PENERAPAN WORD N-GRAM DAN METODE NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN REVIEW STARBUCKS

Indri Syafitri, Jojor Putri Pasaribu
Ilmu Komputer, Universitas Negeri Medan
Jalan William Iskandar Kec. Percut Sei Tuan Medan, Indonesia
indrisyafitri78@gmail.com

ABSTRAK

Penelitian mengenai kepuasan pelanggan Starbucks dirancang untuk menganalisis tingkat kepuasan pelanggan berdasarkan ulasan yang diberikan. Penelitian ini dilakukan agar perusahaan dapat melakukan evaluasi guna mengidentifikasi faktor-faktor yang memengaruhi kepuasan pelanggan. Dengan demikian, kesetiaan pelanggan dapat terwujud jika mereka merasa puas dengan layanan yang diberikan oleh Perusahaan. Analisis sentimen adalah salah satu bidang penelitian dalam *text mining* yang bertujuan untuk mengelompokkan teks opini berdasarkan sentimen yang terkandung di dalamnya. Metode *Naïve Bayes Classifier* (NBC) memanfaatkan pendekatan statistik untuk melakukan inferensi induktif dalam menangani masalah klasifikasi. Salah satu keunggulan metode ini adalah kemampuannya untuk memerlukan sedikit data pelatihan dalam memperkirakan parameter yang dibutuhkan dalam proses klasifikasi. Berdasarkan hal tersebut, metode ini cocok digunakan dalam penelitian terkait kepuasan pelanggan Starbucks. Setelah melakukan serangkaian pengujian, disimpulkan bahwa *Naïve Bayes Classifier* efektif dalam mengklasifikasikan data teks, khususnya dalam penelitian ini yang terdiri dari ulasan pelanggan Starbucks. Dari 850 ulasan yang dianalisis, setelah proses pembersihan data, tersisa 672 ulasan yang dapat digunakan. Sentimen negatif mendominasi dengan 550 ulasan, sementara sentimen positif hanya sebanyak 122 ulasan. Uji akurasi menunjukkan hasil dengan nilai *Accuracy* 96%. Lalu diperoleh *precision* untuk ulasan positif sebesar 95%, *precision* untuk ulasan negatif sebesar 96%, *recall* untuk ulasan positif sebesar 84%, dan *recall* untuk ulasan positif sebesar 99%, *F1-Score* untuk ulasan posistif sebesar 89% dan *F1-Score* untuk ulasan negative sebesar 98%.

Kata kunci : *Analisis Sentimen, Naïve Bayes Classifier, N-gram*

1. PENDAHULUAN

Peningkatan konsumsi kopi di kalangan masyarakat membuka peluang bisnis bagi para pengusaha yang memasuki pasar industri kafe [1]. Oleh karena itu, potensi bisnis di Indonesia sangat besar, terutama di sektor kafe. Salah satu contohnya adalah Starbucks, yang menjadi peluang bisnis Dengan perkembangan pesat dan tingkat persaingan yang tinggi di era modern saat ini, perubahan gaya hidup membuat masyarakat semakin menganggap bahwa membeli dan menikmati kopi di kafe memberikan kesan lebih trendi dan berkelas, baik di kalangan anak muda maupun orang dewasa.. Pada kuartal III tahun 2021, tercatat terdapat 478 gerai resmi Starbucks di Indonesia [2].

Penelitian mengenai kepuasan pelanggan Starbucks dirancang untuk menganalisis tingkat kepuasan pelanggan berdasarkan ulasan yang di peroleh melalui *website Kaggle*. Maka, perusahaan perlu melakukan evaluasi guna mengidentifikasi faktor-faktor yang memengaruhi kepuasan pelanggan. Dengan demikian, kesetiaan pelanggan dapat terwujud jika mereka merasa puas dengan layanan yang diberikan oleh Perusahaan [3].

