen dan format csv seperti ini:

404,“sobat nusantara, pln nusantara power (pln np) borong 14 penghargaan dalam malam apresiasi kinerja stakeholder energi baru terbarukan dan konservasi energi (ebt) 2024 yang diselenggarakan oleh kementerian esdm.”,ENERGY TECHNOLOGY,Positif,“Melaporkan pencapaian PLN meraih 14 penghargaan energi terbarukan”

405,“menyambut tahun baru dengan menaikan pajak ke 12% dan siap siap membabat hutan dengan barley dan rye. kedua tumbuhan itu sama-sama punya daun, tidak ada yang namanya deforestasi, omon-omon saja”,ENVIRONMENTAL_IMPACT,Negatif,“Sarkasme mengkritik logika bahwa tanaman berdaun tidak menyebabkan deforestasi”

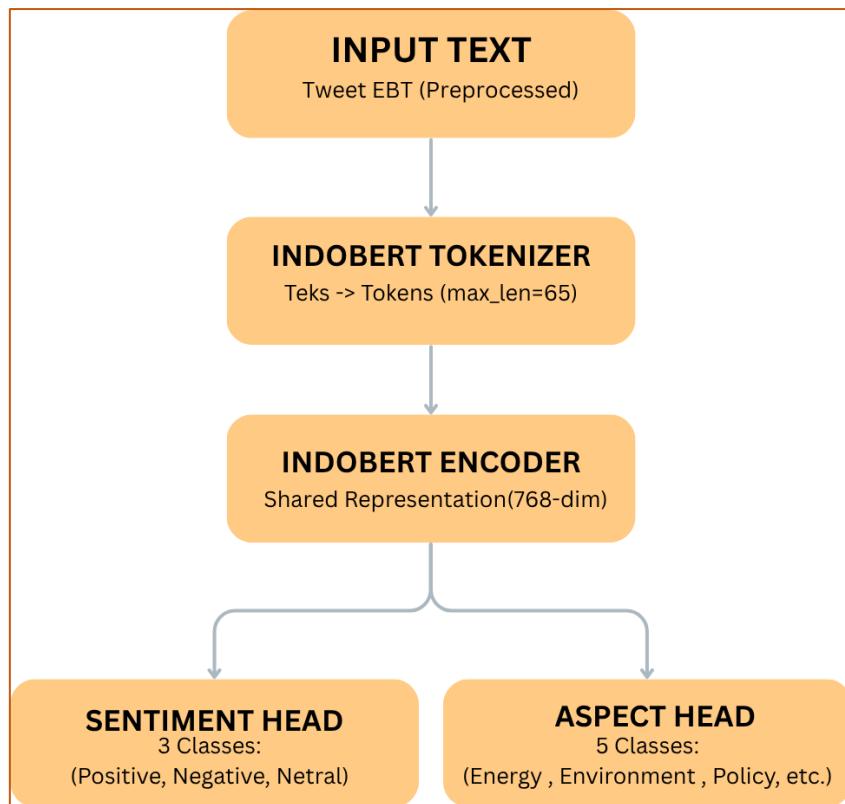
###Data untuk dianalisis adalah...

E. Arsitektur Multi-Task IndoBERT

Penelitian ini mengimplementasikan arsitektur multi-task learning menggunakan IndoBERT sebagai shared backbone untuk melakukan klasifikasi sentimen dan aspek secara bersamaan. Pendekatan ini memungkinkan model untuk mempelajari representasi bersama yang dapat digunakan untuk kedua tugas, meningkatkan efisiensi dan potensi akurasi dibandingkan dengan *training* model terpisah.

Arsitektur multi-task IndoBERT dimulai dengan input berupa tweet EBT yang telah melalui tahap preprocessing, kemudian diproses melalui IndoBERT tokenizer untuk mengkonversi teks menjadi format numerik dengan maksimal panjang 65 token berdasarkan analisis distribusi data. Tokenizer menghasilkan input_ids, attention_mask, dan token_type_ids yang kemudian diteruskan ke IndoBERT encoder sebagai shared backbone. IndoBERT encoder menggunakan model indolem/indobert-base-uncased dengan 12 transformer layers yang menghasilkan shared representation berupa contextualized embeddings dengan dimensi 768 dari [CLS] token, yang berfungsi sebagai representasi semantik bersama untuk kedua tugas klasifikasi.

