

Analisis Sentimen Masyarakat Terhadap Fenomena *Sandwich Generation* Pada Aplikasi X Menggunakan Metode *Naïve Bayes* dan Teknik Pelabelan *Lexicon Inset*

Annisa Nur Hidayati¹, Yuni Yamasari²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

¹annisa.21047@mhs.unesa.ac.id

²yuniyamasari@unesa.ac.id

Abstrak - Fenomena *sandwich generation* merupakan salah satu fenomena sosial yang sudah banyak diperbincangkan masyarakat Indonesia. Salah satu *platform* digital yang banyak membahas isu ini adalah X atau Twitter. *Sandwich generation* merupakan posisi seseorang yang memiliki beban ganda dalam menanggung kebutuhan orang tua dan anak secara bersamaan. Kondisi ini menimbulkan pro dan kontra di masyarakat. Sebagian pihak menganggap kondisi ini sebagai bentuk kewajiban dan bakti anak kepada orang tua. Sementara itu, sebagian lainnya menganggap kondisi ini sebagai beban. Hal ini dikarenakan mereka tidak mampu mencapai tujuan dan prioritas hidupnya sendiri akibat terhalang oleh tanggung jawab ganda tersebut. Oleh karena itu, penelitian ini melakukan analisis sentimen dengan *Naïve Bayes* pada dataset yang memiliki 3 kategori sentimen yaitu positif, negatif, dan netral. Teknik pelabelan yang digunakan pada penelitian ini adalah teknik pelabelan manual dan *Lexicon Inset*. Eksplorasi dilakukan dengan membandingkan *hyperparameter*, teknik *balancing* data (SMOTE, ROS, RUS, Tomek Links dan komposisi *split* data (50:50, 60:40, 70:30, 80:20, 90:10). Hal ini dilakukan untuk mendapatkan akurasi tertinggi dari model yang dibangun. Hasil Evaluasi menunjukkan bahwa model *Naïve Bayes* dapat mencapai performa terbaiknya yaitu *Accuracy* 0.757, *Precision* 0.7540, *Recall* 0.757, *F1-Score* 0.7550. Performa terbaik tersebut dicapai ketika menggunakan dataset pelabelan *Lexicon Inset*, varian *MultinomialNB*, komposisi data 90:10 dan parameter dengan nilai *Alpha* = 0.1, dan *Fit prior* = False.

Kata Kunci: *Sandwich generation*, analisis sentimen, *Naïve Bayes*, *Lexicon INSET*

I. PENDAHULUAN

Perkembangan teknologi memberikan dampak yang signifikan dalam kehidupan sosial masyarakat Indonesia, terutama kemudahan dalam mengakses informasi global secara efektif dan efisien melalui media sosial [1]. Salah satu *platform* komunikasi digital yang banyak digunakan oleh masyarakat Indonesia adalah X (sebelumnya dikenal sebagai Twitter) [2]. *Platform* ini dikenal sebagai media yang cepat dalam menyebarkan informasi sekaligus menjadi sarana untuk menyampaikan aspirasi masyarakat terkait isu sosial yang sedang terjadi [3].

X merupakan jenis *platform microblogging*. Aplikasi ini memberikan fasilitas kepada penggunanya untuk menulis dan mempublikasikan sebuah informasi [4]. Aplikasi ini banyak menyimpan data opini masyarakat dalam bentuk teks singkat yang disebut "*tweet*" [5]. Setiap *tweet* memiliki batas maksimal penggunaan karakter sebanyak 280 [6]. Sehingga mendorong pengguna untuk menyampaikan pendapat langsung pada inti permasalahan. Karena karakteristik ini, X menjadi *platform*

yang kaya akan data opini publik, yang sangat potensial untuk dijadikan sebagai sumber data..

Salah satu isu sosial yang cukup sering menjadi perbincangan di X adalah fenomena *sandwich generation*. *Sandwich generation* merupakan istilah kondisi sosial seseorang, yang posisinya terhimpit diantara dua generasi, yaitu generasi di atasnya atau orang tuanya dan generasi di bawahnya atau anaknya [7]. Fenomena ini umum dialami oleh kelompok usia produktif. Di Indonesia jumlah *sandwich generation* ini cukup tinggi. Berdasarkan hasil survei yang dilakukan oleh Jakpat tahun 2020, dari 1.343 responden, diketahui bahwa sebanyak 48% responden termasuk dalam kategori *sandwich generation* yang memiliki rentang usia 20 - 44 tahun[8].

Posisi *sandwich generation* ini menjadi pro dan kontra di kalangan masyarakat. Dimana sebagian masyarakat menganggap kondisi ini sebagai bentuk tanggung jawab moral dan bakti kepada orang tua. Sedangkan sebagian lagi merasa kondisi ini sebagai beban emosional dan finansial yang menghambat pencapaian tujuan hidup pribadi [9]. Pendapat positif maupun negatif masyarakat banyak ditemukan pada X yang dikenal sebagai wadah untuk bebas berkeluh kesah [10]. Hal tersebut menjadi sebuah potensi sumber data yang dapat dilakukan analisis dan diolah sehingga menghasilkan sebuah informasi yang dapat membantu lembaga sosial untuk mengetahui persepsi masyarakat terhadap isu tersebut menjadi lebih efektif serta dapat mengambil langkah strategis untuk penanganannya dengan teknik analisis sentimen .

Analisis sentimen merupakan salah satu teknik komputasi. Teknik ini digunakan untuk melakukan analisis perasaan dan emosi yang diungkapkan dalam sebuah teks[11]. Teknik mengklasifikasikan opini ke dalam tiga kategori utama, yaitu positif, negatif, dan netral. Hal ini dilakukan untuk memberikan gambaran umum mengenai persepsi publik terhadap suatu isu. Dalam implementasinya, analisis sentimen dapat dilakukan melalui berbagai pendekatan, antara lain pendekatan berbasis *lexicon*, *machine learning*, maupun *deep learning*, yang masing-masing memiliki kelebihan dalam konteks dan jenis data tertentu [12].

Naïve Bayes merupakan salah satu teknik *machine learning* yang termasuk kedalam model *supervised learning*. Teknik ini banyak digunakan untuk melakukan analisis sentimen. Dalam Analisis sentimen teknik ini berfungsi melakukan prediksi atau klasifikasi yang hasilnya akan digunakan untuk menentukan keputusan sentimen dari teks yang dianalisis [13] .