Analisis sentimen adalah salah satu bidang penelitian dalam *text mining* yang berperan dalam mengklasifikasikan dokumen teks berdasarkan opini dan sentimen [4]. Proses analisis ini dapat dilakukan menggunakan metode *machine learning*, seperti *Naïve Bayes Classifier*. Sentimen analisis, yang juga dikenal sebagai analisis subjektivitas atau penambangan opini, adalah teknik *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi pola dan fitur penting dari kumpulan teks. Dengan menggunakan NLP, analisis sentimen dapat mengekstraksi pola pengetahuan yang signifikan dari data tekstual yang besar. Teknik ini mengekstraksi perasaan penulis berdasarkan subjektivitas (objektif dan subjektif), polaritas (negatif, positif, dan netral), serta emosi seperti marah, bahagia, terkejut, sedih, cemburu, atau campuran emosi [5].

Beberapa penelitian sebelumnya yang menggunakan *Naïve Bayes Classifier* Analisis sentimen pada ulasan aplikasi Amazon Shopping di *Google Play Store* dilakukan untuk mempercepat proses analisis, sehingga dapat menghemat waktu dan usaha dalam mengklasifikasikan ulasan [6]. Penelitian tersebut menunjukkan hasil dengan rata-rata *accuracy* 82,15%, *precision* 72,25%, *recall* 83,49%, dan *f1-score* 77,41%. Dari keempat algoritma *Naive Bayes* yang diterapkan, Multinomial NB memberikan akurasi tertinggi sebesar 86,74%, dengan nilai *precision* 78,82%, *recall* 85,90%, dan *f1-score* 82,21%. Selanjutnya penelitian analisis sentimen tentang pemindahan ibu kota negara. Penelitian tersebut menghasilkan akurasi pada dataset yang berjumlah 606 data dengan menggunakan perbandingan data latih dan data uji sebesar 9:1 serta pembobotan kata menggunakan *TF-IDF* dan n-gram

(n=2), diperoleh hasil *accuracy* sebesar 77,05%, *precision* 77,05%, *recall* 100%, dan *f1-score* 87,04%. Penelitian lain tentang Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini dari Twitter [7] menunjukkan bahwa klasifikasi data uji menggunakan algoritma *Naïve Bayes Classifier* mencapai akurasi sebesar 71%. Akurasi untuk sentimen positif tercatat sebesar 71%, sedangkan untuk sentimen negatif mencapai 70%.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah kajian komputasi dalam bidang pengolahan bahasa alami dan linguistik komputasional yang bertujuan untuk menganalisis opini, penilaian, sikap, serta emosi terhadap suatu entitas, seperti produk, layanan, individu, organisasi, atau topik tertentu. Ulasan produk online sering digunakan sebagai data pengujian, di mana emosi seperti sedih, bahagia, atau marah diekspresikan, sehingga memungkinkan untuk mengidentifikasi apakah ulasan tersebut memiliki sentimen positif atau negatif [8].

2.2. N-Gram

N-Gram merupakan deretan n-item berturut-turut dalam suatu data, umumnya berupa teks. Nilai N dapat disesuaikan dengan kebutuhan, mulai dari satu hingga sepanjang teks yang dianalisis. Setiap posisi dalam n-gram berikutnya ditentukan berdasarkan posisi awal yang bergeser sesuai dengan *offset* yang telah ditentukan [9]. N-Gram dapat digambarkan menajdi beberapa klasifikasi yang didasarkan pada jumlah segmen kata atau *substring* yang dihasilkan. Klasifikasi N-Gram ini mencakup *Unigram, Bigram, Trigram*, dan sebagainya yang sesuai dengan bilangan bulat n dalam kerangka N-Gram. *Unigram, Bigram, dan Trigram* mewakili klasifikasi N-Gram yang sering digunakan dalam domain pemrosesan bahasa alami. *Unigram* terdiri dari kata tunggal, sedangkan *Bigram* terdiri dari dua kata, dan *Trigram* mencakup satu set tiga kata. Ketiga klasifikasi N-Gram ini secara sistematis digunakan dalam menjalankan analisis bahasa alami untuk menjelaskan interkoneksi antara kata-kata dalam teks tertentu. Jumlah segmen N-Gram ini dipastikan dengan jumlah gram yang diperlukan. Berikut ini contoh dari hasil penggunaan N-Gram pada kalimat "kami merasa tidak perlu mengatakannya":

a. *Unigram*
   {"kami","merasa","tidak","perlu","mengatakannya"}
b. *Bigram*
   {"kami merasa","merasa tidak", "perlu mengatakannya"}
c. *Trigram*
   {"kami merasa tidak","merasa tidak perlu","tidak perlu mengatakannya"}

