

Perbandingan Algoritma Naïve Bayes dan LSTM untuk Analisis Sentimen Terhadap Opini Masyarakat Tentang Sandwich Generation

Naufal Rizqi Ramadhan*, Nirwana Hendrastuty

Fakultas Teknik dan Ilmu Komputer, Informatika, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}naufal_rizqi_ramadhan@teknokrat.ac.id, ²nirwanahendrastuty@teknokrat.ac.id

Email Penulis Korespondensi: naufal_rizqi-ramadhan@teknokrat.ac.id

Submitted: 02/12/2024; Accepted: 18/12/2024; Published: 18/12/2024

Abstrak—*Sandwich Generation* merupakan sebuah sebutan untuk sekelompok orang yang memiliki orang tua yang sudah berumur dan anak-anak, sehingga mereka harus merawat kedua generasi tersebut. Opini tentang fenomena ini menimbulkan berbagai tanggapan di media sosial twitter, yang memerlukan analisis mendalam. Penelitian ini mengidentifikasi permasalahan yaitu belum adanya penelitian yang membandingkan kinerja algoritma Naïve Bayes dan LSTM dalam menganalisis sentimen opini masyarakat tentang sandwich generation, kompleksitas analisis data media sosial dengan karakteristik bahasa informal, singkatan, dan simbol yang sulit dianalisis manual, kebutuhan mengeksplorasi kemampuan algoritma dalam mengklasifikasikan sentimen, dan menentukan metode paling akurat untuk menganalisis sentimen opini masyarakat. Analisis sentimen digunakan untuk mengevaluasi opini, umpan balik, dan emosi dengan mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral. Hasil yang diperoleh dari penelitian ini yaitu metode LSTM memiliki kinerja lebih baik bila dibandingkan dengan Naive Bayes. Metode LSTM menghasilkan nilai akurasi, presisi dan recall sebesar 91.85%. sedangkan metode Naive Bayes memiliki nilai akurasi sebesar 83%, presisi 90% dan recall 82%.

Kata Kunci: Analisis Sentimen; LSTM; Naive Bayes; Sandwich Generation; SMOTE; Twitter

Abstract—*Sandwich Generation* is a term for a group of people who have elderly parents and children, so they have to take care of both generations. Opinions about this phenomenon have elicited various responses on social media twitter, which requires in-depth analysis. This study identifies the problems of the lack of research comparing the performance of Naïve Bayes and LSTM algorithms in analyzing public opinion sentiment about the sandwich generation, the complexity of social media data analysis with the characteristics of informal language, abbreviations, and symbols that are difficult to analyze manually, the need to explore the algorithm's ability to classify sentiment, and determine the most accurate method to analyze public opinion sentiment. Sentiment analysis is used to evaluate opinions, feedback, and emotions by classifying texts into positive, negative, or neutral categories. The results obtained from this study are that the LSTM method has better performance when compared to Naive Bayes. The LSTM method produced an accuracy, precision and recall value of 91.85%. while the Naive Bayes method has an accuracy value of 83%, precision of 90% and recall of 82%.

Keywords: Sentiment Analysis; LSTM; Naive Bayes; Sandwich Generation; SMOTE; Twitter

1. PENDAHULUAN

Peningkatan demografi di Indonesia mencerminkan perubahan signifikan dalam struktur penduduk, di mana jumlah penduduk usia produktif terus bertambah dan mendominasi populasi. Berdasarkan hasil Sensus Penduduk 2020 yang dirilis oleh Badan Pusat Statistik, dari total populasi di Indonesia 270,20 juta jiwa, sekitar 70,72% tergolong dalam usia produktif (15-64 tahun). Namun, kelompok produktif ini dihadapkan pada tantangan besar karena harus menopang kebutuhan hidup kelompok yang belum produktif (generasi Z dan *post-gen Z*) serta kelompok tidak lagi produktif (generasi *pre-boomer* dan *baby boomer*) yang mencapai 52,25% dari total populasi [1]. Kondisi ini menjadi salah satu faktor yang mempertegas fenomena *sandwich generation* di Indonesia.

Sandwich generation merujuk pada individu-individu yang berada dalam masa produktif tetapi harus menanggung beban finansial dan emosional untuk generasi di atas (orang tua) dan generasi di bawah (anak-anak) mereka secara bersamaan [2]. Istilah ini pertama kali diperkenalkan oleh Dorothy Miller pada tahun 1981 untuk menggambarkan wanita berusia 30-40 tahun yang berada dalam kondisi “terjepit” antara tanggung jawab terhadap keluarga dan pekerjaan [3]. Jika dilihat dari definisinya, *sandwich generation* adalah sebuah sebutan yang digunakan untuk sekelompok orang yang memiliki orang tua yang sudah berumur dan juga anak-anak, sehingga mereka harus merawat anak-anak dan orang tua mereka [4]. Seiring berjalannya waktu, istilah ini berkembang dan kini digunakan untuk menggambarkan pria maupun wanita yang menghadapi kondisi serupa.

Fenomena *sandwich generation* ini semakin kompleks ketika dibahas dalam konteks sosial media. Salah satu sosial media yang populer untuk menyalurkan pendapat bebas maupun aspirasi adalah Twitter. Di mana masyarakat Indonesia sering menyalurkan opini, pengalaman, hingga keluhan mereka. Twitter menjadi *platform* yang menarik untuk dianalisis karena sifatnya yang *real-time* dan sering digunakan untuk menyalurkan isu-isu sosial. Namun, data yang dihasilkan di media sosial ini memiliki karakteristik unik, seperti bahasa informal, singkatan, serta penggunaan simbol atau *emoticon* yang menyulitkan analisis manual. Oleh karena itu analisis sentimen berbasis algoritma menjadi penting untuk memahami persepsi publik secara menyeluruh.

Berdasarkan hal tersebut, penelitian ini mengidentifikasi beberapa permasalahan, antara lain belum adanya penelitian yang secara spesifik membandingkan kinerja algoritma Naive Bayes dan LSTM dalam analisis sentimen terhadap opini masyarakat mengenai fenomena *sandwich generation* di Indonesia, kompleksitas dalam menganalisis data media sosial yang ditandai dengan penggunaan bahasa informal, singkatan, simbol, dan *emoticon* yang

menyulitkan analisis manual, perlunya eksplorasi kemampuan algoritma Naive Bayes dan LSTM dalam mengklasifikasikan sentimen (positif, negatif, netral) pada komentar dan unggahan *sandwich generation*, serta kebutuhan untuk menentukan metode yang paling tepat dan akurat dalam menganalisis sentimen opini masyarakat teentang *sandwich generation*.

