

Sentimen Analisis terhadap Artis yang Terdampak karena Boikot Produk Minuman yang Dipromosikannya Menggunakan BERT

Firbaya Mutiara Ashar, Imam Much Ibnu Subroto

Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Sultan Agung

Correspondence Author: tiaraas28@std.unissula.ac.id

Abstrak

Media sosial, khususnya Instagram, telah menjadi wadah utama bagi masyarakat untuk mengekspresikan opini, termasuk dalam merespons fenomena budaya populer seperti Korean Wave dan Kpop. Baru-baru ini, keputusan grup Kpop NCT untuk berkolaborasi dengan Starbucks memicu beragam reaksi dari penggemar, terutama di tengah gerakan boikot terhadap brand tersebut. Penelitian ini bertujuan untuk menganalisis sentimen komentar penggemar NCT terkait kolaborasi ini menggunakan model Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT). Eksperimen dilakukan pada data yang tidak seimbang dan data yang telah diseimbangkan menggunakan metode Synthetic Minority Over-sampling Technique (SMOTE). Hasil penelitian menunjukkan bahwa data yang tidak seimbang menyebabkan model mengalami overfitting dengan akurasi sebesar 83%, precision 84%, recall 83%, dan f1-score 83%. Setelah dilakukan balancing data menggunakan SMOTE, performa model meningkat dengan akurasi, precision, recall, dan f1-score sebesar 86%. Hal ini membuktikan bahwa balancing data berperan penting dalam meningkatkan performa model dalam analisis sentimen.

Keyword : Analisis sentimen, IndoBERT, SMOTE, Kpop, Boikot, Instagram

1. PENDAHULUAN

Media sosial saat ini telah menjadi bagian fundamental dalam kehidupan sehari-hari, tidak hanya untuk keperluan pribadi tetapi juga bisa digunakan untuk berbisnis, serta berbagai aktivitas lainnya. Media sosial ini menghasilkan banyak data termasuk gambar, komentar teks atau emoticon, video dan lain-lain, sehingga memungkinkan masyarakat untuk bebas beropini. Dengan adanya analisis sentimen terhadap opini yang berkembang di media sosial, dapat menghasilkan data dan informasi yang bermanfaat [1].

Instagram adalah salah satu platform media sosial yang memungkinkan penggunaannya untuk mengirim dan menerima informasi dalam bentuk gambar, video dan cerita dalam berbagai format, termasuk unggahan *feed*, Instagram *stories*, *reels* dan siaran langsung (Instagram *live*). Instagram merupakan salah satu media sosial yang paling cepat berkembang saat ini karena memungkinkan penggunaannya untuk berinteraksi dengan orang lain kapan saja dan dimana saja dari komputer maupun perangkat selular. Instagram cukup efektif sebagai media promosi dalam memberikan rangsangan melalui konten-konten yang diunggah untuk meningkatkan pengetahuan konsumen terkait produk yang ditawarkan[2]. Instagram telah menjadi salah satu platform media sosial terbesar di dunia, dengan lebih dari 2 miliar pengguna aktif per bulan. Karena fiturnya yang luas dan pengguna yang besar, serta kemampuan untuk berbagi pendapat tentang isu-isu terhangat dan kontroversial, instagram telah menjadi pilihan yang populer dibandingkan dengan *platform* sosial media lainnya [3].

Pada awal peluncuran, Instagram hanya memiliki pengguna yang berjumlah 100 ribu orang. Dalam waktu sekitar 2,5 bulan, jumlah pengguna Instagram meningkat pesat menjadi 1 juta pengguna. Instagram telah menjadi salah satu platform media sosial terbesar di dunia dengan lebih dari 2 miliar pengguna aktif bulanan. Cepatnya penyebaran informasi melalui Instagram dapat memfasilitasi pertukaran budaya, seperti budaya korea yang dikenal sebagai Korean Wave atau Hallyu, yang mencakup drama Korea dan musik Kpop. Budaya Korea ini mulai dikenal di Indonesia sejak tahun 2002 dan terus berkembang hingga sekarang, menarik minat remaja maupun orang dewasa. Kpop sebagai salah satu aspek budaya Korea yang paling digemari, telah menjadi kunci kesuksesan dari Korean Wave [4].

Namun belakangan ini salah satu *agency* dari boygrup Kpop menjadi sorotan netizen di Indonesia dikarenakan kerjasamanya dengan *brand coffee shop* yang telah diboikot oleh beberapa negara. Berawal dari ketegangan yang terjadi di Timur Tengah, Gerakan *Free Palestine* mewakili perjuangan yang kompleks untuk kemerdekaan negara Palestina dan hak asasi manusia. Seruan boikot terhadap merek dagang starbucks telah

ramai diserukan sejak tahun lalu. Dampak dari boikot ini sudah dirasakan secara global oleh Starbucks, yang mengumumkan penurunan saham sebesar 17% di awal tahun 2024. Ditengah gerakan boikot ini, keputusan SM dan NCT untuk berkolaborasi dengan Starbucks menimbulkan beragam tanggapan dari para penggemar, yang menyebabkan penurunan jumlah pengikut media sosial NCT sebanyak 500 ribu [5].