2.3. Starbucks

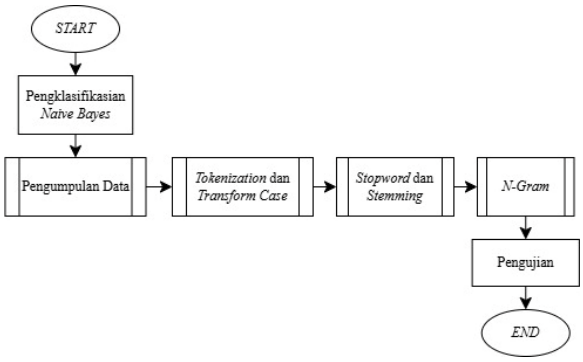
Starbucks Corporation merupakan perusahaan kopi asal Amerika yang beroperasi secara internasional dengan jaringan kedai kopi di seluruh dunia, dan kantor pusatnya berada di Seattle, Washington, D.C. Di Indonesia, PT Sari Coffee Indonesia, pemegang lisensi Starbucks, membuka gerai pertamanya pada tahun 2002 di Mall Plaza Indonesia. Para pelaku bisnis perlu terus memantau dan memperhatikan keputusan pembelian konsumen untuk menjaga keberlanjutan bisnis mereka di tengah persaingan yang semakin ketat.

2.4. *Naïve Bayes Classifier*

Naïve Bayes (NBC) merupakan pengklasifikasi probabilistik yang sederhana, yang menghitung probabilitas berdasarkan frekuensi serta kombinasi nilai dari data yang tersedia [10]. Metode ini menggunakan pendekatan statistik untuk melakukan inferensi induktif dalam menyelesaikan masalah klasifikasi [11]. Salah satu kelebihan dari metode *Naïve Bayes* adalah hanya memerlukan sedikit data pelatihan (*Training Data*) untuk memperkirakan parameter yang diperlukan dalam proses klasifikasi [12]. Persamaan teorema *Naïve Bayes* dinyatakan pada persamaan 1.

$$P(c|X) = \frac{P(X|c).P(c)}{P(X)} \dots\dots\dots (1)$$

Deskripsi:
$P(c|X)$: Probabilitas posterior
$P(c)$: Probabilitas sebelumnya
$P(X|c)$: Probabilitas kondisional yang terkait dengan hipotesis
$P\ (X)$: Probabilitas keseluruhan peristiwa c.
Berikut ini merupakan *Flowchart* dari penerapan algoritma *Naïve Bayes* yang dapat dilihat pada gambar 1.

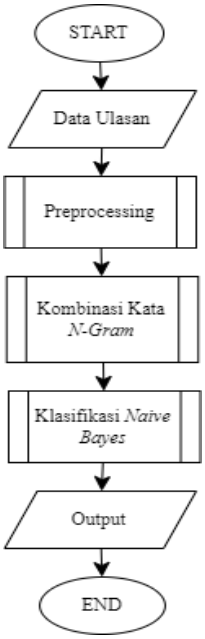


Gambar 1. *Flowchart* metode *Naïve Bayes*

3. METODE PENELITIAN

Metode penelitian ini meliputi pengolahan data yang merupakan tahapan bagaimana data diproses mulai dari input data hingga menghasilkan output yang diharapkan. Pada proses pengolahan data dilakukan analisis data menggunakan bantuan *software Google Colab* mulai dari melakukan *Preprocessing data*, serta proses kombinasi kata N-

gram dan penerapan klasifikasi *Naïve Bayes classifier*. Proses alur penelitian ditunjukkan pada Gambar 2.



