Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 4001~4009
e-ISSN: 2548-6861

Performance Evaluation of Word2Vec and FastText Embeddings in a CNN-BiLSTM Model for Sentiment Classification of the LPDP Alumni Controversy

Dwi Erzalianti [1]*, Joice Junansi Tandirerung [2]*, Cici Suhaeni [3]*, Bagus Sartono [4]*
* Statistika dan Sains Data, IPB University
dwierzalianti@apps.ipb.ac.id [1], junansitrjoice@apps.ipb.ac.id [2], cici_suhaeni@apps.ipb.ac.id [3], bagusco@gmail.com [4]

Article Info

Article history:

Received …
Revised …
Accepted …

Keyword:

Sentiment Analysis,
CNN-BiLSTM,
Imbalanced Data,
Word2Vec,
FastText.

ABSTRACT

This study aims to analyze public sentiment toward the LPDP alumni controversy on social media using a deep learning approach. The research data consist of YouTube user comments related to the LPDP issue, which were processed through text preprocessing and automatically labeled using IndoBERT into three sentiment classes: negative, neutral, and positive. This study compares two text representation methods, namely Word2Vec and FastText, implemented within a hybrid CNN–BiLSTM architecture. In addition, data imbalance was addressed using class weighting and undersampling scenarios, while TF-IDF-based Logistic Regression was used as the baseline model. The results show that the baseline achieved an accuracy of 0.83 but was strongly biased toward the negative class as the majority class. The CNN–BiLSTM model improved the ability to detect minority classes. Under the class weighting scenario, FastText demonstrated more stable performance with an accuracy of 0.77 and a macro F1-score of 0.62. Under the undersampling scenario, Word2Vec was more stable, achieving an accuracy of 0.68 and a macro F1-score of 0.67. These findings indicate that both text representation and imbalance-handling strategies substantially affect sentiment classification performance.

I. PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan signifikan dalam pola komunikasi masyarakat, terutama dalam menyampaikan pandangan terhadap isu sosial, politik, ekonomi, dan kebijakan publik. Media sosial tidak lagi hanya berfungsi sebagai sarana berbagi informasi, tetapi juga menjadi ruang diskursus publik yang memungkinkan masyarakat mengekspresikan opini secara cepat, terbuka, dan luas. *Platform* seperti YouTube, X, Instagram, dan TikTok menghasilkan data tekstual dalam jumlah besar yang merepresentasikan respons masyarakat terhadap suatu peristiwa atau isu tertentu. Oleh karena itu, data media sosial dapat dimanfaatkan sebagai sumber informasi untuk memahami kecenderungan opini publik secara lebih sistematis dan berbasis data.

Salah satu pendekatan yang banyak digunakan untuk menganalisis opini publik pada data teks adalah analisis sentimen. Analisis sentimen merupakan bagian dari *Natural Language Processing* yang bertujuan mengidentifikasi orientasi opini, sikap, atau emosi dalam teks, kemudian mengklasifikasikannya ke dalam kategori tertentu, seperti positif, negatif, atau netral. Analisis sentimen telah diterapkan pada berbagai domain, seperti isu pemilihan kepala daerah [1], layanan transportasi *daring* [2], dan aplikasi publik berbasis digital [3]. Selain itu, analisis sentimen juga banyak digunakan untuk mengkaji komentar pengguna pada *platform* media sosial [4], sehingga pendekatan ini memiliki peran penting dalam memahami persepsi masyarakat terhadap isu yang berkembang di ruang digital.

Pada tahap awal pengembangannya, analisis sentimen banyak menggunakan pendekatan *machine learning* konvensional dengan representasi teks berbasis frekuensi, seperti *Term Frequency–Inverse Document Frequency*. Pendekatan tersebut masih relevan digunakan sebagai model pembanding karena mampu merepresentasikan bobot kata

berdasarkan frekuensi kemunculan dan tingkat kekhasannya dalam dokumen. Namun, representasi berbasis frekuensi memiliki keterbatasan dalam menangkap hubungan semantik dan konteks antarkata. Seiring berkembangnya penelitian klasifikasi teks, pendekatan *deep learning* semakin banyak digunakan karena memiliki kemampuan mempelajari pola data secara otomatis dan menangkap struktur bahasa yang lebih kompleks [5], [6].

Model *deep learning* seperti *Convolutional Neural Network*, *Long Short-Term Memory*, *Bidirectional Long Short-Term Memory*, serta kombinasi dengan *attention layer* telah menunjukkan kinerja yang baik dalam klasifikasi sentimen [6]. *Convolutional Neural Network* memiliki kemampuan dalam mengekstraksi fitur lokal atau pola n-gram penting dari teks, sedangkan *Long Short-Term Memory* dan *Bidirectional Long Short-Term Memory* mampu mempelajari hubungan sekuensial antarkata [7]. Arsitektur *hybrid* CNN–BiLSTM menjadi salah satu pendekatan yang relevan karena mengombinasikan kemampuan CNN dalam mengekstraksi fitur lokal dan kemampuan BiLSTM dalam memahami konteks sekuensial [8]. Kombinasi tersebut memungkinkan model mengidentifikasi pola penting dalam teks sekaligus memahami keterkaitan antarkata secara lebih komprehensif. Penerapan CNN–BiLSTM juga telah digunakan dalam klasifikasi emosi berbasis Word2Vec [9] serta klasifikasi sentimen berbasis arsitektur Word2Vec [10].

