EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer **DOI:** https://doi.org/10.47065/explorer.v6i1.2856

Klasifikasi Sentimen Teks *Code-Mixed* Indonesia–Inggris Non-Formal Pada X Menggunakan Model *Fine-Tuned* DistilBERT

Syifa Arifah Nurbayani[1*], Dian Sa'adillah Maylawati[2], Aldy Rialdy Atmadja[3]
[1*,2,3] Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
[1*]syifaarifahnrb@gmail.com , [2]diansm@uinsgd.ac.id , [3]aldyrialdy@uinsgd.ac.id
[*)]syifaarifahnrb@gmail.com

Abstrak-Perkembangan media sosial mendorong meningkatnya penggunaan bahasa campuran (code-mixed) Indonesia–Inggris dalam komunikasi digital, khususnya pada media sosial X. Karakteristik teks media sosial yang non-formal, seperti penggunaan slang, singkatan, emoji, dan perpindahan bahasa dalam satu kalimat, menyebabkan proses analisis sentimen menjadi lebih kompleks dibandingkan teks monolingual. Penelitian ini bertujuan melakukan klasifikasi sentimen teks code-mixed dengan melakukan pengukuran performa model transformer ringan DistilBERT dan membandingkannya dengan model IndoBERTweet. Penelitian menggunakan metodologi CRISP-DM yang meliputi tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Dataset yang digunakan terdiri dari data primer hasil crawling media sosial X periode 2022–2026 sebanyak 1.108 data dan dataset sekunder dari penelitian sebelumnya sebanyak 5.048 data. Penelitian ini menerapkan tiga skenario pengujian, yaitu translasi ke Bahasa Indonesia, translasi ke Bahasa Inggris, dan raw data tanpa translasi. Hasil penelitian menunjukkan bahwa DistilBERT memperoleh performa terbaik pada skenario translasi Bahasa Inggris dengan akurasi sebesar 73,66% pada dataset primer dan 72,00% pada dataset sekunder. Sementara itu, IndoBERTweet menunjukkan performa terbaik pada skenario translasi Bahasa Indonesia dengan akurasi mencapai 79,01%. Hasil tersebut menunjukkan bahwa kesesuaian bahasa data dengan karakteristik pretraining model berpengaruh terhadap performa klasifikasi sentimen pada teks code-mixed non-formal.

Kata Kunci: Klasifikasi Sentimen, *Code-Mixed*, DistilBERT, Media Sosial, Transformer Ringan

Abstract-The rapid growth of social media has increased the use of Indonesian–English code-mixed language in digital communication, particularly on social media X. The non-formal characteristics of social media text, such as slang, abbreviations, emojis, and language switching within a single sentence, make sentiment analysis more challenging than monolingual text. This study aims to perform sentiment classification on code-mixed text by evaluating the performance of the lightweight Transformer model DistilBERT and comparing it with IndoBERTweet. The study adopts the CRISP-DM methodology, which consists of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment stages. The dataset comprises 1,108 primary data collected from social media X between 2022 and 2026 and 5,048 secondary data obtained from previous research. Three experimental scenarios were applied: translation into Indonesian, translation into English, and raw data without translation. The results show that DistilBERT achieved its best performance on the English translation scenario, with accuracies of 73.66% on the primary dataset and 72.00% on the secondary dataset. Meanwhile, IndoBERTweet obtained the highest performance on the Indonesian translation scenario, achieving an accuracy of 79.01%. These findings indicate that the alignment between the language of the input data and the pre-training characteristics of the model significantly affects sentiment classification performance on non-formal code-mixed text.

Keywords: Sentiment Classification, Code-Mixed, DistilBERT, Social Media, Lightweight Transformer

1. PENDAHULUAN

Perkembangan media sosial telah mengubah pola komunikasi masyarakat secara signifikan, termasuk meningkatnya penggunaan bahasa campuran (*code-mixing*) dalam menyampaikan opini maupun emosi secara daring [1], [2]. Fenomena ini banyak ditemukan pada platform media sosial seperti X (Twitter), di mana pengguna sering mencampurkan Bahasa Indonesia dan Bahasa Inggris dalam satu kalimat [3]. Penggunaan bahasa campuran dipengaruhi oleh dinamika sosial masyarakat multibahasa dan perkembangan budaya digital di kalangan generasi muda [4]. Selain sebagai bentuk komunikasi, *code-mixing* juga sering digunakan sebagaFi penanda identitas sosial, khususnya pada masyarakat perkotaan seperti Jakarta Selatan yang banyak terpapar budaya internasional [5]. Kondisi tersebut menjadikan media sosial sebagai sumber data penting dalam penelitian analisis sentimen, namun sekaligus menghadirkan tantangan dalam bidang *Natural Language Processing* (NLP). Analisis sentimen merupakan proses untuk mengidentifikasi, mengekstraksi, dan menginterpretasikan opini, sikap, atau emosi yang diekspresikan dalam sebuah teks [6]. Analisis sentimen memiliki peran penting dalam berbagai domain seperti

bisnis, politik, dan pemantauan media sosial karena mampu mengungkap opini publik serta mendukung pengambilan keputusan berbasis data[7]. Analisis sentimen pada teks *code-mixed* memiliki kompleksitas lebih tinggi dibandingkan teks monolingual karena mengandung variasi bahasa tidak baku, slang, singkatan, perpindahan bahasa, serta berbagai *noise* seperti emoji dan simbol [9],[9], [10]. Karakteristik tersebut menyebabkan model klasifikasi sentimen kesulitan memahami konteks linguistik secara optimal [11]. Oleh karena itu, diperlukan strategi praproses yang mampu mengurangi *noise* tanpa menghilangkan karakteristik non-formal dari data media sosial. Tahap praproses (*preprocessing*) menjadi salah satu komponen penting dalam meningkatkan kualitas data sebelum digunakan pada proses pelatihan model. Penelitian Astuti et al. menunjukkan bahwa normalisasi kata tidak baku, konversi emoji, serta penanganan variasi bahasa mampu meningkatkan performa model analisis sentimen pada data campuran Indonesia–Inggris [1]. Namun, proses normalisasi yang terlalu agresif berpotensi menghilangkan konteks emosional dan pola bahasa khas pengguna media sosial [12]. Oleh karena itu, penelitian ini menerapkan praproses secara moderat melalui *cleaning text*, normalisasi kata tertentu, penanganan simbol tidak relevan, serta konversi emoji.