Analisis sentimen adalah metode yang digunakan untuk mengevaluasi sentimen, opini, umpan balik, serta emosi terkait produk, layanan, atau kondisi tertentu. Tujuan utama dari analisis sentimen adalah mengelompokkan kata, frasa, atau kalimat yang mengandung sentimen atau opini ke dalam kategori positif, netral, atau negatif. Analisis sentimen, atau yang juga dikenal sebagai *opinion mining*, merupakan proses yang melibatkan pemrosesan bahasa alami, komputasi linguistik, dan *text mining*. Tujuannya adalah untuk mengevaluasi *opini*, sentimen, penilaian, sikap, serta emosi seseorang baik pembicara maupun penulis terhadap topik, produk, layanan, organisasi, individu, atau aktivitas tertentu [5]. Oleh karena itu, diperlukan analisis sentimen terhadap kumpulan komentar dan unggahan tersebut untuk memahami perasaan yang disampaikan terkait topik atau objek tertentu di media sosial [6].

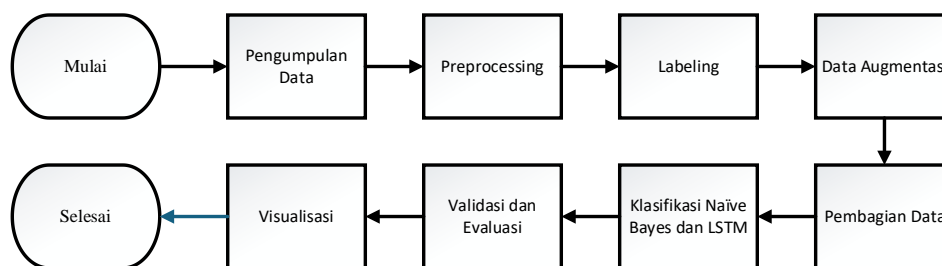
Beragam metode digunakan dalam analisis sentimen, seperti *Support Vector Machine* (SVM) [7], Naive Bayes [8], dan *Long Short Term Memory* (LSTM)[9][10]. Berdasarkan hasil tinjauan literatur, belum ditemukan penelitian yang membahas perbandingan antara metode LSTM dan *Naive Bayes* dalam analisis sentimen terkait opini masyarakat terhadap *sandwich generation*.

Berdasarkan hasil penelitian Apriyani & Kurniati (2020) dalam membandingkan metode *Naive Bayes* dan *Support Vector Machine* (SVM) untuk klasifikasi penyakit diabetes melitus [11]. Hasilnya menunjukkan bahwa algoritma SVM dengan kernel Polinomial lebih unggul sebesar 96% dibandingkan *Naive Bayes* yang mencapai 92%. Namun, pada penelitian Devita et al., 2018 membandingkan *Naive Bayes* dengan K-Nearest Neighbor (KNN) untuk klasifikasi artikel berbahasa Indonesia. *Naive Bayes* menunjukkan kinerja lebih baik dengan akurasi 70%, dibandingkan KNN yang hanya mencapai 40%. Penelitian tersebut merekomendasikan penambahan dataset dan peningkatan tahapan preprocessing, seperti stemming kata, untuk penelitian selanjutnya [12]. Sementara itu, dalam penelitian yang dilakukan oleh (Elisabeth, 2018) membahas perbandingan algoritma LSTM dan *Naive Bayes* dalam analisis sentimen terhadap merek Indomie [13] Mengungkapkan bahwa LSTM menunjukkan kinerja yang lebih unggul dibandingkan *Naive Bayes*.

Setelah melakukan kajian pustaka dan tinjauan literatur, diperoleh informasi bahwa belum ditemukan penelitian yang membandingkan metode LSTM dengan *Naive Bayes* untuk analisis sentimen terkait opini masyarakat terhadap *sandwich generation*. Hal ini memperkuat permasalahan penelitian yang telah diidentifikasi sebelumnya. Oleh karena itu, penelitian ini akan membahas perbandingan antara kedua metode LSTM dan *Naive Bayes* dengan tujuan untuk menentukan metode yang memiliki kinerja terbaik dengan akurasi tertinggi dalam menganalisis sentimen opini masyarakat terhadap *sandwich generation*.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, fokus utama ditujukan pada dua metode klasifikasi, yaitu *Naive Bayes* dan LSTM, meskipun ada berbagai metode klasifikasi lain yang tersedia. Proses untuk mendapatkan hasil yang optimal melibatkan beberapa tahapan, mulai dari pengumpulan dan pelabelan data, hingga preprocessing. Setelah data diproses, langkah selanjutnya adalah menerapkan algoritma *Naive Bayes* dan LSTM untuk klasifikasi. Gambar 1 merupakan tahapan-tahapan penelitian.



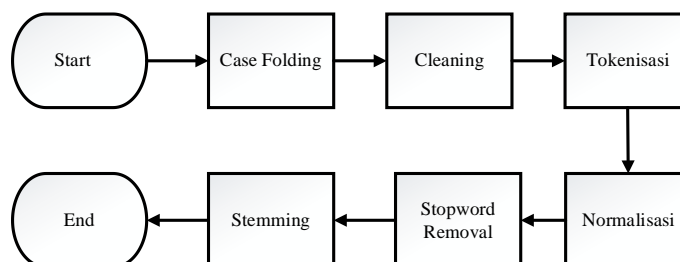
Gambar 1. Alir Diagram Penelitian

2.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan menggunakan metode *crawling* data, yaitu proses otomatisasi pengambilan data dari internet dengan menggunakan perangkat lunak khusus yang dikenal sebagai *web crawler* atau *spider* [14]. Data yang didapatkan bersumber dari *crawling* data dari Twitter. Dataset diperoleh melalui proses *crawling* data pada Twitter dalam rentang waktu sejak 1 Januari 2019 hingga 13 Oktober 2024, dengan total 3.769 data yang berhasil dikumpulkan. Data tersebut diambil menggunakan kata kunci "*sandwich generation*". Peneliti mengumpulkan data dengan memanfaatkan metode yang disediakan oleh API Twitter dan menyimpannya dalam format CSV.

2.2 Pre-Processing

Pemrosesan data atau *preprocessing* teks memiliki peran penting dalam *Text Mining* dan *Natural Language Processing* (NLP). *Preprocessing* adalah proses yang mengubah data mentah menjadi format yang dapat diproses lebih lanjut. Dalam konteks ini, langkah-langkah *preprocessing* meliputi berbagai tahap yang disesuaikan dengan tujuan penelitian, seperti *case folding*, *tokenizing*, penghapusan *stop word*, penanganan *slang word*, normalisasi, dan *stemming*. Tahapan *preprocessing* data dapat dilihat pada Gambar 2.