Selain arsitektur model, representasi kata merupakan aspek penting dalam menentukan kualitas klasifikasi sentimen. *Word embedding* digunakan untuk mengubah kata menjadi vektor numerik yang mengandung informasi semantik berdasarkan konteks kemunculannya. Word2Vec merupakan salah satu metode *embedding* yang mempelajari hubungan antarkata berdasarkan distribusi kemunculan kata dalam korpus [10]. Sementara itu, FastText memiliki keunggulan karena memanfaatkan informasi *subword* atau karakter n-gram, sehingga lebih adaptif dalam menangani variasi kata, kesalahan penulisan, singkatan, dan bentuk tidak baku yang umum ditemukan pada teks media sosial [4]. Penerapan FastText dan CNN telah digunakan dalam analisis sentimen pengguna *platform* digital [11], sedangkan FastText juga dapat digunakan sebagai representasi *embedding* pada analisis sentimen berbasis RNN dan LSTM [12]. Relevansi FastText dalam model *deep learning* juga ditunjukkan pada analisis sentimen aplikasi publik [3] dan komentar video berbahasa Indonesia di media sosial [4].

Karakteristik data media sosial yang tidak terstruktur menjadi tantangan tersendiri dalam analisis sentimen. Komentar pengguna sering kali memuat bahasa informal, kata tidak baku, singkatan, campuran bahasa, kesalahan ketik, serta ekspresi emosional. Kondisi tersebut dapat memengaruhi kualitas representasi teks dan performa model klasifikasi, khususnya ketika data yang digunakan berasal dari komentar publik yang spontan dan tidak mengikuti kaidah bahasa formal. Dalam konteks penelitian ini, polemik alumni Lembaga Pengelola Dana Pendidikan (LPDP) menjadi isu yang memunculkan respons publik secara luas di media sosial. Isu ini tidak hanya berkaitan dengan individu penerima beasiswa, tetapi juga menyentuh aspek yang lebih luas, seperti tanggung jawab penerima dana negara, pengelolaan anggaran publik, nasionalisme, dan keadilan sosial. Sebagai salah satu program beasiswa yang didanai oleh pemerintah, LPDP memiliki peran strategis dalam meningkatkan kualitas sumber daya manusia Indonesia. Oleh karena itu, berbagai polemik yang melibatkan alumni LPDP berpotensi memengaruhi tingkat kepercayaan masyarakat terhadap kredibilitas program serta persepsi publik mengenai efektivitas pengelolaan dana pendidikan. Oleh karena itu, analisis sentimen terhadap komentar publik mengenai polemik alumni LPDP diperlukan untuk memperoleh gambaran yang lebih objektif mengenai kecenderungan persepsi masyarakat. Hasil analisis tersebut diharapkan dapat memberikan manfaat praktis bagi pengelola LPDP sebagai bahan evaluasi dalam menyusun strategi komunikasi publik dan pengambilan kebijakan, serta memberikan kontribusi akademis dalam pengembangan penelitian analisis sentimen pada isu kebijakan publik di Indonesia.

Tantangan lain dalam klasifikasi sentimen adalah ketidakseimbangan distribusi kelas. Pada data opini publik, salah satu kelas sentimen sering kali memiliki jumlah data yang lebih dominan dibandingkan kelas lainnya, sehingga model dapat lebih condong mengenali kelas mayoritas dan kurang mampu mengidentifikasi kelas minoritas [13]. Kondisi tersebut dapat menyebabkan nilai akurasi terlihat tinggi, tetapi performa model pada kelas minoritas menjadi rendah [14]. Untuk mengatasi permasalahan tersebut, beberapa strategi dapat digunakan, antara lain *class weighting* dan *undersampling*. Kajian mengenai data tidak seimbang menunjukkan bahwa teknik *resampling*, baik *undersampling* maupun *oversampling*, dapat digunakan untuk memperbaiki distribusi kelas dalam proses klasifikasi [15], [16]. Berdasarkan uraian tersebut, penelitian ini berfokus pada evaluasi dua metode representasi kata, yaitu Word2Vec dan FastText, pada arsitektur CNN–BiLSTM dengan menerapkan dua strategi penanganan ketidakseimbangan data, yaitu *class weighting* dan *undersampling*. Evaluasi dilakukan untuk mengidentifikasi kombinasi metode representasi kata dan strategi penanganan ketidakseimbangan data yang memberikan performa terbaik dalam klasifikasi sentimen komentar terkait polemik alumni LPDP.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional. Pendekatan ini dipilih karena penelitian berfokus pada pengujian performa model klasifikasi sentimen berdasarkan variasi representasi teks dan strategi penanganan ketidakseimbangan data. Model utama yang digunakan adalah CNN–BiLSTM dengan dua metode representasi kata, yaitu Word2Vec dan FastText. Sebagai pembanding, digunakan *Logistic Regression* dengan representasi TF-IDF untuk memperoleh performa *baseline* sebelum penerapan model *deep learning* [5].

Data penelitian berupa komentar pengguna YouTube yang berkaitan dengan polemik alumni LPDP. Data dikumpulkan melalui proses *scraping* dari enam video YouTube yang relevan dengan topik penelitian. Jumlah data awal yang diperoleh sebanyak 11.153 komentar. Setelah melalui proses pembersihan, penghapusan duplikasi, dan seleksi teks, jumlah data yang digunakan untuk pemodelan menjadi 10.565 komentar. Data tersebut kemudian digunakan sebagai dasar dalam proses pelabelan, eksplorasi, representasi teks, pelatihan model, dan evaluasi klasifikasi sentimen. Sumber data penelitian disajikan pada Tabel 1

TABEL 1.
SUMBER DATA YOUTUBE

NO	JUDUL VIDEO	URL
1	BREAKING NEWS - Kontroversi Alumni LPDP, Menkeu Purbaya: Jangan Menghina Negara Anda Sendiri!	https://youtu.be/oLqvVMMFITc
2	Emosi! Perintah Purbaya Skakmat Alumni LPDP Viral: Minta Uang dan Bunganya!	https://youtu.be/N_-PPxGJCzU
3	Penerima LPDP Jadi WNA, Uang Pajak Rakyat Sia-Sia? [Top News]	https://youtu.be/DbZ9muxQp0s
4	LPDP Ungkap Hasil Pembicaraan dengan Suami Dwi Sasetyaningtyas | Sedang Viral	https://youtu.be/xiEDFPrz5WU
5	MIRIS! Alumni Penerima LPDP Bangga Anak Jadi WNA - [Selamat Pagi Indonesia]	https://youtu.be/tAEOMuzYcwc
6	Pasutri Penerima LPDP Viral, Terancam Kembalikan Dana Rp7,7 Miliar | iNews Today (25/02)	https://youtu.be/0XZopkYspoQ