Output dari shared encoder kemudian diteruskan ke dua classification heads terpisah yang berjalan secara paralel untuk menangani tugas yang berbeda. Sentiment head terdiri dari linear layer yang memetakan representasi 768 dimensi menjadi 3 kelas output (Positif, Negatif, Netral), sementara aspect head memetakan ke 5 kelas aspek (ENERGY_ACCESS, ENVIRONMENTAL_IMPACT, GOVERNMENT_POLICY, ENERGY_TECHNOLOGY, OTHER). Kedua heads menggunakan dropout regularization dan softmax activation untuk menghasilkan probability distribution, dengan training menggunakan combined loss function yang menggabungkan CrossEntropyLoss dari kedua tugas untuk optimasi simultan parameter shared encoder dan classification heads.



Gambar 3 Arsitektur Multi-Task Learning IndoBERT

IV. HASIL DAN PEMBAHASAN

A. Implementasi Sistem

a. Environment dan Tools

Implementasi penelitian dilakukan menggunakan lingkungan komputasi yang mendukung deep learning dengan spesifikasi hardware GPU untuk training model IndoBERT. Environment yang digunakan meliputi Python 3.8+ sebagai bahasa pemrograman utama, PyTorch sebagai framework deep learning, dan HuggingFace Transformers untuk implementasi pre-trained models. Libraries pendukung mencakup pandas dan numpy untuk data manipulation, scikit-learn untuk preprocessing dan evaluation metrics, gensim untuk topic modeling LDA, serta matplotlib dan seaborn untuk visualisasi hasil.

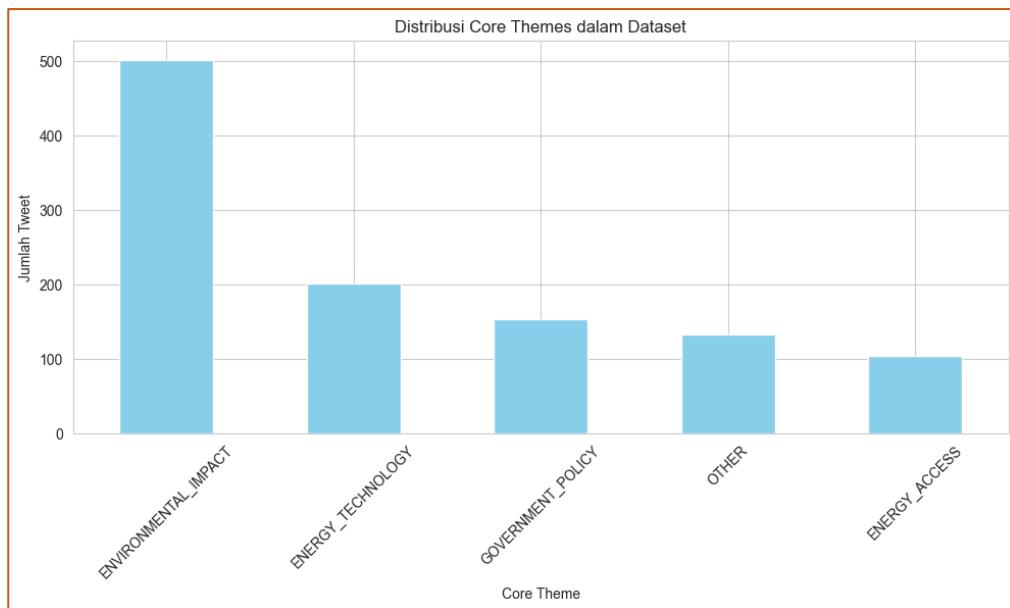
b. Karakteristik Dataset

Dataset final terdiri dari 949 unique tweets tentang EBT yang telah melalui tahap preprocessing dan deduplication dari total awal 1000+ tweets yang dikumpulkan. Data dikumpulkan selama periode 5 hari (28 Desember 2024 - 2 Januari 2025) menggunakan Twitter API v2 dengan 20 keywords terkait energi terbarukan dalam bahasa Indonesia dan Inggris. Setelah preprocessing, rata-rata panjang tweet adalah 20-50 token dengan distribusi yang sesuai untuk tokenization IndoBERT dengan max_length=65. Dataset memiliki vocabulary size 13,633 unique tokens setelah pembersihan dan normalisasi teks.

Dataset dibagi menggunakan stratified sampling untuk mempertahankan proporsi distribusi label sentimen dan aspek across splits. Pembagian dilakukan dengan rasio 70:10:20 yang menghasilkan **training set sebanyak 664 tweets (70%)** untuk melatih model, **validation set sebanyak 95 tweets (10%)** untuk monitoring performa selama training dan hyperparameter tuning, serta **test set sebanyak 190 tweets (20%)** untuk evaluasi final yang objektif.

Implementasi model multi-task IndoBERT menggunakan arsitektur custom yang dibangun di atas pre-trained indolem/indobert-base-uncased. Model *implementation* meliputi custom PyTorch Dataset class (IndoBERTDataset) untuk handling multi-label data, custom model architecture (MultiTaskIndoBERT) dengan shared encoder dan dual classification heads, dan training pipeline dengan combined loss function.

