

Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma *Naive Bayes* dan Algoritma Genetika

Arif Rahman^{1*}, Ema Utami², Sudarmawan³
^{1,2,3}Magister Teknik Informatika, Universitas Amikom Yogyakarta
*email: arif.1996@students.amikom.ac.id

DOI: <https://doi.org/10.31603/komtika.v5i1.5188>

Date received: 22-06-2021, Last Revised: 10-07-2021, Accepted: 13-07-2021

ABSTRACT

Several studies have been carried out to determine the sentiment analysis of applications on the Google Playstore. Of the 4 studies that have been reviewed by the author, there are several shortcomings, namely the average test carried out only once so that the accuracy value obtained is still not well tested and it is not stated how many datasets are used in each of these studies. From these shortcomings, the author will improve the quality of the research that the author has reviewed. The research that the author built started with several processes. Starting with the dataset collection process, dataset classification, oversampling process, pre-processing process, *Naive Bayes* algorithm process, and genetic algorithm. From the research process, the accuracy values for each application were obtained, namely Shopee with an accuracy value of 96.53%, Ruangguru with an accuracy value of 95.54%, Tokopedia with an accuracy value of 96.87%, and Gojek with an accuracy value of 96.54%.

Keywords: Google Playstore, *Naive Bayes* Algorithm, Genetic Algorithm

ABSTRAK

Ada beberapa penelitian yang sudah dilakukan untuk mengetahui sentimen analisis aplikasi pada google playstore. Dari 4 penelitian yang sudah penulis *review* terdapat beberapa kekurangan yaitu pengujian yang dilakukan rata-rata hanya 1 kali sehingga nilai akurasi yang didapat masih belum teruji dengan baik dan tidak disebutkan berapa jumlah dataset yang digunakan pada masing-masing penelitian tersebut. Dari kekurangan tersebut penulis akan meningkatkan kualitas penelitian yang telah penulis *review*. Penelitian yang penulis bangun dimulai dengan beberapa proses. Dimulai dengan proses pengumpulan dataset, pengkelasan dataset, proses *oversampling*, proses *pre-processing*, proses algoritma *Naive Bayes* dan proses algoritma genetika. Dari proses penelitian tersebut didapatkan nilai akurasi pada masing-masing aplikasi yaitu Shopee dengan nilai akurasi 96,53%, Ruangguru dengan nilai akurasi 95,54%, Tokopedia dengan nilai akurasi 96,87% dan Gojek dengan nilai akurasi 96,54%.

Kata-kata kunci: Google Playstore, Algoritma *Naive Bayes*, Algoritma Genetika

PENDAHULUAN

Sentimen analisis merupakan suatu ilmu untuk mengekstrak teks untuk mendapatkan emosi seseorang [1]. Untuk itu sentimen analisis memiliki banyak manfaat salah satunya untuk mengetahui apakah para pelanggan memiliki tanggapan yang baik atau tidak terhadap produk dan ini bisa menjadi masukan untuk perkembangan bisnis produk tersebut dimasa depan [2]. Karena manfaat ini, banyak bidang yang menggunakan sentimen analisis, salah satunya para penyedia aplikasi di google playstore guna mengetahui kualitas aplikasi dengan melihat *review-review* yang diberikan terhadap aplikasi tersebut [3]. Sehingga banyak penelitian-penelitian yang dibuat guna mengetahui sentimen analisis aplikasi di *google playstore*.

Dimulai dari penelitian yang dilakukan oleh irfani dan Faizal Fakhri ini bertujuan untuk menganalisa sentimen *review* aplikasi ruangguru. Data penelitian ini diambil dari google

playstore dengan jenis data yang diambil berupa teks dengan jumlah *review* yang diambil sebanyak 2000 *review*. Lalu algoritma klasifikasi yang digunakan adalah algoritma *Support Vector Machine*. Dataset yang telah dikumpulkan akan dibagi menjadi dua jenis yaitu untuk data latih dan data uji dengan pembagian 2 jenis data ini yaitu 60% untuk latih dan 40% untuk data uji. Lalu metode uji algoritmanya menggunakan metode *K-Fold Cross-Validation*. Tingkat akurasi yang didapat dari pengujian *K-Fold Cross-Validation* rata-rata 90% [4].