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data teks sebelum digunakan dalam pemodelan. Proses ini mencakup *case folding*, penghapusan URL, *mention*, *hashtag*, angka, tanda baca, simbol, emoji, serta karakter khusus yang tidak relevan. Pada data media sosial, pembersihan teks diperlukan karena komentar pengguna umumnya memuat variasi bahasa informal, singkatan, dan ekspresi nonformal yang dapat memengaruhi kualitas representasi teks [4]. Selain itu, normalisasi kata tidak baku menggunakan kamus *slang*, penghapusan data duplikat, dan penyaringan komentar yang terlalu pendek dilakukan agar data yang digunakan lebih konsisten dan informatif. Tahapan ini penting karena variasi bentuk kata, kesalahan pengetikan, serta karakteristik bahasa tidak baku pada komentar media sosial dapat memengaruhi performa model klasifikasi sentimen [3], [11].

Pelabelan sentimen dilakukan secara otomatis menggunakan IndoBERT dengan tiga kategori, yaitu negatif, netral, dan positif. Setelah data memperoleh label, dilakukan eksplorasi data untuk memahami karakteristik awal dataset. Eksplorasi data mencakup analisis distribusi kelas sentimen dan visualisasi *word cloud*. Analisis distribusi sentimen digunakan untuk mengidentifikasi adanya ketidakseimbangan kelas, sedangkan *word cloud* digunakan untuk melihat kata-kata dominan yang muncul dalam komentar pengguna [2], [4]. Untuk memvalidasi hasil pelabelan otomatis, sebanyak 300 komentar dipilih secara acak untuk diverifikasi secara manual oleh peneliti. Hasil validasi menunjukkan bahwa 279 dari 300 komentar memiliki label yang sesuai dengan hasil pelabelan IndoBERT, sehingga diperoleh tingkat kesesuaian sebesar 93,0%. Temuan ini menunjukkan bahwa pelabelan otomatis menggunakan IndoBERT cukup konsisten dengan anotasi manual.

Data yang telah melalui tahap pra-pemrosesan dan pelabelan kemudian dibagi menjadi data latih dan data uji menggunakan metode *train-test split* dengan rasio 80:20. Pembagian data dilakukan secara acak menggunakan `random_state=42` untuk memastikan hasil eksperimen dapat direproduksi. Pendekatan ini digunakan agar proporsi setiap kelas sentimen tetap terwakili secara proporsional pada masing-masing subset data. Pada tahap representasi teks, model *baseline* menggunakan TF-IDF, sedangkan model utama menggunakan Word2Vec dan FastText. Kedua metode embedding tersebut digunakan untuk membentuk *embedding matrix* sebagai masukan pada model CNN–BiLSTM [12].

TABEL 2.
KONFIGURASI WORD2VEC DAN FASTTEXT

Parameter	Word2Vec	FastText
Sumber korpus	Komentar YouTube terkait polemik alumni LPDP	Komentar YouTube terkait polemik alumni LPDP
Jenis *embedding*	*Trained from scratch*	*Trained from scratch*
Arsitektur	*Skip-gram*	*Skip-gram*
Dimensi vektor (*vector size*)	300	300
Window size	7	7
Minimum word frequency (min_count)	2	3
Jumlah *epoch*	30	30

Perbedaan nilai *min_count* disesuaikan dengan karakteristik masing-masing metode. Word2Vec menggunakan *min_count* sebesar 2 untuk mempertahankan lebih banyak kata berfrekuensi rendah, sedangkan FastText menggunakan *min_count* sebesar 3 untuk mengurangi *noise* tanpa mengurangi kualitas representasi kata berkat mekanisme *subword*.

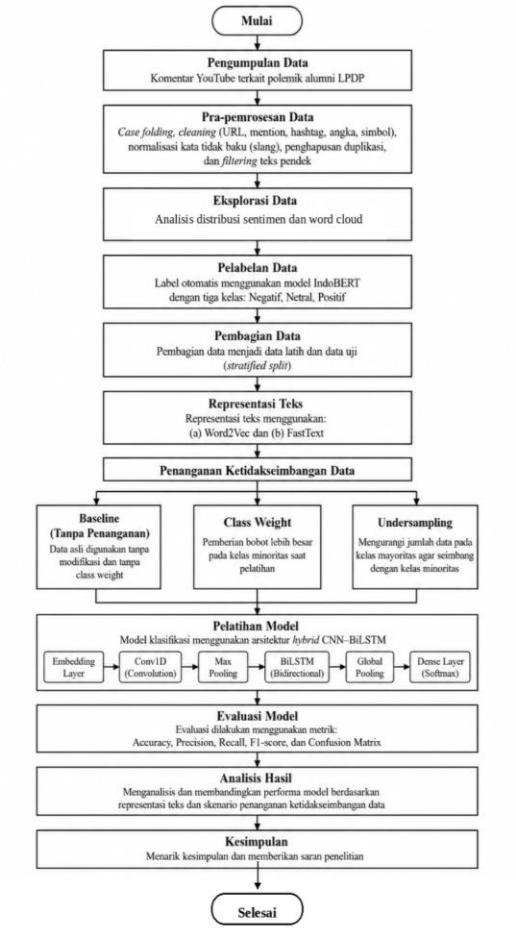
Arsitektur CNN–BiLSTM yang digunakan dalam penelitian ini terdiri atas *embedding layer*, Conv1D, MaxPooling1D, Bidirectional LSTM, GlobalMaxPooling1D, *dense layer*, *dropout*, dan *output layer*. Lapisan Conv1D berperan dalam mengekstraksi pola lokal atau fitur penting pada teks, sedangkan *Bidirectional* LSTM digunakan untuk mempelajari konteks kata dari dua arah sehingga hubungan sekuensial dalam kalimat dapat dipahami secara lebih komprehensif [8]. *Output layer* menggunakan aktivasi Softmax karena penelitian ini melibatkan klasifikasi

