

THE EFFECT OF UNIGRAM AND BIGRAM IN THE NAÏVE BAYES MULTINOMIAL FOR ANALYZING OF COMMENT SENTIMENT OF GOJEK APPLICATION IN GOOGLE PLAY STORE

Adrian Dwinanda Adyatma¹, Lasmedi Afuan^{*2}, Eddy Maryanto^{*3}

^{1,2,3}Informatics Department, Engineering Faculty, Universitas Jenderal Soedirman, Indonesia
Email: ¹adrianadyatma21@gmail.com, ²lasmedi.afuan@unsoed.ac.id, ³eddy.maryanto@unsoed.ac.id

(Article received: August 09, 2023; Revision: December 21, 2023; published: Desember 23, 2023)

Abstract

In sentiment classification systems that use Naïve Bayes Classifier, a commonly used feature extraction method is TF-IDF with unigram and bigram, where the two is used separately. In the reality, most of texts contain single or composed word, so it is needed to use the combination of unigram and bigram to maximize the accuracy of the classification results. In this research, the impact and performance improvement between classification systems using unigram or bigram solely and those using a combination of both are studied. Using 1000 data of reviews with ratings 1 (negative) and 5 (positive) from Gojek users on the Google Play Store, and performing performance validation with K-Fold at K=10, the system that uses the combined TF-IDF feature extraction of unigrams and bigrams achieves the best performance among the three systems with an accuracy of 0.84, however the accuracy of the system that uses unigrams solely has accuracy of 0.83, and 0.7 for the system that uses bigram. From the results of the research, it can be concluded that the use of the combination of unigram and bigram can increase the accuracy of the classification result.

Keywords: Google Play Store, Naïve Bayes, N-Gram, sentiment classification, TF-IDF.

PENGARUH UNIGRAM DAN BIGRAM DALAM IMPLEMENTASI MULTINOMIAL NAÏVE BAYES UNTUK ANALISIS SENTIMEN ULASAN APLIKASI GOJEK PADA GOOGLE PLAY STORE

Abstrak

Metode ekstraksi fitur yang banyak digunakan pada sistem klasifikasi sentimen yang menggunakan Naïve Bayes Classifier adalah TF-IDF unigram dan bigram yang digunakan secara terpisah. Pada kenyataannya, teks yang ada dapat mengandung sekaligus kata tunggal maupun kata majemuk, sehingga untuk memaksimalkan akurasi hasil klasifikasi, unigram dan bigram harus digunakan secara bersamaan. Tujuan dari penelitian ini adalah melakukan investigasi tentang pengaruh dan besarnya peningkatan performansi antara sistem klasifikasi yang hanya menggunakan unigram atau bigram, dengan sistem yang menggunakan gabungan dari keduanya. Pada penelitian ini digunakan 1000 data ulasan pengguna Gojek dengan rating 1 (negatif) dan 5 (positif) yang diambil dari *Google Play Store*, dan validasi performa dengan K-Fold dengan K=10. Hasil penelitian menunjukkan bahwa sistem yang menggunakan ekstraksi fitur TF-IDF gabungan unigram dan bigram menghasilkan performansi paling baik dengan akurasi sebesar 0.84, sistem yang menggunakan unigram saja memiliki akurasi sebesar 0.83, sedangkan sistem yang menggunakan bigram saja memiliki akurasi sebesar 0,7. Berdasarkan hasil penelitian dapat disimpulkan bahwa penggunaan kombinasi unigram dan bigram dapat meningkatkan akurasi hasil klasifikasi.

Kata kunci: Google Play Store, klasifikasi sentimen, Naïve Bayes, N-Gram, TF-IDF.

1. PENDAHULUAN

1.1. Layanan Transportasi Online

Perkembangan teknologi yang pesat membawa manfaat dan perubahan gaya hidup yang cukup signifikan bagi manusia. Salah satu pengaruh dari perkembangan teknologi adalah maraknya penggunaan layanan transportasi *online* dalam

beberapa tahun terakhir. Pertumbuhan pangsa pasar transportasi *online* di Indonesia mengalami pertumbuhan sangat besar, pada tahun 2015 bernilai US\$900 juta, dan meningkat di tahun 2018 sebesar US\$12,7 miliar dan di prediksi mencapai US\$40 miliar pada 2025 [1].