Salah satu *marketplace* populer di masyarakat salah satunya adalah shopee. Pengguna shopee memberikan keluhan terhadap aplikasi salah satunya menggunakan google playstore. Tanggapan yang sudah diberikan di google playstore ini akan dianalisa sentimen negatif ataupun positifnya. Sehingga sentimen yang telah didapatkan bisa digunakan sebagai sumbangan pemikiran untuk pengembangan peneliti dimasa depan. Lalu algoritma klasifikasi yang digunakan adalah *Naive Bayes*. Maka nilai akurasi dari penelitian yang telah dilakukan ini adalah sebesar 96,667% [5].

Salah satu aplikasi penyedia layanan transportasi *online* yang paling populer di Indonesia adalah Go-Jek. Agar kualitas aplikasi Go-Jek tetap terjaga kualitasnya maka harus bisa menganalisa sentimen yang diberikan pelanggan Go-Jek dan sentimen ini paling banyak diberikan di google playstore lalu kelas sentimen yang digunakan adalah positif, negatif atau netral [6].

Pada saat ini pertumbuhan belanja melalui *E-commerce* terus menunjukkan peningkatan sehingga ini menjadikan sentimen publik terhadap produk atau layanan menjadi penting. Sehingga dibutuhkan proses pemantauan dan pengklasifikasian terhadap sentimen publik tersebut. Tetapi ini bukanlah pekerjaan mudah karena jumlah sentimen sangat banyak jika proses yang dilakukan secara manual. Sehingga peneliti ini melakukan penelitian untuk menyelesaikan masalah ini menggunakan algoritma klasifikasi *Support Vector Machine*. Kemudian klasifikasi dari sentimen ini dibagi menjadi dua jenis yaitu positif dan negatif. Proses uji algoritma menggunakan *Cross Validation* dengan jumlah K yang digunakan sebesar 10. Lalu jumlah ulasan yang digunakan adalah 3000 dengan 1500 menjadi data training dan 1500 lagi akan dijadikan data testing Dan hasil akurasi yang didapat berdasarkan penelitian yang telah dilakukan adalah 90.67% untuk Tokopedia [7].

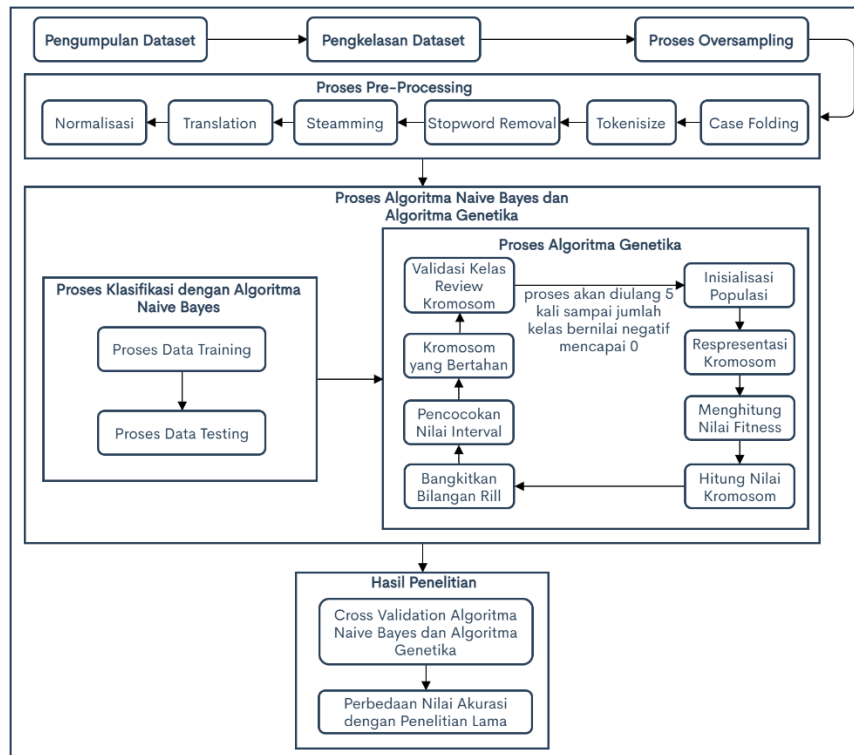
Berdasarkan *review* penelitian yang telah penulis kumpulkan, masih banyak kekurangan dari penelitian-penelitian tersebut, yaitu nilai akurasi yang didapat masih belum diuji dengan baik karena pegujian yang dilakukan rata-rata hanya 1 kali dan rata-rata dataset yang dikumpulkan tidak disebutkan jumlahnya berapa setiap aplikasi. Selain itu, tidak dijelaskan rata-rata jumlah dataset setiap kelas itu berapa sehingga ini menjadi faktor tingkat akurasi yang kurang baik pada beberapa penelitian yang telah *direview*.