multikelas, yaitu sentimen negatif, netral, dan positif. Untuk mengatasi ketidakseimbangan data, penelitian ini menerapkan dua strategi, yaitu *class weighting* dan *undersampling*. *Class weighting* memberikan bobot lebih besar pada kelas minoritas agar model lebih sensitif terhadap kelas yang jumlah datanya lebih sedikit [15]. Sementara itu, *undersampling* dilakukan dengan mengurangi jumlah data pada kelas mayoritas sehingga distribusi kelas menjadi lebih seimbang dalam proses pelatihan model [16]. Parameter model disajikan pada Tabel 3.

TABEL 3.
PARAMETER MODEL CNN-BILSTM

Parameter	Konfigurasi
Embedding layer	*Trainable* (Word2Vec/FastText), 300 dimensi
CNN	Conv1D, 128 *filters*, *kernel size* = 3, ReLU
Pooling	*MaxPooling1D* (*pool size* = 2)
BiLSTM	64 units
Global pooling	*GlobalMaxPooling1D*
Dense layer	128 (ReLU), 64 (ReLU)
Dropout	0.4
Output layer	3 *neurons*, Softmax
Optimizer	Adam
Learning rate	0.0005
Loss function	*Sparse Categorical Crossentropy*
Batch size	64
Epoch	30

Eksperimen dilakukan melalui tiga skenario utama. Skenario pertama menggunakan *Logistic Regression* dengan representasi TF-IDF sebagai model *baseline* untuk memperoleh gambaran performa awal sebelum penerapan model *deep learning*. Skenario kedua menerapkan CNN–BiLSTM dengan representasi Word2Vec dan FastText menggunakan strategi *class weighting*, sedangkan skenario ketiga menerapkan CNN–BiLSTM dengan representasi Word2Vec dan FastText menggunakan strategi *undersampling*. Evaluasi model dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *macro F1-score*, dan *confusion matrix*. Penggunaan metrik evaluasi yang beragam diperlukan agar performa model dapat dianalisis secara lebih komprehensif, terutama pada data dengan distribusi kelas tidak seimbang [13]. Selain itu, penggunaan *macro F1-score* dan *confusion matrix* penting untuk menilai kemampuan model dalam mengenali setiap kelas sentimen, khususnya kelas minoritas yang sering kurang terdeteksi pada data tidak seimbang [14]. Alur penelitian ditunjukkan pada Gambar 1.

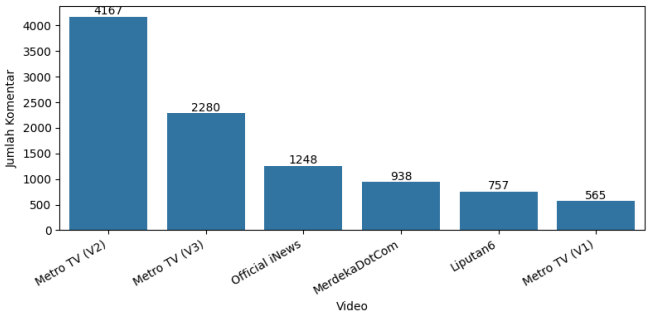


Gambar 1. Diagram alir penelitian

III. HASIL DAN PEMBAHASAN

A. *Exploratory Data Analysis (EDA)*

Dataset yang digunakan merupakan komentar YouTube terkait polemik alumni LPDP yang telah melalui pra-pemrosesan dan pelabelan otomatis. Setelah proses pembersihan dan penghapusan duplikasi, diperoleh 10.565 data yang siap digunakan untuk klasifikasi.



Gambar 2. Distribusi Komentar Setiap Video Youtube

Berdasarkan Gambar 2, video Metro TV (V2) menghasilkan komentar terbanyak (4.167 komentar), sedangkan Metro TV (V1) memiliki komentar paling sedikit

(565 komentar). Hal ini menunjukkan adanya variasi tingkat interaksi pengguna pada setiap video yang menjadi sumber data penelitian.

TABEL 4.
CONTOH DATA SEBELUM DAN SESUDAH PREPROCESSING

Data Asli	Hasil Preprocessing
kalo otak tidak di pakai untuk berbuat,berbicara dll... tamat lah sudah al kisah kamiðŸ˜,ðŸ˜,ðŸ˜,	kalo otak tidak di pakai untuk berbuatberbicara dll tamat lah sudah al kisah kami
" Menurut keyakinan saya " Kasus LPDP ini bakal di perpanjang buat pengalihan isu " Tarif dagang TRUMP "	menurut keyakinan saya kasus lpdp ini bakal di perpanjang buat pengalihan isu tarif dagang trump
BALIKIN DUIT RAKYAT RI DAN JANGAN INJAK INDONESIA LAGI !!!!! MEMALUKAN !!!!!!	balikin duit rakyat ri dan jangan injak indonesia lagi memalukan

Berdasarkan Tabel 4 terlihat bahwa komentar yang semula mengandung huruf kapital, tanda baca, emoji, dan kata-kata yang kurang informatif secara bertahap disederhanakan hingga menghasilkan kumpulan kata dasar yang lebih representatif terhadap isi komentar. Proses ini membantu meningkatkan kualitas data sebelum dilakukan ekstraksi fitur dan klasifikasi sentimen.