Analisis sentimen adalah teknik untuk menganalisis pikiran, perasaan, dan penilaian pengguna. Menurut beberapa sudut pandang, itu digunakan untuk menentukan apa yang diyakini pengguna berdasarkan informasi seperti opini tertulis. Karena perkembangan era digital yang tidak dapat dihindari, orang lebih sering mengungkapkan dan mengunggah ide-ide mereka di media sosial [6]. Proses sentimen analisis terdiri dari dua bagian yaitu *Sentimen Extraction* yang telah dievaluasi dan dianalisis, dan *Sentimen Classification* adalah aspek yang bermakna positif, negatif atau netral. Selanjutnya, proses penentuan perspektif mengenai masing-masing aspek tersebut [7].

Text mining menambang data dalam bentuk teks, dengan sumber data biasanya berasal dari data, *text mining* bertujuan untuk menemukan kata-kata yang dapat menunjukkan isi data, sehingga dapat dilakukan analisis hubungan antar data.[8]

Text mining menggunakan kumpulan teks yang berformat tidak beraturan, terstruktur atau minimal semi-terstruktur untuk mendapatkan informasi berguna dari sekumpulan data. *Text mining* melakukan dua tugas khusus yaitu mengkategorikan teks (*text categorization*) dan pengelompokan teks (*text clustering*). Proses *text mining* menggunakan konsep dan teknik data mining untuk menemukan pola dalam teks, khususnya dalam proses menganalisis teks, untuk menemukan informasi yang berguna untuk tujuan tertentu. Proses ini memerlukan beberapa langkah awal untuk mempersiapkan dasar agar teks dapat dianalisis [9]. *Text mining* dan *data mining* berbeda karena preprocessingnya. Data mining berkonsentrasi pada penomoran (*indexing*) dan normalisasi data, sedangkan *text mining* berkonsentrasi pada identifikasi dan ekstraksi fitur [10].

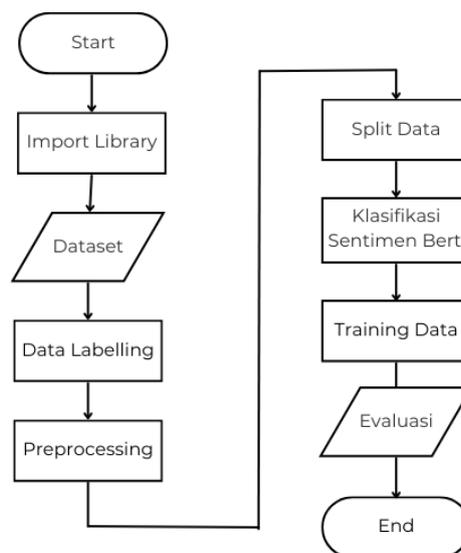
Salah satu penelitian yang telah dilakukan dibidang sentimen analisis pada e- marketplace yaitu oleh [11] yang melakukan sentimen analisis pada ulasan produk mainan di shopee. Menggunakan data sebanyak 1000, yang dibagi menjadi 700 data latih dan 300 data uji dengan kategori positif, negatif dan netral. Kemudian memperoleh akurasi sebesar 79,3333%, dimana akurasi tersebut diperoleh dengan cara pembobotan data menggunakan TF-IDF dan mengklasifikasi data uji menggunakan metode *K-Nearest Neighbor*(K-NN).

Penelitian selanjutnya yaitu tentang penerapan analisis sentimen pada pengguna twitter yang dilakukan oleh [7] dengan jumlah data set sebanyak 2000 data berbahasa Indonesia yang terbagi menjadi dua kategori positif dan negatif. Tahapan selanjutnya yaitu melakukan klasifikasi data dengan menggunakan metode *K-Nearest Neighbor*(K-NN) dan memperoleh nilai akurasi terbesar yaitu 67,2%.

Berdasarkan uraian dari tinjauan pustaka diatas, maka pada penelitian ini akan dibuat penelitian yang menggunakan teknik sentimen analisis dengan kategori positif, negatif dan netral dengan menerapkan model IndoBERT.

2. METODE PENELITIAN

Dalam penelitian ini sistem yang digunakan yaitu sistem sentimen analisis terhadap penggemar NCT mengenai *endorsement* dengan Starbucks yang masuk daftar hitam menggunakan model IndoBERT.