Gojek merupakan salah satu layanan transportasi *online* yang berasal dari Indonesia.

Jumlah unduh aplikasi Gojek mencapai 142 juta, dengan layanan pesan-antar makanan mencapai 400 ribu mitra di 370 kota di Indonesia, dan bermitra dengan 28 institusi keuangan [2].

Berdasarkan data dari laman Gojek di Google Play Store yang dikumpulkan pada 19 Januari 2023, Gojek memiliki rating 4,4 dari 5 yang diberikan oleh 5,25 juta pemberi ulasan [3]. *Review* pengguna berisi dua bagian, yaitu nilai rating dan komentar secara tekstual. Nilai rating menunjukkan evaluasi keseluruhan pengalaman pengguna menggunakan skala numerik, namun komentar tekstual mampu bercerita lebih mendalam.

Fungsi analisis sentimen adalah menggolongkan tiap kutub teks pada berbagai sumber internet maupun media sosial yang berbentuk dokumen atau kalimat kemudian ditentukan apakah kata tersebut termasuk kategori positif, netral ataupun negatif [4]. Data ulasan dari Google Play Store dapat dijadikan acuan untuk penilaian kinerja dari aplikasi Gojek dan sebagai sumber masukan untuk klasifikasi sentimen karena menilai dari: objektivitas data, kemudahan dalam mendapatkan data, kualitas data itu sendiri, sedikitnya jumlah data yang tidak berhubungan, dan kemudahan dalam mengidentifikasi sentimen pengguna. Karena manfaat ini, banyak bidang yang menggunakan sentimen analisis guna mengetahui kualitas aplikasi dengan melihat ulasan yang diberikan terhadap aplikasi terkait[5].

Dalam menyelesaikan permasalahan analisis sentimen, salah satu pendekatan yang paling banyak digunakan adalah menggunakan algoritma Multinomial Naïve Bayes. Melalui penelusuran di Google Scholar, terdapat 20.800 penelitian dengan kata kunci “sentimen naïve bayes” [6].

Multinomial Naïve Bayes banyak digunakan karena memiliki tingkat akurasi baik walaupun dengan masukan data yang tidak terlalu banyak. Pada sebuah penelitian analisis sentimen dari ulasan aplikasi e-government menggunakan Multinomial Naïve Bayes dengan 577 data latih dan 65 data uji mampu menghasilkan tingkat akurasi sebesar 89% [7].

1.2. Klasifikasi Multinomial Naïve Bayes

Berdasarkan latar belakang yang telah diuraikan di muka, rumusan masalah yang ingin dipecahkan dalam penelitian ini yaitu: bagaimanakah performa kinerja Gojek sebagai salah satu penyedia layanan transportasi *online* jika dinilai dari sentimen ulasan penggunaannya di laman Google Play Store. Salah metode klasifikasi yang dapat digunakan untuk menganalisis sentimen pengguna Gojek adalah Klasifikasi Multinomial Naïve Bayes. Berdasarkan hasil penelitian terdahulu terdapat pengaruh dari penggunaan metode unigram dan bigram terhadap kinerja ekstraksi fitur dari sistem Klasifikasi Multinomial Naïve Bayes, sehingga perlu diteliti bagaimanakah pengaruh dari metode unigram,

bigram dan kombinasi dari keduanya, terhadap kinerja ekstraksi fitur dari sistem klasifikasi Multinomial Naïve Bayes untuk memaksimalkan kinerja sistem. *Sentiment analysis* atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [8]. Teknik yang lebih umum dari *sentiment analysis* terkait dengan pemrosesan data tekstual adalah *text mining*. *Text mining* atau penambangan teks digunakan untuk mengekstrak fakta dan hubungan dalam bentuk terstruktur yang dapat digunakan untuk membuat anotasi pada basis data khusus, mentransfer pengetahuan antar domain, dan secara lebih umum dalam intelijen bisnis untuk mendukung pengambilan keputusan operasional dan strategis dari suatu organisasi [9].