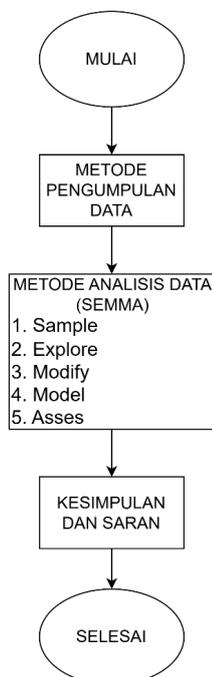
Penelitian yang dilakukan oleh Fadiah dan Arita yang melakukan klasifikasi sentimen masyarakat terhadap Presiden

Joko Widodo di akhir masa jabatannya. Penelitian ini menggunakan teknik *Naïve Bayes* dengan kombinasi teknik *balancing* data SMOTE dan teknik evaluasi model split data dengan kombinasi 80:20 dan 90:10. Model terbaik dicapai dengan komposisi split data 90:10 dengan nilai *accuracy* 88%. Sedangkan hasil dari kombinasi split data 80:20 memiliki *accuracy* 81% [14]. Penelitian yang dilakukan oleh Fadlan dan Achmad yang membahas analisis emosi masyarakat terhadap 3 bakal calon presiden pada youtube Mata Najwa. Penelitian ini menggunakan metode *Multinomial Naïve Bayes* dengan teknik pelabelan *Lexicon Inset* dengan fitur ekstraksi TF-IDF. Penelitian ini mampu mencapai kinerja terbaik dengan nilai *accuracy* 90.67% [15].

Berdasarkan beberapa penelitian yang sudah dilakukan sebelumnya, hasil penelitian menunjukkan bahwa penggunaan kombinasi metode *machine learning Naïve Bayes*, Pelabelan *Lexicon Inset*, dan komposisi pembagian dataset mampu meningkatkan hasil akurasi model. Penulis tertarik untuk melakukan analisis sentimen pada media sosial X terkait fenomena social *sandwich generation*. Tujuan dari penelitian ini adalah untuk mengetahui nilai performa yang dihasilkan dari penggunaan metode *Machine Learning* khususnya *Naïve Bayes*. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi faktor apa saja yang mempengaruhi performa tersebut dengan judul penelitian “Analisis Sentimen Masyarakat Terhadap Fenomena *Sandwich Generation* Pada Aplikasi X Menggunakan Metode *Naïve Bayes* dan Teknik Pelabelan *Lexicon Inset*”.

II. METODE PENELITIAN

A. Alur Penelitian



Gbr 1. Alur Penelitian

Alur penelitian merupakan tahapan dari proses penelitian yang memberikan panduan langkah-langkah agar penelitian dapat berjalan dari awal hingga selesai. Alur dari penelitian ini dijelaskan pada Gbr 1.

B. Metode Pengumpulan Data

1) Studi Literatur

Peneliti melakukan kajian literatur untuk mengumpulkan informasi dan data yang relevan dengan topik penelitian ini. Sumber literatur berasal dari penelitian terdahulu baik website jurnal nasional dan internasional, skripsi, tesis, serta artikel-artikel terkait.

2) Sumber Dataset

Penelitian ini menggunakan jenis data primer yang dikumpulkan langsung oleh peneliti menggunakan teknik *crawling* data pada aplikasi X. *Crawling* data dilakukan menggunakan *tweet harvest* API Key akun X untuk mendapatkan akses dari aplikasi.

C. Metode Analisis Data (SEMMA)

Proses analisis data pada penelitian ini menggunakan metode SEMMA. SEMMA merupakan salah satu metode data mining yang mudah dipahami dan mampu memberikan acuan alur penelitian yang jelas. Metode SEMMA terdiri dari 5 tahapan, yaitu *Sample*, *Explore*, *Modify*, *Model*, dan *Asses*. Berikut merupakan penjelasan dari masing-masing tahapan SEMMA.

1) Sample

Tahap *sample* merupakan tahap pengumpulan dataset yang berkaitan dengan topik penelitian. Dataset diperoleh melalui teknik *crawling* data menggunakan *library tweet harvest* dengan pemrograman python pada platform google colab. Proses *crawling* data memanfaatkan *auth token* akun untuk mendapatkan akses API aplikasi X. Dari proses *crawling* data diperoleh dataset dengan jumlah 4498 yang sudah difilter dan dilabeli manual oleh peneliti.

2) Explore

Tahap *explore* merupakan tahap untuk melakukan penggalan informasi terkait isi dari dataset. Penggalan informasi yang dilakukan adalah eksplorasi tipe data, jumlah data, duplikasi, distribusi label, dan kata yang sering muncul dalam dataset.

3) Modify

Pada tahap *modify* dilakukan *encoding* label, yaitu merubah label *string* menjadi *integer* (0= positif, 1 = negatif, 2= netral). Selanjutnya akan dilakukan proses *pre-processing*, berikut merupakan penjelasan tahapan dari *pre-processing* :

1. Cleaning

Dalam proses *cleaning* data diberikan beberapa perintah untuk membersihkan teks seperti menghapus url, menghapus mentions, menghapus hashtags, menghapus angka, menghapus karakter khusus, menghapus emoji,

menghapus huruf berulang, dan menghapus spasi di awal/akhir.

2. *Case folding*

Proses *case folding* dilakukan untuk melakukan standarisasi format huruf dari huruf yang bercampur antara *lowercase* dan *uppercase* menjadi huruf kecil semua.

3. *Tokenization*

Proses *tokenization* merupakan proses untuk memecah teks menjadi potongan yang lebih kecil (token kata tunggal). Kata yang dipecah merupakan setiap kata yang dipisahkan oleh spasi.

4. Normalisasi

Normalisasi teks merupakan proses perubahan kata yang tidak baku, baik singkatan atau kata alay menjadi kata yang sesuai dengan standar atau baku.

5. *Stop removal*

Proses *stop removal* merupakan proses penghapusan kata umum yang tidak memiliki makna penting dan sering muncul dalam teks dataset.

6. *Stemming*

Stemming merupakan proses transformasi dari bentuk kata yang memiliki imbuhan ke dalam bentuk kata dasar.

Setelah tahapan *pre-processing* data selesai selanjutnya dilakukan proses pelabelan dataset menggunakan metode *Lexicon InSet*.