Gambar 1. Flowchart Alur Penelitian

Pada gambar 1, Alur penelitian ini dimulai dari mengimport *library* dan memuat dataset, penulis akan memuat dataset yang didapat dari proses scraping data pada komentar postingan instagram. Setelah itu, dataset diberi pelabelan positif, negatif dan netral yang dilambangkan dengan [0, 1, 2]. Setelah itu, dataset masuk ke tahap pre-processing. Pada tahap ini menyiapkan data yang sebelumnya tidak terstruktur menjadi data yang lebih terstruktur melalui proses *data cleaning* dan *tokenizing*. Lalu, data yang dihasilkan dari tahap ini dibagi menjadi data latih dan data uji. Dataset kemudian dilatih dengan menggunakan model BERT setelah melewati tahapan tersebut. Selanjutnya, data diklasifikasikan menjadi tiga sentimen untuk menentukan ke dalam kategori positif, negatif, atau netral. Setelah melewati tahap pelatihan dan klasifikasi, data kemudian dievaluasi untuk memastikan keakuratannya.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini didapatkan melalui *web scraping* pada kolom komentar postingan kerjasama antara NCT dan starbucks di instagram. Data tersebut berupa komentar Penggemar NCT mengenai *endorsement* nya dengan starbucks yang masuk daftar hitam. Tujuan dari *web scraping* ialah untuk mendapatkan data kemudian melakukan ekstraksi informasi yang dimiliki data tersebut.

2.2 Pelabelan

Pelabelan kelas sentimen data komentar penggemar dikelompokkan menjadi tiga kelas yaitu positif, negatif dan netral. Berdasarkan pada kelas negatif bermakna ejekan, keluhan dan penolakan terhadap keputusan yang diambil oleh pihak NCT dan *agency*, sedangkan kelas netral teks yang bermakna tidak menyatakan perasaan atau penilaian emosional positif dan negatif, lalu yang terakhir kelas positif bermakna pujian, dukungan dan menyatakan setuju.

2.3 Preprocessing

Tahapan selanjutnya yaitu perlu melakukan pembersihan data dengan tujuan supaya data dapat digunakan pada tahap selanjutnya. Adapun tahapan yang dilakukan untuk penelitian yaitu sebagai berikut :

a. Data Cleaning

Pada tahap ini, data cleaning digunakan untuk menghilangkan tanda baca, huruf besar, angka, *symbol*, URL, *Username*, *hashtag*, spasi berlebih dan pengulangan karakter pada dataset.

b. Tokenizing

Pada tahap ini merupakan proses untuk memecah kalimat menjadi per kata sehingga lebih mudah untuk diolah secara terpisah.

2.4 Split Data

Setelah tahap tokenisasi, dataset dibagi menjadi 30% data uji dan 70% data latih. *Training*/latihan membantu model mengenali pola dalam data. Sedangkan *testing*/pengujian memastikan bahwa model yang telah dilatih mampu memprediksi label yang belum dipelajari dengan baik.

2.5 Klasifikasi Metode IndoBERT

Sistem pada laporan tugas akhir ini menerapkan metode sentimen analisis *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT). Sistem ini akan mengidentifikasi sentimen positif, negatif atau netral dari komentar penggemar NCT pada instagram dengan menggunakan model IndoBERT-base-pl. Model tersebut merupakan model dari IndoBERT dengan arsitekturnya yaitu *Base* yang memiliki 12 *layer encoder*, 768 hidden nodes dan 124 juta parameter. *Library* yang digunakan pada tugas akhir ini yaitu *Transformer* yang disediakan oleh HuggingFace. Sistem pada tugas akhir ini nantinya akan berbasis *Streamlit*.

2.6 Evaluasi

Setelah proses train model selesai, dataset masuk ke tahap evaluasi. Proses evaluasi yang sangat penting dilakukan untuk menentukan kualitas program berdasarkan *accuracy*, *precision*, *recall* dan *f1-score*. yang dihasilkan dengan menggunakan *confusion metrics*. Kemudian simpan model untuk digunakan nantinya.

3. HASIL DAN ANALISA

3.1. Deskripsi Dataset

Pada penelitian ini, data diambil dari komentar pada aplikasi Instagram menggunakan *Web scraping*, teknik ini digunakan untuk mengambil informasi dari berbagai situs web. File *download* dalam bentuk *excel*, selanjutnya diberi label secara manual dengan membaca kalimat atau komentar satu persatu, jumlah data yang diambil 1084 dikategorikan positif 70, negatif 808 dan netral 206.

Tabel 1. Informasi *Dataset*

No	Comment	Label
0	udh pernah nyoba ga enak amis rasa darah	Negatif
1	Percayalah enakkan nutrisari jeruk peras 🙄🙄	Negatif
2	BOIKOT MEREKA	Negatif
3	Kenapa harus nct, kenapa harus nct sih...	Netral
4	ini udah selesai kan plis collabnya??	Netral
1079	ga sabarrrr 🙌🥰	Positif

No	Comment	Label
1080	ini bakalan ada diindo ga kak	Netral
1081	Kecewa bgt:)))	Netral
1082	Wkwkwkwkwkwkwkwkwk 🤔😂	Netral
1083	Delete our comments? 🗑️🗑️🗑️🗑️	Netral

Tabel 1 memperlihatkan informasi *dataset* untuk memberikan wawasan mengenai struktur dan fitur *dataset* yang digunakan dalam pelatihan model pada penelitian ini. Dataset yang diambil terdiri dari kolom Comment dan label dengan data yang berjumlah 1084.