Secara umum ada beberapa langkah dalam proses analisis sentimen Naïve Bayes yaitu

1. Text Preprocessing

Text Preprocessing adalah bagian dari *Text Mining* yang dilakukan untuk menghapus *noise* pada kalimat. *Text Preprocessing* bertujuan untuk menghindari data yang kurang sempurna, gangguan pada data dan data yang tidak konsisten. Sebelum membuat model, *dataset* yang akan digunakan harus dilakukan pembersihan terlebih dahulu dari kata yang tidak relevan dengan sentimen seperti tanda baca, angka, *link url* dan simbol-simbol [10].

2. Naïve Bayes Classifier

Naïve Bayes Classifier merupakan sebuah metode pengklasifikasian dengan menggunakan probabilitas sederhana yang berakar pada Teorema Bayes dan memiliki asumsi ketidaktergantungan (*independent*) yang tinggi dari masing-masing kondisi atau kejadian [11].

3. TF-IDF

Term Frequency (TF) dan *Inverse Document Frequency (IDF)* merupakan sebuah metode untuk melakukan vektorisasi data, yaitu pemberian nilai dari tiap fitur yang dalam kasus ini berupa sebuah kata. Pembobotan ini dilakukan agar sistem dapat menganalisa pola yang ada dari suatu teks terhadap suatu kelas [12].

4. K-Fold Cross Validation

Data hasil proses klasifikasi Naïve Bayes dan pengujian *confusion matrix* akan divalidasi menggunakan teknik *K-Fold Cross Validation*. *K-Fold* merupakan salah satu metode *cross validation* yang digunakan untuk melakukan evaluasi kinerja dari sistem *classifier* dengan memecah data ke beberapa bagian dan mengulangi eksperimen sebanyak *k*. Validasi ini digunakan untuk mendapatkan nilai performa dari sistem klasifikasi [13].

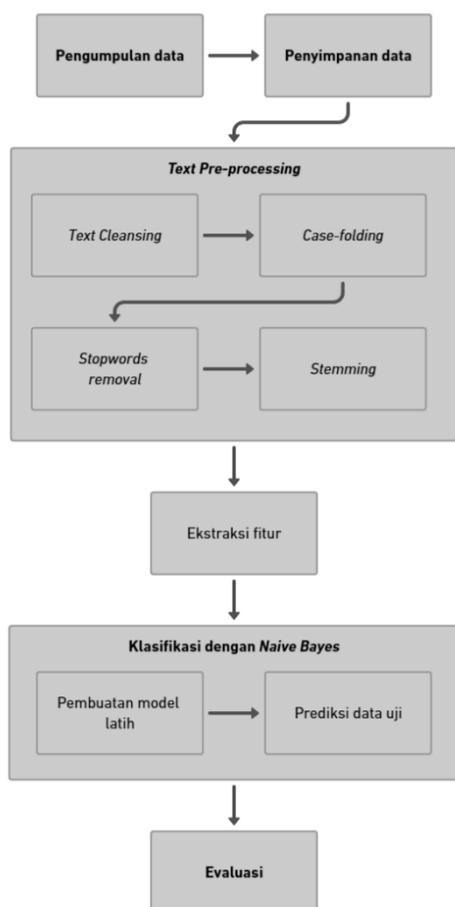
2. METODE PENELITIAN

Pelaksanaan penelitian dibagi menjadi beberapa tahapan: studi pustaka, pengembangan sistem

analisis sentimen, evaluasi hasil, kemudian diakhiri dengan penyusunan laporan.

Pengembangan sistem dan evaluasi hasil akan dilakukan menggunakan Google Colab, sebuah platform development interaktif berbasis web. Penggunaan platform ini dipilih dikarenakan kelengkapan dan keandalannya untuk melakukan komputasi ilmiah dan data-based research. Karena berbasis web, Google Colab juga bisa diakses dari beragam perangkat.