TABEL 5.
CONTOH KOMENTAR DAN LABEL SENTIMEN

Komentar	Label	Keterangan
info media sosial mahluk itu donk biar netizen indo yang ramein	0	Negatif
menurut keyakinan saya kasus lpdp ini bakal di perpanjang buat pengalihan isu tarif dagang trump	1	Netral
semoga baikbaik tentram selama nanti tinggal di inggris semoga kdpn tidak dpt perlakuan rasial sbg pendatang minoritas asia	2	Positif

Tabel 5 menunjukkan contoh komentar beserta label sentimen yang diberikan. Pelabelan dilakukan ke dalam tiga kategori, yaitu negatif netral, dan positif sesuai dengan konteks dan opini yang terkandung dalam komentar.

TABEL 6.
DISTRIBUSI SENTIMEN PUBLIK TERHADAP POLEMIK ALUMNI LPDP

Kelas	Data Asli	*Class Wieghting*	*Undersampling*
Negatif	8.144	8.144	755
Netral	1.056	1.056	755
Positif	755	755	755

Berdasarkan Tabel 6, distribusi sentimen pada dataset menunjukkan adanya ketidakseimbangan kelas yang cukup signifikan. Sebelum dilakukan penanganan ketidakseimbangan data, kelas negatif mendominasi dengan 8.144 komentar (81,81%), sedangkan kelas netral dan positif masing-masing berjumlah 1.056 (10,61%) dan 755 komentar (7,58%). Pada skenario *class weighting*, distribusi data tetap sama karena metode ini tidak mengubah jumlah data pada setiap kelas, melainkan memberikan bobot yang lebih besar pada kelas minoritas selama proses pelatihan model. Sementara itu, pada skenario *undersampling*, jumlah data pada kelas mayoritas dikurangi hingga sama dengan jumlah data pada kelas minoritas, sehingga masing-masing kelas terdiri atas 755 komentar. Penyeimbangan distribusi tersebut bertujuan untuk mengurangi bias model terhadap kelas mayoritas dan meningkatkan kemampuan model dalam mengenali kelas minoritas.

Gambar 3. *WordCloud* Komentar

Visualisasi *WordCloud* pada Gambar 3 menunjukkan bahwa kata yang sering muncul meliputi “indonesia”, “negara”, “orang”, “anak”, dan “beasiswa”. Kemunculan kata “koruptor”, “pejabat”, dan “uang rakyat” menunjukkan adanya kritik terhadap pengelolaan dana publik dan isu keadilan sosial. Pola ini memperkuat temuan bahwa komentar publik terhadap isu LPDP didominasi oleh opini kritis.

TABEL 7.
KATA-KATA DOMINAN PADA SETIAP KELAS SENTIMEN

Kelas Sentimen	Kata Dominan
Negatif	orang, Indonesia, negara, uang, beasiswa, pajak, rakyat, pejabat, korupsi
Netral	Indonesia, LPDP, penerima, beasiswa, negara, pemerintah, audit, kembali, uang
Positif	Indonesia, anak, bangga, baik, beasiswa, hidup, Inggris, uang

Berdasarkan Tabel 7, setiap kelas sentimen memiliki karakteristik kata dominan yang berbeda. Pada sentimen negatif, kata-kata seperti “negara”, “uang”, “pajak”, “rakyat”, “pejabat”, dan “korupsi” menunjukkan bahwa komentar masyarakat banyak menyoroti isu penggunaan dana publik, akuntabilitas penerima beasiswa, serta transparansi pengelolaan program LPDP. Hal ini menunjukkan bahwa polemik alumni LPDP tidak hanya dipandang sebagai

permasalahan individu, tetapi juga dikaitkan dengan tanggung jawab terhadap dana negara.

Sentimen netral didominasi oleh kata-kata seperti "LPDP", "penerima", "beasiswa", "pemerintah", dan "audit". Komentar pada kategori ini cenderung berupa penyampaian informasi, pertanyaan, maupun diskusi mengenai kebijakan LPDP tanpa menunjukkan kecenderungan sentimen yang kuat.

Sementara itu, pada sentimen positif muncul kata-kata seperti "bangga", "baik", "Indonesia", dan "beasiswa". Hal ini menunjukkan bahwa sebagian masyarakat tetap memberikan apresiasi terhadap program LPDP dan berharap program tersebut terus memberikan manfaat bagi peningkatan kualitas sumber daya manusia Indonesia meskipun sedang muncul polemik.

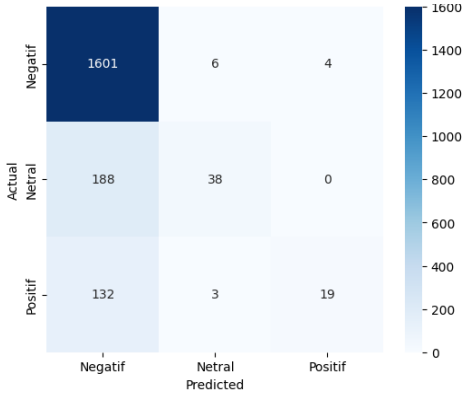
B. *Hasil Model Baseline*

Model baseline menggunakan *Logistic Regression* dengan representasi TF-IDF. Model ini dilatih pada data asli tanpa penanganan ketidakseimbangan sehingga dapat menjadi pembanding awal terhadap model CNN–BiLSTM. Hasil evaluasi *baseline* ditunjukkan pada Tabel 8.