Gambar 2. Diagram alur penelitian

3.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah data ulasan toko Starbucks yang merupakan jaringan kedai kopi terkenal yang diambil dari situs *web www.kaggle.com* dan disimpan pada file dengan ekstensi .csv. Dataset yang ada dibagi menjadi data *training* dan data *testing* dengan jumlah data *training* sebanyak 60%. Dataset terdiri 850 ulasan dengan 6 variabel yaitu *name, location, Date, Rating, Review*, dan *Image_Links*. Adapun definisi dari variabel yang digunakan dalam penelitian ini, yaitu:

Tabel 1. Definisi Variabel

Variabel	Definisi Variabel
name	Nama pelanggan yang memberikan *review* / ulasan
Location	Lokasi pelanggan yang memberikan *review* / ulasan
Date	Tanggal pemberian *review* / ulasan
Rating	Penilaian yang diberikan pelanggan kepada Starbucks dengan rasio 1 - 5
Review	Ulasan yang diberikan pelanggan
Image_Links	Tautan gambar yang berisi gambar/foto yang diberikan pelanggan

3.2. *Preprocessing*

Proses *Preprocessing* dilakukan untuk mengolah data yang beragam, karena data ini dapat secara signifikan mempengaruhi tahapan *training* dan *testing* nantinya. Prosedur *Preprocessing* menggunakan *Library Natural Language Toolkit* (NLTK). Langkah awal dalam proses *training* adalah *case folding*, yang berupaya mengubah semua teks menjadi huruf kecil. Langkah selanjutnya melibatkan tokenisasi, yang dirancang untuk mengelompokkan kalimat dan menghapus karakter seperti *apostrof* ('), titik (.), titik dua (:), dan lain-lain. Langkah ketiga meliputi *stopword removal*, yang dimaksudkan untuk membuang kata-kata yang sering muncul yang tidak memiliki makna atau arti. Langkah terakhir adalah *stemming*, yang bertujuan untuk menjadikan kata-kata yang diproses menjadi kata dasar.

3.3. Perancangan Algoritma

Pada tahap awal, dataset akan diklasifikasi atau diberi label positif dan negatif. Selanjutnya, data ulasan akan menjalani *Preprocessing* dengan tujuan mengubah data ulasan menjadi *term index*. Proses *Preprocessing* mencakup *case folding, stemming, tokenizing dan filtering*. Setelah ini, proses *filtering* akan dilakukan yang meliputi proses *stopword removal*. Data yang diproses sebelumnya akan menghasilkan *term* yang selanjutnya akan menjalani kombinasi kata menggunakan algoritma N-gram [13]. Metodologi pembobotan kata kemudian akan memfasilitasi klasifikasi dokumen menjadi sentimen positif dan negatif melalui algoritma *Naïve Bayes Classifier*.

Data yang digunakan pada proses klasifikasi ini dibagi sebesar 60% untuk data *training* dan 40% untuk data *testing* yang dapat dilihat pembagiannya pada tabel 2. Setelah melewati proses *Preprocessing* dan seleksi fitur oleh algoritma N-gram, data *training* yang ada akan digunakan sebagai bahan pembelajaran dalam menentukan klasifikasi pada data *testing*.

Tabel 2. Pembagian Data *Training* dan Data *Testing*

	Data *Training*	Data *Testing*
Rasio	60%	40%
Jumlah	403	269

3.4. *Confusion Matrix*

Confusion Matrix merupakan pendekatan sistematis yangdigunakan untuk menilai kemanjuran model prediktif. Dalam matriks ini, setiap baris menggambarkan kelas data asli, sementara setiap kolom mewakili kelas yang telah diperkirakan oleh model [14].

Tabel 3. Confusion Matrix

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	*True* Negatif (TN)	*False* Positif (FP)
Aktual Positif	*False* Negatif (FN)	*True* Positif (TP)

a. *True Positif* = Jumlah contoh yang termasuk dalam kelas positif dari data aktual yang modelnya secara akurat memprediksi kepositifan.
b. *True Negatif* = Jumlah *instance* milik kelas negatif dari data aktual yang modelnya secara akurat memprediksi negativitas.
c. *False Positif* = Jumlah contoh milik kelas negatif dari data aktual yang modelnya secara keliru memprediksi kepositifan.

d. *False Negatif* = Jumlah contoh milik kelas positif dari data aktual yang modelnya secara keliru memprediksi negativitas.