Untuk itu penulis akan meningkatkan kualitas penelitian yang telah penulis *review*, agar pengetahuan pada bidang sentimen analisis aplikasi di google playstore bisa lebih meningkat. Untuk algoritma klasifikasi penulis memutuskan untuk menggunakan algoritma *Naive Bayes* karena algoritma *Naive Bayes* mudah digunakan tetapi hasil akurasi yang didapat tetap tinggi dengan syarat dataset sudah dibersihkan secara optimal dengan kelas negatif dan positif dari dataset seimbang dan karena beberapa penelitian yang telah penulis *review* nilai akurasi *Naive Bayes* nya adalah 70-90% maka penulis akan menggunakan algoritma genetika untuk mengoptimasi algoritma *Naive Bayes* sehingga nilai akurasinya bisa jauh lebih baik dari pada penelitian yang sebelumnya. Objek aplikasi yang penulis gunakan adalah Go-Jek, Shopee,

Tokopedia dan Ruangguru, alasan penulis memiliki objek ini agar menyamakan dengan objek penelitian sebelumnya.

METODE

Metode penelitian yang penulis lakukan dibagi menjadi beberapa tahapan seperti yang ditunjukkan oleh Gambar 1.



Gambar 1. Tahapan Penelitian

Penelitian dimulai dari mengumpulkan dataset menggunakan aplikasi web scrapper dari google chrome. Tahap selanjutnya adalah melakukan pengelasan pada masing-masing *review* yang telah dikumpulkan. Proses pengelasan dilakukan secara manual menggunakan excel. Kemudian dilakukan proses oversampling agar jumlah kelas yang telah dikelaskan bisa lebih seimbang sehingga kinerja algoritma *Naive Bayes* bisa lebih optimal. salah satu teknik yang penulis gunakan untuk proses *oversampling* adalah *RWO-Sampling* dengan persamaan yang digunakan sebagai berikut [8] .

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (a_i(j) - \mu_i)^2}, i \in \{1,2,\dots,m\}, j \in \{1,2,\dots,n\} \quad (1)$$

Keterangan :

σ_i = varian ke i

μ_i = rata-rata nilai i

$\sigma_i(j)$ = nilai atribut ke i sampai ke j

n = jumlah sampel

Selanjutnya dilakukan proses *Pre-Processing* yang dibagi menjadi *case folding* yang berfungsi untuk merubah semua kata pada dataset menjadi huruf kecil [9]. *Tokenisize* berguna untuk membagi setiap kata menjadi token. *Stopword removal* berfungsi untuk menghilangkan titik, koma, tanda baca dan juga menghilangkan kata-kata dasar yang tidak memiliki makna [10]. Lalu *stemming* untuk menghilangkan kata imbuhan menjadi kata dasar agar semua kata dasar yang sama tapi memiliki imbuhan berbeda bisa menjadi sama [9]. *Translate* berguna untuk mengganti kata yang berbahasa asing kedalam bahasa indonesia [4]. Dan terakhir adalah normalisasi yang berfungsi untuk menghilangkan *noise* seperti merubah kata tidak baku menjadi baku [11]. Kemudian masuk ketahap klasifikasi menggunakan algoritma *naïve bayes*, hal pertama yang dilakukan adalah mencari nilai *probability* masing-masing kata dataset menggunakan persamaan matematik tertentu seperti pada persamaan 2. [12]

$$P(w_i|c_j) = \frac{\text{count}(w_i, c_j) + 1}{(\sum_{w \in V} \text{count}(w_i, c_j)) + |V|} \quad (2)$$

Keterangan :

$\text{count}(w_i, c_j) + 1$ adalah total jumlah kemunculan kata dalam setiap kelas ditambah nilai 1 untuk menghindari nilai 0.

$(\sum \text{count}(w, j_i))$ merupakan total jumlah kemunculan seluruh kata pada masing-masing kelas.

$|V|$ merupakan total jumlah kata unik yang muncul didalam seluruh kelas.