Gambar 2. Alir Diagram Preprocessing

- Case folding* adalah proses mengubah huruf kapital menjadi huruf kecil, yang bertujuan untuk memudahkan perbandingan atau komparasi teks [15].
- Cleaning* membersihkan teks dari elemen-elemen yang tidak relevan.
- Tokenisasi* memecah teks menjadi unit-unit kecil yang disebut token atau kata per kata.
- Normalisasi* mengubah teks menjadi bentuk yang baku atau standar.
- Stopword Removal* menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti “dan”, “yang”, “di”.
- Stemming* mengubah kata ke bentuk dasarnya

2.3 Labeling

Pelabelan dilakukan dengan mengelompokkan data ke dalam tiga kategori, yaitu memberikan label 1 untuk data netral, label 2 untuk data positif, dan label 0 untuk data negatif. Arsitektur model transformer memungkinkan pemberian label secara otomatis. Transformer menggunakan berbagai mekanisme perhatian untuk membangun arsitektur yang secara signifikan lebih unggul dalam tugas terjemahan mesin canggih, tanpa memerlukan lapisan berulang [16].

2.4 Data Augmentasi

Data augmentasi adalah suatu teknik yang digunakan untuk meningkatkan keberagaman dan jumlah data pelatihan dengan membuat variasi sintesis pada dataset asli. Dilakukan optimisasi berupa agumentasi data karena data yang tidak banyak sehingga dengan data yang sedikit maka variasi data pelatihan akan bertambah [17]. Pada penelitian ini, data augmentasi diterapkan untuk meningkatkan performa model dengan menghasilkan variasi data yang lebih banyak, mengurangi ketergantungan pada data yang terbatas, serta mengurangi risiko *overfitting* pada data pelatihan.

2.5 Pembagian Data

Setelah melalui proses *preprocessing*, pelabelan, dan memastikan keseimbangan dataset, dataset siap untuk diproses lebih lanjut. Langkah berikutnya adalah membagi dataset menjadi dua bagian, yaitu data *training* dan data *testing*. Data *testing* digunakan untuk mengukur kinerja model, sementara data *training* digunakan untuk melatih dan mengembangkan model.

2.6 Klasifikasi Menggunakan Algoritma *Naïve Bayes* dan LSTM

Pada penelitian ini dilakukan klasifikasi menggunakan dua algoritma yaitu *Naive Bayes* dan LSTM.

2.7 *Naïve Bayes*

Naïve Bayes adalah algoritma yang populer dalam melakukan klasifikasi, khususnya dalam analisis sentimen. Algoritma ini berlandaskan pada Teorema Bayes [18]. Penelitian ini menggunakan metode MultinomialNB, sebuah varian dari *Naïve Bayes Classifier* yang dirancang khusus untuk menangani data kategori atau berbentuk frekuensi, seperti pada analisis teks. MultinomialNB mengasumsikan bahwa fitur dalam model mengikuti distribusi Multinomial, yang umum digunakan dalam klasifikasi teks, di mana fitur berupa kata atau token dalam dokumen dan representasi data berupa frekuensi kemunculan kata-kata tersebut.

2.8 *Long-Short Term Memory* (LSTM)

Metode LSTM (*Long Short-Term Memory*) dikembangkan oleh Sepp Hochreiter dan Jurgen Schmidhuber pada tahun 1997 sebagai variasi dari *Recurrent Neural Network* (RNN). LSTM memiliki beberapa keunggulan, antara lain kemampuannya untuk menyimpan informasi jangka panjang, membaca dan memperbarui informasi sebelumnya, serta mengatasi masalah vanishing gradient yang sering terjadi pada varian RNN lainnya. Secara umum, LSTM terdiri dari



cell, *input gate*, *output gate*, dan *forget gate* [19]. Salah satu keunggulan utama LSTM adalah kemampuannya dalam mengingat informasi dari urutan data jangka panjang, yang sulit dicapai oleh metode tradisional dengan fitur yang lebih sederhana [20]. Kemampuan ini didukung oleh sistem gating yang ada dalam LSTM, di mana *cell state* berfungsi sebagai memori penyimpanan yang dapat dimanipulasi melalui *gating system* tersebut.

2.9 Validasi dan Evaluasi

Pada tahap ini, data yang telah dilakukan SMOTE untuk penyeimbangan kelas akan divalidasi menggunakan metode *Hold-Out*. Dengan metode ini, data dibagi menjadi dua bagian, yaitu data pelatihan dan data validasi. *train_test_split* digunakan untuk membagi data hasil SMOTE menjadi 80% untuk data pelatihan dan 20% untuk data validasi. Data pelatihan digunakan untuk melatih model, sementara data validasi digunakan untuk menguji kinerja model. Evaluasi dilakukan dengan melakukan pengujian menggunakan data yang sudah dibagi menjadi data testing maupun data *training*. Selanjutnya dijalankan pengujian tingkat performa dengan *confusion matrix*. Performa diukur dengan *matrix* seperti *Batch Size*, *Validation Loss*, *Validation Accuracy*, *Precision*, *Recall*, dan *F1-score*. Evaluasi dilakukan tidak hanya untuk setiap algoritma secara terpisah, tetapi juga melakukan pengujian berdasarkan *batch size*, nilai *neuron*, dan nilai *epoch*.

2.10 Visualisasi

Visualisasi data merupakan proses mengubah data mentah menjadi bentuk grafis, seperti grafik atau diagram, yang memudahkan pemahaman, analisis, dan komunikasi informasi. Proses ini memungkinkan identifikasi pola, tren, dan anomali dalam data, serta mendukung proses pengambilan keputusan yang berbasis pada data.

3. HASIL DAN PEMBAHASAN

3.1 Data Awal

Penelitian ini menggunakan data yang diperoleh melalui proses *crawling* menggunakan Google Colaboratory. *Dataset* dikumpulkan dari Twitter dalam periode antara 1 Januari 2019 hingga 13 Oktober 2024, dengan total 3.769 data yang berhasil dihimpun. Data tersebut diambil dengan kata kunci "*sandwich generation*" dan disimpan dalam format CSV menggunakan metode yang disediakan oleh API Twitter. Penelitian ini membandingkan kinerja algoritma *Naive Bayes* dengan LSTM untuk mengukur perbandingan dalam hal *accuracy*, *precision*, dan *recall*. Beberapa tahapan yang dilakukan mencakup *preprocessing*, validasi, dan evaluasi, dengan tujuan untuk memberikan label sentimen pada data tweet yang dikelompokkan menjadi tiga kategori: positif, negatif, dan netral. Tabel 1 menunjukkan hasil *crawling* data.

Tabel 1. Ulasan Komentar *Sandwich Generation*

No	Tweet
1	Kaluna di film Home Sweet Loan mah bukan sandwich generation
2	Sandwich generation tuh hal yang benar apa ngga si sebenarnya? Soalnya menurut aku salah
3	uang 1 7 blm ada seminggu tinggal sisa 500 pdhl ga nyicil apk apapun. ya allah sandwich generation gini amat ya:
4	Gue sandwich generation dan harus selalu naruh keluarga diatas kebahagiaan dan masa muda gue sendiri.
5	aku sebagai sandwich generation sangat terharu menonton video ini terima kasih bang Gerald videonya sangat menginspirasi
...	...
3769	Klo gue lebih demen cewe karir yg kerja mulu karena Sandwich Generation #FamilyFirst daripada cewe sosialita yang nongkrong & amp

3.2 Preprocessing

Proses ini melibatkan berbagai teknik, seperti pembersihan data, *case folding*, tokenisasi, penghapusan *stopword*, normalisasi, dan *stemming*. Hasil dari penerapan teknik-teknik tersebut akan dijelaskan secara rinci dalam bagian ini.

a. Case Folding

Pada langkah ini, semua huruf kapital diubah menjadi huruf kecil. Tanpa *case folding*, dua string yang tampak serupa bisa dianggap berbeda. Namun, setelah proses *case folding*, kedua string tersebut dianggap identik karena keduanya telah dikonversi menjadi huruf kecil. Rincian proses *case folding* dapat dilihat pada Tabel 2.