Tahap-tahap pengembangan sistem analisis sentimen pada penelitian ini secara garis besar dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Pelaksanaan Penelitian

Data yang digunakan berupa ulasan pengguna Gojek berasal dari laman Gojek di Google Play Store, di laman tersebut tersedia secara publik ulasan-ulasan yang dibuat oleh pengguna aplikasi Gojek dari perangkat Android.

Ulasan yang dikumpulkan adalah ulasan yang berkategori paling relevan, dan memiliki skor 1 atau 5 saja, dengan masing-masing 500 data.

Data yang didapatkan adalah kumpulan ulasan dengan informasi sebagai berikut: tanggal pembuatan ulasan, ID ulasan, teks ulasan, penulis ulasan, skor/rating untuk aplikasi tersebut, jumlah vote, dan balasan dari penerbit/pengembang.

Data mentah selanjutnya dipilah dan disimpan dalam format berkas CSV berisikan teks ulasan dan

skor/rating. Adapun struktur data yang menampilkan teks dan polaritas, dapat dilihat pada Tabel 2.

Tabel 1. Struktur Penyimpanan Data

Teks	Skor
Aplikasi yg sangat membantu dan sangat menguntungkan. Makanya selalu ku gunakan setiap hari. Setelah sekian lama menggunakan gofood (mungkin keseringan pesan), akibatnya promo resto dihilangkan, sehingga jadi males pesan makanan. Semoga bisa kembali seperti semula.	5
Ini aplikasi gojek tampilan nya kenapa pula, tidak ada tampilan saldo dan mau lihat riwayat transaksi terakhir aja tidak bisa!!! Kocak ini aplikasi lama"!!! Ini pas di pilih bantuan dan laporan notif nya selalu server padat merayap!!! Hadeh...	1

Teks ulasan yang sudah didapatkan, kemudian akan dilakukan pre-processing.

Text cleansing melakukan pembersihan karakter yang tidak digunakan seperti angka, simbol-simbol, dan tanda baca (contoh: !, @, #, \$, %, ^, & = + * ?).

Kemudian, dilakukan *case-folding*, yaitu pengubahan seluruh karakter huruf menjadi huruf kecil. Ini dilakukan agar dua kata yang sama namun memiliki kapitalisasi yang berbeda tidak dibaca sebagai kata yang berbeda (contoh: 'MantaP', dan 'mantap') di proses vektorisasi fitur nantinya.

Langkah ketiga adalah melakukan penghapusan *stopwords*, sistem akan menyaring kata-kata yang kurang berguna agar tidak disertakan dalam dokumen.

Tahap terakhir pada tahap *text pre-processing* adalah stemming, yaitu mengubah kata berimbuhan menjadi kata dasar, sebagai contoh 'menyenangkan' menjadi 'senang', 'memakan' menjadi 'makan'.

Sebelum data bisa diklasifikasi, data teks harus diubah atau ditransformasi terlebih dahulu menjadi data numerik atau fitur numerikal agar bisa diproses oleh komputer. Proses ini disebut juga sebagai ekstraksi fitur yang merupakan sebuah proses vektorisasi, yang mana pada penelitian ini proses vektorisasi ini dilakukan dengan menggunakan metode TF-IDF.

Tahap berikutnya adalah proses klasifikasi dengan menggunakan Naïve Bayes Classifier. Pada proses klasifikasi, data dibagi menjadi dua bagian yaitu data latihan dan data uji.

Naïve Bayes Classifier akan melakukan pelatihan dengan mencocokkan tiap fitur dengan kelas sesuai bobot berdasarkan data latihan yang mana setiap data latihan memiliki fitur dan kelas (dalam penelitian ini, kelas menggunakan rating dari tiap data). Data dengan kelas=1, akan disebut juga sebagai kelas negatif, dan data dengan kelas=5, akan disebut juga sebagai kelas positif.

Penelitian ini menggunakan K-Fold *Cross Validation* dengan nilai K=10. Dengan menggunakan nilai K=10 maka sistem akan melakukan iterasi sebanyak 10 kali dengan data uji dan latihan yang bergantian. Skema K-Fold *Cross Validation* dengan

1000 data pada penelitian ini dapat dilihat pada Tabel 3.