4) Model

Dalam membangun model *Machine Learning* yang memiliki performa baik, digunakan 2 skenario pada dataset pelabelan manual dan pelabelan *Lexicon Inset* dengan kombinasi teknik diantaranya :

1. Parameter Terbaik

Eksperimen akan dilakukan terhadap setiap tipe *Naïve Bayes* dengan melakukan eksplorasi parameter untuk mendapatkan kombinasi yang memberikan performa terbaik terhadap dataset. Pada model *Naïve Bayes* dilakukan eksplorasi kombinasi parameter menggunakan metode *GridSearchCV* pada library python scikit-learn, dengan *hyperparameter* seperti yang dijelaskan pada Tabel 1.

Tabel 1. *Hyperparameter Tuning*

Metode	Hyperparameter	
<i>MultinomialNB</i>	<i>Alpha</i>	0.1, 0.5, 1, 2, 5
	<i>Fit prior</i>	True dan False
<i>GaussianNB</i>	<i>Var Smoothing</i>	1e-9, 1e-8, 1e-7, 1e-6
<i>BernoulliNB</i>	<i>Alpha</i>	0.1, 0.5, 1, 5, 10

	<i>Fit prior</i>	True dan False
--	------------------	----------------

2. Teknik Balancing

Kondisi dataset yang tidak seimbang akan memberikan pengaruh pada nilai performa model *machine learning*. Untuk menangani kasus ini maka diperlukan teknik *balancing* data untuk menyeimbangkan kategori kelas pada dataset yang digunakan. Pada penelitian ini diterapkan 4 teknik balancing data terhadap metode *Naïve Bayes* yaitu *Synthetic Minority Over-sampling Technique* (SMOTE), *Tomek Links*, *Random Over Sampling* (ROS), *Random Under Sampling* (RUS).

Pada pemodelan *Machine Learning* proses pemisahan data dibagi menjadi data latih dan data uji. Data latih digunakan untuk membangun dan menyesuaikan parameter model, sedangkan data uji digunakan untuk melakukan evaluasi terhadap performa model yang dibangun terhadap data baru yang belum pernah dilihat sebelumnya oleh model. Proses pembagian data dilakukan sebelum data di balancing untuk menghindari kebocoran data uji dan membuat hasil evaluasi performa model menjadi tidak valid. Pada penelitian ini diterapkan 5 skenario pembagian data latih dan data uji seperti pada Tabel 2.

Tabel 2. Skenario Split Data Latih dan Data Uji

Data Latih	Data Uji
50	50
60	40
70	30
80	20
90	10

5) Asses

Pengujian model ditampilkan dalam bentuk *confusion matrixs* yang menunjukkan jumlah prediksi benar dan prediksi salah. Dari *confusion matrix* dapat dihitung matrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-Score* untuk mengukur performa model dalam analisis sentimen. Penelitian ini menggunakan teknik evaluasi model *split* data, yaitu mengukur kinerja model menggunakan skenario pembagian data latih dan data uji. Selain itu juga disertai perbandingan penerapan teknik pelabelan dataset manual dan *Lexicon InSet* serta penerapan teknik *balancing*.

Dari *confusion matrixs* dapat dihitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-Score* Berikut merupakan penjelasannya :

1. *Accuracy*

Accuracy digunakan untuk mengukur seberapa banyak model melakukan klasifikasi dengan benar.

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

2. Precision

Precision digunakan untuk mengukur seberapa banyak model melakukan klasifikasi positif yang benar dari seluruh klasifikasi yang diprediksi positif.

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

3. Recall

Recall digunakan untuk mengukur seberapa banyak data positif yang berhasil dideteksi sebagai positif.

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

4. F1-Score

F1-Score digunakan untuk menyeimbangkan *precision* dan *recall*.

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

D. Kesimpulan dan Saran

Hasil penelitian penelitian yang mencakup temuan utama, pencapaian tujuan penelitian serta mampu menjawab semua permasalahan yang ditemukan dalam penelitian dirangkum pada bagian kesimpulan akhir. Dari kesimpulan penelitian kemudian dapat dilakukan rekomendasi atau saran-saran perbaikan untuk penelitian atau pengembangan sistem selanjutnya.

III. HASIL DAN PENELITIAN

A. Sample

Dataset primer yang digunakan untuk penelitian ini didapatkan dengan teknik *crawling* disimpan dengan nama "sandwich generation dataset.csv". Data hasil *crawling* disimpan setiap kategori bulan Januari - Desember 2024 dalam format csv. Tabel merupakan contoh dataset kotor dari Januari-Desember 2024.

Tabel 3. Contoh Dataset Kotor Hasil *Crawling* Dari Januari-Desember 2024

No	Tweet
1	pengen ngabarin orang tua tapi nanti disangka lagi banyak duit . Gini amat jadi sandwich generation.
2	@asumsico Emang nasib jadi orang miskin lebih effort buat bisa naik kelas. Masuk negeri biar bisa kuliah murah. Eh UKTnya naik mulu ampe gak bisa bayar. Malah disuruh ngutang. Nanti lulus2 udah susah cari kerja gaji pas UMR jadi sandwich generation + bayari utang dan bunganya. Nasib
3	yang kena imbasnya kakak-kakaknya. Jadi sandwich generation karena usia orangtuanya udah ga produktif tapi masih ada anak-anaknya yang harus dibiayain :)
4	Cari kerja susah. Sandwich generation. Bulanan ngasih ortu. Bayar listrik wifi. Gaji UMR Jogja. Kuliah pake student loan 😊

5	yaa tapi semoga Tuhan permudah rezeki kita sebagai sandwich generation pasti bisa
---	---

Dataset yang sudah dikumpulkan dan digabung selanjutnya dilakukan tahap filtering untuk pelabelan secara manual oleh peneliti dan penghapusan duplikasi data sehingga dataset yang dihasilkan bebas dari duplikasi data dan berlabel sesuai dengan kategori sentimen.