Tabel 2. Data Hasil *Case Folding*

No	Tweet	Case Folding
1	@yhaterus Semangat ia sandwich generation	@yhaterus semangat ia sandwich generation
2	Sandwich generation tuh hal yang benar apa ngga si sebenarnya? Soalnya menurut aku salah	sandwich generation tuh hal yang benar apa ngga si sebenarnya? Soalnya menurut aku salah



3	5 Drakor Bertema Sandwich Generation Haru dan Menginspirasi! https://t.co/VJWCKVY1LN	5 drakor bertema sandwich generation haru dan menginspirasi! https://t.co/VJWCKVY1LN
4	Roti apa yang jarang jajan? sandwich generation	roti apa yang jarang jajan? sandwich generation
5	Maaf mau Nanya Ini beneran Sandwich Generation seberat ini ?	maaf mau nanya Ini beneran sandwich generation seberat ini ?

b. *Cleaning*

Pada tahap ini, data teks dibersihkan dari elemen-elemen yang tidak relevan untuk analisis. Proses ini mencakup penghapusan tanda baca, simbol khusus (seperti @, #, dan RT), serta emotikon. Hasil dari tahap *cleansing* data dapat dilihat pada Tabel 3.

Tabel 3. Data Hasil *Cleansing*

No	Case Folding	Cleaning
1	@yhaterus semangat ia sandwich generation	semangat ia sandwich generation
2	sandwich generation tuh hal yang benar apa ngga si sebenarnya? Soalnya menurut aku salah	sandwich generation tuh hal yang benar apa ngga si sebenarnya soalnya menurut aku salah
3	5 drakor bertema sandwich generation haru dan menginspirasi! https://t.co/VJWCKVY1LN	drakor bertema sandwich generation haru dan menginspirasi
4	roti apa yang jarang jajan? sandwich generation	roti apa yang jarang jajan sandwich generation
5	maaf mau nanya Ini beneran sandwich generation seberat ini ?	maaf mau nanya Ini beneran sandwich generation seberat ini

c. *Tokenisasi*

Dalam tahap ini, angka, tanda baca, dan karakter lain yang tidak relevan juga dihapus karena tidak memiliki pengaruh signifikan terhadap analisis teks. Contoh penggunaan tokenisasi dapat dilihat pada Tabel 4.

Tabel 4. Data Hasil *Tokenisasi*

No	Cleaning	Tokenisasi
1	semangat ia sandwich generation	['semangat', 'ia', 'sandwich', 'generation']
2	sandwich generation tuh hal yang benar apa ngga si sebenarnya soalnya menurut aku salah	['sandwich', 'generation', 'tuh', 'hal', 'yang', 'benar', 'apa', 'ngga', 'si', 'sebenarnya', 'soalnya', 'menurut', 'aku', 'salah']
3	drakor bertema sandwich generation haru dan menginspirasi	['drakor', 'bertema', 'sandwich', 'generation', 'haru', 'dan', 'menginspirasi']
4	roti apa yang jarang jajan sandwich generation	['roti', 'apa', 'yang', 'jarang', 'jajan', 'sandwich', 'generation']
5	maaf mau nanya Ini beneran sandwich generation seberat ini	['maaf', 'mau', 'nanya', 'ini', 'beneran', 'sandwich', 'generation', 'seberat', 'ini']

d. *Normalisasi*

Proses ini dilakukan dengan menyusun kamus data yang berisi koreksi terhadap kesalahan ejaan umum, serta menambahkan entri khusus untuk memperbaiki kesalahan yang ditemukan dalam dataset. Contoh proses normalisasi dapat dilihat pada Tabel 5.

Tabel 5. Data Hasil *Normalisasi*

No	Tokenisasi	Normalisasi
1	['semangat', 'ia', 'sandwich', 'generation']	['semangat', 'ia', 'sandwich', 'generation']
2	['sandwich', 'generation', 'tuh', 'hal', 'yang', 'benar', 'apa', 'ngga', 'si', 'sebenarnya', 'soalnya', 'menurut', 'aku', 'salah']	['sandwich', 'generation', 'itu', 'hal', 'yang', 'benar', 'apa', 'tidak', 'sih', 'sebenarnya', 'soalnya', 'menurut', 'aku', 'salah']
3	['drakor', 'bertema', 'sandwich', 'generation', 'haru', 'dan', 'menginspirasi']	['drama korea', 'bertema', 'sandwich', 'generation', 'harus', 'dan', 'menginspirasi']
4	['roti', 'apa', 'yang', 'jarang', 'jajan', 'sandwich', 'generation']	['roti', 'apa', 'yang', 'jarang', 'jajan', 'sandwich', 'generation']
5	['maaf', 'mau', 'nanya', 'ini', 'beneran', 'sandwich', 'generation', 'seberat', 'ini']	['maaf', 'mau', 'tanya', 'ini', 'benaran', 'sandwich', 'generation', 'seberat', 'ini']

e. *Stopword Removal*

Tahap ini kata-kata yang termasuk dalam *stopwords* akan dihapus biasanya mencakup kata depan, kata ganti, dan kata penghubung, seperti "dan," "atau," "di," dan "ke." Hasil dari proses *stopwords removal* dapat dilihat pada Tabel 6.

Tabel 6. Data Hasil Stopword Removal

No	Normalisasi	Stopword Removal
1	['semangat', 'ia', 'sandwich', 'generation']	['semangat', 'sandwich', 'generation']
2	['sandwich', 'generation', 'itu', 'hal', 'yang', 'benar', 'apa', 'tidak', 'sih', 'sebenarnya', 'soalnya', 'menurut', 'aku', 'salah']	['sandwich', 'generation', 'salah']
3	['drama korea', 'bertema', 'sandwich', 'generation', 'harus', 'dan', 'menginspirasi']	['drama korea', 'bertema', 'sandwich', 'generation', 'menginspirasi']
4	['roti', 'apa', 'yang', 'jarang', 'jajan', 'sandwich', 'generation']	['roti', 'jarang', 'jajan', 'sandwich', 'generation']
5	['maaf', 'mau', 'tanya', 'ini', 'benaran', 'sandwich', 'generation', 'seberat', 'ini']	['maaf', 'benaran', 'sandwich', 'generation', 'seberat']

f. *Stemmi*ng

Dengan menggunakan pustaka Python Sastrawi, modul *StemmerFactory* digunakan dalam proses *stemming* untuk mengubah kata berimbuhan menjadi kata dasar. Hasil *stemming* ditampilkan pada Tabel 7.