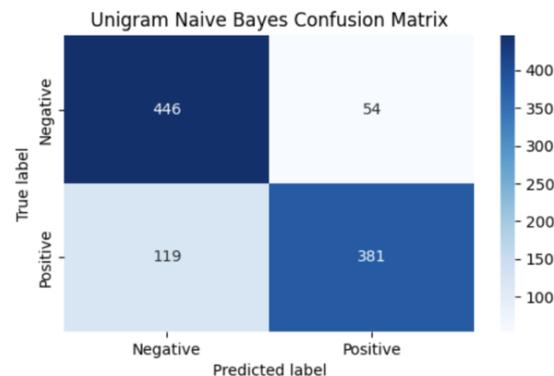
Tabel 2. Skema K-Fold Cross Validation

K	Data	Run									
		1	2	3	4	5	6	7	8	9	10
1	1-100	Uji									
2	101-200		Uji								
3	210-300			Uji							
4	301-400				Uji						
5	401-500					Uji					
6	501-600						Uji				
7	601-700							Uji			
8	701-800								Uji		
9	801-900									Uji	
10	901-1000										Uji

Pada tahap Evaluasi yang merupakan langkah terakhir, dilakukan analisis dengan menggunakan *confusion matrix*, mencari tahu tingkat akurasi, *precision*, *recall*, dan *F1-score* dari hasil yang didapatkan. Pada tahap ini dilakukan perbandingan performansi hasil klasifikasi yang menggunakan ekstraksi fitur dengan parameter yang berbeda yaitu unigram, bigram, dan kombinasi unigram dan bigram.

3. HASIL DAN PEMBAHASAN

Bentuk *confusion matrix* serta penghitungan *accuracy*, *precision*, *recall*, *f1-score* dari klasifikasi menggunakan vektorisasi data unigram, dapat dilihat pada Gambar 2.



```

Unigram classification report:
precision  recall  f1-score  support
1          0.79    0.89     0.84     500
5          0.88    0.76     0.81     500

accuracy   0.83
macro avg  0.83
weighted avg 0.83
    
```

Gambar 2. Hasil Klasifikasi Unigram

Hasil klasifikasi unigram mendapatkan akurasi sebesar 0.83. Kelas negatif (*rating* 1) menghasilkan *precision* sebesar 0.79, dan *recall* sebesar 0.89. Kelas positif (*rating* 5) menghasilkan *precision* sebesar 0.88, dan *recall* sebesar 0.76.

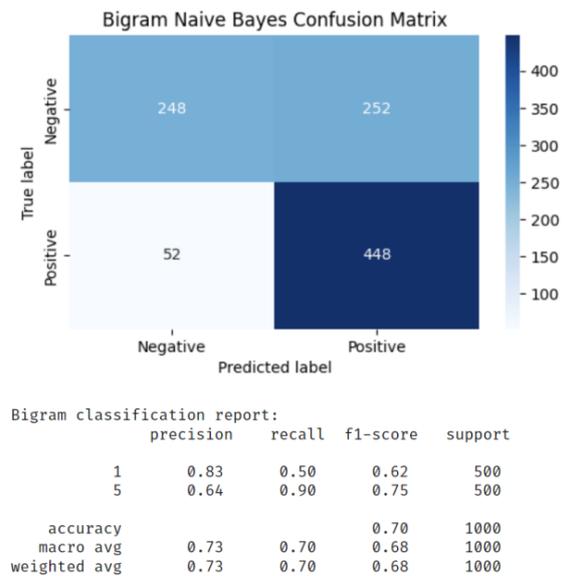
Dari 500 data negatif, 446 berhasil diprediksi dengan benar, menyisakan 54 atau 10,8% dari

keseluruhan data negatif diprediksi salah sebagai positif. Di sisi lain, sistem memprediksi 565 data sebagai negatif, dengan 119 atau 21.1% dari prediksi tersebut sebenarnya adalah data positif.

Dari 500 data positif, 381 berhasil diprediksi benar positif, menyisakan 119 data atau 23.8% salah diprediksi sebagai negatif. Di sisi lain, dari 500 data positif, sistem memprediksi 435 sebagai positif, dengan 54 data atau 12.4% dari prediksi tersebut sebenarnya adalah data negatif.