Berbagai penelitian sebelumnya telah menerapkan model berbasis Transformer untuk mengatasi permasalahan analisis sentimen pada teks media sosial. Astuti et al.[12] menggunakan IndoBERTweet pada dataset Twitter campuran Indonesia–Inggris dan menunjukkan bahwa model tersebut mampu memberikan performa lebih baik dibandingkan beberapa model Transformer lainnya pada tugas klasifikasi sentimen *code-mixed*. Penelitian Hidayatullah [3] pada dataset Indonesia–Jawa–Inggris juga menunjukkan bahwa model monolingual seperti IndoBERT dan IndoBERTweet mampu mengungguli model multibahasa seperti mBERT, DistilmBERT, dan XLM-RoBERTa dalam memahami karakteristik bahasa lokal. Penelitian lain oleh Koto et al. [13] menunjukkan bahwa model berbasis BERT memiliki kemampuan representasi konteks yang baik untuk berbagai tugas NLP Bahasa Indonesia, termasuk klasifikasi teks dan analisis sentimen. Selain itu, penelitian Setiono et al. [1] menerapkan pendekatan translasi pada data *code-mixed* dan menunjukkan bahwa translasi dapat membantu meningkatkan kualitas representasi teks pada model Transformer. Sementara itu, penelitian Devlin et al. [14] memperkenalkan BERT sebagai model Transformer dengan performa tinggi pada berbagai tugas NLP, namun model tersebut memiliki jumlah parameter besar dan membutuhkan sumber daya komputasi yang tinggi.

Meskipun berbagai penelitian telah menerapkan model Transformer untuk analisis sentimen pada data multilingual dan code-mixed, sebagian besar penelitian berfokus pada model berukuran besar seperti BERT, IndoBERT, dan IndoBERTweet. Evaluasi terhadap model Transformer ringan seperti DistilBERT pada data code-mixed Indonesia–Inggris non-formal masih relatif terbatas. Selain itu, pengaruh strategi translasi terhadap performa DistilBERT pada berbagai ukuran dataset belum banyak dikaji secara mendalam. Oleh karena itu, diperlukan penelitian yang mengevaluasi kemampuan DistilBERT dalam menangani karakteristik bahasa campuran dan non-formal pada media sosial.

Berdasarkan permasalahan dan research gap yang telah diidentifikasi, penelitian ini menggunakan DistilBERT sebagai model Transformer ringan untuk klasifikasi sentimen teks *code-mixed* Indonesia–Inggris pada media sosial X. DistilBERT dipilih karena memiliki jumlah parameter yang lebih sedikit dibandingkan model Transformer konvensional, namun tetap mampu mempertahankan kemampuan representasi bahasa yang baik [15]. Selain itu, penelitian ini menerapkan pendekatan *fine-tuning* dengan dukungan tahapan praproses teks seperti *cleaning text*, normalisasi kata tidak baku, penanganan simbol tidak relevan, dan konversi emoji untuk meningkatkan kualitas representasi data. Pendekatan *fine-tuning* memungkinkan model pra-latih digunakan kembali pada tugas tertentu tanpa perlu melakukan pelatihan dari awal, sehingga kebutuhan data dan komputasi menjadi lebih efisien [16]. Penelitian ini juga mengevaluasi pengaruh skenario translasi dan variasi ukuran dataset terhadap performa model dalam melakukan klasifikasi sentimen. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kemampuan model secara komprehensif. Tujuan dari penelitian ini adalah menganalisis performa DistilBERT pada klasifikasi sentimen teks *code-mixed* Indonesia–Inggris dengan mempertimbangkan aspek akurasi dan efisiensi komputasi. Penelitian ini memberikan kontribusi dalam pengembangan model analisis sentimen yang lebih ringan, efisien, dan adaptif terhadap karakteristik bahasa non-formal pada media sosial.

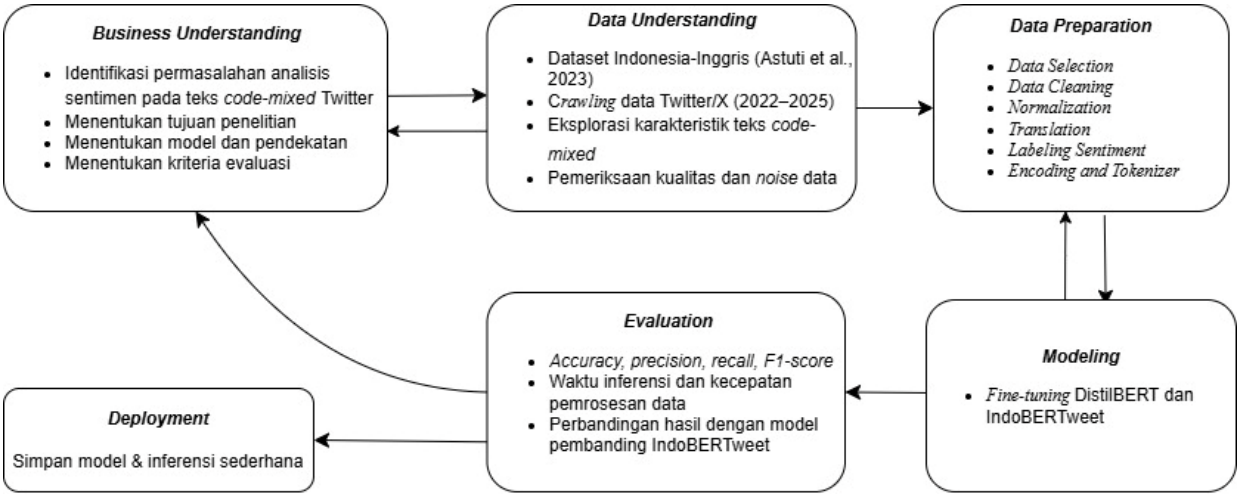
2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahapan utama, yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment* [17]. Metode ini dipilih karena memberikan alur kerja yang sistematis dan fleksibel dalam pengembangan model pembelajaran mesin, khususnya pada tugas analisis sentimen berbasis Transformer. Tahapan penelitian dimulai dari identifikasi permasalahan analisis sentimen pada teks *code-mixed* Indonesia–Inggris di media sosial X, dilanjutkan dengan pengumpulan dan pemahaman dataset, proses praproses

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer **DOI:** https://doi.org/10.47065/explorer.v6i1.2856

data, pelatihan model menggunakan teknik *fine-tuning*, evaluasi performa model, hingga implementasi model untuk prediksi sentimen. Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Menggunakan CRISP-DM

Pada tahap *Business Understanding* dilakukan identifikasi permasalahan penelitian, tujuan yang ingin dicapai, serta kebutuhan sistem dalam melakukan klasifikasi sentimen pada teks *code-mixed* Indonesia–Inggris. Tahap *Data Understanding* mencakup proses pengumpulan data dari platform X dan analisis karakteristik data, seperti distribusi sentimen, variasi bahasa, serta bentuk-bentuk bahasa non-formal yang digunakan pengguna. Selanjutnya, tahap *Data Preparation* dilakukan untuk meningkatkan kualitas data melalui *cleaning text*, normalisasi kata tidak baku, penanganan simbol yang tidak relevan, konversi emoji, translasi sesuai skenario pengujian, pelabelan sentimen, dan tokenisasi. Pada tahap *Modeling* dilakukan proses *fine-tuning* model DistilBERT dan IndoBERTweet menggunakan dataset yang telah dipersiapkan. Tahap *Evaluation* dilakukan untuk mengukur performa model menggunakan metrik evaluasi yang telah ditentukan. Tahap terakhir yaitu *Deployment* dilakukan dengan mengimplementasikan model terbaik untuk melakukan prediksi sentimen pada data baru.