Setelah nilai *probability* ditemukan maka proses selanjutnya adalah melakukan pencocokan antara kata yang dicari dengan data nilai *probability* yang telah ditemukan, sehingga akan ditemukan hasil akhir berupa kelas klasifikasi. Kemudian tahap selanjutnya adalah memasukan data hasil klasifikasi *naïve bayes* kedalam algoritma genetika, proses algoritma genetika penulis membaginya menjadi 8 proses. Dimulai dengan inialisasi populasi yang berfungsi membentuk gen dan kromosom awal. Lalu masuk keproses representasi kromosom, disini penulis menggunakan Teknik permutasi. Kemudian masuk ketahap menghitung nilai fitness, untuk menghitung nilai fitness, penulis menggunakan 1 proses dari 7 proses untuk menghitung nilai fitness dalam kasus teks mining yaitu menggunakan proses menghitung panjang kata pada setiap kromosom dengan persamaan yang digunakan [13]. Persamaan 3 adalah rumusan matematik perhitungannya.

$$\text{Score} = \frac{\text{Jumlah kata}}{\text{kata unik dalam dokumen}} \quad (3)$$

Setelah nilai fitness pada masing-masing kromosom sudah ditemukan, maka tahap selanjutnya adalah menghitung nilai kromosom proses ini adalah proses menyusun nilai fitness setiap kromosom sehingga terbentuk sebuah susunan nilai interval. Proses selanjutnya adalah membangkitkan bilangan rill menggunakan metode *roda rollate*. Setelah itu melakukan pencocokan antara nilai bilangan rill yang sudah ditemukan dengan nilai interval yang sudah ditemukan. Kemudian diproses selanjutnya dapat diketahui kromosom yang bertahan. Proses selanjutnya adalah melihat kelas kromosom tersebut apakah masih terdapat kromosom yang

bernilai negatif, jika masih ada proses akan diulang dari awal sebanyak 5 kali sampai ditemukan hasil yang maksimal [14].

Tahap selanjutnya adalah masuk kedalam proses *Cross Validation* dengan jumlah K yang digunakan adalah 10. Tahap terakhir adalah melihat perbedaan nilai akurasi dari penelitian lama dengan penelitian yang dikerjakan oleh penulis. Proses untuk mencari perbedaan ini akan dituangkan pada proses Analisa hasil penelitian.

HASIL DAN PEMBAHASAN

Pengumpulan Dataset

Tahap pertama yang dilakukan adalah mengumpulkan dataset dari *review* aplikasi Gojek, Shopee, Tokopedia dan Ruangguru. Lalu Pengumpulan dataset ini menggunakan *tools* yang ada pada *google chrome* yaitu *web scrapper* dengan tanggal pengambilan masing-masing aplikasi akan dibedakan. Tabel 1 menunjukkan hasil serta jumlah dataset yang telah dikumpulkan melalui aplikasi *web scrapper*.

Tabel 1. Hasil dan Jumlah Dataset

No	Aplikasi	Jumlah Dataset	Tanggal Pengambilan Review
1	Gojek	1.231 Review	1 Januari – 1 Maret 2021
2	Shopee	376 Review	1 Maret – 15 Maret 2021
3	Tokopedia	3.378 Review	1 Januari – 1 Maret 2021
4	Ruangguru	2112 Review	1 Desember 2020 – 1 Maret 2021

Pengkelasan Dataset

Setelah berhasil mengumpulkan semua dataset. Maka tahap selanjutnya adalah pengkelasan dataset. Jenis kelas yang digunakan, penulis samakan dengan penelitian sebelumnya yaitu kelas positif dan negatif. Lalu pengkelasan dataset ini penulis melakukan secara manual melalui aplikasi excel. Tabel 2 menunjukkan hasil pengkelasan dataset yang sudah dilakukan.