Memanfaatkan empat kumpulan data yang disebutkan di atas, perhitungan dapat diturunkan untuk mengevaluasi kemanjuran model, khususnya :

a. *Accuracy* = Rasio prediksi akurat terhadap total dataset.

$$\frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots (2)$$

b. *Precision* = Frekuensi prediksi positif yang dibuat oleh model memang benar.

$$\frac{TP}{TP + FP} \dots\dots\dots (3)$$

c. *Recall* = Frekuensi model memprediksi kepositifan ketika kelas sebenarnya positif.

$$\frac{TP}{TP + FN} \dots\dots\dots (4)$$

4. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan menggunakan *software Google Colab* dan hardware Laptop HP 14s-cf2xxx Processor Intel Core i3 dengan RAM 4GB. Pengklasifikasian *Naïve Bayes* pada penelitian ini mencakup pelaksanaan pemrosesan tekstual. [15]. Tahapan-tahapan dalam pengklasifikasian *Naïve Bayes* yaitu sebagai berikut.

4.1. Pengumpulan Data

Pengumpulan data yang didapatkan dari situs https://www.kaggle.com/datasets/harshalhonde/starbucks-reviews-dataset yang kemudian dilakukan penyusunan hingga menjadi dataset yang rapih yang contohnya dapat dilihat pada tabel 4.

Tabel 4. Data Review Starbucks

Name	Date	Rating	Review	Sentimen
Helen	Reviewed Sept. 13, 2023	5	*Amber and LaDonna at the Starbucks on Southwest Parkway are always so warm and welcoming…*	Positif
Alyssa	Reviewed Sept. 14, 2023	1	*We had to correct them on our order 3 times. They never got….*	Negatif
Taylor	Reviewed May. 26, 2023	5	*Me and my friend were at Starbucks and my card didnâ€™t work. Thankful the worker there…*	Positif

4.2. *Tokenization*

Tahapan ini bertujuan untuk memecahkan kalimat ulasan menjadi token-token kata dan sekaligus menghilangkan jumlah tanda baca yang tidak diperlukan, dengan menghasilkan total frekuensi kata yang muncul.

4.3. *Transform Case*

Pada tahapan ini akan akan dilakukan pengubahan huruf kapital menjadi huruf kecil (*lower case)* oleh *Rapidminer*, yang bertujuan untuk menghindari dobel kata.

4.4. *Stopword*

Tahapan ini memiliki tujuan untuk menghilangkan kata yang sering digunakan atau yang banyak muncul walaupun kata tersebut mencerminkan makna pada *review* tersebut. Daftar kata yang akan dihilangkan seperti kata *"don", "don't", "ain", "aren", "aren't"* dan lainnya.

4.5. *Stemming*

Pada tahap *stemming* ini kata yang berimbuhan akan diubah menjadi kata dasar.

4.6. *N-gram*

Penerapan *word* N-gram dengan jumlah data sebanyak 850 ulasan menghasilkan *attribute* sebanyak 5390 *attribute* pada *uni-gram*, 32540 *attribute* pada *Bi-gram*, dan 50510 *attribute* pada *Tri-gram*. Berikut ini ditampilkan contoh dari penerapan *word N-Gram* yang dilakukan pada penelitian ini.