2.2 *Business Understanding*

Tahap *Business Understanding* dilakukan untuk memahami kebutuhan penelitian dalam analisis sentimen pada teks *code-mixed* Indonesia–Inggris yang memiliki karakteristik tidak baku, seperti penggunaan slang, singkatan digital, emoji, dan perpindahan bahasa dalam satu kalimat. Permasalahan tersebut membutuhkan model klasifikasi sentimen yang tidak hanya akurat, tetapi juga efisien secara komputasi. Pada tahap ini ditentukan tujuan penelitian, yaitu menerapkan model DistilBERT untuk klasifikasi sentimen pada data media sosial X menggunakan pendekatan *fine-tuning.* Selain itu, ditentukan pula metrik evaluasi yang digunakan, meliputi *accuracy, precision, recall*, dan *F1-score*. Pada tahap ini ditentukan tujuan penelitian, yaitu menerapkan model DistilBERT untuk klasifikasi sentimen pada data media sosial X menggunakan pendekatan *fine-tuning*. Selain itu, ditentukan pula metrik evaluasi yang digunakan, meliputi *accuracy, precision, recall*, dan *F1-score* untuk mengukur performa model dalam melakukan klasifikasi sentimen.

2.3 *Data Understanding*

Tahap *Data Understanding* dilakukan dengan mengumpulkan dan mempelajari karakteristik dataset yang digunakan dalam penelitian. Dataset terdiri dari data primer dan data sekunder. Data primer diperoleh melalui proses *crawling* pada *platform* X menggunakan *tools* Tweet Harvest dengan dua metode pengambilan data, yaitu *by keyword* dan *by user*. Pada metode *by keyword*, pengambilan data didasarkan pada fenomena sosiolinguistik bahasa "Jaksel" yang identik dengan penggunaan campuran Bahasa Indonesia dan Bahasa Inggris di media sosial [18]. Proses *crawling* menggunakan kombinasi kata kunci seperti *literally*, *actually*, *worth it*, *stress*, *capek*, *prefer*, dan berbagai istilah non-formal lainnya dengan parameter *lang:id*, *-is:retweet*, dan *-is:reply*. Sementara itu, metode *by user* dilakukan dengan mengambil data dari beberapa *public figure* dan akun media sosial yang dikenal sering menggunakan bahasa campuran, seperti Cinta Laura, Vidi Aldiano, Raymond Chins, dan akun lainnya.

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

Tabel 1. Contoh Hasil *Crawling By Keywords*

No	***Keyword***	**Teks *Code-Mixed***
1	*Worth It*	*the most* ***worth it*** *things to do during day off is* "tidur siang" gatau knp sih *at least* sejam dua jam
2	*Prefer*	aku lebih ***prefer*** buat *take a rest*, bangun pagi / subuh terus lanjut belajar
3	*literally*	***Literally*** baru aja lewat *memories* ig wktu w iconic bbrp tahun lalu

Tabel 2. Contoh Hasil *Crawling By User*

No	***Keyword***	**Teks *Code-Mixed***
1	*@xcintakiehlx*	Kamu yang menentukan nilai dirimu. *Walk your worth because you're worth it.*
2	*@vidialdiano*	Demi merayakan juga bias gue punya social media! *Are you ready for this?*
3	*@ raymondchins*	*Doomscrolling is probably the most underrated thing* yang banyak bikin orang ga sukses2

Proses *crawling* dilakukan pada periode Oktober 2022 hingga Maret 2026 dan menghasilkan 1.108 data setelah tahap praproses. Selain data primer, penelitian ini juga menggunakan data sekunder berupa dataset *Indonglish Code-Mixed Sentiment Analysis* dari penelitian Astuti *et al.* dengan periode data Agustus 2020 hingga September 2022 yang berjumlah 5.048 data. Penelitian ini juga menggunakan dataset gabungan antara data primer dan data sekunder untuk menganalisis pengaruh variasi dataset terhadap performa model. Pada tahap *data understanding*, dilakukan analisis awal data meliputi pemeriksaan data duplikat, *missing value*, jumlah data, tipe data, serta karakteristik bahasa non-formal seperti penggunaan *slang*, singkatan digital, emoji, dan variasi penulisan pada teks *code-mixed*. Hasil analisis digunakan sebagai dasar dalam menentukan strategi praproses dan pemodelan pada tahap selanjutnya

Tabel 3. Jenis Dataset

Jenis Dataset	Jumlah Data
Primer	1.108
Sekunder	5.048
Gabungan (Primer dan Sekunder)	6.156

2.4 *Data Preparation*

Pada Tahap *Data Preparation* bertujuan menghasilkan dataset yang siap digunakan pada proses pelatihan model. Tahapan pertama yang dilakukan adalah *data selection*, yaitu melakukan penyaringan data berdasarkan bahasa dengan memilih teks berlabel *lang: in*. Selanjutnya dilakukan proses identifikasi teks *code-mixed* menggunakan dua tahap, yaitu pendekatan *dictionary-based* untuk mendeteksi campuran Bahasa Indonesia dan Bahasa Inggris secara otomatis, kemudian dilanjutkan dengan proses pemeriksaan manual untuk memastikan kesesuaian data *code-mixed*. Setelah proses seleksi data, dilakukan *data cleaning* yang meliputi penanganan data duplikat, perbaikan *encoding* emoji, serta konversi emoji menjadi representasi teks untuk mempertahankan informasi sentimen pada data media sosial. Tahap berikutnya adalah *data transformation* melalui normalisasi sederhana pada kata tidak baku serta penerapan translasi ke Bahasa Indonesia dan Bahasa Inggris sesuai skenario penelitian. Selanjutnya dilakukan tahap *labeling* berupa pemberian label sentimen untuk menentukan polaritas suatu pendapat bersifat positif, negatif, atau netral [19], [20]. Pemberian label sentimen menggunakan pendekatan *pseudo labeling* dengan bantuan model *pretrained* XLM-RoBERTa [21]. Hasil label kemudian dilakukan proses *encoding* menjadi bentuk numerik sebelum dilakukan tokenisasi menggunakan *tokenizer* bawaan BERT agar format data sesuai dengan arsitektur model Transformer. Seluruh tahapan praproses dilakukan untuk mengurangi *noise* dan meningkatkan kualitas representasi teks sebelum digunakan pada proses pelatihan model.