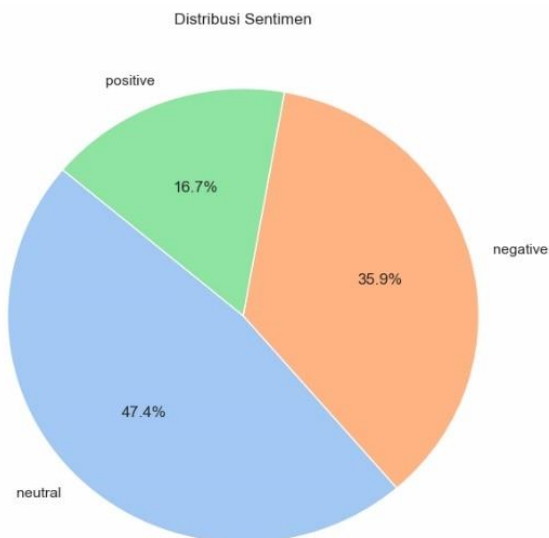
Tabel 7. Data Hasil *Stemming*

No	Stopword Removal	Stemming
1	['semangat', 'sandwich', 'generation']	semangat sandwich generation
2	['sandwich', 'generation', 'salah']	sandwich generation salah
3	['drama korea', 'bertema', 'sandwich', 'generation', 'menginspirasi']	drama korea tema sandwich generation inspirasi
4	['roti', 'jarang', 'jajan', 'sandwich', 'generation']	roti jarang jajan sandwich generation
5	['maaf', 'benaran', 'sandwich', 'generation', 'seberat']	maaf benar sandwich generation berat

3.3 Labeling Data

Setelah melalui tahap *preprocessing*, data yang telah dikumpulkan akan diberi label dengan mengelompokkan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Jumlah data yang diperoleh dari hasil *preprocessing* mencapai 3109 data.

```
label
neutral    1474
negative   1116
positive    519
Name: count, dtype: int64
```



Gambar 3. Diagram Labeling Data

Pelabelan data dan pengelompokan menjadi tiga kategori label yaitu positif, negatif, dan netral, diperoleh dataset dengan total 3109 data. Dataset ini terdiri dari 1.474 data berlabel netral, 1.116 data berlabel negatif, dan 519 data berlabel positif. Distribusi data yang telah dilabeli ditampilkan dalam diagram pada Gambar 3.

3.4 Data Augmentasi

Data augmentasi digunakan dalam penelitian ini untuk meningkatkan kinerja model dengan menghasilkan lebih banyak variasi, mengurangi ketergantungan pada jumlah data yang lebih kecil, dan mengurangi risiko overfitting selama pelatihan.

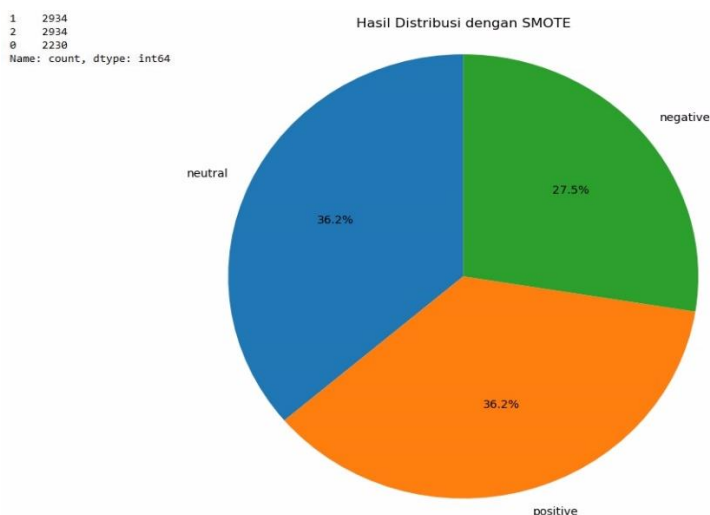
		Stemming	label
0	kaluna film rumah sweet loan mah sandwich gene...		neutral
1	uang minggu tinggal sisa nyicil apk apa iya al...		neutral
2	harga tiket segitu rencana ajak teman otomatis...		neutral
3	moga allah lancar rezeki sandwich generation a...		positive
4	kembang stuck situ dftr kuliah dukung improve ...		positive
...	
6189	gampang nabungnya tanggung Sousse sandwich gen...		negative
6190	sandwich generation milik dukung keluarga masy...		neutral
6191	tantang hadap sandwich generation berat bagi e...		negative
6192	sandwich generation kelompok orangutan jepit t...		neutral
6193	suka gadis karir kerja sandwich generation gad...		neutral

Gambar 4. Augmentasi Data

Pada Gambar 4 menunjukkan hasil augmentasi data, di mana jumlah data awal sebanyak 3.109 meningkat menjadi 6.193 setelah proses augmentasi..

3.5 Oversampling Data

Metode *random oversampling* digunakan dalam penelitian ini untuk menyeimbangkan distribusi kelas dengan menambah jumlah sampel ke kelas minoritas. Tujuan dari langkah ini adalah untuk mencegah dominasi kelas mayoritas dalam model dengan meningkatkan representasi kelas minoritas. Namun, menambah sampel ke kelas minoritas dapat menyebabkan *overfitting* pada data pelatihan, yang dapat mengganggu kinerja model pada data uji. Teknik *oversampling* sintesis minoritas (SMOTE) dipilih untuk mengatasi kelemahan *oversampling random* karena mampu meningkatkan representasi kelas minoritas tanpa sekadar menduplikasi sampel yang ada tetapi juga menghasilkan data sintesis baru. SMOTE pertama kali digunakan oleh Nithes V. Chawla [21] dan bekerja dengan mencari k-nearest neighbors untuk setiap sampel kelas minoritas, kemudian menghasilkan data sintesis berdasarkan persentase duplikasi yang tersedia. Hasil dari proses *random oversampling* dengan SMOTE dapat dilihat pada Gambar 5.



Gambar 5. SMOTE

3.6 Pembagian Data Latih dan Data Uji

Dalam penelitian ini, setelah data melalui proses SMOTE (*Synthetic Minority Oversampling Technique*), data yang telah diimbangi dibagi menjadi dua bagian, yaitu data pelatihan ('*X_train*' dan '*y_train*') dan data validasi ('*X_val*' dan '*y_val*'). Pembagian ini dilakukan menggunakan fungsi *train_test_split* dari *library Scikit-learn*, yang mengalokasikan 80% data untuk pelatihan dan 20% data untuk validasi. Parameter *random_state* digunakan untuk memastikan pembagian data yang konsisten setiap kali kode dijalankan.