Tabel 4. Dataset *Sandwich Generation* Berlabel

No	Tweet	Kategori
1	yaa tapi semoga Tuhan permudah rezeki kita sebagai sandwich generation pasti bisa	Positif
2	Tapi aku punya target menaikkan income karena aku adalah sandwich generation	Positif
3	Pusing juga jadi Sandwich generation mengejar cita - cita sambil menopang ekonomi keluarga semoga ada rezekinya aja ya. Semangat!!!	Positif
4	Ini adalah kesempatan berbakti dan saya rasa bodoh sekali seorang anak menyesali hidupnya sebagai sandwich generation. Apalagi sampai menyalahkan orang tuanya.	Positif
5	Selamat hari Senin. Selamat berjuang untuk sandwich generation. Semoga nasib baik selalu bersama kalian sehingga kalian bisa memutus rantai sandwich itu.	Positif

B. Explore

Dataset berlabel memiliki 3 kolom yaitu Tanggal, Teks, dan Kategori yang sudah tersimpan dalam format csv dengan nama file dataset.csv. Berikut merupakan tahapan-tahapan eksplorasi yang dilakukan pada dataset, diantaranya :

1. Eksplorasi Tipe Data

No	Nama Kolom	Type
1	Tanggal	object
2	Teks	object
3	Kategori	object

2. Jumlah Data

Dataset yang akan digunakan pada penelitian ini terdiri dari 3 kolom dan 4498 baris data dengan satu kategori unik (positif, negatif, dan netral) pada setiap baris datanya.

3. Data Kosong

Dataset yang akan digunakan pada penelitian ini tidak memiliki data kosong di setiap kolom Tanggal, Teks, dan Kategori.

4. Duplikasi Data

Data ganda pada dataset akan menyebabkan proses analisis data menjadi berat dan menghasilkan akurasi yang kurang akurat karena model bias

	Teks	tokenized_tweet
0	yaa tapi semoga Tuhan permudah rezeki kita seb...	[yaa, tapi, semoga, tuhan, permudah, rezeki, k...
1	Tapi aku punya target menaikkan income karena ...	[tapi, aku, punya, target, menaikkan, income, ...
2	Pusing juga jadi Sandwich generation mengejar ...	[pusing, juga, jadi, sandwich, generation, men...
3	Ini adalah kesempatan berbakti dan saya rasa b...	[ini, adalah, kesempatan, berbakti, dan, saya...
4	Selamat hari Senin. Selamat berjuang untuk san...	[selamat, hari, senin, selamat, berjuang, untu...
5	kerja karena gue mau ga mau bakalan jadi sandw...	[kerja, karena, gue, mau, ga, mau, bakalan, ja...
6	@worksfess 24 nder sandwich generation kunciny...	[nder, sandwich, generation, kuncinya, frugal...
7	sbg anak pertama yg tdk jadi sandwich generati...	[sbg, anak, pertama, yg, tdk, jadi, sandwich, ...
8	Sebagai seseorang yg pernah jadi sandwich gene...	[sebagai, seseorang, yg, pernah, jadi, sandwic...
9	Sudah disini lagi semangat sandwich generation...	[sudah, disini, lagi, semangat, sandwich, gene...

Gbr 7. Hasil Tokenization Data

4. Normalisasi

Tahapan ini dilakukan perubahan kata yang tidak baku menjadi kata baku menggunakan kamus normalisasi yang berisi kata kata slang yang tersimpan pada file slang.txt. Hasil Normalisasi ditampilkan pada Gbr 8.

	Teks	normalized_tweet
0	yaa tapi semoga Tuhan permudah rezeki kita seb...	[ya, tapi, semoga, tuhan, permudah, rezeki, ki...
1	Tapi aku punya target menaikkan income karena ...	[tapi, aku, punya, target, menaikkan, income, ...
2	Pusing juga jadi Sandwich generation mengejar ...	[pusing, juga, jadi, sandwich, generation, men...
3	Ini adalah kesempatan berbakti dan saya rasa b...	[ini, adalah, kesempatan, berbakti, dan, saya...
4	Selamat hari Senin. Selamat berjuang untuk san...	[selamat, hari, senin, selamat, berjuang, untu...
5	kerja karena gue mau ga mau bakalan jadi sandw...	[kerja, karena, saya, mau, tidak, mau, bakalan...
6	@worksfess 24 nder sandwich generation kunciny...	[nder, sandwich, generation, kuncinya, frugal...
7	sbg anak pertama yg tdk jadi sandwich generati...	[sebagai, anak, pertama, yang, tidak, jadi, sa...
8	Sebagai seseorang yg pernah jadi sandwich gene...	[sebagai, seseorang, yang, pernah, jadi, sandw...
9	Sudah disini lagi semangat sandwich generation...	[sudah, disini, lagi, semangat, sandwich, gene...

Gbr 8. Hasil Normalisasi Data

5. Stop Removal

Kata- kata umum yang tidak memiliki informasi penting (*stopword*) akan dihapus untuk meningkatkan performa model. Penghapusan kata-kata ini menggunakan *library* sastrawi Indonesia serta penghapusan daftar kata tambahan. Hasil dari proses *stopword removal* ditampilkan pada Gbr 9.

	Teks	stopword_removed_tweet
0	yaa tapi semoga Tuhan permudah rezeki kita seb...	[semoga, tuhan, permudah, rezeki, sandwich, ge...
1	Tapi aku punya target menaikkan income karena ...	[target, menaikkan, income, sandwich, generation]
2	Pusing juga jadi Sandwich generation mengejar ...	[pusing, sandwich, generation, mengejar, cita...
3	Ini adalah kesempatan berbakti dan saya rasa b...	[kesempatan, berbakti, bodoh, anak, menyesali...
4	Selamat hari Senin. Selamat berjuang untuk san...	[selamat, senin, selamat, berjuang, sandwich, ...
5	kerja karena gue mau ga mau bakalan jadi sandw...	[kerja, sandwich, generation, wisuda]
6	@worksfess 24 nder sandwich generation kunciny...	[nder, sandwich, generation, kuncinya, frugal...
7	sbg anak pertama yg tdk jadi sandwich generati...	[anak, sandwich, generation, disyukuri, privil...
8	Sebagai seseorang yg pernah jadi sandwich gene...	[sandwich, generation, pelajari, om, termgiang...
9	Sudah disini lagi semangat sandwich generation...	[semangat, sandwich, generation]

Gbr 9. Hasil Stopword Removal Data

6. Stemming

Kata dalam dataset diubah menjadi bentuk kata dasar menggunakan kamus sastrawi Indonesia. Perubahan kata menjadi kata dasar misalnya “kesempatan” menjadi “sempat”. Hasil dari tahapan *stemming* ditampilkan pada Gbr 10.