a. *Unigram: [('Amber',), ('and',), ('LaDonna',), ('at',), ('the',), ('Starbucks',), ('on',), ('Southwest',), ('Parkway',), ('are',), ('always',), ('so',), ('warm',)*
b. *Bigram: [('Amber', 'and'), ('and', 'LaDonna'), ('LaDonna', 'at'), ('at', 'the'), ('the', 'Starbucks'), ('Starbucks', 'on'), ('on', 'Southwest'), ('Southwest', 'Parkway'), ('Parkway', 'are'), ('are', 'always'), ('always', 'so'), ('so', 'warm')*
c. *Trigram: [('Amber', 'and', 'LaDonna'), ('and', 'LaDonna', 'at'), ('LaDonna', 'at', 'the'), ('at', 'the', 'Starbucks'), ('the', 'Starbucks', 'on'), ('Starbucks', 'on', 'Southwest'), ('on', 'Southwest', 'Parkway'), ('Southwest', 'Parkway', 'are'), ('Parkway', 'are', 'always'), ('are', 'always', 'so'), ('always', 'so', 'warm'),*

Hasil dari pengklasifikasian *Naïve Bayes* tercermin dalam *volume* data dan atribut yang digambarkan dalam Tabel 5 berikut.

Tabel 5**.** Hasil Pengolahan Data Awal

Proses	Jumlah Data	Jumlah *Attribute*
Tokenization	70150	329292
Transform	38933250	27888753
Stopword	7365750	29898
Stemming	30407	4038
Uni Gram	70154	5390
Bi Gram	69482	32540
Tri Gram	68810	50510

4.7. Wordcloud

Wordcloud adalah representasi grafis di mana kata-kata yang digunakan secara berulang atau sering digunakan dalam setiap ulasan akan muncul sebagai gambar, dengan kata-kata yang menonjol atau berukuran besar mewakili yang sering digunakan, sedangkan kata-kata kecil menunjukkan kata ulasan yang digunakan dengan frekuensi yang lebih rendah.

Gambar 3. Wordcloud

Dari hasil *text_processing* yang sudah dilakukan didapatkan total 672 ulasan dari total 850 data ulasan yang ada dan menghasilkan sentimen negatif lebih dominan dengan jumlah sentimen sebanyak 550 ulasan, dan sentimen positif sebanyak 122 ulasan, yang ditunjukkan seperti pada tabel 6.

Tabel 6. Hasil Analisis Sentimen

Positively Rated	***Count***
0	550
1	122

4.8. Confusion Matrix

Pada titik ini, dilakukan *Confusion Matrix* untuk memastikan *accuracy*, *precision*, *recall*, dan *F1-Score* menggunakan data *Training* sebanyak 403 data dan data *Testing* sebanyak 269 data, dan didapat hasilnya seperti yang ditunjukkan pada gambar 4.

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.99	0.98	110
1	0.95	0.84	0.89	25
accuracy			0.96	135
macro avg	0.96	0.92	0.94	135
weighted avg	0.96	0.96	0.96	135

Gambar 4. Hasil *Confusion Matrix*

Dari hasil pengujian didapatkan hasil *accuracy* sebesar 96%. Lalu diperoleh *precision* untuk ulasan positif sebesar 95%, *precision* untuk ulasan negatif sebesar 96%, *recall* untuk ulasan positif sebesar 84%, dan *recall* untuk ulasan positif sebesar 99%, *F1-Score* untuk ulasan posistif sebesar 89% dan *F1-Score* untuk ulasan negative sebesar 98%.

5. KESIMPULAN DAN SARAN

Setelah pelaksanaan beberapa rangkaian evaluasi pada sistem, dapat disimpulkan bahwa metodologi *Naïve Bayes Classifier* mampu mengkategorikan data dalam bentuk tekstual, terutama dalam konteks penyelidikan ini, yang berfokus pada teks yang berasal dari ulasan pelanggan yang berkaitan dengan perusahaan Starbucks. Hasil klasifikasi sentimen berdasarkan 850 ulasan, setelah proses pembersihan data yang menghasilkan total 672 ulasan, menunjukkan bahwa sentimen negatif sebagian besar diwakili dengan 550 ulasan, dibandingkan dengan total 122 ulasan yang mencerminkan sentimen positif. Temuan dari penilaian akurasi menghasilkan skor *accuracy* 96%, Lalu diperoleh *precision* untuk ulasan positif sebesar 95%, *precision* untuk ulasan negatif sebesar 96%, *recall* untuk ulasan positif sebesar 84%, dan *recall* untuk ulasan positif sebesar 99%, *F1-Score* untuk ulasan posistif sebesar 89% dan *F1-Score* untuk ulasan negative sebesar 98%. Dengan didapat hasil *accuracy* tersebut maka dapat disimpulkan bahwa model yang dibangun pada penelitian ini dapat dinyatakan berhasil. Saran yang dapat dilakukan oleh peneliti selanjutnya yaitu sebaiknya dilakukan pengujian menggunakan metode lain untuk membandingkan tingkat keakuratan dengan metode yang telah dilakukan dalam penelitian ini.