Sistem klasifikasi unigram cenderung menghasilkan prediksi negatif walaupun dengan *dataset/support* yang seimbang. Ketidakeimbangan ini menghasilkan nilai *recall* yang baik untuk kelas negatif, karena dari 500 data negatif, 446 data benar diprediksi sebagai negatif. Namun, sistem lebih presisi dalam memprediksi kelas positif, karena dari 435 data yang diprediksi positif, hanya 54 prediksi salah yang sebenarnya adalah data negatif.

Bentuk *confusion matrix* serta penghitungan *accuracy*, *precision*, *recall*, *f1-score* dari klasifikasi menggunakan vektorisasi data bigram, dapat dilihat pada Gambar 3.



Gambar 3. Hasil Klasifikasi Bigram

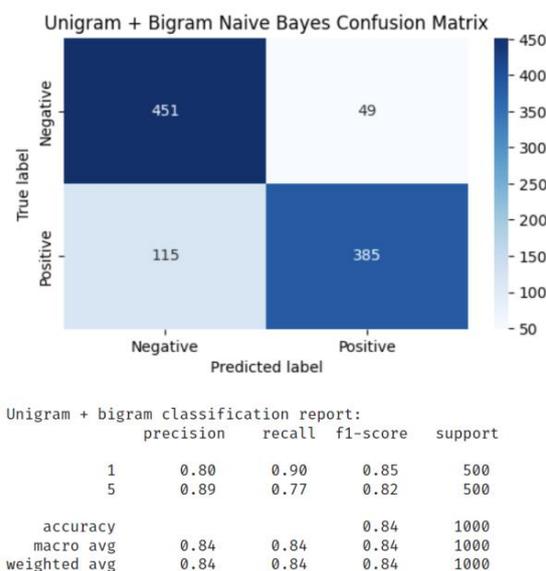
Hasil klasifikasi bigram mendapatkan akurasi terburuk di antara ketiga metode ekstraksi n-gram, di angka 0.70.

Dari 500 data negatif, hanya 248 diprediksi benar sebagai negatif, sedangkan sisanya sejumlah 252 salah diprediksi sebagai data positif.

Sistem klasifikasi bigram cenderung menghasilkan prediksi positif, dengan menghasilkan 700 prediksi data positif. Ketidakeimbangan ini menghasilkan nilai *recall* kelas positif yang sangat tinggi di angka 0.90 karena sistem berhasil memprediksi 448 data sebagai positif dari 500 data yang benar positif, menyisakan 52 data salah diprediksi sebagai negatif. Namun, karena tidak seimbang cenderung memprediksi positif, sistem klasifikasi menghasilkan *error False Positive* yang

banyak di angka 252 atau 36% dari 700 prediksi data positif sebenarnya adalah negatif.

Bentuk *confusion matrix* serta penghitungan *accuracy*, *precision*, *recall*, *f1-score* dari klasifikasi menggunakan ekstraksi fitur n-gram gabungan unigram dan bigram, dapat dilihat pada Gambar 4.



Gambar 4. Hasil Klasifikasi Unigram dan Bigram

Hasil klasifikasi sistem gabungan unigram dan bigram menghasilkan akurasi sebesar 0.84, dan secara keseluruhan merupakan peningkatan dari sistem klasifikasi unigram saja.

Dari 500 data positif dan 500 data negatif, sistem menghasilkan prediksi 566 prediksi negatif dan 434 prediksi positif, menunjukkan sistem klasifikasi *balanced* dengan sedikit bias ke arah negatif sebesar 6.67%.

Nilai *recall* dan *precision* tiap kelas mendapatkan peningkatan 0.01 poin dari sistem klasifikasi unigram. Peningkatan ini merupakan hasil penambahan parameter bigram yang cenderung memprediksi kelas positif ke sistem unigram yang sudah berjalan baik.

Dari 500 data positif, 385 data diprediksi benar sebagai positif, menyisakan 115 data atau 23% salah diprediksi sebagai negatif. Angka ini merupakan peningkatan sebesar 0.8% dari sistem unigram yang menghasilkan *error* sebesar 23.8% pada kelas yang sama.