TABEL 8.
HASIL EVALUASI *BASELINE*

Kelas	Precision	Recall	F1-Score
Negatif	0.83	0.99	0.91
Netral	0.81	0.17	0.28
Positif	0.83	0.12	0.21
Accuracy	0.83		

Baseline menghasilkan akurasi sebesar 0,83. Namun, performa tersebut terutama berasal dari kemampuan model mengenali kelas negatif sebagai kelas mayoritas dengan *F1-score* 0,91. Kelas netral dan positif memiliki *F1-score* masing-masing 0,28 dan 0,21, dengan *recall* yang sangat rendah. *Confusion matrix* pada Gambar 4 memperlihatkan bahwa sebagian besar data netral dan positif salah diklasifikasikan sebagai negatif. Hal ini menunjukkan adanya bias kuat terhadap kelas mayoritas.



Gambar 4. *Confusion Matrix Baseline*

C. *Hasil Model CNN-BilSTM dengan Class Weight*

Pada skenario *class weighting*, model CNN–BiLSTM dilatih menggunakan data tidak seimbang, tetapi bobot kelas minoritas diperbesar selama proses pelatihan. Hasil evaluasi Word2Vec dan FastText ditunjukkan pada Tabel 9 dan 10.

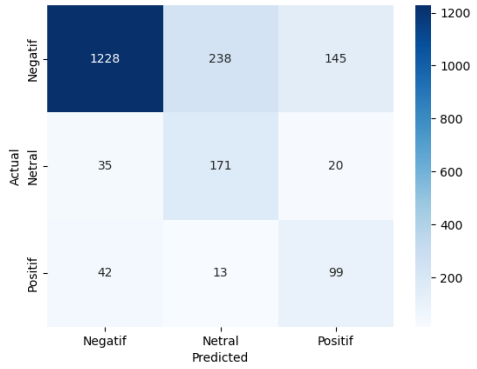
TABEL 9.
PERBANDINGAN HASIL EVALUASI PER KELAS CNN-BILSTM (*CLASS WEIGHT*)

Kelas	Word2Vec			FastText		
	Prec	*Rec*	F1	*Prec*	*Rec*	F1
Negatif	0.94	0.76	0.84	0.93	0.80	0.86
Netral	0.41	0.76	0.53	0.46	0.65	0.54
Positif	0.38	0.64	0.47	0.37	0.66	0.47

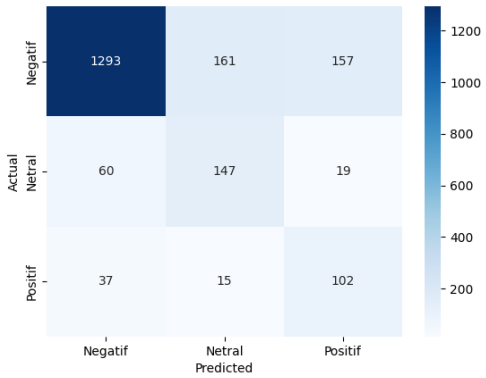
TABEL 10.
RINGKASAN PERBANDINGAN PERFORMA CNN-BILSTM (*CLASS WEIGHT*)

Embedding	*Accuracy*	*Macro F1*
Word2Vec	0.75	0.61
FastText	0.77	0.62

Pada *class weighting*, Word2Vec memperoleh akurasi 0,75 dan *macro F1-score* 0,61. Model ini meningkatkan *recall* kelas netral dan positif masing-masing menjadi 0,76 dan 0,64, tetapi *precision* pada kelas minoritas masih rendah. FastText memperoleh akurasi 0,77 dan *macro F1-score* 0,62 sehingga sedikit lebih stabil daripada Word2Vec. Nilai *recall* kelas positif pada FastText mencapai 0,66, sementara kelas negatif tetap memiliki *F1-score* tinggi sebesar 0,86. Dengan demikian, *class weighting* meningkatkan sensitivitas model terhadap kelas minoritas, meskipun terdapat *trade-off* berupa penurunan *precision*.



Gambar 5(a). *Confusion Matrix* Word2Vec dengan *Class Weight*



Gambar 5(b). *Confusion Matrix* FastText dengan *Class Weight*

Confusion matrix pada Gambar 5 menunjukkan bahwa Word2Vec dan FastText mampu mengklasifikasikan kelas negatif dengan cukup baik. FastText menghasilkan jumlah prediksi benar yang lebih tinggi pada kelas negatif, sedangkan Word2Vec lebih mampu mengenali kelas netral. Pada kelas positif, FastText menunjukkan sedikit keunggulan melalui jumlah prediksi benar yang lebih tinggi dan kesalahan klasifikasi yang lebih terkendali.

D. Hasil Model CNN-BiLSTM dengan Undersampling

Skenario berikutnya menggunakan *undersampling*, yaitu menyeimbangkan distribusi kelas dengan mengurangi jumlah data pada kelas mayoritas. Hasil evaluasi Word2Vec dan FastText pada data yang telah diseimbangkan disajikan pada Tabel 11 dan Tabel 12.

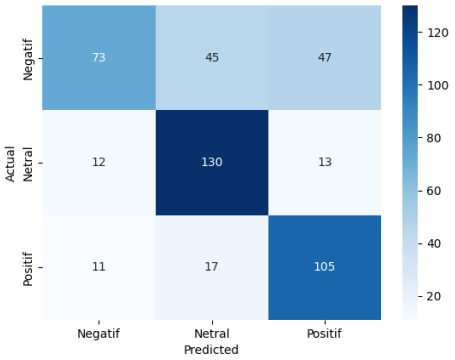
TABEL 11.
PERBANDINGAN HASIL EVALUASI PER KELAS CNN-BILSTM (*UNDERSAMPLING*)

Kelas	Word2Vec			FastText		
	Prec	***Rec***	**F1**	***Prec***	***Rec***	**F1**
Negatif	0.76	0.44	0.56	0.60	0.39	0.48
Netral	0.68	0.84	0.75	0.64	0.85	0.73
Positif	0.64	0.79	0.70	0.58	0.60	0.59