DAFTAR PUSTAKA

[1] A. S. Herlambang and E. Komara, "Pengaruh Kualitas Produk , Kualitas Pelayanan , Dan Kualitas Promosi Terhadap Kepuasan Pelanggan (Studi kasus pada Starbucks Coffee Reserve Plaza Senayan)," *J. Ekon. Manaj. dan Perbank.*, vol. 8114, 2021.

[2] J. P. Putra, T. Janji, and R. Sitinjak, "Pengaruh kualitas produk dan harga terhadap kepuasan pelanggan Kopi Starbucks di Summarecon Mall Kelapa Gading 3," *J. Ilm. Akunt. dan Keuang.*, vol. 4, no. 8, pp. 3462–3470, 2022.

[3] H. S. Hopipah, R. Mayasari, U. S. Karawang, T. Timur, and J. Barat, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-Nearest Neighbor (k-NN) dan Naïve Bayes," *Technomedia J.*, vol. 6, no. 1, pp. 99–110, 2021.

[4] E. Engineering, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, pp. 32–35, 2023.

[5] M. M. Lamba, *Text Mining for Information Professionals*. 2022.

[6] E. Hasibuan and E. Allistair, "ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY," *J. Tek. dan Sci.*, vol. 1, no. 3, pp. 13–24, 2022.

[7] K. Zuhri, N. Adha, and O. Saputri, "Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Comput. Inf. Syst. Ampera*, vol. 1, no. 3, pp. 185–199, 2020.

[8] D. A. Agustina, S. Subanti, E. Zukhronah, P. S. Statistika, and U. S. Maret, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," *Indones. J. Appl. Stat.*, vol. 3, no. 2, pp. 109–122, 2020.

[9] C. Supriadi, H. D. Purnomo, and I. Sembiring, "Sensitivitas Sistem Pencarian Artikel Bahasa Indonesia Menggunakan Metode n-gram Dan Tanimoto Cosine," *TRANSFORMTIKA*, vol. 18, no. 1, pp. 63–70, 2020.

[10] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Citec J.*, vol. 2, no. 3, pp. 207–217, 2022.

[11] F. Liantoni and H. Nugroho, "KLASIFIKASI DAUN HERBAL MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DAN K- NEAREST NEIGHBOR," *J. SimanteC*, vol. 5, no. 1, pp. 9–16, 2022.

[12] L. Syafie, D. Indra, A. Djamalilleil, H. Hamrul, S. Anraeni, and L. B. Ilmawan, "Comparison of Artificial Neural Network and Gaussian Naïve Bayes in Recognition of Hand- Writing Number," *2018 2nd East Indones. Conf. Comput. Inf. Technol.*, no. 1, pp. 276–279, 2022.

[13] V. Nava, M. Kusman, V. Metayani, and O. Karnalim, "Prediksi Analisis Sentimen Data Debat Pemilihan Presiden 2024 Menggunakan Support Vector Machine (SVM) Prediction of Sentiment Analysis of 2024 Presidential Election Debate Data Using Support Vector Machine (SVM)," *J. KEILMUAN DAN Apl. Tek. INFROMATIKA*, vol. 3489, pp. 1–5, 2024.

[14] S. K. Wardani and Y. A. Sari, "Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5582–5590, 2021.

[15] M. N. Fadillah and R. N. Sukmana, "PENERAPAN ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN KELANGKAAN MINYAK GORENG PADA," *J. infotronik*, vol. 7, no. 2, pp. 82–90, 2022, doi: 10.32897/infotronik.2022.7.2.1716.