	Teks	stemmed_tweet
0	yaa tapi semoga Tuhan permudah rezeki kita seb...	[moga, tuhan, mudah, rezeki, sandwich, generat...
1	Tapi aku punya target menaikkan income karena ...	[target, naik, income, sandwich, generation]
2	Pusing juga jadi Sandwich generation mengejar ...	[pusing, sandwich, generation, kejar, cita, ci...
3	Ini adalah kesempatan berbakti dan saya rasa b...	[sempat, bakti, bodoh, anak, sali, hidup, sand...
4	Selamat hari Senin. Selamat berjuang untuk san...	[selamat, senin, selamat, juang, sandwich, gen...
5	kerja karena gue mau ga mau bakalan jadi sandw...	[kerja, sandwich, generation, wisuda]
6	@worksfess 24 nder sandwich generation kunciny...	[nder, sandwich, generation, kunci, frugal, li...
7	sbg anak pertama yg tdk jadi sandwich generati...	[anak, sandwich, generation, syukur, privilege...
8	Sebagai seseorang yg pernah jadi sandwich gene...	[sandwich, generation, ajar, om, ngiang, punya...
9	Sudah disini lagi semangat sandwich generation...	[semangat, sandwich, generation]

Gbr 10. Hasil Tahapan Stemming Data

Setelah tahapan *pre-processing* data selesai selanjutnya dilakukan proses pelabelan dataset menggunakan metode pendekatan kamus bahasa indonesia *Lexicon InSet*. Kamus *Lexicon Inset* terdiri atas 6609 kata negatif dan 3609 kata positif. Proses labeling menggunakan metode ini berdasarkan jumlah pembobotan kata berdasarkan bobot kata dalam kamus *Lexicon Inset* yang memiliki rentang nilai -5 sampai dengan 5. Label sentimen ditetapkan berdasarkan total bobot dalam sebuah kalimat, jika bobot kurang dari 0 maka teks berlabel negatif, jika bobot lebih dari 0 maka teks berlabel positif, dan jika bobot sama dengan 0 maka teks berlabel netral.

Hasil dari proses pelabelan dataset yang sudah di *preprocessing* menggunakan metode pendekatan kamus bahasa indonesia *Lexicon InSet*, teks sentimen memiliki label otomatis berdasarkan jumlah skor kata-kata di dalamnya sesuai dengan yang ada pada kamus *lexicon*. Hasil dari proses pelabelan *Lexicon Inset* ditampilkan pada Tabel 5.

Tabel 5. Contoh Dataset *Preprocessing* Berlabel Dengan *Lexicon InSet*.

No	stemmed _ tweet	tokenized	polarity _ score	Kategori
1	moga tuhan mudah rezeki sandwich generatio n	['moga', 'tuhan', 'mudah', 'rezeki', 'sandwich', 'generation']	10	0
2	target naik income sandwich generatio n	['target', 'naik', 'income', 'sandwich', 'generation']	1	0
3	pusing sandwich generatio n kejar cita cita topang	['pusing', 'sandwich', 'generation', 'kejar', 'cita', 'cita', 'topang',	1	0

No	stemmed - tweet	tokenized	polarity _score	Kategori
	ekonomi keluarga moga rezeki semangat	'ekonomi', 'keluarga', 'moga', 'rezeki', 'semangat']		
4	sempat bakti bodoh anak sali hidup sandwich generatio n salah orang tua	['sempat', 'bakti', 'bodoh', 'anak', 'sali', 'hidup', 'sandwich', 'generation', 'salah', 'orang', 'tua']	-9	1

D. Model

1. Parameter Terbaik

a) Naïve Bayes Pelabelan Manual

Hasil dari *hyperparameter tuning* dataset pelabelan manual ditampilkan pada Tabel 6.

Tabel 6. Parameter Terbaik Naive Bayes Data Pelabelan Manual

Variasi Naïve Bayes	Rasio	Parameter Terbaik	Metrik Evaluasi	
MultinomialNB	50 : 50	Alpha = 0.5 Fit Prior = False	Accuracy	0.6127
			Precision	0.6033
			Recall	0.6127
			F1-Score	0.5982
	60 : 40	Alpha = 1 Fit Prior = False	Accuracy	0.6267
			Precision	0.6236
			Recall	0.6267
	70 : 30	Alpha = 1 Fit Prior = False	Accuracy	0.6422
			Precision	0.6382
			Recall	0.6422
	80 : 20	Alpha = 0.5 Fit Prior = False	Accuracy	0.6433
			Precision	0.6333
Recall			0.6433	
90 : 10	Alpha = 0.1 Fit Prior = True	Accuracy	0.6600	
		Precision	0.6669	
		Recall	0.6600	
GaussianNB	50 : 50	Var Smoothing = 1e-06	Accuracy	0.4108
			Precision	0.4653
			Recall	0.4108
	60 : 40	Var Smoothing = 1e-06	Accuracy	0.4178
			Precision	0.4727
			Recall	0.4178
	70 : 30	Var Smoothing = 1e-06	Accuracy	0.4242
			Precision	0.4133
			Recall	0.4242
	80 : 20	Var Smoothing = 1e-06	Accuracy	0.4133
			Precision	0.4745
			Recall	0.4133
90 : 10	Var Smoothing = 1e-06	Accuracy	0.4207	
		Precision	0.4078	
		Recall	0.4078	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.4078	
		Precision	0.4724	
		Recall	0.4078	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.4140	
		Precision	0.4000	
		Recall	0.4140	

Variasi Naïve Bayes	Rasio	Parameter Terbaik	Metrik Evaluasi	
BernoulliNB	50 : 50	Alpha = 0.1 Fit Prior = True	Accuracy	0.5914
			Precision	0.5947
			Recall	0.5914
			F1-Score	0.5926
	60 : 40	Alpha = 0.5 Fit Prior = False	Accuracy	0.6083
			Precision	0.6138
			Recall	0.6083
	70 : 30	Alpha = 0.5 Fit Prior = True	Accuracy	0.6185
			Precision	0.6252
			Recall	0.6185
	80 : 20	Alpha = 0.5 Fit Prior = True	Accuracy	0.6124
			Precision	0.6478
Recall			0.6478	
90 : 10	Alpha = 0.5 Fit Prior = True	Accuracy	0.6420	
		Precision	0.6526	
		Recall	0.6420	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.6533	
		Precision	0.6533	
		Recall	0.6533	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.6465	
		Precision	0.6465	
		Recall	0.6465	

b) Naïve Bayes Pelabelan Lexicon Inset

Hasil dari *hyperparameter tuning* dataset dengan pelabelan *Lexicon Inset* ditampilkan pada Tabel 7.