Tabel 2. Hasil Pengkelasan Dataset

Aplikasi	Jumlah Dataset Sebelum Pengkelasan	Jumlah Dataset Setelah Pengkelasan	Jumlah Dataset Kelas Positif	Jumlah Dataset Kelas Negatif
Gojek	1.231 Review	1.093 Review	793	300
Shopee	376 Review	312 Review	281	31
Tokopedia	3.378 Review	3.097 Review	2.198	899
Ruangguru	2112 Review	2008 Review	1.231	777

Proses Oversampling

Berdasarkan Tabel 2. Jumlah dataset kelas positif jauh lebih banyak dari pada jumlah dataset kelas negatif, hal ini bisa membuat algoritma *Naive Bayes* cenderung mengklasifikasikan kelas positif sehingga proses oversampling dilakukan agar jumlah dataset kelas positif dan negatif menjadi seimbang

Tabel 3. Hasil *Oversampling*

Aplikasi	Jumlah Review Kelas Positif	Jumlah Review Kelas Negatif	Jumlah Review Kelas Negatif Setelah Oversampling	Total Dataset
Gojek	793	300	623	1.416
Shopee	281	31	183	464
Tokopedia	2.198	899	1.676	3.874
Ruangguru	1.231	777	1094	2.325

Proses Pre-Processing

Setelah kelas dataset menjadi seimbang maka tahap selanjutnya adalah melakukan pembersihan terhadap dataset tersebut, proses pembersihan meliputi *case folding*, *tokenize*, *stopword removal*, *stemming*, *translation* dan normalisasi. Tabel 4 menunjukkan hasil *case folding*, Tabel 5 menunjukkan hasil *tokenize*, Tabel 6 menunjukkan hasil *stopword removal*, Tabel 7 menunjukkan hasil *stemming*, Tabel 8 menunjukkan hasil *translation*, dan terakhir Tabel 9 menunjukkan hasil normalisasi.

Tabel 4. Hasil *Case Folding*

Review Sebelum Perubahan	Review Setelah Perubahan
Gojek kok makin lama makin ngeselin yaaa! Aplikasi gk bisa dibuka udh di instal ulang, hapus data, paksa berhenti, tetepp aja begitu tiap hari.	gojek kok makin lama makin ngeselin yaaa! aplikasi gk bisa dibuka udh di instal ulang, hapus data, paksa berhenti, tetepp aja begitu tiap hari.

Tabel 5. Hasil *Tokenize*

Review Sebelum Perubahan	Review Setelah Perubahan
gojek kok makin lama makin ngeselin yaaa! aplikasi gk bisa dibuka udh di instal ulang, hapus data, paksa berhenti, tetepp aja begitu tiap hari.	[gojek][kok][makin][lama][makin][ngeselin][yaaa!] [aplikasi][gk][bisa][dibuka][udh][di][instal][ulang,] [hapus][data,][paksa][berhenti,][tetepp][aja][begitu] [tiap][hari.]

Tabel 6. Hasil *Stopword Removal*

Review Sebelum Perubahan	Review Setelah Perubahan
[gojek][kok][makin][lama][makin][ngeselin][yaaa!][aplikasi][gk][bisa][dibuka][udh][di][instal][ulang,][hapus][data][paksa][berhenti,][tetepp][aja][begitu][tiap][hari.]	[gojek][kok][makin][lama][makin][ngeselin][aplikasi][gk][bisa][dibuka][udh][instal][ulang][hapus][data][paksa][berhenti][tetepp][aja][begitu][tiap][hari]

Tabel 7. Hasil *Stemming*

Review Sebelum Perubahan	Review Setelah Perubahan
[gojek][kok][makin][lama][makin][ngeselin][aplikasi][gk][bisa][dibuka][udh][instal][ulang][hapus][data][paksa][berhenti][tetepp][aja][begitu][tiap][hari]	[gojek][kok][makin][lama][makin][ngeselin][aplikasi][gk][bisa][buka][udh][instal][ulang][hapus][data][paksa][henti][tetepp][aja][begitu][tiap][hari]

Tabel 8. Hasil *Translation*

Review Sebelum Perubahan	Review Setelah Perubahan
[gojek][kok][makin][lama][makin][ngeselin][aplikasi][gk][bisa][buka][udh][instal][ulang][hapus][data][paksa][henti][tetepp][aja][begitu][tiap][hari]	[gojek][kok][makin][lama][makin][ngeselin][aplikasi][gk][bisa][buka][udh][instal][ulang][hapus][data][paksa][henti][tetepp][aja][begitu][tiap][hari]

Tabel 9. Normalisasi

Review Sebelum Perubahan	Review Setelah Perubahan
[gojek][kok][makin][lama][makin][ngeselin][aplikasi][gk][bisa][buka][udh][instal][ulang][hapus][data][paksa][henti][tetepp][aja][begitu][tiap][hari]	[gojek][kok][makin][lama][makin][ngeselin][aplikasi][tidak][bisa][buka][udh][install][ulang][hapus][data][paksa][henti][tetap][aja][begitu][tiap][hari]

Proses Algoritma Naïve Bayes dan Algoritma Genetika

Pada tahap ini memasukan dataset hasil *pre-processing* kedalam proses data training algoritma *naïve bayes* sehingga didapat nilai *probability* masing-masing kelas dataset tersebut. Kemudian nilai *probability* tersebut akan masuk kedalam proses *data testing*. Tabel 10 menunjukkan *sample nilai probability* dataset dan Tabel 11 menunjukkan proses *data testing*.