3.7 Evaluasi Model

3.7.1 Pengujian LSTM Berdasarkan *Batch Size*

Pengujian ukuran batch dilakukan dengan mengubah ukuran batch dalam interval tertentu untuk menemukan nilai terbaik. Jumlah sampel data yang diproses pada *neural network* dalam satu iterasi disebut sebagai ukuran batch. Model untuk melakukan pengujian data dengan melihat perbandingan tiap *batch* dengan ketentuan menggunakan *neuron* 100 dan *epoch* 50 untuk menguji pada ukuran batch 32, 64, 128. Pengujian ini dilakukan sebagai bagian dari

hyperparameter tuning untuk memastikan model belajar secara efisien, mencapai generalisasi yang baik, dan memaksimalkan akurasi dalam waktu yang wajar. Hasil akhirnya adalah menemukan ukuran *batch* yang optimal untuk dataset dan model yang digunakan, serta perangkat keras yang tersedia, agar pelatihan menjadi lebih efektif dan efisien.

Tabel 8. Hasil Pengujian *Batch Size*

BS	VL (%)	VA (%)	PR (%)	RE (%)	FS (%)
32	0.444943	0.880247	0.879317	0.880247	0.878648
64	0.451938	0.880864	0.882705	0.880864	0.879416
128	0.444512	0.870370	0.869927	0.870370	0.868879

Dari hasil pengujian yang ditunjukkan pada Tabel 8, dapat dilihat bahwa *batch size* 64 memberikan performa terbaik dengan akurasi validasi sebesar 88%, *precision* sebesar 88.27%, *recall* sebesar 88%, dan *f1-score* sebesar 87.94%. Secara keseluruhan, *batch size* yang lebih besar (64) memberikan hasil yang sedikit lebih baik dibandingkan dengan ukuran *batch* lainnya, yaitu 32 dan 128.

3.7.2 Pengujian LSTM Berdasarkan Nilai *Neuron*

Pengujian jumlah *neuron* dilakukan dengan mengubah nilai *neuron* dalam rentang tertentu untuk mencari nilai optimal. Dalam pengujian ini, *batch size* ditetapkan sebesar 64 dan *epoch* sebanyak 50, karena berdasarkan hasil pengujian sebelumnya, akurasi tertinggi diperoleh pada *neuron* 100. Oleh karena itu, Pengujian data untuk melihat perbandingan dari *neuron* 75, 100, 125 dengan mengambil *batch size* terbaik dari pengujian sebelumnya. Pengujian ini bertujuan untuk menguji performa model LSTM dengan jumlah *neuron* berbeda. Hasil eksperimen ini akan membantu memilih konfigurasi terbaik (jumlah *neuron* yang optimal) berdasarkan performa model pada data validasi. Hasil pengujian tersebut dapat dilihat pada Tabel 9. Tabel ini menunjukkan bahwa N merupakan jumlah *neuron* yang diuji, VL adalah *Validation Loss*, VA adalah Akurasi, PR dan FS adalah presisi dan *F1-Score*.

Tabel 9. Hasil Pengujian *Neuron*

N	VL (%)	VA (%)	PR (%)	RE (%)	FS (%)
75	0.440879	0.875309	0.875315	0.875309	0.873382
100	0.452847	0.873457	0.872506	0.873457	0.871497
125	0.460865	0.866049	0.865893	0.866049	0.863722

Berdasarkan hasil di Tabel 9, dapat disimpulkan bahwa jumlah *neuron* 75 menghasilkan performa terbaik pada semua metrik evaluasi (*validation loss*, *validation accuracy*, *precision*, *recall*, dan *F1-score*). Sedangkan jumlah *neuron* 125 cenderung memberikan performa yang lebih rendah, menunjukkan bahwa penambahan *neuron* lebih banyak tidak selalu meningkatkan performa model dalam kasus ini.

3.7.3 Pengujian LSTM Berdasarkan Nilai *Epoch*

Melakukan pengujian untuk melihat perbandingan dari tiap *epoch* yang terdiri dari 75, 100, 150 *epoch* dengan *neuron* dan *batch size* terbaik dari hasil sebelumnya. Pengujian *epoch* bertujuan untuk mengatur jumlah siklus pelatihan model agar cukup untuk mempelajari pola dalam data tanpa menyebabkan *overfitting*. Dengan pengujian ini, model dapat mencapai akurasi tinggi pada data pelatihan dan, yang lebih penting, generalisasi yang baik pada data baru atau data validasi. Dari pengujian tersebut didapatkan hasil yang ditunjukkan pada Tabel 10.

Tabel 10. Hasil Pengujian *Epoch*

Epochs	VL (%)	VA (%)	PR (%)	RE (%)	FS (%)
75	0.385411	0.903086	0.902554	0.903086	0.901946
100	0.425250	0.882716	0.882062	0.882716	0.881256
150	0.371738	0.918519	0.917909	0.918519	0.917695

Berdasarkan hasil di Tabel 10, dapat disimpulkan bahwa *Epoch* 150 memberikan performa terbaik pada semua metrik evaluasi (*validation loss*, *validation accuracy*, *precision*, *recall*, dan *F1-Score*). Meskipun *Epoch* 75 menghasilkan akurasi yang cukup baik, *Epoch* 150 menunjukkan performa yang lebih baik dalam hal mengurangi *error (loss)* dan meningkatkan ketepatan serta deteksi data positif (*precision* dan *recall*). Sebaliknya, *Epoch* 100 tidak memberikan hasil yang optimal dibandingkan dengan *epoch* lainnya.

3.7.4 Pengujian Algoritma *Naïve Bayes*

Pada pengujian Algoritma *Naïve Bayes*, model Multinomial *Naïve Bayes* digunakan untuk klasifikasi teks setelah data diproses dan diseimbangkan menggunakan SMOTE. *Hyperparameter* alpha diuji melalui *GridSearchCV* dengan nilai

yang bervariasi antara 0.1 hingga 2.0, dan hasil terbaik digunakan untuk melatih model. Setelah itu, model diuji pada data validasi yang telah diproses dengan representasi TF-IDF dan dievaluasi menggunakan *classification report* untuk mengukur akurasi, *precision*, *recall*, dan *F1-score*. *Confusion matrix* digunakan untuk memvisualisasikan kesalahan klasifikasi, sementara *cross-validation* pada data pelatihan memberikan gambaran kestabilan model. Hasil pengujian menunjukkan performa model dalam memprediksi sentimen positif, negatif, dan netral dengan metrik evaluasi yang mencerminkan kinerja secara keseluruhan. Pada pengujian ini didapatkan hasil yang ditunjukkan pada Tabel 11.