3.2. Labeling

Pada proses *labeling*, dilakukan secara manual dengan membaca kalimat atau komentar satu persatu. Kategori pelabelan meliputi kalimat positif, negatif dan netral. Setelah mengganti nilai, mengkonversi tipe data pada kolom "label" menjadi integer. Sehingga menjadi positif (0), negatif (1) dan netral (2).

	Comment	label
0	udh pernah nyoba ga enak amis rasa darah	1
1	Percayalah enakkan nutrisari jeruk peras 🙄🙄	1
2	BOIKOT MEREKA	1
3	Kenapa harus nct,kenapa harus nct sih..	2
4	ini udah selesai kan plis collabnya??	2
...
1079	ga sabarrrr 🙄😂	0
1080	ini bakalan ada diindo ga kak	2
1081	Kecewa bgt:)))	2
1082	wkwkwkwkwkwkwkwkwk 🤔😂	2
1083	Delete our comments? 🗑️🗑️🗑️🗑️	2

[1084 rows x 2 columns]

Gambar 2. Hasil *Labeling dataset*

Gambar merupakan hasil dari perubahan pada kolom "label" dari tipe data string menjadi tipe data *integer*.

3.3. Data Cleaning

Proses *data cleaning* yang dilakukan adalah membersihkan data yang akan dijadikan input model yaitu data pada kolom komentar dengan menghapus karakter selain huruf, angka serta spasi. Selain itu, diterapkan juga *lowercase* untuk mengkonversi semua huruf dalam setiap baris di kolom komentar menjadi huruf kecil. Pembersihan data ini berfungsi untuk memberikan data yang konsisten untuk pelatihan model, sehingga model dapat memahami pola teks dengan baik serta dapat meningkatkan akurasi dan kinerjanya. Pada proses ini, memberikan hasil pada *dataset* menjadi seperti pada Gambar 3.

	Comment	label
0	udh pernah nyoba ga enak amis rasa darah	1
1	percayalah enakkan nutrisari jeruk peras	1
2	boikot mereka	1
3	kenapa harus nctkenapa harus nct sih..	2
4	ini udah selesai kan plis collabnya	2
...
1079	ga sabarrrr	0
1080	ini bakalan ada diindo ga kak	2
1081	kecewa bgt	2
1082	wkwkwkwkwkwkwkwkwk	2
1083	delete our comments	2

[1084 rows x 2 columns]

Gambar 3. Hasil *data cleaning*

Untuk mengetahui perbandingan data sebelum dan sesudah dilakukan *data cleaning*, dapat dilihat dari gambar 3, untuk lebih jelasnya lagi ditampilkan dua sampel hasilnya pada tabel berikut.

Tabel 2. Perbandingan sebelum dan sesudah *data cleaning*

Sebelum	Sesudah
BOIKOT MEREKA	boikot mereka
ini udah selesai kan plis collabnya??	ini udah selesai kan plis collabnya

Sebelum	Sesudah
Percayalah enakkan nutrisari jeruk peras 🙄🙄 ga sabarrrr 🙌🥰 Kecewa bgt:)))	Percayalah enakkan nutrisari jeruk peras ga sabarrrr kecewa bgt

Berdasarkan Tabel 2 dapat diketahui bahwa proses *data cleaning* yang digunakan berhasil membersihkan teks komentar dengan mengkonversi semua huruf menjadi huruf kecil seperti “BOIKOT MEREKA”. Kemudian, tanda baca pada kolom komentar dihapus seperti tanda baca “??”.

3.4. Eksperimen Model

Dikarenakan data yang didapatkan mengalami inbalancing, maka pada penelitian ini dilakukan beberapa eksperimen. *Hyperparameter* yang digunakan yaitu ditunjukkan pada Tabel 3.

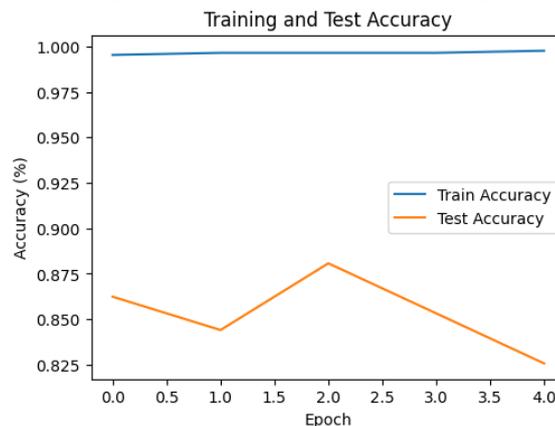
Tabel 3. *Hyperparameter* eksperimen model

No	<i>Hyperparameter</i>	Ukuran
1.	Batch Size	16
2.	Learning Rate	5e-5
3.	Epoch	20
4.	Max_length	256

Dari *hyperparameter* yang digunakan menghasilkan perbandingan antara data yang inbalance dan data yang sudah mengalami proses balancing.

3.4.1. Data Inbalance

Hasil eksperimen pada *dataset* yang mengalami *inbalance* ini dilakukan dengan menggunakan *hyperparameter batch size 32, learning rate 5e-5, epoch 5, serta max length 256* dapat dilihat pada gambar 4.