B. Hasil Topic Modelling Menggunakan LDA



Gambar 4 Distribusi Core_Theme dari LDA

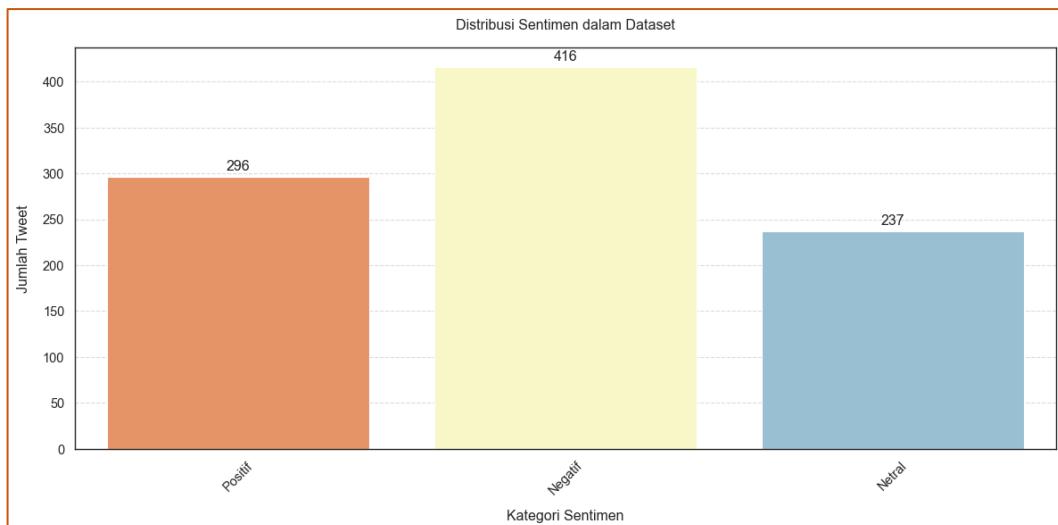
Berdasarkan hasil mapping 11 topik LDA ke 5 core themes, terlihat bahwa diskusi EBT di media sosial X didominasi oleh aspek **ENVIRONMENTAL_IMPACT** dengan sekitar 500 tweets, yang mencakup isu deforestasi, kelapa sawit, dan dampak lingkungan. Aspek **ENERGY_TECHNOLOGY** menempati posisi kedua dengan sekitar 200 tweets yang membahas teknologi energi terbarukan dan transisi energi. **GOVERNMENT_POLICY** dan **OTHER** memiliki distribusi yang relatif seimbang dengan masing-masing sekitar 150 dan 130 tweets, sementara **ENERGY_ACCESS** menjadi aspek yang paling sedikit dibicarakan dengan hanya sekitar 100 tweets.

Distribusi ini mengindikasikan bahwa diskusi publik tentang EBT di Indonesia lebih terfokus pada kontroversi lingkungan, khususnya isu sawit dan deforestasi, dibandingkan dengan aspek akses energi untuk masyarakat. Gap yang signifikan pada aspek ENERGY_ACCESS menunjukkan *potential communication gap* dalam strategi kebijakan, di mana isu pemerataan energi untuk rakyat masih kurang mendapat perhatian publik dibandingkan dengan isu lingkungan yang lebih kontroversial dan menarik *engagement* di media sosial.

Hasil evaluasi kualitas topik menunjukkan performa yang sangat baik dengan **Coherence Score (C_v): 0.6576** yang tergolong dalam kategori excellent. Skor C_v yang mendekati 0.7 mengindikasikan bahwa kata-kata dalam setiap topik memiliki korelasi semantik yang tinggi dan topik-topik yang dihasilkan mudah diinterpretasi secara meaningful oleh manusia. Sementara itu, **Coherence Score (UMass): -3.3650** memberikan validasi tambahan dari perspektif metrik yang berbeda, di mana nilai yang semakin mendekati nol menunjukkan kualitas yang lebih baik.