Tabel 10. *Sample Nilai Probability* Dataset

Nilai <i>Probability</i>				Kelas
maaf 0.02	kakak 0.02	saya 0.02	ingin 0.02	Positif
menanyakan 0.02	kenapa 0.02	gopay 0.02	tidak 0.03	
bisa 0.03	transfer 0.03	aplikasi 0.03	bagus 0.02	
tapi 0.02	sayang 0.02	seluruh 0.02	indonesia 0.02	
barang 0.02	antar 0.02	padahal 0.02	aku 0.02	
maaf 0.01	kakak 0.008	saya 0.03	ingin 0.008	
menanyakan 0.008	kenapa 0.008	gopay 0.008	tidak 0.008	
bisa 0.008	transfer 0.008	aplikasi 0.01	bagus 0.01	
tapi 0.008	sayang 0.008	seluruh 0.008	indonesia 0.008	
barang 0.008	antar 0.008	padahal 0.008	aku 0.008	

Tabel 11. Proses *Data Testing*

kata	bersyukur	tidak	sudah	mudah	kita	Nilai <i>Probabilitas Testing</i>
Positif	1	0,03	0,01	1	1	0.0003
Negatif	1	0,008	0,01	1	1	0.00008
Hasil Klasifikasi						Positif

Maka *data testing* “beryukur tidak sudah mudah kita” diklasifikasi dengan kelas positif, kemudian hasil dari kelas yang telah diklasifikasi ini akan diproses Kembali pada algoritma genetika.

Proses Algoritma Genetika

Hasil klasifikasi dari algoritma *Naive Bayes* tersebut akan masuk kedalam 8 proses algoritma genetika yang direpresentasikan pada Tabel 12 yang merupakan informasi terkait inialisasi populasi, Tabel 13 yang menunjukkan representasi kromosom, Tabel 14 menunjukkan nilai fitness dari masing-masing kromosom, Tabel 15 menunjukkan proses hitung nilai kromosom, Tabel 16 menunjukkan proses pembangkitan bilangan rill, dan Tabel 17 menunjukkan pencocokan nilai interval.

Tabel 12. Inialisasi Populasi

Review	Pembentukan Gen dan Kromosom		
kenapa gojek padahal nomor saya udah terdaftar lama tapi mau masuk pakai nomor kok tidak bisa	Kromosom		
	gen1 kenapa	gen2 gojek	gen3 padahal
	gen4 nomor	gen5 saya	gen6 udah
	gen7 terdaftar	gen8 lama	gen9 tapi
	gen10 mau	gen11 masuk	gen12 pakai
	gen14 nomor	gen15 kok	gen16 tidak
	gen17 bisa		

Tabel 13. Respresentasi Kromosom

Review	Respresentasi Kromosom
kenapa gojek padahal nomor saya udah terdaftar lama tapi mau masuk pakai nomor kok tidak bisa	Kromsom1
saya pesan gofood terjadi kendala masalah aplikasi selesai orderan jadi kembar driver tidak bisa batal	Kromsom2
memudahkan para kaum masyarakat dalam segala hal mulai dari antar jemput pesan makan kirim barang	Kromsom3
bersyukur tidak sudah mudah kita	Kromsom4

Tabel 14. Nilai Fitness Masing-Masing Kromosom

Kromsom1	Kromosom2	Koromsom3	Kromosom4
108	124	115	83

Tabel 15. Proses Hitung Nilai Kromosom

Kromsom1	Kromosom2	Koromsom3	Kromosom4
0,00 – 0,15	0,15 – 0,31	0,31 – 0,78	0,78 - 1

Tabel 16. Pembangkitan Bilangan Rill

Daftar Pembangkitan	Nilai Bilangan Rill
Pembangkitan 1	0,85
Pembangkitan 2	0,24
Pembangkitan 3	0,37
Pembangkitan 4	0,30