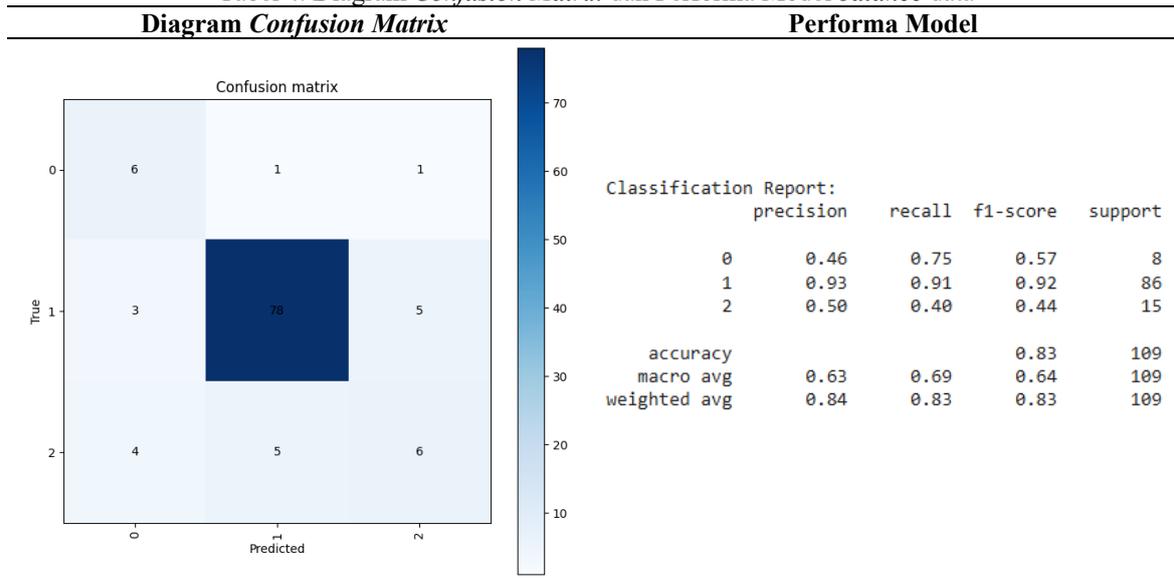
Gambar 4. Grafik *Accuracy Inbalance Dataset*

Pada gambar 4 menunjukkan bahwa selama proses pelatihan, akurasi *training* tetap sangat tinggi, mendekati 100%, tanpa perubahan yang signifikan. Hal ini mengindikasikan bahwa model sangat baik dalam menyesuaikan diri dengan data *training*. Namun, akurasi testing justru mengalami *overfitting* dan menurun setelah awal pelatihan. Pada grafik *epoch 0.0*, akurasi testing berada di sekitar 86%, kemudian sempat meningkat pada *epoch 2.0*, tetapi menurun kembali hingga sekitar 82% pada *epoch 4.0*.

Hal ini dapat terjadi karena model terlalu kompleks untuk *dataset* yang digunakan, karena jumlah data *training* yang tidak mencukupi untuk menangkap pola yang lebih umum.

Setelah proses pelatihan dilakukan, tahap selanjutnya yaitu proses menguji performa model. Hasil dari implementasi model terhadap data *training* dan data *testing* menunjukkan hasil *accuracy, precision, recall* dan *f1-score* dari *dataset inbalance* yang diperoleh ditunjukkan pada Tabel 4.

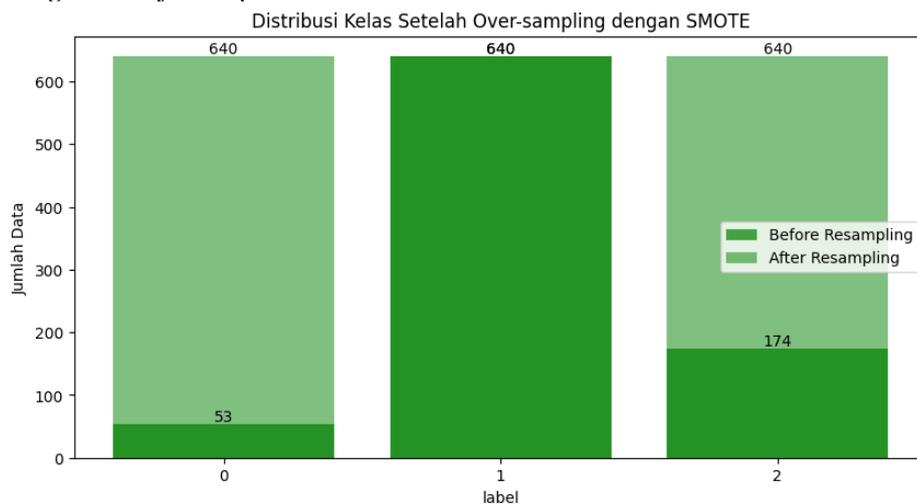
Tabel 4. Diagram *Confusion Matrix* dan Performa Model *balance* data



Pada Tabel 4 menunjukkan performa model yang dihasilkan berdasarkan data yang tidak seimbang, sehingga memperoleh *accuracy* sebesar 83%, *precision* 84%, *recall* 83%, *f1-score* 83%. Terdapat diagram *confusion matrix* yang menunjukkan performa model klasifikasi dengan tiga kelas yang diberi label 0, 1 dan 2. Nilai diagonal utama (6, 78, 6) menunjukkan jumlah prediksi yang benar untuk masing-masing kelas, dimana model berhasil mengklasifikasikan *true positif* 6 data, *true negative* 78 data, dan *true netral* 6 data.