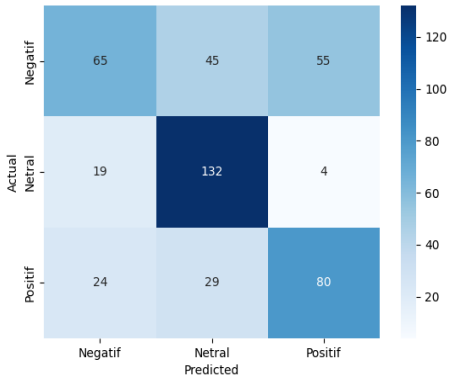
TABEL 12.
RINGKASAN PERBANDINGAN PERFORMA CNN-BILSTM (*UNDERSAMPLING*)

Embedding	***Accuracy***	***Macro F1***
Word2Vec	0.68	0.67
FastText	0.61	0.60

Pada *undersampling*, Word2Vec memperoleh akurasi 0,68 dan *macro F1-score* 0,67. Nilai *recall* kelas netral dan positif relatif tinggi, masing-masing 0,84 dan 0,79, yang menunjukkan bahwa model mampu mengenali kelas minoritas dengan lebih baik. Akan tetapi, *recall* kelas negatif turun menjadi 0,44 akibat berkurangnya data mayoritas dalam pelatihan. FastText memperoleh akurasi 0,61 dan *macro F1-score* 0,60. Model ini sangat baik dalam mengenali kelas netral dengan *recall* 0,85, tetapi performa kelas negatif dan positif lebih rendah dibandingkan Word2Vec.



Gambar 6(a). *Confusion Matrix* Word2Vec dengan *Undersampling*



Gambar 6(b). *Confusion Matrix* FastText dengan *Undersampling*

Perbandingan kedua *embedding* menunjukkan bahwa FastText lebih unggul pada skenario *class weighting* karena mampu mempertahankan akurasi dan *macro F1-score* yang sedikit lebih tinggi. Keunggulan ini selaras dengan karakter FastText yang memanfaatkan informasi *subword*, sehingga mampu menghasilkan representasi kata yang lebih baik, termasuk pada kata-kata yang jarang muncul, tanpa mengubah distribusi data. Sebaliknya, pada skenario *undersampling*, Word2Vec menunjukkan performa yang lebih stabil dengan akurasi dan *macro F1-score* lebih tinggi dibandingkan FastText. Hal ini mengindikasikan bahwa distribusi data yang lebih seimbang membuat representasi kata berbasis kata utuh (*whole word*) pada Word2Vec menjadi lebih optimal selama proses pelatihan, sedangkan FastText cenderung lebih sensitif terhadap perubahan distribusi data akibat pengurangan data pada kelas mayoritas.

Secara keseluruhan, strategi penanganan ketidakseimbangan data memberikan pengaruh yang signifikan terhadap performa model CNN–BiLSTM. Tanpa penanganan, model cenderung bias terhadap kelas mayoritas. Penerapan *class weighting* meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengubah distribusi data, sedangkan *undersampling* menghasilkan distribusi performa yang lebih seimbang dengan konsekuensi berkurangnya sebagian informasi dari kelas mayoritas. Temuan ini menunjukkan bahwa efektivitas strategi penanganan ketidakseimbangan data dipengaruhi oleh karakteristik metode *embedding* yang digunakan. Oleh karena

itu, pemilihan kombinasi metode *embedding* dan strategi penanganan ketidakseimbangan data perlu disesuaikan dengan tujuan analisis, apakah memprioritaskan akurasi keseluruhan atau kemampuan model dalam mengidentifikasi kelas minoritas.

Selain membandingkan performa metode *embedding*, penelitian ini juga menunjukkan bahwa *Logistic Regression* sebagai model *baseline* mampu menghasilkan akurasi yang lebih tinggi pada beberapa skenario dibandingkan model CNN–BiLSTM. Hasil ini menunjukkan bahwa model yang lebih kompleks tidak selalu memberikan performa terbaik. Karakteristik komentar YouTube yang relatif pendek serta representasi TF-IDF masih mampu menangkap informasi penting untuk proses klasifikasi. Di sisi lain, model CNN–BiLSTM umumnya memerlukan data yang lebih besar dan lebih beragam agar kemampuannya dalam mempelajari hubungan kontekstual dapat dimanfaatkan secara optimal.

Temuan-temuan tersebut menunjukkan bahwa analisis sentimen terhadap komentar publik di media sosial dapat dimanfaatkan sebagai salah satu sumber informasi untuk memahami persepsi masyarakat terhadap program LPDP. Informasi yang diperoleh dapat menjadi bahan evaluasi bagi pengelola LPDP dalam menyusun strategi komunikasi publik, meningkatkan transparansi program, serta merespons isu yang berkembang di masyarakat secara lebih efektif.

Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Pelabelan sentimen dilakukan secara otomatis menggunakan IndoBERT sebelum proses pelatihan model CNN–BiLSTM. Meskipun sebagian data telah diverifikasi secara manual dan menunjukkan tingkat kesesuaian yang baik dengan anotasi peneliti, kualitas label yang dihasilkan tetap bergantung pada kemampuan IndoBERT dalam memahami konteks bahasa Indonesia. Oleh karena itu, kesalahan pelabelan yang masih mungkin terjadi berpotensi memengaruhi proses pembelajaran serta hasil evaluasi model yang dikembangkan. Penelitian selanjutnya dapat mempertimbangkan penggunaan anotasi manual oleh lebih dari satu anotator atau mengombinasikan pelabelan otomatis dengan validasi manual yang lebih luas untuk meningkatkan reliabilitas data. Selain itu, penelitian ini hanya menggunakan komentar YouTube sebagai sumber data sehingga hasil analisis belum sepenuhnya merepresentasikan opini masyarakat secara keseluruhan. Karakteristik pengguna YouTube serta pola interaksi pada *platform* tersebut dapat berbeda dengan populasi umum, sehingga hasil penelitian perlu diinterpretasikan sesuai dengan konteks sumber data yang digunakan.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa model CNN–BiLSTM mampu melakukan klasifikasi sentimen pada komentar media sosial terkait polemik alumni LPDP dengan cukup baik. Model *baseline* TF-IDF dan *Logistic Regression* memperoleh akurasi tinggi, tetapi cenderung bias terhadap kelas negatif sebagai kelas mayoritas. Penerapan *class weighting* pada CNN–BiLSTM meningkatkan kemampuan deteksi kelas minoritas, dengan FastText memberikan performa lebih stabil pada skenario tersebut. Sementara itu, *undersampling* menghasilkan performa yang lebih seimbang antar kelas, tetapi menurunkan akurasi keseluruhan karena jumlah data mayoritas berkurang; pada skenario ini Word2Vec menunjukkan performa lebih stabil. Dengan demikian, pemilihan representasi teks dan strategi penanganan data tidak seimbang menjadi faktor penting dalam meningkatkan kualitas klasifikasi sentimen, khususnya pada data media sosial yang memiliki distribusi kelas tidak proporsional.