Tabel 17. Pencocokan Nilai Interval

kromosom	Nilai Bilangan Rill	Pencocokan Nilai Kromsوم	Hasil kromosom
Kromosom1	0,85	0,78 - 1	Kromosom4
Kromosom2	0,24	0,15 - 0,31	Kromosom2
Kromosom3	0,37	0,31 - 0,78	Kromosom3
Kromosom4	0,30	0,15 - 0,31	Kromosom2

Setelah proses pencocokan nilai. Maka kromosom yang bertahan adalah kromosom2, kromosom3 dan kromosom4. Karena kelas kromosom2 bernilai negatif, maka proses algoritma genetika akan diulang Kembali sampai ditemukan hasil yang optimal. Jika hasil algoritma genetika sudah optimal maka selanjutnya melakukan proses *Cross Validation* dengan hasil yang didapat seperti pada Tabel 18 terkait aplikasi Gojek, Tabel 19 menunjukkan hasil pada aplikasi Shopee, Tabel 20 menunjukkan hasil pada aplikasi tokopedia, Tabel 21 menunjukkan hasil pada aplikasi Ruangguru, dan Tabel 22 menunjukkan hasil Rata-Rata Nilai Keseluruhan.

Tabel 18. Aplikasi Gojek

	Nilai K				Nilai rata-rata
Akurasi	k1	k2	k3	k4	96,54%
	97,32%	93,91%	95,10%	96,12%	
	k5	k6	k7	k8	
	97,05%	94,90%	98,32%	97,56%	
Jumlah prediksi salah	k9	k10			49 review
	97,00%	98,21%			
	k1	k2	k3	k4	
	38 review	86 review	69 review	55 review	
Jumlah prediksi salah	k5	k6	k7	k8	13 review
	42 review	72 review	24 review	35 review	
	k9	k10			
	42 review	25 review			

Tabel 19. Aplikasi Shopee

	Nilai K				Nilai rata-rata
Akurasi	k1	k2	k3	k4	97,15%
	98,20%	97,16%	95,40%	99,12%	
	k5	k6	k7	k8	
	96,65%	97,26%	98,21%	96,32%	
Jumlah prediksi salah	k9	k10			13 review
	98,28%	94,90%			
	k1	k2	k3	k4	
	8 review	13 review	21 review	4 review	
Jumlah prediksi salah	k5	k6	k7	k8	13 review
	16 review	13 review	8 review	17 review	
	k9	k10			
	8 review	24 review			

Tabel 20. Aplikasi Tokopedia

	Nilai K				Nilai rata-rata
	Akurasi	k1	k2	k3	k4
97,00%		98,83%	98,12%	95,00%	
k5		k6	k7	k8	
99,21%		94,94%	97,21%	96,21%	
k9		k10			
	94,15%	98,20%			
Jumlah prediksi salah	k1	k2	k3	k4	121 Review
	116 review	45 review	73 review	194 review	
	k5	k6	k7	k8	
	31 review	196 review	108 review	147 review	
	k9	k10			
	227 review	70 review			

Tabel 21. Aplikasi Ruangguru

	Nilai K				Nilai rata-rata
	Akurasi	k1	k2	k3	k4
95,83%		97,32%	95,20%	93,75%	
k5		k6	k7	k8	
97,02%		98,12%	96,16%	97,44%	
k9		k10			
	98,93%	94,54%			
Jumlah prediksi salah	k1	k2	k3	k4	83 review
	97 review	62 review	112 review	145 review	
	k5	k6	k7	k8	
	69 review	44 review	89 review	60 review	
	k9	k10			
	25 review	127 review			

Tabel 22. Rata-Rata Nilai Keseluruhan

Aplikasi	Jumlah Dataset	Rata-Rata Akurasi	Rata-Rata Jumlah Prediksi Salah
Shopee	464 review	97,15%	13 review (2.85%)
Ruangguru	2.325 Review	96,43%	83 review (3.57%)
Tokopedia	3.874 Review	96,87%	121 review (3.13%)
Gojek	1.416 Review	96,54%	49 review (3.46%)
Total Rata-Rata Jumlah Prediksi Salah Dalam Percent			2.6%