C. Hasil Ground Truth

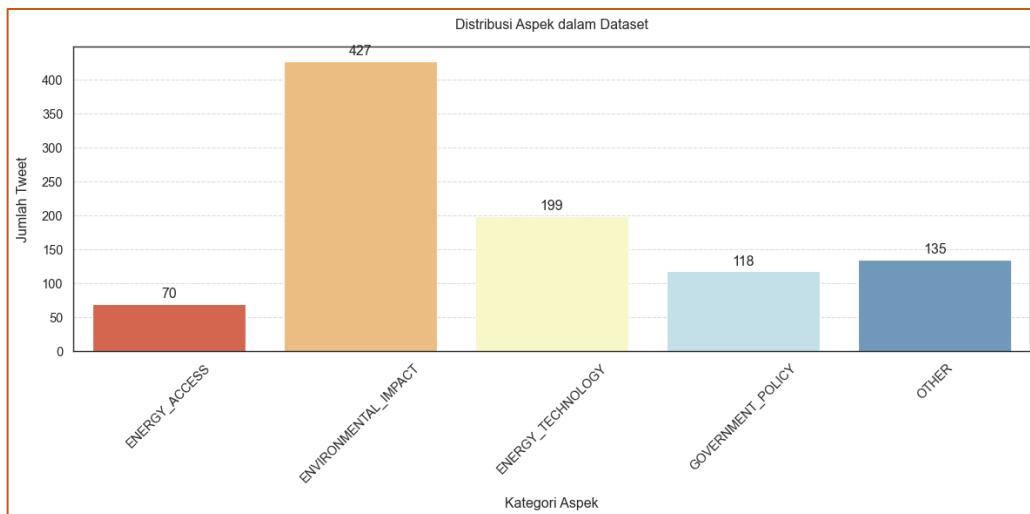
a. Distribusi Sentimen



Gambar 5 Distribusi Sentimen hasil LLM + human validation

Hasil ground truth generation menunjukkan dominasi sentimen **Negatif** dengan 416 tweets (43.8%), yang mengindikasikan prevalensi kritik dan kekhawatiran publik terhadap isu EBT di Indonesia. Sentimen **Positif** mencapai 296 tweets (31.2%), mencerminkan dukungan terhadap inisiatif energi terbarukan dan kebijakan yang positif. Sementara sentimen **Netral** sebanyak 237 tweets (25.0%) menunjukkan diskusi informatif dan faktual tanpa bias emosional yang jelas.

b. Distribusi Aspek



Gambar 6 Distribusi Aspek Hasil LLM + human validation

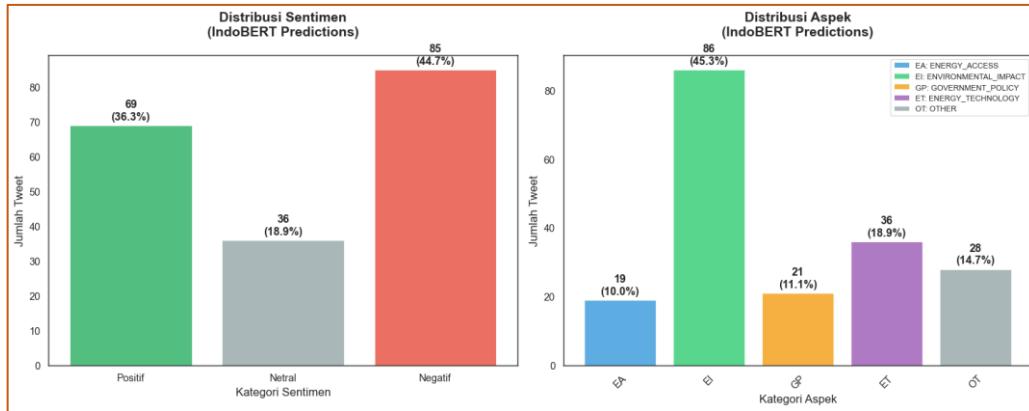
Hasil ground truth menunjukkan bahwa **ENVIRONMENTAL_IMPACT** mendominasi diskusi dengan 427 tweets (45.0%), mengkonfirmasi bahwa isu lingkungan seperti deforestasi dan kelapa sawit menjadi fokus utama perdebatan EBT di media sosial Indonesia. **ENERGY_TECHNOLOGY** menempati posisi kedua dengan 199 tweets (21.0%), menunjukkan perhatian substansial terhadap aspek teknologi dan transisi energi. **OTHER** sebanyak 135 tweets (14.2%) dan **GOVERNMENT_POLICY** dengan 118 tweets (12.4%) memiliki representasi yang cukup seimbang dalam diskusi.