Dari sisi negatif, 451 data dari 500 data positif benar diprediksi sebagai negatif, menyisakan 49 data atau 9.8% salah diprediksi sebagai positif. Angka ini menunjukkan peningkatan performa dari sistem klasifikasi unigram yang menghasilkan *error* 12.4% pada kelas yang sama.

4. DISKUSI

Pada penelitian ini skala rating yang digunakan sama dengan skala rating yang digunakan oleh Asri et al. [14] yaitu 1 sampai dengan 5. Jumlah sampel dan

perbandingan split data juga sama masing-masing berturut-turut 1000 dan 90:10.

Nilai performansi *precision*, *recall*, *f1-score*, dan akurasi ketiga parameter n-gram dirangkum pada Tabel 4, nilai terendah untuk tiap parameter disorot dengan warna merah, dan nilai tertinggi untuk tiap parameter disorot dengan warna hijau.

Tabel 3. Performa Sistem Klasifikasi dengan Perbedaan N-Gram

N-gram	Kelas	Precision	Recall	F1-score	Akurasi
1	Negatif	0.79	0.89	0.84	0.83
	Positif	0.88	0.76	0.81	
2	Negatif	0.83	0.50	0.62	0.70
	Positif	0.64	0.90	0.75	
1, 2	Negatif	0.80	0.90	0.85	0.84
	Positif	0.89	0.77	0.82	

Sistem klasifikasi menggunakan gabungan unigram dan bigram mendapatkan akurasi terbaik sebesar 0.84, mengalahkan unigram dengan selisih hanya 0.01. Sedangkan sistem klasifikasi yang menggunakan bigram saja mendapatkan akurasi terburuk di angka 0.70. Secara umum tingkat akurasi yang diperoleh belum maksimal, hal ini mungkin disebabkan oleh adanya asumsi bahwa antar atribut tidak terdapat hubungan [15].

Setelah diteliti lebih dalam, sistem klasifikasi gabungan unigram & bigram memiliki peningkatan parameter performa sebesar 0.01 di semua sisi terhadap sistem klasifikasi yang menggunakan unigram saja.

Sistem klasifikasi gabungan juga memiliki mayoritas angka performansi tertinggi dari ketiga sistem. Dari 7 total parameter, hanya 2 parameter yang lebih buruk dari sistem klasifikasi bigram, yaitu: *precision* kelas negatif, dan *recall* kelas positif. Namun sistem klasifikasi bigram memiliki nilai performansi terburuk di antara ketiga sistem secara keseluruhan, dengan empat parameter performansi terburuk dihasilkan dari sistem ini.

5. KESIMPULAN

Sistem klasifikasi Naïve Bayes dari data ulasan Google Play Store dapat membantu pelaku industri mendapatkan wawasan yang lebih luas mengenai pendapat penggunanya. Tentu saja hal ini harus diimbangi dengan sistem yang menghasilkan performansi tinggi agar tidak mendapatkan informasi yang salah.

Sistem klasifikasi unigram menghasilkan performansi yang cukup baik dengan akurasi 0.83. Sistem klasifikasi bigram memiliki akurasi 0.70 yang merupakan sistem dengan performansi terburuk diantara ketiga sistem.

Sistem kombinasi unigram dan bigram dapat meningkatkan performansi sebesar 0.01 di semua parameter, dan menjadi sistem klasifikasi terbaik di antara ketiga sistem ekstraksi fitur n-gram pada penelitian ini.

Penggunaan dan manfaat *stemming* pada penelitian klasifikasi sentimen serupa bisa diteliti lebih dalam lagi. Dalam penelitian ini, proses yang paling memakan waktu adalah *preprocessing* pada tahapan *stemming* dengan library Sastrawi. Perlu diteliti penggunaan *stemming* pada sistem klasifikasi sentimen apakah dapat meningkatkan performa, dan jika memang dapat meningkatkan performa, apakah peningkatan performa tersebut sepadan dengan waktu proses yang menjadi lebih lama.