DAFTAR PUSTAKA

[1] D. A. Firdaus and E. B. Setiawan, "Analisis Sentimen pada X Terhadap Pilkada 2024 Menggunakan Ekspansi Fitur FastText dan CNN dengan Optimasi Bat Algorithm," *eProceedings of Engineering*, 2025.

[2] A. R. Gunawan and R. F. Alfa Aziza, "Sentiment Analysis Using LSTM Algorithm Regarding Grab Application Services in Indonesia," *Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 322–332, Mar. 2025, doi: 10.30871/jaic.v9i2.8696.

[3] Handoko, Junadhi, Triyani Arita Fitri, and Agustin, "Optimization of Deep Learning with FastText for Sentiment Analysis of the SIREKAP 2024 Application," *The Indonesian Journal of Computer Science*, vol. 14, no. 2, Apr. 2025, doi: 10.33022/ijcs.v14i2.4809.

[4] D. Ariyus, D. Manongga, and I. Sembiring, "Enhancing Sentiment Analysis of Indonesian Tourism Video Content Commentary on TikTok: A FastText and Bi-LSTM Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18020–18028, Dec. 2024, doi: 10.48084/etasr.8859.

[5] M. E. Alzahrani, T. H. H. Aldhyani, S. N. Alsubari, M. M. Althobaiti, and A. Fahad, "Developing an Intelligent System with Deep Learning Algorithms for Sentiment Analysis of E-Commerce Product Reviews," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, May 2022, doi: 10.1155/2022/3840071.

[6] A. S. Mirdan, S. Buyrukoglu, and M. R. Baker, "Advanced deep learning techniques for sentiment analysis: combining Bi-LSTM, CNN, and attention layers," *International Journal of Advances in Intelligent Informatics*, vol. 11, no. 1, p. 55, Feb. 2025, doi: 10.26555/ijain.v11i1.1848.

[7] L. Khan, A. Amjad, K. M. Afaq, and H.-T. Chang, "Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media," *Applied Sciences*, vol. 12, no. 5, p. 2694, Mar. 2022, doi: 10.3390/app12052694.

[8] V. R. Prasetyo, M. F. Naufal, and K. Wijaya, "Sentiment Analysis of ChatGPT on Indonesian Text using Hybrid CNN and Bi-LSTM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 2, pp. 327–333, Apr. 2025, doi: 10.29207/resti.v9i2.6334.

[9] Merinda Lestandy and Abdurrahim, "Effect of Word2Vec Weighting with CNN-BiLSTM Model on Emotion Classification," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 12, no. 1, pp. 99–107, Mar. 2023, doi: 10.23887/janapati.v12i1.58571.

[10] L. N. Aqilla, Y. Sibaroni, and S. S. Prasetiyowati, "Word2vec Architecture in Sentiment Classification of Fuel Price Increase Using CNN-BiLSTM Method," *Sinkron*, vol. 8, no. 3, pp. 1654–1664, Jul. 2023, doi: 10.33395/sinkron.v8i3.12639.

[11] Muhammad Taufik Maulana, Lailil Muflikhah, and Tirana Noor Fatyanosa, "Analisis Sentimen Pengguna Indodax Menggunakan FastText dan Convolutional Neural Network (CNN)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK)*, vol. 9, no. 6, Jun. 2025.

[12] A. S. Nugroho and K. Nugroho, "Comparison of RNN and LSTM Algorithms Based on Fasttext Embeddings in Sentiment Analysis

on the Merdeka Mengajar Platform," *sinkron*, vol. 9, no. 1, pp. 117–128, Jan. 2025, doi: 10.33395/sinkron.v9i1.14296.

[13] A. Saekhu, B. Berlilana, and D. I. S. Saputra, "Comparative Analysis of Data Balancing Techniques for Machine Learning Classification on Imbalanced Student Perception Datasets," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 627–640, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.4286.

[14] C. Sintiya, G. H. Hutagaol, D. Bate`e, and S. Irviantina, "Evaluasi Teknik Resampling untuk Class Balancing dalam Analisis Sentimen Kesehatan Mental Berbasis Bi-LSTM," *Jurnal Sifo Mikroskil*, vol. 26, no. 2, Oct. 2025, doi: 10.55601/jsm.v26i2.1799.

[15] Y. Yang, H. A. Khorshidi, and U. Aickelin, "A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems," *Front. Digit. Health*, vol. 6, Jul. 2024, doi: 10.3389/fdgth.2024.1430245.

[16] M. Carvalho, A. J. Pinho, and S. Brás, "Resampling approaches to handle class imbalance: a review from a data perspective," *J. Big Data*, vol. 12, no. 1, p. 71, Mar. 2025, doi: 10.1186/s40537-025-01119-4.