Tabel 4. Contoh Hasil Translasi dan Label Sentimen

Teks *code-mixed*	***Translation Indonesian***	***Translation English***	**Label Sentimen**
Merayakan minggu terakhir bulan puasa dengan bermain dan berbagi dengan anak-	Merayakan minggu terakhir bulan puasa dengan bermain dan berbagi dengan anakanak baik ini sore hari yang	*celebrate the last week of the fasting month by playing and sharing with these good children an afternoon well spent id like to*	

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

anak baik ini. *An afternoon well spent.. I'd like to extend a special thank you to everyone that helped make this happen!*	dihabiskan dengan baik id ingin menyampaikan terima kasih khusus kepada semua orang yang membantu mewujudkan ini.	*extend a special thank you to everyone who helped make this happen*	Positif
underrated fact paylater is toxic mending kartu kredit terlalu gampang bikin amp *habit* konsumtif nya ngerusak mental blm lagi bunga nya bs lebih *think twice*	fakta yang diremehkan paylater adalah racun memperbaiki kartu kredit terlalu mudah bikin amp kebiasaan konsumtif nya ngerusak mental blm lagi bunga nya bs lebih berpikir dua kali	*underrated fact paylater is toxic, it's better if credit cards are too easy to make & the consumptive habit is mentally damaging, not to mention the interest can be more, think twice*	Negatif
Rekomendasi parfum yg *affordable* nih guys	rekomendasi parfum yang terjangkau nih guys	*Recommended affordable perfume, guys*	Netral

2.5 *Modeling*

Tahap *Modeling* dilakukan dengan menerapkan teknik *fine-tuning* pada model DistilBERT untuk tugas klasifikasi sentimen teks *code-mixed*. DistilBERT dipilih karena merupakan model Transformer ringan hasil distilasi BERT yang memiliki jumlah parameter lebih sedikit namun tetap mampu mempertahankan kemampuan representasi bahasa yang baik [15]. Selain DistilBERT sebagai model utama, penelitian ini juga menggunakan IndoBERTweet sebagai model pembanding untuk mengevaluasi performa model ringan terhadap model Transformer monolingual yang telah dilatih menggunakan domain media sosial Twitter berbahasa Indonesia [13]. Dataset yang telah melalui tahap praproses kemudian dibagi menjadi data *training*, *validation*, dan *testing* [12].Proses pelatihan model dilakukan menggunakan teknik *fine-tuning* dengan beberapa parameter pelatihan seperti *learning rate*, *batch size*, dan jumlah *epoch* yang disesuaikan dengan kebutuhan penelitian. Pelatihan dilakukan selama tiga *epoch* [14] disertai proses validasi pada setiap *epoch* untuk memperoleh model dengan nilai *validation loss* terbaik. Analisis terhadap *training loss* dan *validation loss* digunakan untuk melihat kemampuan model dalam mempelajari data serta mengidentifikasi kemungkinan terjadinya *overfitting* atau *underfitting*. Model terbaik hasil validasi selanjutnya digunakan pada tahap pengujian menggunakan data *testing* yang telah melalui tahapan praproses yang sama dengan data pelatihan. Untuk menganalisis pengaruh translasi dan variasi dataset terhadap performa model, penelitian ini menerapkan beberapa skenario eksperimen yang dijelaskan pada bagian Skenario pengujian.

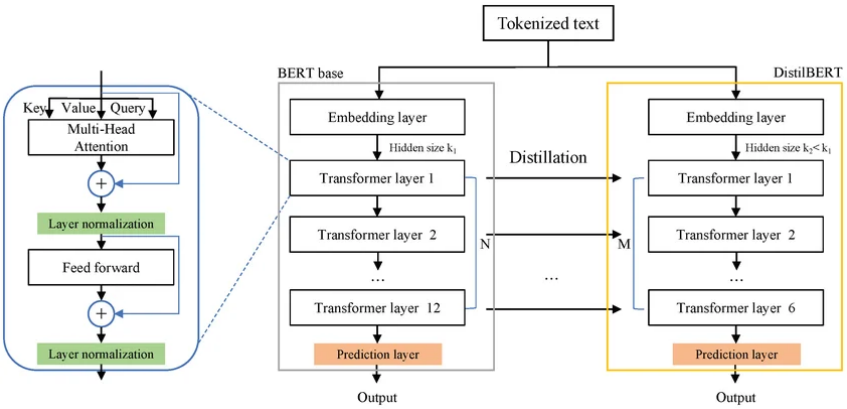
Tabel 5. Skenario Pengujian

Jenis Dataset	Tahapan Proses
Skenario 1	1. *Data Selection* 2. *Data Cleaning* 3. *Emoji conversion* 4. *Normalization* 5. *Translation to English* 6. *Labeling Sentiment* 7. *Training Model DistilBERT and IndoBERTweet* 8. *Validation* 9. *Testing and Evaluation*
Skenario 2	1. *Data Selection* 2. *Data Cleaning* 3. *Emoji conversion* 4. *Normalization* 5. *Translation to Indonesia* 6. *Labeling Sentiment* 7. *Training Model DistilBERT and IndoBERTweet* 8. *Validation* 9. *Testing and Evaluation*
Skenario 3	1. *Data Selection* 2. *Labeling Sentiment* 3. *Training Model DistilBERT and IndoBERTweet* 4. *Validation* 5. *Testing and Evaluation*

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

2.5.1 Model DistilBERT

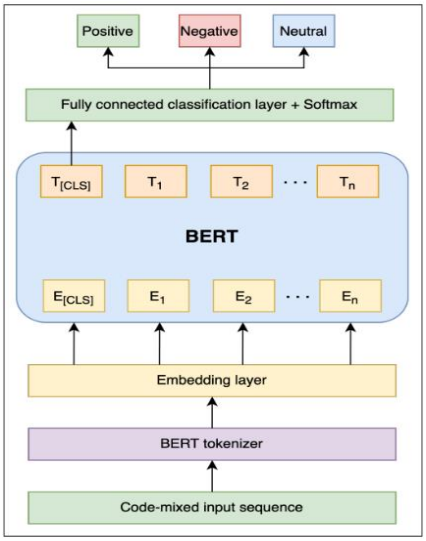
DistilBERT merupakan model Transformer ringan hasil proses distilasi dari BERT yang dikembangkan untuk mengurangi kompleksitas komputasi tanpa menghilangkan kemampuan representasi bahasa secara signifikan [15]. Model ini memiliki jumlah parameter lebih sedikit dibandingkan BERT konvensional sehingga lebih efisien dalam penggunaan memori dan waktu pelatihan. DistilBERT tetap mempertahankan kemampuan contextual embedding yang baik pada berbagai tugas Natural Language Processing (NLP), termasuk klasifikasi teks dan analisis sentimen. Pada penelitian ini, DistilBERT digunakan sebagai model utama untuk melakukan klasifikasi sentimen pada teks code-mixed Indonesia–Inggris melalui pendekatan fine-tuning. Gambar 2 menunjukkan proses distillation pada DistilBERT dari model BERT base. DistilBERT menggunakan jumlah layer Transformer yang lebih sedikit dibandingkan BERT sehingga mampu mengurangi kompleksitas komputasi dan mempercepat proses inferensi tanpa mengurangi performa secara signifikan [22].