3.4.2. Data Balance

Dikarenakan terdapat ketidakseimbangan distribusi data untuk setiap kelas pada dataset, dapat berpengaruh pada kinerja model. Model yang dilatih dengan data yang tidak seimbang akan cenderung mengklasifikasikan kelas mayoritas secara berlebihan dan mengabaikan kelas minoritas, sehingga akan mengakibatkan banyak kesalahan klasifikasi. Selain itu, distribusi data yang tidak seimbang juga dapat mengakibatkan *overfitting*. *Overfitting* merupakan kondisi dimana model memiliki kinerja yang baik terhadap data *training*, namun tidak pada data *testing*. Oleh karena itu, pada penelitian ini dilakukan penyeimbangan data *training* menggunakan SMOTE sebelum data digunakan untuk pelatihan model. SMOTE melakukan *oversampling* untuk kelas minoritas agar jumlah datanya menjadi seimbang dengan kelas mayoritas. Hasil dari proses *balancing* ini ditunjukkan pada Gambar 5.



Gambar 5. Jumlah data hasil *balancing*

Gambar 5 memperlihatkan distribusi data *training* setelah dilakukan *resampling* menggunakan SMOTE. Warna hijau tua menunjukkan jumlah data asli sebelum di seimbangkan, dan warna hijau muda menunjukkan jumlah data setelah dilakukan *resampling* dengan SMOTE. Jumlah data kelas minoritas diseimbangkan dengan kelas mayoritas sehingga jumlah datanya menjadi 640 untuk setiap kelas. Hasil eksperimen berdasarkan dataset *resampling* dapat dilihat pada gambar 6.

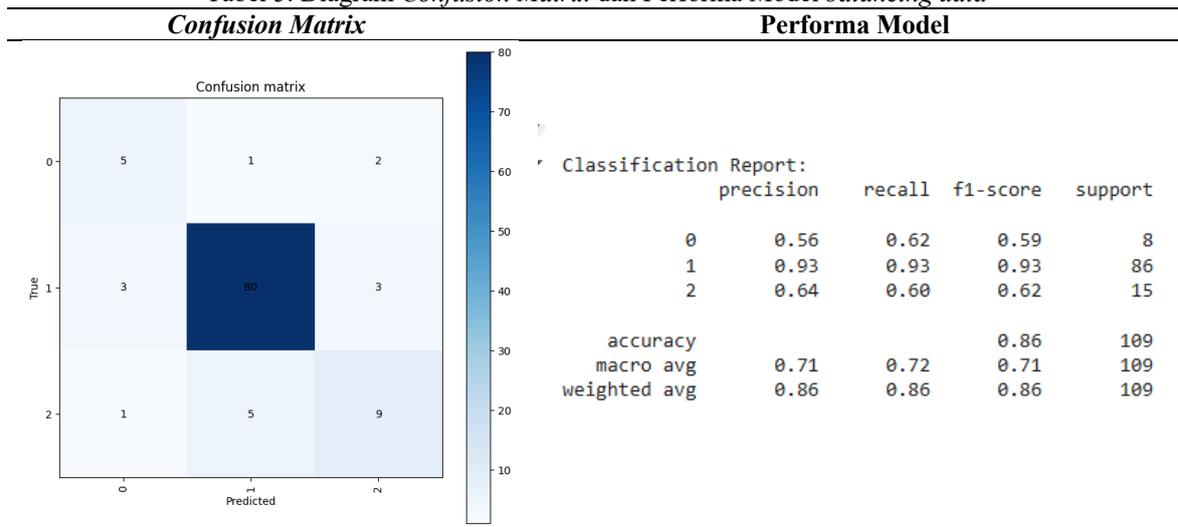


Gambar 6. Grafik *Accuracy Balancing Dataset*

Gambar 6 menunjukkan bahwa akurasi training tetap sangat tinggi dan hampir mendekati 100% di sepanjang epoch. Namun, akurasi test mengalami *overfitting*, dengan train awal meningkat hingga epoch 2.0, kemudian menurun pada epoch 3.0, dan kembali meningkat pada epoch 4.0. Selain itu, perbedaan besar antara akurasi training dan akurasi test menandakan bahwa model lebih banyak menghafal data training daripada memahami pola yang dapat diterapkan ke data baru.

Setelah proses pelatihan dilakukan, tahap selanjutnya yaitu proses menguji performa model. Hasil dari implementasi model terhadap data training dan data testing menunjukkan hasil *accuracy*, *precision*, *recall* dan *f1-score* dari dataset inbalance yang diperoleh ditunjukkan pada Tabel 5.