DAFTAR PUSTAKA

- [1] R. Wahyudi dan G. Kusumawardhana, "Analisis Sentimen pada Review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine," *Jurnal Informatika.*, vol. 8, no. 2, pp. 200–207, 2021, doi: [10.31294/ji.v8i2.9681](https://doi.org/10.31294/ji.v8i2.9681).
- [2] N. Fitriyah, B. Warsito, dan 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.9.3.376-390](https://doi.org/10.14710/j.gauss.9.3.376-390).
- [3] "Gojek - Aplikasi di Google Play," 2023. <https://play.google.com/store/apps/details?id=com.gojek.app&hl=id&gl=US> (accessed January 19, 2023).
- [4] S. M. Salsabila, A. Alim Murtopo, dan N. Fadhilah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," *Jurnal Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: [10.33395/jmp.v11i2.11640](https://doi.org/10.33395/jmp.v11i2.11640).
- [5] A. Rahman, E. Utami, dan S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," *Jurnal Komtika Komputasi Dan Inform.*, vol. 5, no. 1, pp. 60–71, 2021, doi: [10.31603/komtika.v5i1.5188](https://doi.org/10.31603/komtika.v5i1.5188).
- [6] "Sentimen Naive Bayes - Google Scholar," 2022. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=sentimen+naive+bayes&btnG= (accessed June 3, 2022).
- [7] A. I. Tanggraeni dan M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI Jurnal Teknik Informatika Dan Sistem Informasi*, vol. 9, no. 2, pp. 785–795, 2022, doi: [10.35957/jatisi.v9i2.1835](https://doi.org/10.35957/jatisi.v9i2.1835).
- [8] T. Yusnitasari, D. Ikasari, E. E. S. Pratiwi, dan N. Syahri, "Analisis Sentimen Terhadap Review Restoran Fish Streat pada Aplikasi Zomato Menggunakan Stemming Nazief Adriani dan Naive Bayes Classifier," *Prosiding Sentrinov*, 2017, vol. 3, pp. 12, url: <http://proceeding.sentrinov.org/index.php/sentrinov/article/view/181/160>.
- [9] Raksaka Indra Alhaqq, I Made Kurniawan Putra, dan Yova Ruldeviyani, "Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store," *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, vol. 11, no. 2, pp. 105–113, 2022, doi: [10.22146/jnteti.v11i2.3528](https://doi.org/10.22146/jnteti.v11i2.3528).
- [10] A. Kusuma dan H. N. Irmanda, "Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (Senamika)*, 2022, pp. 563–574, url: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2214/1761>.
- [11] F. V. Sari dan A. Wibowo, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, pp. 681-686, 2019, url: <https://jurnal.umk.ac.id/index.php/simet/article/view/3487/1883>.
- [12] A. H. Al-Atho'illah, E. D. Wahyuni, dan A. A. Arifyanti, "Model Klasifikasi Teks Produk Terlarang Menggunakan Algoritma Campuran (Hybrid) SVM dan Leksikon," *Jurnal Informatika dan Sistem Informasi*, vol. 1, no. 2, pp. 10, 2020, url: <https://garuda.kemdikbud.go.id/documents/detail/1757899>.
- [13] D. N. N. Husnina, D. E. Ratnawati, dan B. Rahayudi, "Analisis Sentimen Pengguna Aplikasi RedBus berdasarkan Ulasan di Google Play Store menggunakan Metode Naive Bayes," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 2, pp. 737–743, 2023, url: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12297/5593>.
- [14] Y. Asri, W. N. Suliyanti, D. Kuswardani, dan M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile: Analisis Sentimen," *PETIR*, vol. 15, no. 2, pp. 264–275, 2022, doi: [10.33322/petir.v15i2.1733](https://doi.org/10.33322/petir.v15i2.1733).
- [15] N. Salmi and Z. Rustam, "Naïve Bayes Classifier Models for Predicting the Colon Cancer," *IOP Conf. Series: Materials Science and Engineering 546 (2019) 052068*, pp. 1–8, 2019, doi:[10.1088/1757-899X/546/5/052068](https://doi.org/10.1088/1757-899X/546/5/052068).