Gambar 2. Arsitektur Model DistilBERT

2.5.2 Model IndoBERTweet

IndoBERTweet merupakan model *pretrained Transformer* berbasis BERT yang dilatih menggunakan korpus media sosial berbahasa Indonesia, khususnya data Twitter [13]. Model ini dirancang untuk memahami karakteristik bahasa non-formal pada media sosial, seperti penggunaan slang, singkatan, emoji, dan variasi penulisan informal lainnya. Pada penelitian ini, IndoBERTweet digunakan sebagai model pembanding untuk mengevaluasi performa DistilBERT dalam klasifikasi sentimen teks *code-mixed* Indonesia–Inggris. Penggunaan IndoBERTweet bertujuan untuk melihat pengaruh kesesuaian domain bahasa terhadap performa model dalam memahami karakteristik data media sosial. Gambar 3 menunjukkan arsitektur IndoBERTweet berbasis BERT encoder untuk klasifikasi sentimen. Teks diproses melalui tokenizer, embedding layer, dan encoder BERT, kemudian representasi token [CLS] digunakan pada classification layer dan softmax untuk menentukan kelas sentimen. IndoBERTweet dilatih menggunakan korpus Twitter berbahasa Indonesia sehingga lebih sesuai untuk data media sosial non-formal [3].



Gambar 3. Arsitektur Klasifikasi Sentimen Berbasis BERT pada IndoBERTweet

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

2.6 *Evaluation*

Tahap evaluasi dilakukan untuk mengukur performa model dalam melakukan klasifikasi sentimen pada teks code-mixed Indonesia–Inggris. Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score [1]. Selain performa klasifikasi, penelitian ini juga menganalisis efisiensi komputasi melalui waktu inferensi dan kecepatan pemrosesan data (throughput). Hasil evaluasi digunakan untuk membandingkan performa DistilBERT dan IndoBERTweet pada berbagai skenario pengujian.
Accuracy mengukur proporsi data yang berhasil diklasifikasikan dengan benar, yaitu gabungan antara true positive dan true negative, dibandingkan dengan keseluruhan sampel.

$$Accuracy = \frac{TP+TN}{Total} \quad (1)$$

Precision menggambarkan tingkat ketepatan model dalam memprediksi suatu kelas tertentu, yang dihitung berdasarkan rasio antara *true positive* dan *false positive.*

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall atau *sensitivity* mengukur kemampuan model dalam mengenali seluruh sampel yang benar untuk suatu kelas, yakni perbandingan antara *true positive* dan *false negative.*

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score, yang merupakan *harmonic mean* dari *precision* dan *recall*, digunakan untuk menyeimbangkan *trade-off* antara kedua metrik tersebut, terutama ketika dataset memiliki distribusi kelas yang tidak seimbang [1]

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Labeling

Hasil proses labeling menghasilkan tiga kelas sentimen, yaitu positif, negatif, dan netral. Distribusi data pada masing-masing kelas dapat dilihat pada Tabel 6.

Tabel 6. Distribusi Sentimen Dataset

Dataset	Positif	Negatif	Netral	Total
Primer	266	453	389	1.108
Sekunder	1.570	1.987	1.491	5.048
Gabungan (Primer dan Sekunder)	1.836	2.440	1.880	6.156

Berdasarkan hasil distribusi sentimen, dataset terdiri atas tiga kelas yaitu positif, negatif, dan netral pada dataset primer, sekunder, dan gabungan. Pada seluruh dataset, kelas negatif memiliki jumlah data yang sedikit lebih tinggi dibandingkan kelas lainnya, namun secara umum distribusi antar kelas masih relatif seimbang.

3.2 Hasil Modeling

Tahap *Evaluation* dilakukan untuk mengukur performa model dalam melakukan klasifikasi sentimen. Evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* [1] Selain performa klasifikasi, penelitian ini juga menganalisis efisiensi komputasi melalui waktu inferensi dan kecepatan pemrosesan data. Hasil evaluasi digunakan untuk membandingkan performa DistilBERT dan IndoBERTweet dalam menangani data *code-mixed* Indonesia–Inggris pada media sosial X.

3.2.1 Hasil Pengujian pada Dataset Primer

Hasil pengujian pertama dilakukan menggunakan Dataset Primer dengan tiga skenario data, yaitu hasil translasi bahasa Indonesia, data hasil translasi bahasa Inggris, dan data mentah (*raw data*). Hasil *accuracy* yang diperoleh dapat dilihat pada Tabel 7.

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

Tabel 7. Hasil Pengujian pada Dataset Primer

Skenario	Model	*Precision*	*Recall*	*F1-Score*	*Accuracy*
Skenario 1 (*Translate to Indonesia*)	DistilBERT	0.3728	0.4910	0.4236	0.4910
	IndoBERTweet	0.8051	0.7901	0.7908	**0.7901**
Skanrio 2 (*Translate to English*)	DistilBERT	0.5912	0.7366	0.5510	**0.7366**
	IndoBERTweet	0.7058	0.7053	0.7054	0.7053
Skeanrio 3 (*Raw Data*)	DistilBERT	0.5944	0.6052	0.5735	0.6052
	IndoBERTweet	0.7497	0.7456	0.7397	**0.7456**

Berdasarkan Tabel 7, hasil pengujian menunjukkan bahwa skenario translasi memengaruhi performa model dalam klasifikasi sentimen teks code-mixed Indonesia–Inggris. Pada Skenario 1 (Translate to Indonesia), IndoBERTweet memperoleh accuracy tertinggi sebesar 79.01%, sedangkan DistilBERT hanya mencapai 49.10%. Sebaliknya, pada Skenario 2 (Translate to English), performa DistilBERT meningkat menjadi 73.66%, yang menunjukkan bahwa model lebih optimal pada teks berbahasa Inggris sesuai karakteristik pre-training-nya. Pada Skenario 3 (Raw Data), kedua model mengalami penurunan performa akibat kompleksitas teks code-mixed yang masih mengandung campuran bahasa dan variasi bahasa non-formal. Secara keseluruhan, hasil tersebut menunjukkan bahwa kesesuaian bahasa data dengan karakteristik model berpengaruh terhadap performa klasifikasi sentimen

3.2.2 Hasil Pengujian pada Dataset Sekunder

Hasil pengujian dilakukan menggunakan Dataset Sekunder dengan tiga skenario data, yaitu hasil translasi bahasa Indonesia, data hasil translasi bahasa Inggris, dan data mentah (*raw data*). Hasil *accuracy* yang diperoleh dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pengujian pada Dataset Sekunder