Tabel 7. Parameter Naive Bayes Data Pelabelan Lexicon Inset

Variasi Naïve Bayes	Rasio	Parameter Terbaik	Metrik Evaluasi	
MultinomialNB	50 : 50	Alpha = 0.5 Fit Prior = False	Accuracy	0.7488
			Precision	0.7444
			Recall	0.7488
			F1-Score	0.7357
	60 : 40	Alpha = 0.5 Fit Prior = False	Accuracy	0.7528
			Precision	0.7472
			Recall	0.7528
	70 : 30	Alpha = 0.5 Fit Prior = False	Accuracy	0.7403
			Precision	0.7489
			Recall	0.7420
	80 : 20	Alpha = 1 Fit Prior = False	Accuracy	0.7489
			Precision	0.7385
Recall			0.7489	
90 : 10	Alpha = 0.1 Fit Prior = False	Accuracy	0.7556	
		Precision	0.7577	
		Recall	0.7556	
GaussianNB	50 : 50	Var Smoothing = 1e-09	Accuracy	0.7389
			Precision	0.7578
			Recall	0.7578
	60 : 40	Var Smoothing = 1e-09	Accuracy	0.7500
			Precision	0.7500
			Recall	0.7500
	70 : 30	Var Smoothing = 1e-06	Accuracy	0.5322
			Precision	0.5584
			Recall	0.5322
	80 : 20	Var Smoothing = 1e-06	Accuracy	0.5409
			Precision	0.5206
			Recall	0.5206
90 : 10	Var Smoothing = 1e-06	Accuracy	0.5315	
		Precision	0.5030	
		Recall	0.5030	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.5030	
		Precision	0.5592	
		Recall	0.5030	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.5168	
		Precision	0.4911	
		Recall	0.4911	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.5614	
		Precision	0.4911	
		Recall	0.4911	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.5081	
		Precision	0.4689	
		Recall	0.4689	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.5499	
		Precision	0.4839	
		Recall	0.4839	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.4839	
		Precision	0.4839	
		Recall	0.4839	
50 : 50	Var Smoothing = 1e-06	Accuracy	0.7443	
		Precision	0.7443	
		Recall	0.7443	

Variasi Naïve Bayes	Rasio	Parameter Terbaik	Metrik Evaluasi	
BernoulliNB		Alpha = 0.5 Fit Prior = False	Precision	0.7389
			Recall	0.7443
			F1-Score	0.7401
	60 : 40	Alpha = 0.5 Fit Prior = True	Accuracy	0.7461
			Precision	0.7445
			Recall	0.7461
	70 : 30	Alpha = 0.5 Fit Prior = True	Accuracy	0.7541
			Precision	0.7502
			Recall	0.7541
	80 : 20	Alpha = 0.5 Fit Prior = True	Accuracy	0.7567
			Precision	0.7563
			Recall	0.7567
	90 : 10	Alpha = 0.5 Fit Prior = True	Accuracy	0.7578
			Precision	0.7620
			Recall	0.7578
		F1-Score	0.7502	

2. Teknik Balancing Data

Dari tahapan sebelumnya diketahui pada kedua dataset, memiliki performa terbaik ketika menggunakan varian *MultinomialNB*. Kombinasi parameter terbaik pada data dengan pelabelan manual adalah $\alpha = 0.1$ dan $fit_prior = True$ dan pelabelan *lexicon inset* adalah $\alpha = 0.1$ dan $fit_prior = False$. Teknik balancing data diterapkan pada *MultinomialNB* diantaranya adalah SMOTE, *Tomek Links*, ROS, dan RUS. Berikut merupakan hasil penerapan teknik *balancing* pada masing-masing dataset :

a) Naïve Bayes Pelabelan Manual

Teknik Balancing	Accuracy	Precision	Recall	F1-Score
SMOTE	0.6400	0.6436	0.6400	0.6391
<i>Tomek Links</i>	0.6489	0.6551	0.6489	0.6308
ROS	0.6289	0.6278	0.6289	0.6271
RUS	0.5822	0.6065	0.5822	0.5856

b) Naïve Bayes Pelabelan Lexicon Inset

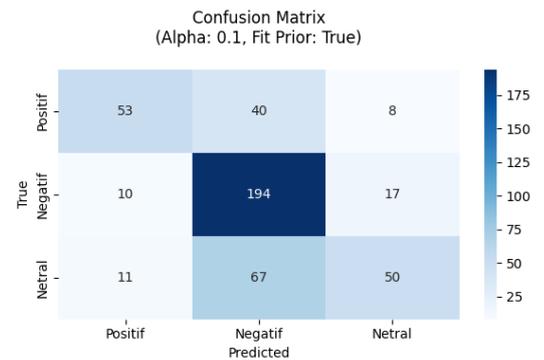
Teknik Balancing	Accuracy	Precision	Recall	F1-Score
SMOTE	0.7333	0.7297	0.7333	0.7311
<i>Tomek Links</i>	0.7489	0.7441	0.7489	0.7456
ROS	0.7289	0.7254	0.7289	0.7267
RUS	0.6756	0.6957	0.6756	0.6821

E. Asses

1. Parameter Terbaik

a) Naïve Bayes Pelabelan Manual

Setelah dilakukan *hyperparameter tuning* model *Naïve Bayes* pada dataset dengan pelabelan manual menggunakan teknik *Grid Search*, Didapatkan hasil terbaik kombinasi parameter adalah *MultinomialNB* dengan komposisi data 90:10, nilai $\alpha = 0.1$, dan $Fit\ prior = True$.