Tabel 11. Tabel Hasil Klasifikasi Naive Bayes

	Precision (%)	Recall (%)	F1-score (%)	Support
Negatif	0.90	0.82	0.86	446
Netral	0.86	0.83	0.85	587
Positive	0.64	0.82	0.72	206
Accuracy			0.83	1239

Secara keseluruhan, *accuracy* model *Naive Bayes* adalah 83%, yang berarti model berhasil mengklasifikasikan 83% data dengan benar, meskipun *performa* terbaik tercapai pada kategori negatif dan netral, sementara kategori positif lebih sulit untuk diprediksi dengan akurasi tinggi.

3.7.5 Perbandingan Hasil Metode

dilakukan perbandingan hasil evaluasi antara metode LSTM dan Naïve Bayes berdasarkan beberapa metrik utama, yaitu akurasi (VA), presisi (PR), recall (RE), dan F1-score (FS)

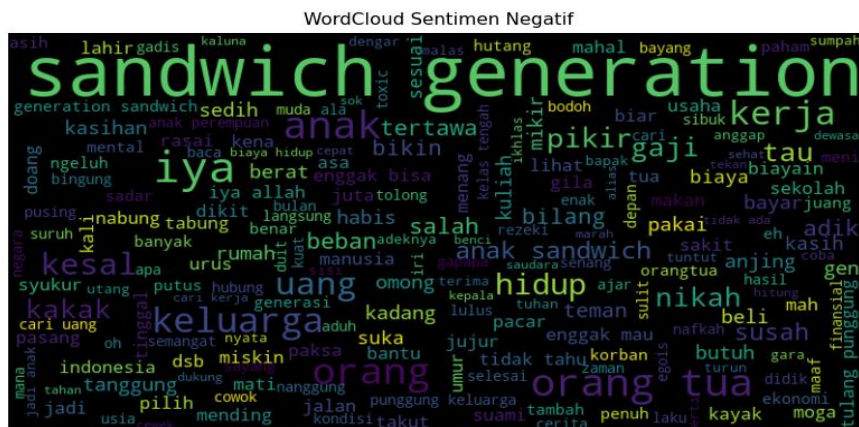
Tabel 12. Perbandingan Hasil Metode

M	VA (%)	PR (%)	RE (%)	FS (%)
LSTM	91.85	91.79	91.85	91.76
Naïve Bayes	83	90	82	86

Tabel 12 menampilkan hasil pengujian untuk setiap metode yang digunakan. Berdasarkan nilai akurasi, LSTM menunjukkan kinerja yang lebih unggul dengan akurasi sebesar 91,85%, sementara *Naïve Bayes* memperoleh akurasi sebesar 83%. Ini menunjukkan bahwa LSTM memberikan hasil yang lebih baik dibandingkan *Naïve Bayes* dalam hal akurasi.

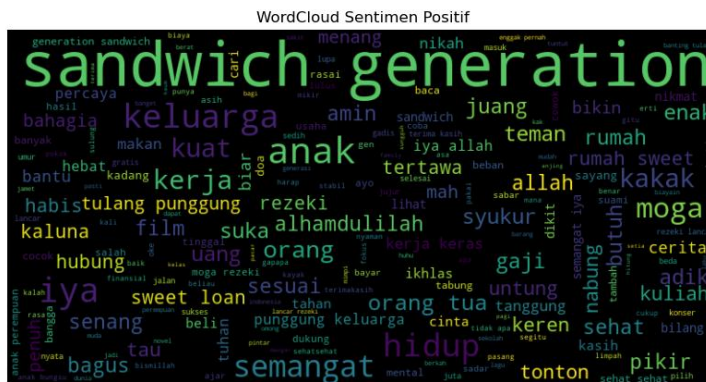
3.8 Visualisasi Sentimen Ulasan

Pada tahap ini, digunakan teknik *Wordclouds* untuk menggambarkan distribusi kata-kata yang sering muncul dalam ulasan atau *opini* yang dianalisis. *Wordclouds* adalah representasi grafis dari kata-kata di mana ukuran masing-masing kata mencerminkan frekuensi kemunculannya dalam dataset. Semakin besar ukuran suatu kata, semakin sering kata tersebut muncul dalam ulasan yang dianalisis. Hasil visualisasi kata pada data tweet mengenai *sandwich generation* dapat dilihat pada gambar 7.



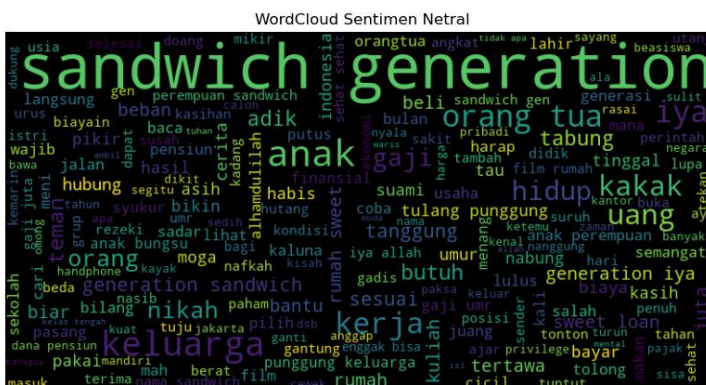
Gambar 6. Wordcloud Analisis Sentiment Sandwich Generation Negatif

Beberapa kata yang muncul lebih dominan dalam visualisasi ini, seperti "sandwich", "generation", "kerja", "gaji", "uang", "beban", "keluarga", dan "hidup". Kata-kata ini mengindikasikan bahwa ulasan dengan sentimen negatif cenderung berfokus pada masalah yang dihadapi oleh generasi *sandwich*, seperti beban finansial, masalah pekerjaan, tanggung jawab terhadap keluarga, serta tantangan dalam memenuhi kebutuhan hidup.



Gambar 7. Wordcloud Analisis Sentiment Sandwich Generation Positif

Visualisasi *Wordcloud* sentimen positif dapat dilihat pada Gambar 8. Terdapat kata yang dominan dalam *Wordcloud* ini, seperti "sandwich", "generation", "keluarga", "anak", "kerja", "hidup", "syukur", "semangat", "gaji", "rezeki", dan "sehat", mencerminkan tema utama dari ulasan dengan sentimen positif. Kata-kata ini menunjukkan fokus pada kebahagiaan, dukungan keluarga, serta keberhasilan dalam menghadapi tantangan sebagai bagian dari generasi *sandwich*. Topik yang sering dibicarakan meliputi rasa syukur terhadap keluarga, rezeki, dan semangat untuk bekerja keras serta mencapai kehidupan yang lebih baik.



Gambar 8. Wordcloud Analisis Sentiment Sandwich Generation Netral

Gambar 9 menunjukkan *Wordcloud* sentimen netral, dengan kata-kata dominan seperti "sandwich", "generation", "anak", "keluarga", "kerja", "gaji", "uang", "hidup", "tabung", dan "orang tua". Kata-kata ini mengindikasikan fokus pada aspek kehidupan generasi *sandwich*, seperti tantangan finansial, pekerjaan, dan tanggung jawab keluarga, tanpa menonjolkan perasaan positif atau negatif, sehingga mencerminkan sentimen netral.

4. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan, dapat disimpulkan bahwa metode *Long-Short Term Memory* (LSTM) memberikan performa yang lebih baik dibandingkan dengan metode *Naïve Bayes* dalam analisis sentimen pada data ulasan. Metode LSTM berhasil mencapai nilai akurasi sebesar 91.85%, dengan *presisi* 91.79%, *recall* 91.85%, dan *F1-score* 91.76%. Hal ini menunjukkan bahwa LSTM mampu memprediksi sentimen ulasan dengan lebih akurat dan konsisten. Sementara itu, metode *Naïve Bayes* memperoleh nilai akurasi 83%, *presisi* 90%, *recall* 82%, dan *F1-score* 86%. Meskipun *Naïve Bayes* memiliki nilai presisi yang cukup tinggi, namun nilai akurasinya lebih rendah dibandingkan dengan LSTM. Perbandingan antara kedua metode menunjukkan bahwa LSTM memiliki keunggulan dalam menangani kompleksitas data dan memanfaatkan informasi urutan kata dalam ulasan, sehingga lebih efektif dalam menangani teks panjang yang kaya akan konteks. Di sisi lain, *Naïve Bayes* meskipun lebih sederhana dan lebih cepat dalam proses pelatihan, tetapi kurang mampu menangkap hubungan yang lebih kompleks antar kata dalam teks, yang mengarah pada kinerja yang lebih rendah dalam hal akurasi dan *recall*. Secara keseluruhan, hasil ini menunjukkan bahwa LSTM lebih unggul dalam klasifikasi sentimen dibandingkan *Naïve Bayes*. Untuk penelitian selanjutnya, pengoptimalan parameter pada metode *Naïve Bayes* atau penggabungan keduanya dalam teknik *ensemble* bisa dipertimbangkan untuk meningkatkan performa lebih lanjut.

REFERENCES

- [1] rani, "Mengenal Sandwich Generation," psychology.binus. Accessed: Nov. 25, 2024. [Online]. Available: <https://psychology.binus.ac.id/2022/11/29/mengenal-sandwich-generation/>



- [2] A. M. Kristi, “5 Hal Menarik Tentang Sandwich Generation, Kamu Termasuk?,” IDN Times. Accessed: Nov. 25, 2024. [Online]. Available: <https://www.idntimes.com/life/inspiration/agata-melinda-kristi/5-hal-menarik-tentang-sandwich-generation-kamu-termasuk-c1c2?page=all>
- [3] V. Diller, “The ‘Over-Stuffed’ Sandwich Generation,” Psychology Today. Accessed: Nov. 25, 2024. [Online]. Available: <https://www.psychologytoday.com/intl/blog/face-it/201210/the-over-stuffed-sandwich-generation>
- [4] Cambridge Dictionary, “sandwich generation,” sandwich generation. Accessed: Nov. 25, 2024. [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/sandwich-generation>
- [5] L. Ardiani, H. Sujaini, and T. Tursina, “Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak,” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- [6] Wahyu Sejati, Ankur Singh Bist, and Amirsyah Tambunan, “Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis,” *Jurnal MENTARI: Manajemen, Pendidikan dan Teknologi Informasi*, vol. 2, no. 1, pp. 95–103, 2023, doi: 10.33050/mentari.v2i1.377.
- [7] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm),” *Jurnal Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [8] F. Ratnawati, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *INOVTEK Polbeng - Seri Informatika*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [9] C. K. Poetra, S. F. Pane, and N. S. Fatonah, “Meningkatkan Akurasi Long-Short Term Memory (LSTM) pada Analisis Sentimen Vaksin Covid-19 di Twitter dengan Glove,” *Jurnal Telematika*, vol. 16, no. 2, pp. 85–90, 2022, doi: 10.61769/telematika.v16i2.400.
- [10] M. A. Nurrohmat and A. SN, “Sentiment Analysis of Novel Review Using Long Short-Term Memory Method,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 3, p. 209, 2019, doi: 10.22146/ijccs.41236.
- [11] H. Apriyani and K. Kurniati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” *Journal of Information Technology Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issuue3.year2020.page133-143.
- [12] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, “Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 4, pp. 427–434, 2018, doi: 10.25126/jtiik.201854773.
- [13] Elisabeth;, “Perbandingan sentimen analisis terhadap brand Indomie menggunakan naive bayes & kong short term memory (LSTM),” 2018, Accessed: Dec. 09, 2024. [Online]. Available: http://sllims.umn.ac.id/index.php?p=show_detail&id=21183
- [14] D. E. Saputra and A. R. Isnain, “IMPLEMENTASI ALGORITMA CONVOLUTIONAL NEURAL NETWORK UNTUK ANALISIS SENTIMEN BACAPRES 2024 PADA KOLOM KOMENTAR YOUTUBE MATA NAJWA,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1431–1441, Aug. 2024, doi: 10.29100/jupi.v9i3.5420.
- [15] I. Mawanta, T. S. Gunawan, and W. Wanayumini, “Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 726, 2021, doi: 10.30865/mib.v5i2.2935.
- [16] K. Mohiuddin *et al.*, “Retention Is All You Need,” *International Conference on Information and Knowledge Management, Proceedings*, no. Nips, pp. 4752–4758, 2023, doi: 10.1145/3583780.3615497.
- [17] S. I. Pradika, B. Nugroho, and E. Y. Puspaningrum, “Pengenalan Tulisan Tangan Huruf Hijaiyah Menggunakan Convolution Neural Network Dengan Augmentasi Data,” *Prosiding Seminar Nasional Informatika Bela Negara*, vol. 1, pp. 129–136, Nov. 2020, doi: 10.33005/santika.v1i0.35.
- [18] R. Afandi, M. Afdal, and R. Novita, “Analisis Sentimen Masyarakat Terhadap Pinjaman Online di Twitter Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor,” vol. 6, no. 2, pp. 596–605, 2024, doi: 10.47065/bits.v6i2.5300.
- [19] A. Khumaidi, R. Raafi’udin, and I. P. Solihin, “Pengujiian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung,” *Jurnal Telematika*, vol. 15, no. 1, pp. 13–18, 2020, doi: 10.61769/telematika.v15i1.340.
- [20] L. Zaman, S. Sumpeno, and M. Hariadi, “Analisis Kinerja LSTM dan GRU sebagai Model Generatif untuk Tari Remo,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 2, p. 142, 2019, doi: 10.22146/jnteti.v8i2.503.
- [21] R. Peranganing, E. J. G. Harianja, I. K. Jaya, and B. Rumahorbo, “Penerapan Algoritma Safe-Level-Smote Untuk Peningkatan Nilai G-Mean Dalam Klasifikasi Data Tidak Seimbang,” *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 4, no. 1, pp. 67–72, 2020, doi: 10.46880/jmika.vol4no1.pp67-72.