Skenario	Model	*Precision*	*Recall*	*F1-Score*	*Accuracy*
Skenario 1 (*Translate to Indonesia*)	DistilBERT	0.6035	0.6027	0.6030	0.6027
	IndoBERTweet	0.7686	0.7692	0.7687	**0.7692**
Skanrio 2 (*Translate to English*)	DistilBERT	0.7188	0.7200	0.7151	0.7200
	IndoBERTweet	0.7472	0.7418	0.7394	**0.7418**
Skeanrio 3 (*Raw Data*)	DistilBERT	0.6560	0.6567	0.6561	0.6567
	IndoBERTweet	0.7396	0.7374	0.7358	**0.7478**

Berdasarkan hasil pengujian pada Dataset Astuti, model DistilBERT memperoleh performa terbaik ketika menggunakan data hasil translasi bahasa Inggris dengan *accuracy* sebesar 72.00%. Hasil tersebut menunjukkan bahwa DistilBERT lebih optimal dalam memproses teks berbahasa Inggris dibandingkan teks berbahasa Indonesia. Sementara itu, model IndoBERTweet menunjukkan performa terbaik pada data hasil translasi bahasa Indonesia dengan *accuracy* sebesar 76.56%. Selain itu, performa IndoBERTweet juga relatif stabil pada seluruh skenario pengujian dibandingkan DistilBERT. Perbedaan performa antara kedua model terlihat cukup signifikan pada data translate Indonesia dengan selisih *accuracy* sekitar 15%. Akan tetapi, pada data translate Inggris, selisih performa kedua model menjadi lebih kecil, yaitu sekitar 2%. Hal tersebut menunjukkan bahwa penggunaan bahasa Inggris membantu meningkatkan performa DistilBERT sehingga mendekati performa IndoBERTweet.

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

3.2.3 Hasil Pengujian pada Dataset Gabungan Primer dan Sekunder

Pengujian kedua dilakukan menggunakan Dataset Gabungan Primer dan Sekunder untuk mengetahui pengaruh penambahan data dengan distribusi bahasa yang lebih beragam terhadap performa model. Hasil pengujian dapat dilihat pada Tabel 9.

Tabel 9. Hasil Pengujian pada Dataset Gabungan Primer dan Sekunder

Skenario	Model	*Precision*	*Recall*	*F1-Score*	*Accuracy*
Skenario 1 (*Translate to Indonesia*)	DistilBERT	0.6083	0.6117	0.6088	0.6117
	IndoBERTweet	0.7710	0.7692	0.7700	**0.7692**
Skanrio 2 (*Translate to English*)	DistilBERT	0.7025	0.7021	0.6995	0.7021
	IndoBERTweet	0.7465	0.7457	0.7453	**0.7457**
Skeanrio 3 (*Raw Data*)	DistilBERT	0.6541	0.6529	0.6479	0.6529
	IndoBERTweet	0.7220	0.7183	0.7146	**0.7183**

Berdasarkan hasil pengujian pada Dataset Gabungan Primer dan Sekunder, DistilBERT kembali menunjukkan performa terbaik pada data hasil translasi bahasa Inggris dengan *accuracy* sebesar 70.21%. Namun, jika dibandingkan dengan Dataset Astuti sebelumnya, *accuracy* DistilBERT mengalami penurunan. Sebaliknya, IndoBERTweet tetap mempertahankan performa terbaiknya pada data berbahasa Indonesia dengan *accuracy* sebesar 76.92%. Bahkan setelah dilakukan penambahan data baru, model IndoBERTweet masih mampu menghasilkan performa yang stabil dan cenderung meningkat. Hasil tersebut menunjukkan bahwa IndoBERTweet memiliki kemampuan generalisasi yang lebih baik terhadap variasi data dibandingkan DistilBERT. Penambahan data baru tidak memberikan penurunan performa yang signifikan pada IndoBERTweet, sedangkan DistilBERT mengalami penurunan *accuracy* pada hampir seluruh skenario pengujian.

3.2.4 Analisis Efisiensi Komputasi

Analisis efisiensi komputasi dilakukan untuk membandingkan kinerja DistilBERT dan IndoBERTweet dari aspek waktu inferensi dan kecepatan pemrosesan data. Pengujian dilakukan menggunakan skenario terbaik dari masing-masing model berdasarkan hasil evaluasi klasifikasi sentimen.

Tabel 10. Hasil Analisis Efisiensi Komputasi

Model	Accuracy	F1-Score	Runtime	Samples/sec
DistilBERT	73.66%	73.49%	0.9763	229.437
IndoBERTweet	79.01%	79.08%	1.4588	153.551

Berdasarkan Tabel 10, IndoBERTweet memperoleh performa klasifikasi terbaik dengan *accuracy* sebesar 79.01% dan *F1-score* sebesar 79.08%. Namun, DistilBERT menunjukkan efisiensi komputasi yang lebih baik dibandingkan IndoBERTweet. DistilBERT memiliki waktu inferensi yang lebih cepat, yaitu 0.9763 detik, sedangkan IndoBERTweet membutuhkan waktu 1.4588 detik. Selain itu, DistilBERT juga mampu memproses data lebih cepat dengan *throughput* sebesar 229.437 samples/sec dibandingkan IndoBERTweet sebesar 153.551 samples/sec. Hasil tersebut menunjukkan bahwa DistilBERT lebih ringan dan efisien secara komputasi meskipun performa klasifikasinya masih berada di bawah IndoBERTweet. Dengan demikian, DistilBERT memiliki potensi untuk diterapkan pada lingkungan dengan sumber daya komputasi terbatas.

3.3 Hasil Deployment

Tahap *Deployment* dilakukan dengan menyimpan model terbaik beserta tokenizer yang digunakan selama proses pelatihan. Model kemudian diimplementasikan dalam bentuk skrip inferensi sederhana untuk melakukan prediksi sentimen pada teks baru. Selain itu, tahap ini juga mencakup perencanaan penerapan, pemantauan, serta pemeliharaan model agar tetap relevan dan dapat digunakan secara berkelanjutan [23]. Untuk memverifikasi kemampuan model dalam melakukan prediksi pada data baru, dilakukan pengujian inferensi sederhana menggunakan beberapa contoh teks yang tidak termasuk dalam data pelatihan. Hasil inferensi dari model DistilBERT dan IndoBERTweet ditunjukkan pada Gambar 4 dan Gambar 5.