Tabel 5. Diagram *Confusion Matrix* dan Performa Model *balancing data*



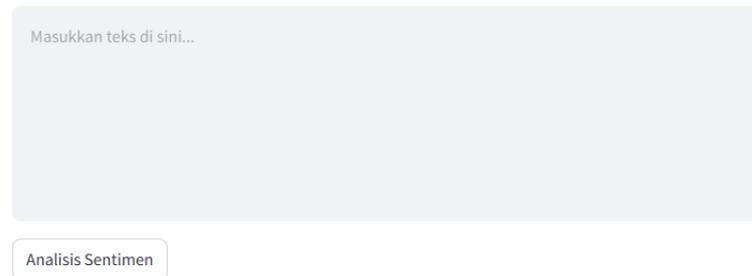
Tabel 5 menunjukkan performa model yang dihasilkan berdasarkan data yang sudah melalui proses *resampling*, sehingga memperoleh *accuracy* sebesar 86%, *precision* 86%, *recall* 86%, dan *f1-score* 86%. Terdapat diagram confusion matrix yang menunjukkan performa model klasifikasi dengan tiga kelas yang diberi label 0, 1 dan 2. Nilai diagonal utama (5, 80, 9) menunjukkan jumlah prediksi yang benar untuk masing-masing kelas, dimana model berhasil mengklasifikasikan *true positif* menghasilkan 5 data, *true negatif* 80 data, dan *true netral* 9 data.

3.5. Hasil Uji Coba Sistem

Hasil dari pemodelan yang sebelumnya sudah dilakukan dengan menggunakan model IndoBERT, selanjutnya akan dibuat sistem yang menggunakan *framework Streamlit*. Sehingga nantinya model ini memiliki sistem yang dapat melakukan inputan yang dilakukan oleh fans NCT. Hal tersebut ditunjukkan pada gambar 7.

Analisis Sentimen Teks dengan IndoBERT

Masukkan teks yang ingin dianalisis sentimennya



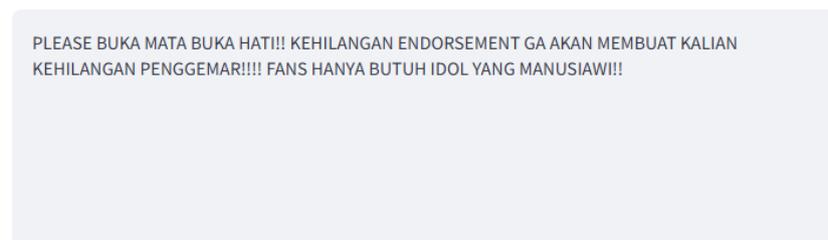
Masukkan teks di sini...

Analisis Sentimen

Gambar 7. Tampilan awal *Streamlit*

Gambar 7 merupakan tampilan awal dari *Streamlit*, sebelumnya agar tampilan awal bisa muncul harus menjalankan perintah `streamlit run namafile.py` di command prompt terlebih dahulu. Setelah tampilan awal muncul, tahap selanjutnya *User* dapat memasukkan pendapat mereka terhadap *Endorsement NCT* dengan Starbucks. Seperti yang terlihat pada gambar 8.

Masukkan teks yang ingin dianalisis sentimennya

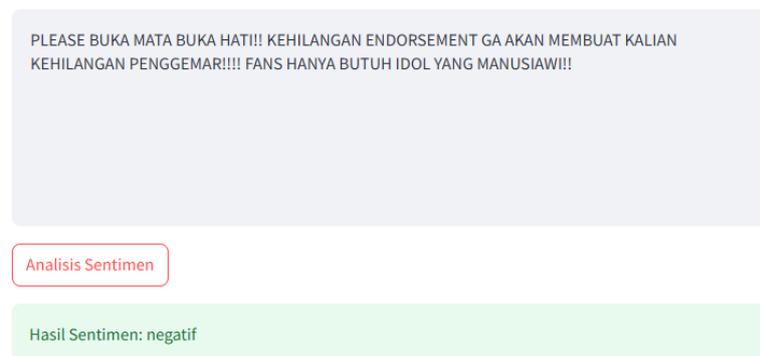


PLEASE BUKA MATA BUKA HATI!!! KEHILANGAN ENDORSEMENT GA AKAN MEMBUAT KALIAN KEHILANGAN PENGGEMAR!!!! FANS HANYA BUTUH IDOL YANG MANUSIAWI!!!

Gambar 8. *Inputan Text* Komentar

Gambar 8 menunjukkan inputan yang telah diisi oleh *User*. Tahap selanjutnya *User* dapat mengklik tombol hasil Analisis Sentimen maka hasil analisisnya nanti akan keluar.

Masukkan teks yang ingin dianalisis sentimennya



PLEASE BUKA MATA BUKA HATI!!! KEHILANGAN ENDORSEMENT GA AKAN MEMBUAT KALIAN KEHILANGAN PENGGEMAR!!!! FANS HANYA BUTUH IDOL YANG MANUSIAWI!!!