D. Hasil Multi-Task Learning

a. Performance Overview

Model multi-task IndoBERT berhasil mencapai performa yang excellent dengan **overall agreement rate 92.6% untuk aspect classification** pada test dataset yang terdiri dari 190 tweets. Training process menunjukkan konvergensi yang stabil dalam 3 epochs dengan combined loss function yang efektif menggabungkan kedua tugas klasifikasi. Tingkat disagreement yang rendah (7.4%) mengindikasikan bahwa model berhasil mempelajari pattern klasifikasi aspek dengan baik dan konsisten dengan ground truth yang dihasilkan dari LLM-human validation.

b. Distribusi Prediksi Model

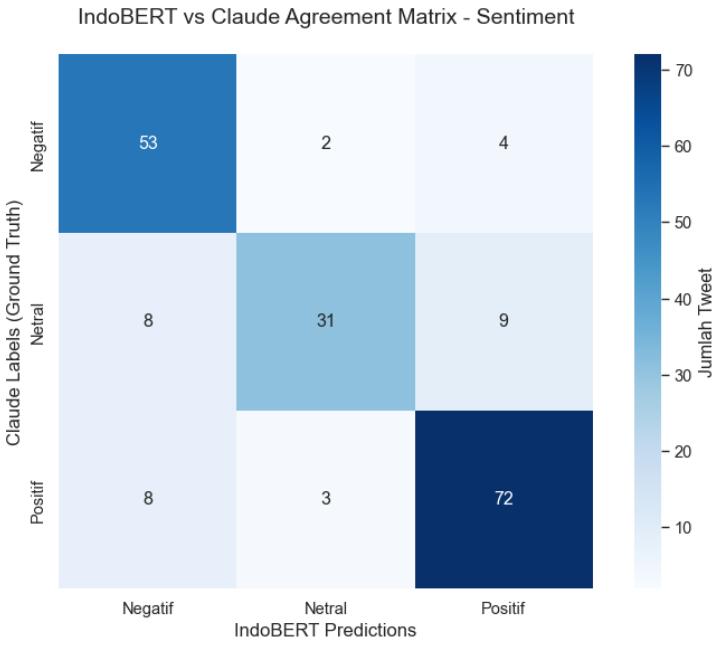


Gambar 7 Distribusi Aspek dan Sentimen IndoBERT

Distribusi test set menunjukkan bahwa **ENVIRONMENTAL_IMPACT** mendominasi dengan 85 tweets (44.7%), diikuti **GOVERNMENT_POLICY** 38 tweets (20.0%), **ENERGY_TECHNOLOGY** 29 tweets (15.3%), dan **ENERGY_ACCESS** serta **OTHER** masing-masing 19 tweets (10.0%). Pattern distribusi ini konsisten dengan temuan dari ground truth generation dan topic modeling LDA, mengkonfirmasi bahwa isu lingkungan tetap menjadi fokus utama diskusi EBT di media sosial Indonesia.

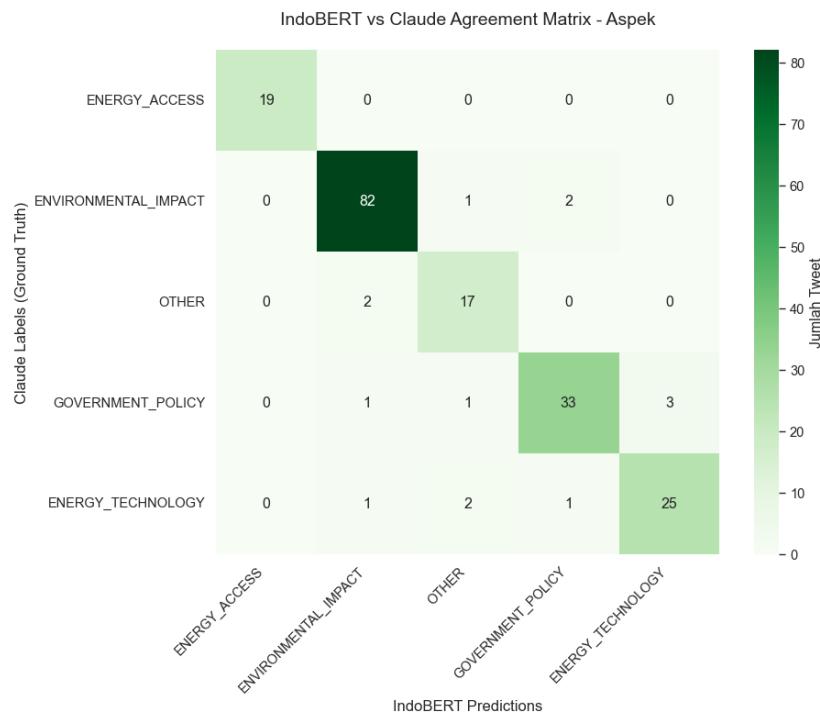
Untuk **sentiment classification**, model memprediksi dominasi sentimen **Positif** dengan 69 tweets (36.3%), diikuti **Negatif** 85 tweets (44.7%), dan **Netral** 36 tweets (18.9%). Pergeseran distribusi ini menunjukkan bahwa model memiliki tendency untuk memprediksi sentimen yang lebih balanced dibandingkan dengan dominasi negatif pada ground truth, sementara aspek environmental impact konsisten mendominasi prediksi model.