EXPLORER
Journal of Computer Science and Information Technology
E-ISSN 2774-4647
Website: https://journal.fkpt.org/index.php/Explorer DOI: https://doi.org/10.47065/explorer.v6i1.2856

```
Masukkan teks: Definitely uninstalling this, the latest update just introduced more bugs lol.
Predicted Sentiment: negative
----------------------------------------
Masukkan teks: Looks like the latest version is already available on the Play Store.
Predicted Sentiment: neutral
----------------------------------------
Masukkan teks: The app feels incredibly smooth after the update, absolutely awesome.
Predicted Sentiment: positive
```

Gambar 4. *Screenshot* hasil inferensi sederhana model DistilBERT

```
Masukkan teks: fix ini salah satu update paling gagal tahun ini
Predicted Sentiment: negative
----------------------------------------
Masukkan teks: yaudah lah dipake dulu aja sambil liat perkembangan
Predicted Sentiment: neutral
----------------------------------------
Masukkan teks: wkwkwk fitur baru nya kepake banget sih, respect buat tim dev
Predicted Sentiment: positive
```

Gambar 5. *Screenshot* hasil inferensi sederhana model IndoBERTweet

3.4 Pembahasan

Berdasarkan seluruh hasil eksperimen, terlihat bahwa variasi jenis data dan skenario translasi memberikan pengaruh terhadap performa model DistilBERT dan IndoBERTweet dalam klasifikasi sentimen teks code-mixed Indonesia–Inggris. DistilBERT menunjukkan performa terbaik pada data hasil translasi Bahasa Inggris, sedangkan IndoBERTweet secara konsisten memperoleh performa terbaik pada data hasil translasi Bahasa Indonesia. Hasil tersebut menunjukkan bahwa kesesuaian bahasa data dengan karakteristik pre-training model berpengaruh terhadap kualitas representasi teks dan performa klasifikasi sentimen. Temuan ini sejalan dengan penelitian Astuti et al. yang menunjukkan bahwa translasi pada data code-mixed dapat membantu meningkatkan kualitas representasi teks pada model Transformer [12]. Selain itu, penelitian Hidayatullah juga menunjukkan bahwa model monolingual seperti IndoBERT dan IndoBERTweet cenderung lebih optimal dalam memahami karakteristik bahasa lokal dibandingkan model multilingual atau model umum lainnya [3].

Pada penelitian ini, IndoBERTweet menunjukkan performa yang relatif lebih stabil terhadap perubahan variasi dataset dibandingkan DistilBERT. Hal tersebut dipengaruhi oleh proses pre-training IndoBERTweet yang menggunakan korpus media sosial berbahasa Indonesia sehingga model lebih memahami struktur kalimat, kosakata informal, singkatan, serta konteks bahasa Indonesia yang umum digunakan pada media sosial. Sementara itu, DistilBERT menunjukkan peningkatan performa yang cukup signifikan ketika menggunakan data hasil translasi Bahasa Inggris. Kondisi tersebut menunjukkan bahwa DistilBERT lebih optimal ketika memproses teks yang sesuai dengan karakteristik korpus bahasa saat proses pre-training dilakukan.

Selain pengaruh translasi, penelitian ini juga menunjukkan bahwa penambahan data dengan distribusi bahasa yang lebih beragam tidak selalu meningkatkan performa model. Pada dataset gabungan, performa DistilBERT mengalami sedikit penurunan dibandingkan pengujian pada dataset sekunder. Penurunan tersebut diduga dipengaruhi oleh meningkatnya keragaman distribusi bahasa pada data media sosial yang berasal dari periode waktu berbeda. Dataset gabungan mencakup variasi penggunaan slang, gaya komunikasi, serta kosakata non-formal yang lebih beragam dibandingkan dataset tunggal. Variasi bahasa tersebut dapat memengaruhi kualitas representasi teks yang dipelajari model, terutama pada model ringan seperti DistilBERT yang memiliki jumlah parameter lebih sedikit dibandingkan model Transformer yang lebih besar. Temuan ini sejalan dengan beberapa penelitian sebelumnya yang menunjukkan bahwa penambahan data tidak selalu menghasilkan peningkatan performa model apabila data tambahan memiliki distribusi bahasa yang berbeda atau mengandung variasi linguistik yang lebih kompleks. Penelitian pada domain media sosial juga menunjukkan bahwa perubahan distribusi bahasa (*language drift*) dan perkembangan bahasa non-formal dapat memengaruhi stabilitas performa model klasifikasi teks pada data yang bersifat dinamis [24], [25], [26], [27].

Dari aspek efisiensi komputasi, DistilBERT menunjukkan kinerja yang lebih ringan dibandingkan IndoBERTweet. Pada pengujian terbaik, DistilBERT memiliki waktu inferensi lebih cepat dengan *runtime* sebesar 0.9763 detik dan kecepatan pemrosesan data sebesar 229.437 *samples/sec*. Sementara itu, IndoBERTweet membutuhkan waktu inferensi sebesar 1.4588 detik dengan kecepatan pemrosesan sebesar 153.551 *samples/sec*. Hasil tersebut menunjukkan bahwa DistilBERT lebih efisien secara komputasi karena memiliki jumlah parameter yang lebih sedikit dibandingkan model *Transformer* konvensional. Meskipun performa klasifikasinya masih berada di bawah IndoBERTweet, DistilBERT tetap menunjukkan performa yang kompetitif dengan efisiensi komputasi yang lebih

baik. Penelitian Sanh et al. menjelaskan bahwa DistilBERT mampu mempertahankan sebagian besar kemampuan representasi bahasa dari BERT dengan jumlah parameter yang lebih sedikit, ukuran model yang lebih kecil, serta waktu inferensi yang lebih cepat dibandingkan model BERT standar [15]. Oleh karena itu, DistilBERT memiliki potensi untuk diterapkan pada lingkungan dengan sumber daya komputasi terbatas.

4. KESIMPULAN

Penelitian ini berhasil menerapkan model *fine-tuned* DistilBERT untuk klasifikasi sentimen teks *code-mixed* Indonesia–Inggris non-formal pada media sosial X. Penelitian dilakukan menggunakan metode CRISP-DM yang mencakup tahapan *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation,* dan *Deployment*. Selain menggunakan DistilBERT sebagai model utama, penelitian ini juga membandingkan performanya dengan IndoBERTweet sebagai model pembanding pada tiga skenario pengujian, yaitu translasi ke Bahasa Inggris, translasi ke Bahasa Indonesia, dan data tanpa translasi. Hasil penelitian menunjukkan bahwa skenario translasi berpengaruh terhadap performa model dalam memahami konteks linguistik pada data *code-mixed*. DistilBERT memperoleh performa terbaik pada skenario translasi ke Bahasa Inggris, sedangkan IndoBERTweet menunjukkan performa terbaik pada skenario translasi ke Bahasa Indonesia. Temuan ini menunjukkan bahwa kesesuaian bahasa data dengan karakteristik pretraining model memiliki peran penting dalam meningkatkan kualitas representasi teks dan performa klasifikasi sentimen. Selain itu, hasil penelitian juga menunjukkan bahwa perbedaan karakteristik dan distribusi dataset dapat memengaruhi kemampuan model dalam melakukan klasifikasi sentimen pada teks non-formal. Secara keseluruhan, DistilBERT mampu memberikan performa yang kompetitif dengan kebutuhan komputasi yang lebih ringan dibandingkan model Transformer berukuran besar, sehingga berpotensi diterapkan pada lingkungan dengan sumber daya komputasi terbatas. Meskipun demikian, penelitian ini memiliki keterbatasan pada jumlah data dan variasi bahasa yang digunakan. Oleh karena itu, penelitian selanjutnya dapat dilakukan dengan memperluas dataset dan melakukan optimasi model untuk meningkatkan performa klasifikasi.