Analisis Sentimen

Hasil Sentimen: negatif

Gambar 9. Tampilan Tombol Klik dan Hasil Analisis Sentimen Kategori Negatif

Pada gambar 9 setelah *User* memasukkan text kemudian *User* dapat meng-klik tombol Analisis Sentimen setelah itu hasil analisisnya akan keluar dan menghasilkan sentimen negatif seperti yang terlihat pada gambar 9.

Analisis Sentimen Teks dengan IndoBERT

Masukkan teks yang ingin dianalisis sentimennya

Cantik sekali, aku ingin membelinya juga haha.

Analisis Sentimen

Hasil Sentimen: positif

Gambar 10. Hasil Sentimen Positif
Percobaan kedua dapat dilihat pada gambar 10 menunjukkan hasil analisis sentimen kategori positif.

Analisis Sentimen Teks dengan IndoBERT

Masukkan teks yang ingin dianalisis sentimennya

ini orang Starbucks liat komen nya ga siiii, kena mentall kata gue mah 🤔🤔

Analisis Sentimen

Hasil Sentimen: netral

Gambar 11. Hasil Sentimen Netral
Percobaan ketiga dapat dilihat pada gambar 11 menunjukkan hasil analisis sentimen kategori netral.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa, komentar para fans NCT terhadap *endorsement* dengan Starbucks memiliki data sentimen yang tidak seimbang yaitu lebih banyak pola sentimen negatif, hal itu dapat dilihat dari hasil eksperimen yang telah dilakukan. Bahwa dari eksperimen data yang tidak seimbang dengan data yang sudah diseimbangkan, akurasi yang dihasilkan lebih baik data yang sudah diseimbangkan daripada data yang tidak seimbang. Sehingga pada eksperimen awal data yang tidak seimbang mengalami overfitting dan performa model yang tidak terlalu baik. Sedangkan pada data yang sudah melalui proses balancing mendapatkan akurasi data dan performa model yang cukup baik, yaitu *accuracy* sebesar 86%, *precision* 86%, *recall* 86%, dan *f1-score* 86%.

DAFTAR PUSTAKA

- [1] P. A. Permatasari, L. Linawati, and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, p. 177, 2021, doi: 10.24843/mite.2021.v20i02.p01.
- [2] A. Ramadan and A. Fatchiya, "Efektivitas Instagram sebagai Media Promosi Produk ' Rendang Uninam ' The Effectiveness of Instagram as a Promotional Media of Products ' Rendang Uninam ,'" vol. 05, no. 01, pp. 64–82, 2021.
- [3] G. D. Pilar, S. B. Isabel, P. M. Diego, and G. Á. José Luis, "A novel flexible feature extraction algorithm for Spanish tweet sentiment analysis based on the context of words," *Expert Systems with Applications*, vol. 212, no. September 2022, 2023, doi: 10.1016/j.eswa.2022.118817.
- [4] N. Q. Rizkina and F. N. Hasan, "Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1136–1144, 2023, doi: 10.47065/josh.v4i4.3803.
- [5] D. Melati, "No Title," *GEMA KETIDAKSETUJUAN: GERAKAN BOIKOT SM ENTERTAINMENT DAN NCT TERKAIT KERJASAMA STARBUCKS*, p. 3, 2024.

- [6] P. Chinnasamy, V. Suresh, K. Ramprathap, B. J. A. Jebamani, K. Srinivas Rao, and M. Shiva Kranthi, "COVID-19 vaccine sentiment analysis using public opinions on Twitter," *Materials Today: Proceedings*, vol. 64, pp. 448–451, 2022, doi: 10.1016/j.matpr.2022.04.809.
 - [7] M. F. Mubaraq and W. Maharani, "Sentiment Analysis on Twitter Social Media towards Climate Change on Indonesia Using IndoBERT Model," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 2426, 2022, doi: 10.30865/mib.v6i4.4570.
 - [8] S. Analisis Pada Twitter CommuterLine, A. Pratama Putra, Y. Pratama, E. Kharisma Krisnadi, I. Purnamasari, and D. Dwi Saputra, "Text Mining untuk Sentimen Analisis dengan Metode Naïve Bayes, SMOTE, N-Gram dan AdaBoost Pada Twitter CommuterLine," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6, no. 2, pp. 961–973, 2022.
 - [9] F. Apri Wenando, "Analisis Opini Publik Terhadap Undang-Undang KUHP Tahun 2022 Menggunakan Algoritma Naïve Bayes Classifier," *Jurnal Fasilkom*, vol. 13, no. 02, pp. 334–339, 2023, doi: 10.37859/jf.v13i02.5670.
 - [10] A. R. Isnain, A. I. Sakti, D. Alita, and N. S. Marga, "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *Jurnal Data Mining dan Sistem Informasi*, vol. 2, no. 1, p. 31, 2021, doi: 10.33365/jdmsi.v2i1.1021.
 - [11] R. Mas, R. W. Panca, K. Atmaja1, and W. Yustanti2, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Jeisbi*, vol. 02, no. 03, p. 2021, 2021.
-