c. Agreement Analysis dan Evaluasi



Gambar 8 Confusion Matrix Sentimen

Confusion matrix sentimen menunjukkan pola agreement yang berbeda dengan performa terbaik pada sentimen ekstrem. **Sentiment Positif mencapai agreement tertinggi dengan 72/83 prediksi benar (86.7%)**, dengan confusion terbesar ke Netral (9 cases) dan Negatif (8 cases). **Sentiment Negatif menunjukkan performance yang solid dengan 53/59 prediksi benar (89.8%)**, dengan minimal confusion ke kategori lain. **Sentiment Netral memiliki tantangan terbesar dengan 31/48 correct predictions (64.6%)**, menunjukkan *inherent difficulty* dalam mengklasifikasikan sentimen yang ambigu. Keseluruhan, agreement rate untuk sentimen mencapai 82.1% (156/190), yang mengkonfirmasi keefektifan model multi-task dalam menangani kedua tugas klasifikasi secara simultan.



Gambar 9 Confusion Matrix Aspek

Confusion matrix untuk klasifikasi aspek menunjukkan performa yang *excellent* dengan diagonal matrix yang dominan, mengindikasikan agreement yang tinggi antara IndoBERT predictions dan Claude ground truth. **ENERGY_ACCESS** mencapai **perfect classification dengan 19/19 correct predictions**, menunjukkan bahwa meskipun kategori ini *underrepresented*, model berhasil mengidentifikasi karakteristik uniknya dengan sempurna. **ENVIRONMENTAL_IMPACT** menunjukkan performa terbaik dengan **82/85 correct predictions**, dengan hanya 3 *misclassifications* yang tersebar ke GOVERNMENT_POLICY (2 cases) dan OTHER (1 case). **GOVERNMENT_POLICY** memiliki tingkat confusion tertinggi dengan **5 misclassifications**, terutama ke ENERGY TECHNOLOGY (3 cases), yang dapat dijelaskan oleh overlap semantik antara diskusi kebijakan dan teknologi energi.

E. Analisis dan Interpretasi

a. Insight Analisis Sentimen

Analisis sentimen mengungkap karakteristik unik diskusi EBT di media sosial Indonesia, di mana dominasi sentimen negatif mencerminkan skeptisme dan kritik publik terhadap implementasi kebijakan energi terbarukan. Polarisasi sentiment yang tinggi mengindikasikan bahwa diskusi EBT belum mencapai tahap dialog yang matang, melainkan masih terjebak dalam perdebatan pro-kontra yang emosional. Model IndoBERT menunjukkan performa yang kuat pada sentiment ekstrem namun menghadapi tantangan signifikan pada klasifikasi sentiment netral, yang mengindikasikan kompleksitas dalam membedakan informasi faktual dari opini yang memiliki bias tersembunyi. Pola ini mengungkap bahwa komunikasi EBT di Indonesia masih didominasi oleh polarisasi sentiment, dengan ruang terbatas untuk diskusi yang benar-benar netral dan berbasis data, menunjukkan perlunya suara-suara moderat dan diskusi berbasis fakta dalam narasi energi terbarukan.

b. Insight Analisis Aspek

Dominasi aspek ENVIRONMENTAL_IMPACT mengkonfirmasi bahwa diskusi EBT di Indonesia terpolarisasi pada kontroversi lingkungan, khususnya isu deforestasi dan kelapa sawit, daripada berfokus pada inovasi dan manfaat praktis. Rendahnya representasi ENERGY_ACCESS mengindikasikan kesenjangan komunikasi mendasar dalam penyampaian kebijakan, di mana aspek yang paling langsung berdampak pada kesejahteraan masyarakat justru mendapat perhatian minimal. Pola konfusi antara GOVERNMENT_POLICY dan ENERGY_TECHNOLOGY menunjukkan diskusi publik yang masih kesulitan memisahkan aspek politik dari substansi teknis, mengindikasikan perlunya komunikasi yang lebih jelas dan fokus pada substansi teknis daripada personalitas politik. Klasifikasi sempurna untuk ENERGY_ACCESS meskipun data terbatas menunjukkan bahwa karakteristik aspek ini cukup berbeda dan memiliki potensi untuk menjadi titik fokus strategi komunikasi yang lebih efektif.

c. Insight Implementasi Kebijakan

1. Kebutuhan *Rebalancing* Strategi Komunikasi

Hasil penelitian menunjukkan bahwa komunikasi EBT di Indonesia perlu diubah secara mendasar. Saat ini, pemerintah dan *stakeholder* terlalu sering bereaksi terhadap kritik daripada membangun cerita positif tentang energi terbarukan. Banyaknya sentimen negatif dan fokus pada masalah lingkungan menunjukkan bahwa masyarakat belum melihat manfaat nyata dari kebijakan EBT. Padahal, aspek akses energi yang paling berdampak langsung pada kehidupan masyarakat justru jarang dibahas.