Tabel 23. Perbedaan Penelitian

Penelitian Lama			
Kode Penelitian	Objek Penelitian	Jumlah Dataset	Nilai Akurasi
[5]	Shopee	200 Review	96,667%
[4]	Ruangguru	2.000 Review	90%
[7]	Tokopedia	3.000 Review	90.67%
[6]	Gojek	1.200 Review	89,90%
Penelitian yang Dikerjakan			
Objek Penelitian	Jumlah Dataset	Nilai Akurasi	
Shopee	464 review	97,15%	
Ruangguru	2.325 Review	96,43%	
Tokopedia	3.874 Review	96,87%	
Gojek	1.416 Review	96.54%	

Setelah hasil *Cross Validation* sudah didapatkan maka perbandingan nilai akurasi dari penelitian yang lama adalah seperti pada Tabel 23 yang menunjukkan nilai perbedaan dari tiap penelitian yang ada.

KESIMPULAN

Nilai akurasi dari 4 aplikasi yang penulis gunakan berdasarkan hasil penelitian jauh lebih baik dari pada penelitian sebelumnya dimana nilai akurasi dari 4 aplikasi diatas 90%. Faktor yang membuat nilai akurasi lebih tinggi dari penelitian sebelumnya karena penulis berhasil mereduksi jumlah kata yang bernilai salah pada masing-masing aplikasi menggunakan algoritma genetika dimana rata-rata jumlah kelas *review* dengan prediksi salah seperti hasil yang ditampilkan pada tabel 22 hanya 2.6% dari total jumlah dataset sehingga bisa dikatakan bahwa algoritma genetika mampu meningkatkan kualitas dari algoritma klasifikasi *Naive Bayes*.

DAFTAR PUSTAKA

- [1] V. A. and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 5–15, 2016, doi: 10.5120/ijca2016908625.
- [2] S. Moghaddam, "Beyond sentiment analysis: Mining defects and improvements from customer feedback," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9022, pp. 400–410, 2015, doi: 10.1007/978-3-319-16354-3_44.
- [3] S. Fransiska and A. Irham Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020.
- [4] F. F. Irfani, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, pp. 258–266, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [5] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012043.
- [6] S. Wahyu Handani, D. Intan Surya Saputra, Hasirun, R. Mega Arino, and G. Fiza Asyrofi Ramadhan, "Sentiment analysis for go-jek on google play store," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, 2019, doi: 10.1088/1742-6596/1196/1/012032.
- [7] M. I. Ahmadi, F. Apriani, M. Kurniasari, and ..., "Sentiment Analysis Online Shop on the Play Store Using Method Support Vector Machine (Svm)," *Semin. Nas. ...*, vol. 2020, no. Semnasif, pp. 196–203, 2020.
- [8] W. Ustyannie and S. Suprpto, "Oversampling Method To Handling Imbalanced Datasets Problem in Binary Logistic Regression Algorithm," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 14, no. 1, p. 1, 2020, doi: 10.22146/ijccs.37415.
- [9] T. Mustaqim, K. Umam, and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," *J. Phys. Conf. Ser.*, vol. 1567, no. 3, pp. 0–7, 2020, doi: 10.1088/1742-6596/1567/3/032024.
- [10] V. Maslej-Krešňáková, M. Sarnovský, P. Butka, and K. Machová, "Comparison of deep

- learning models and various text pre-processing techniques for the toxic comments classification,” *Appl. Sci.*, vol. 10, no. 23, pp. 1–26, 2020, doi: 10.3390/app10238631.
- [11] M. Javed and S. Kamal, “Normalization of unstructured and informal text in sentiment analysis,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 10, pp. 78–85, 2018, doi: 10.14569/IJACSA.2018.091011.
- [12] F. Gunawan, M. A. Fauzi, and P. P. Adikara, “Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile),” *Syst. F., Fauzi, M. A., Adikara, P. P. (2017). Anal. Sentimen Pada Ulas. Apl. Mob. Menggunakan Naive Bayes dan Norm. Kata Berbas. Levenshtein Distance (Studi Kasus Apl. BCA Mobile). Syst. Inf. Syst. Inf.*, vol. 3, no. 2, pp. 1–6, 2017, doi: 10.29080/systemic.v3i2.234.
- [13] J. Pardede, D. Rosmala, and P. J. Artaguna, “Implementasi Algoritma Genetika Pada Aplikasi Peringkat Dokumen Berita Bahasa Indonesia.”
- [14] Zainudin Zuhri, 2013, *Algoritma Genetika Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi*, C.V Andi, Yogyakarta



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