Gbr 11. Confusion Matrix *MultinomialNB*-Manual-Parameter Terbaik

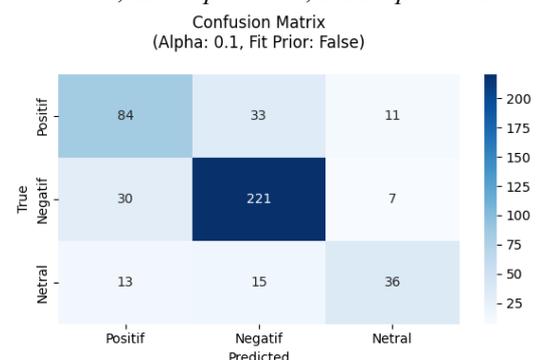
Dari *confusion matrix* pada Gbr 11 diketahui bahwa Diagonal utama merupakan jumlah data yang diprediksi benar sesuai dengan kelasnya. Pada kelas data positif sebanyak 53 data diprediksi benar sebagai data berlabel positif. Pada kelas data negatif sebanyak 194 data diprediksi benar sebagai data berlabel negatif. Pada kelas data netral sebanyak 50 data diprediksi benar sebagai data berlabel netral.

Model bekerja sangat baik pada kelas Negatif, karena mampu memprediksi benar 194 dari total 221 data berlabel negatif (10 + 194 + 17). Model bekerja kurang baik pada kelas Netral, karena hanya mampu memprediksi benar 50 dari total 128 data berlabel netral (11 + 67 + 50).

Dari *confusion matrix* dapat dilakukan perhitungan nilai *accuracy* didapatkan hasil $Accuracy = (53 + 194 + 50) / 450 = 297 / 450 = 0.6600$

b) Naïve Bayes Pelabelan Lexicon Inset

Hasil *hyperparameter tuning* pada dataset pelabelan *Lexicon Inset* pada model *Naïve Bayes* dapat diketahui kombinasi parameter terbaik menggunakan *MultinomialNB* dengan komposisi data 90:10, nilai $\alpha = 0.1$, dan $Fit\ prior = False$.



Gbr 12. Confusion Matrix *MultinomialNB*-Lexicon Inset-Parameter Terbaik

Dari *confusion matrix* pada Gbr 12 diketahui bahwa diagonal utama merupakan jumlah data

yang diprediksi benar sesuai dengan kelasnya. Pada kelas data positif sebanyak 84 data diprediksi benar sebagai data berlabel positif. Pada kelas data negatif sebanyak 221 data diprediksi benar sebagai data berlabel negatif. Pada kelas data netral sebanyak 36 data diprediksi benar sebagai data berlabel netral.

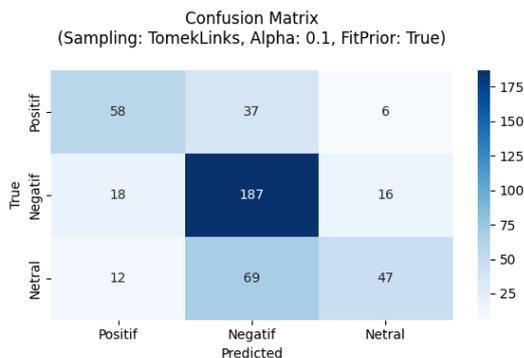
Model bekerja dengan sangat baik pada data kelas negatif, dimana model mampu melakukan prediksi benar terhadap 465 data dari 517 (31+465+21) data berlabel negatif Model bekerja sangat baik pada kelas Negatif, karena mampu memprediksi benar 221 dari total 258 data berlabel negatif (30 + 221 + 7). Model bekerja kurang baik pada kelas Netral, karena hanya mampu memprediksi benar 36 dari total 64 data berlabel netral (13 + 15 + 36).

Dari *confusion matrix* dapat dilakukan perhitungan nilai *accuracy* didapatkan hasil $Accuracy = (84 + 221 + 36) / 450 = 341 / 450 = 0.7578$

2. Teknik Balancing

a) Naïve Bayes Pelabelan Manual

Kombinasi teknik *balancing* terbaik dari parameter terbaik pada model *Naïve Bayes* dengan pelabelan manual menggunakan *MultinomialNB* dengan komposisi data 90:10, nilai *Alpha* = 0.1 dan *Fit prior* = True adalah teknik *balancing Tomek Links*.



Gbr 13. *Confusion Matrix MultinomialNB-Manual-Parameter Terbaik-Tomeklinks*

Dari *confusion matrix* pada Gbr 13 diketahui bahwa diagonal utama merupakan jumlah data yang diprediksi benar sesuai dengan kelasnya. Pada kelas data positif sebanyak 58 data diprediksi benar sebagai data berlabel positif. Pada kelas data negatif sebanyak 187 data diprediksi benar sebagai data berlabel negatif. Pada kelas data netral sebanyak 47 data diprediksi benar sebagai data berlabel netral.

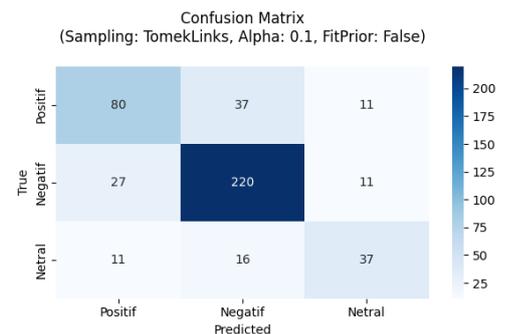
Model bekerja sangat baik pada data kelas negatif, di mana model mampu memprediksi benar 187 data dari total 221 data berlabel negatif (18 + 187 + 16 = 221). Model bekerja kurang baik

pada data kelas netral, di mana model hanya mampu memprediksi benar 47 data dari total 128 data berlabel netral (12 + 69 + 47 = 128).

Teknik *Balancing Tomek Links* merupakan kombinasi teknik *balancing* yang paling optimal. Namun Kombinasi parameter terbaik dengan teknik *balancing* cenderung menurunkan nilai *accuracy* yaitu $Accuracy = (58 + 187 + 47) / 450 = 292 / 450 = 0.6489$

b) Naïve Bayes Pelabelan Lexicon Inset

Kombinasi teknik *balancing* terbaik dari parameter terbaik pada model *Naïve Bayes* dengan pelabelan *Lexicon Inset* menggunakan *MultinomialNB* dengan nilai *Alpha* = 0.1 dan *Fit prior* = False adalah teknik *balancing Tomek Links*.



Gbr 14. *Confusion Matrix MultinomialNB-Lexicon Inset-Parameter Terbaik-Tomeklinks*

Dari *confusion matrix* pada Gbr 14 diketahui bahwa diagonal utama merupakan jumlah data yang diprediksi benar sesuai dengan kelasnya. Pada kelas data positif sebanyak 80 data diprediksi benar sebagai data berlabel positif. Pada kelas data negatif sebanyak 220 data diprediksi benar sebagai data berlabel negatif. Pada kelas data netral sebanyak 37 data diprediksi benar sebagai data berlabel netral.