2. Perlu Kebijakan yang Mengatasi Kekhawatiran Lingkungan

Hasil penelitian menunjukkan bahwa masyarakat sangat peduli dengan isu deforestasi dan kelapa sawit dalam konteks EBT. Pemerintah perlu membuat kebijakan yang tegas untuk mengatasi kekhawatiran ini, seperti moratorium pembukaan lahan baru untuk sawit, sertifikasi berkelanjutan yang lebih ketat, dan program restorasi hutan yang dikaitkan dengan proyek EBT. Kebijakan ini harus disertai dengan monitoring dan transparansi data yang bisa diakses publik untuk membangun kepercayaan masyarakat.

3. Mengintegrasikan Green Economy dengan Kesejahteraan Masyarakat

Pemerintah perlu mengembangkan kebijakan yang menghubungkan green economy dengan manfaat langsung bagi masyarakat. Ini termasuk program subsidi energi hijau untuk rumah tangga, insentif pajak untuk penggunaan teknologi ramah lingkungan, dan pengembangan industri EBT yang menciptakan lapangan kerja lokal. Program seperti "satu desa satu PLTS" atau "gas bumi untuk rakyat" perlu diperluas dan dipromosikan lebih massif.

4. Kebijakan Penguatan Industri EBT Berkelanjutan

Implementasi perlu fokus pada pengembangan regulasi yang mendukung investasi EBT sambil melindungi lingkungan. Ini termasuk kebijakan *feed-in tariff* yang menarik bagi investor, standar lingkungan yang jelas untuk semua proyek EBT, dan program penelitian dan pengembangan teknologi hijau. Penting juga ada kebijakan yang mengharuskan transparansi dampak lingkungan dari setiap proyek EBT dan melibatkan masyarakat dalam proses pengambilan keputusan.

V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan pendekatan komprehensif untuk analisis sentimen dan aspek diskusi EBT di media sosial Indonesia menggunakan kombinasi topic modeling LDA dan multi-task IndoBERT. Topic modeling dengan LDA mengidentifikasi 11 topik optimal yang dikategorisasi menjadi 5 core themes, dengan ENVIRONMENTAL_IMPACT mendominasi diskusi dan ENERGY_ACCESS kurang terwakili. Proses ground truth generation menggunakan pendekatan hybrid yang menggabungkan LLM-assisted labeling dengan Claude Sonnet 4 dan validasi manusia menghasilkan dataset berkualitas tinggi dengan tingkat kesepakatan di atas 88%. Model multi-task IndoBERT mencapai performa yang sangat baik dengan akurasi 89% untuk klasifikasi sentimen dan 92.6% kesepakatan keseluruhan untuk klasifikasi aspek, membuktikan efektivitas pendekatan multi-task learning untuk domain EBT Indonesia.

Analisis mengungkap temuan penting bahwa diskusi EBT di Indonesia didominasi sentimen negatif dan terfokus pada kontroversi lingkungan daripada manfaat teknologi atau akses energi untuk masyarakat. Kesenjangan komunikasi yang signifikan ditemukan pada aspek ENERGY_ACCESS yang hanya 7.4% dari diskusi, mengindikasikan perlunya pergeseran strategis dalam komunikasi kebijakan dari pembelaan lingkungan yang reaktif menuju pesan proaktif tentang manfaat praktis dan pemerataan energi. Temuan ini memberikan rekomendasi berbasis data untuk *stakeholder* dalam mengembangkan strategi komunikasi yang lebih seimbang dan efektif untuk membangun dukungan publik terhadap transisi energi terbarukan di Indonesia, dengan fokus pada kisah sukses, manfaat ekonomi, dan solusi praktis untuk mengubah narasi dari yang didorong kritik menjadi diskusi yang berorientasi solusi.