REFERENSI

[1] Nisrina Hanifa Setiono and Yunita Sari, "Exploring the Impact of Back-Translation on BERT's Performance in Sentiment Analysis of Code-Mixed Language Data ," vol. 19, Apr. 2025, doi: https://doi.org/10.22146/ijccs.104757.

[2] Cuk Tho, Yaya Heryadi, Iman Herwidiana Kartowisastro, and Widodo Budiharto, "Code-Mixed Sentiment Analysis Indonesian-English Using Transformer Model," *ICIC Express Letters*, vol. 17, no. 11, pp. 1295–1302, Feb. 2023, doi: 10.24507/icicel.17.11.1295.

[3] A. F. Hidayatullah, "Code-Mixed Sentiment Analysis on Indonesian-Javanese-English Text Using Transformer Models," in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, IEEE, Aug. 2024, pp. 340–345. doi: 10.1109/ICITISEE63424.2024.10730138.

[4] A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, "Pre-trained language model for code-mixed text in Indonesian, Javanese, and English using transformer," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 30, Mar. 2025, doi: 10.1007/s13278-025-01444-9.

[5] W. Mustikawati and Kundharu Saddhono, "Analisis Penggunaan Bahasa Jaksel Pada Video Berjudul 'Language Barriers, Culture Shock & Tck Identity Crisis Ft. Mella Carli' Di Youtube: Kajian Sosiolinguistik Campur Kode," *BASA Journal of Language & Literature*, vol. 4, no. 2, pp. 89–94, Oct. 2024, doi: https://doi.org/10.33474/basa.v4i2.22031.

[6] G. Singh, "Sentiment Analysis of Code-Mixed Social Media Text (Hinglish)," Feb. 2021, doi: https://doi.org/10.48550/arXiv.2102.12149.

[7] A. K. Jena, A. K. Jena, K. M. Gopal, A. Tripathy, and N. Panda, "Optimized Feature Selection Approach for Semi-Supervised Sentiment Analysis of E-Commerce Feedback," *Journal of Computer Science*, vol. 21, no. 2, pp. 363–379, Feb. 2025, doi: 10.3844/jcssp.2025.363.379.

[8] A. Perera and A. Caldera, "Sentiment Analysis of Code-Mixed Text: A Comprehensive Review," *JUCS - Journal of Universal Computer Science*, vol. 30, no. 2, pp. 242–261, Feb. 2024, doi: 10.3897/jucs.98708.

[9] Q. Wu, P. Wang, and C. Huang, "MeisterMorxrc at SemEval-2020 Task 9: Fine-Tune Bert and Multitask Learning for Sentiment Analysis of Code-Mixed Tweets," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, Stroudsburg, PA, USA: International Committee for Computational Linguistics, 2020, pp. 1294–1297. doi: 10.18653/v1/2020.semeval-1.174.

[10] S. Javdan, T. Shangipour ataei, and B. Minaei-Bidgoli, "IUST at SemEval-2020 Task 9: Sentiment Analysis for Code-Mixed Social Media Text Using Deep Neural Networks and Linear Baselines," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, Stroudsburg, PA, USA: International Committee for Computational Linguistics, 2020, pp. 1270–1275. doi: 10.18653/v1/2020.semeval-1.170.

[11] A. Patil, V. Patwardhan, A. Phaltankar, G. Takawane, and R. Joshi, "Comparative Study of Pre-Trained BERT Models for Code-Mixed Hindi-English Data," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2023, pp. 1–7. doi: 10.1109/I2CT57861.2023.10126273.

[12] L. W. Astuti, Y. Sari, and S. -, "Code-Mixed Sentiment Analysis using Transformer for Twitter Social Media Data," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023, doi: 10.14569/IJACSA.2023.0141053.

[13] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.04607

[14] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: https://github.com/tensorflow/tensor2tensor

[15] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Mar. 2020.

[16] Y. Aliyu, A. Sarlan, K. Usman Danyaro, A. S. B. A. Rahman, and M. Abdullahi, "Sentiment Analysis in Low-Resource Settings: A Comprehensive Review of Approaches, Languages, and Data Sources," *IEEE Access*, vol. 12, pp. 66883–66909, 2024, doi: 10.1109/ACCESS.2024.3398635.

[17] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, Oct. 2021, doi: 10.30871/jaic.v5i2.3200.

[18] A. D. Wijaya and B. Bram, "A Sociolinguistic Analysis Of Indoglish Phenomenon In South Jakarta," *PROJECT (Professional Journal of English Education)*, vol. 4, no. 4, p. 672, Jul. 2021, doi: 10.22460/project.v4i4.p672-684.

[19] C. Tho, Y. Heryadi, L. Lukas, and A. Wibowo, "Code-mixed sentiment analysis of Indonesian language and Javanese language using Lexicon based approach," *J. Phys. Conf. Ser.*, vol. 1869, no. 1, p. 012084, Apr. 2021, doi: 10.1088/1742-6596/1869/1/012084.

[20] D. I. Putri, A. N. Alfian, M. Y. Putra, and P. D. Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," *Journal of Applied Informatics and Computing*, vol. 8, no. 1, pp. 7–12, Jul. 2024, doi: 10.30871/jaic.v8i1.7440.

[21] F. Barbieri, L. E. Anke, and J. Camacho-Collados, "XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond," May 2022, [Online]. Available: http://arxiv.org/abs/2104.12250

[22] H. Adel *et al.*, "Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm," *Mathematics*, vol. 10, no. 3, Feb. 2022, doi: 10.3390/math10030447.

[23] R. A. Casonatto, T. De Pádua Grillo Souza, and A. M. Mariano, "Quality and Risk Management in Data Mining: A CRISP-DM Perspective.," *Procedia Comput. Sci.*, vol. 242, pp. 161–168, 2024, doi: 10.1016/j.procs.2024.08.257.

[24] J. Khan, K. Ahmad, S. K. Jagatheesaperumal, and K. A. Sohn, "Textual variations in social media text processing applications: challenges, solutions, and trends," *Artif. Intell. Rev.*, vol. 58, no. 3, Mar. 2025, doi: 10.1007/s10462-024-11071-z.

[25] N. Calderon *et al.*, "Measuring the Robustness of NLP Models to Domain Shifts," Apr. 2024, [Online]. Available: http://arxiv.org/abs/2306.00168

[26] S. Chen, L. Neves, and T. Solorio, "Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp," Online Workshop, 2021. [Online]. Available: https://github.com/

[27] A. Bansal and I. Gangwani, "Zero-Training Temporal Drift Detection for Transformer Sentiment Models: A Comprehensive Analysis on Authentic Social Media Streams," Nov. 2025, [Online]. Available: http://arxiv.org/abs/2512.20631