Model bekerja sangat baik pada data kelas negatif, di mana model mampu melakukan prediksi benar terhadap 220 data dari total 258 data berlabel negatif (27 + 220 + 11 = 258). Model bekerja kurang baik pada data kelas netral, di mana model hanya mampu memprediksi benar 37 data dari total 64 data berlabel netral (11 + 16 + 37 = 64).

Teknik *Balancing Tomek Links* merupakan kombinasi teknik *balancing* yang paling optimal. Namun Kombinasi parameter terbaik dengan teknik *balancing* cenderung menurunkan nilai *accuracy* yaitu $Accuracy = (80 + 220 + 37) / 450 = 337 / 450 = 0.7489$

IV. KESIMPULAN DAN SARAN

Hasil eksperimen menunjukkan bahwa pelabelan menggunakan *Lexicon Inset* lebih baik dibandingkan pelabelan

manual. *Lexicon Inset* memberikan label secara objektif dan konsisten berdasarkan bobot nilai sentimen tiap kata dalam teks. Sebaliknya, pelabelan manual bergantung pada intuisi peneliti yang bersifat subjektif dan rentan bias, terutama pada teks ambigu.

Hyperparameter Tuning dan rasio split data latih dan data uji mempengaruhi performa model. Model *Naïve Bayes* dapat mencapai performa terbaiknya yaitu *Accuracy* 0.757, *Precision* 0.7540, *Recall* 0.757, *F1-Score* 0.7550. Performa terbaik tersebut dapat dicapai ketika menggunakan dataset pelabelan *Lexicon Inset* dengan menggunakan varian *MultinomialNB* dengan komposisi data 90:10 dan parameter nilai *Alpha* = 0.1, dan *Fit prior* = False.

Penurunan performa pada pembagian data 90:10 setelah dilakukan balancing disebabkan oleh jumlah data uji yang terlalu sedikit, sehingga hasil evaluasi menjadi kurang representatif dan sensitif terhadap kesalahan prediksi kecil.

REFERENSI

- [1] Fenilinas Adi Artanto, "Analisis Sentimen Opini Publik terhadap Fenomena Bunuh Diri Mahasiswa di Twitter Menggunakan Algoritma Naive Bayes," *SATESI: Jurnal Sains Teknologi dan Sistem Informasi*, vol. 4, no. 1, pp. 70–77, Apr. 2024, doi: 10.54259/satesi.v4i1.2908.
- [2] A. Aliyah et al., "Analisis Sentimen Twitter Terhadap Tren Penyebaran Informasi Pelaku Kejahatan Menggunakan Algoritma Naives Bayes," *BRIDGE : Jurnal publikasi Sistem Informasi dan Telekomunikasi*, vol. 2, pp. 85–97, 2024.
- [3] Nurochman and L. Ashiilah, "Perbandingan SVM dan LSTM Untuk Memprediksi Gangguan Kecemasan (Anxiety Disorder) Berdasarkan Cuitan di Platform Aplikasi X (Twitter)," *Jurnal Teknik Komputer AMIK BSI*, vol. 10, no. 2, pp. 147–155, 2024, doi: 10.31294/jtk.v4i2.
- [4] A. Liken Anggoro, L. V. P. Ken, and M. G. Setiawan, "Analisis Media Text Clustering pada Twitter Akan Kasus Selebriti Menggunakan Orange Data Mining," *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 7, no. 1, pp. 189–195, Jan. 2023, doi: 10.33395/remik.v7i1.12001.
- [5] A. M. Rizqiyah, I. Kadek, and D. Nuryana, "Analisis Sentimen Masyarakat terhadap Kebijakan Iuran Tabungan Perumahan Rakyat (Tapera) pada Platform X Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine," *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, 2024.
- [6] Maharduta Adiwijaya and Irwan Dwi Arianto, "Curhat Anonimitas Di Media Sosial Melalui Akun Menfess Twitter S Oleh Remaja Awal Surabaya," *Innovative: Journal Of Social Science Research*, vol. 3, pp. 7053–7065, 2023.
- [7] R. A. Khalil and M. B. Santoso, "Generasi Sandwich: Konflik Peran Dalam Mencapai Keberfungsian Sosial," *Share : Social Work Journal*, vol. 12, no. 1, 2022, doi: 10.24198/share.v12i1.39637.
- [8] JAKPAT Survey Report 2020, "How Indonesian Sandwich Generation Deal with The Economic Shock of COVID-19." [Online]. Available: <https://www.cnbcindonesia.com/news/20200414104733-8-151727/28-juta-pekerja-di-phk-dan-dirumahkan-akibat-covid-19>
- [9] S. Shofiyah, I. Arifin, U. Karimah, L. Yumna, H. Taufiqrohman, and D. A. Nurhovivah, "Generasi Sandwich Dalam Perspektif Pendidikan Islam," *Proceeding International Seminar on Islamic Studies*, vol. 4, 2023.
- [10] R. Mursyid and A. Dwi Indriyanti, "Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter)," *Journal of Informatics and Computer Science*, vol. 06, 2024.
- [11] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 411–423, Aug. 2022, doi: 10.55123/insologi.v1i4.770.
- [12] R. Wesley and R. Gunawan, "Literatur Review: Metode Deep Learning Untuk Analisis Teks," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 5, 2024.
- [13] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 556–562, Feb. 2024, doi: 10.57152/malcom.v4i2.1253.
- [14] F. Nur Salsabilla and A. Witanti, "Analisis Sentimen Akhir Masa Jabatan Presiden Jokowi Pada Media Sosial X Menggunakan Naïve Bayes," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 1, pp. 106–115, 2025, [Online]. Available: <https://t.co/2WBroEW27g>
- [15] F. Amrullah and A. Solichin, "Analisis Emosi Pada Live Chat Youtube 'Mata Najwa: 3 Bacapres Bicara Gagasan' Menggunakan Pendekatan Lexicon dan Algoritma Naive Bayes," *Jurnal TICOM: Technology of Information and Communication*, vol. 12, no. 3, p. 2024, 2024, [Online]. Available: <https://saifmohammad.com/WebPages/NRC-Emotion->