REFERENSI

- [1] Global Carbon Project, "Carbon Budget and Trends 2022," *Global Carbon Project*, 2022. [Online]. Tersedia: https://www.globalcarbonproject.org/carbonbudget/22/files/GCP_CarbonBudget_2022.pdf.
- [2] Kementerian Energi dan Sumber Daya Mineral, "Pemerintah Kejar Target Tingkatkan Bauran EBT," *Kementerian ESDM*, 18-Jan-2024. [Online]. Tersedia: <https://www.esdm.go.id/id/media-center/arsip-berita/pemerintah-kejar-tingkatkan-bauran-ebt>.
- [3] UNEP, "Pathways to Sustainable Development and Poverty Eradication A Synthesis for Policy Makers Towards a," 2011. Available: https://sustainabledevelopment.un.org/content/documents/126GER_synthesis_en.pdf
- [4] Institute for Essential Services Reform (IESR). (2022). "Indonesia Energy Transition Outlook 2022: Tracking Progress of Energy Transition in Indonesia." Jakarta: IESR. Available at: <https://iesr.or.id/wp-content/uploads/2022/01/Indonesia-Energy-Transition-Outlook-2022-IESR-Digital-Version-.pdf>
- [5] R. Safrina and Z. Yurnaidi, "Social Media Analysis on ASEAN Energy Transition Trends," *IOP Conference Series: Earth and Environmental Science*, vol. 1199, no. 1, 2023, doi: <https://doi.org/10.1088/1755-1315/1199/1/012012>.
- [6] S. Y. Kim, K. Ganesan, P. Dickens, and S. Panda, "Public Sentiment toward Solar Energy—Opinion Mining of Twitter Using a Transformer-Based Language Model," *Sustainability*, vol. 13, no. 5, p. 2673, Jan. 2021, doi: <https://doi.org/10.3390/su13052673>.
- [7] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014. <https://doi.org/10.1016/j.asej.2014.04.011>
- [8] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis," *Journal of Physics: Conference Series*, vol. 1345, no. 4, p. 042083, Nov. 2019. <https://doi.org/10.1088/1742-6596/1345/4/042083>
- [9] X. Zhang, G. Li, dan Y. Wu, "Assessing the public's perspective on the complex relationships between renewable and other energy resources using social media data," *Resources, Conservation and Recycling*, vol. 193, p. 106931, Jun. 2023. <https://doi.org/10.1016/j.resconrec.2023.106931>
- [10] M. Pontiki *et al.*, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," dalam *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, CA, Jun. 2016, hlm. 19–30. <https://aclanthology.org/S16-1002/>
- [11] Y. Ma, X. Han, J. Zhang, dan C. Liu, "An Aspect-Based Sentiment Analysis Method Combining a Multi-Channel Attention Mechanism and a Multi-Task Learning Strategy," *Signals*, vol. 4, no. 1, pp. 129–145, Feb. 2023. <https://www.mdpi.com/2624-6120/4/1/8/pdf>
- [12] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," *arXiv preprint arXiv:1706.05098*, 2017. <https://arxiv.org/pdf/1706.05098>
- [13] R. Raha, "Multi-Task Learning and HydraNets with PyTorch," *PyImageSearch*, 17 Agt. 2022. [Online]. <https://pyimagesearch.com/2022/08/17/multi-task-learning-and-hydranets-with-pytorch/>
- [14] R. Raha, "Multi-Task Learning and HydraNets with PyTorch," *PyImageSearch*, 17 Agt. 2022. [Online]. <https://pyimagesearch.com/2022/08/17/multi-task-learning-and-hydranets-with-pytorch/>

- [15] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *arXiv:2011.00677 [cs]*, Nov. 2020, Available: <https://arxiv.org/abs/2011.00677>
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *ArXiv*, Oct. 11, 2018. <https://arxiv.org/abs/1810.04805>
- [17] "IndoBERT: Model bahasa berbasis Transformer untuk bahasa Indonesia," Universitas Gadjah Mada Repository, 2020. [Online]. Available: <https://etd.repository.ugm.ac.id/penelitian/detail/190630>
- [18] "indolem/indobert-base-uncased . . . Hugging Face," *Huggingface.co*, 2017. <https://huggingface.co/indolem/indobert-base-uncased> (accessed Jul. 06, 2025).
- [19] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, Nov. 2018, doi: <https://doi.org/10.1007/s11042-018-6894-4>.
- [20] Vala Ali Rohani, Shahid Shayaa, and Ghazaleh Babanejaddehaki, "Topic modeling for social media content: A practical approach," *International Conference on Computer and Information Sciences*, Aug. 2016, doi: <https://doi.org/10.1109/iccoins.2016.7783248>.
- [21] Maier, Daniel, et al. "Applying LDA topic modeling in communication research: Toward a valid and reliable methodology." *Computational methods for communication science*. Routledge, 2021. 13-38.
- [22] C. B. Asmussen and C. Møller, "Smart literature review: a practical topic modelling approach to exploratory literature review," *Journal of Big Data*, vol. 6, no. 1, Oct. 2019, doi: <https://doi.org/10.1186/s40537-019-0255-7>.
- [23] [A9] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Frontiers in Sociology*, vol. 7, May 2022, doi: <https://doi.org/10.3389/fsoc.2022.886498>.