

Klasifikasi Sentimen Opini Pada Aplikasi Gitar Tuner

Christian Moris Dachi^{1*}, Sunneng Sandino Berutu², Jatmika³
 Informatika, Universitas Kristen Immanuel, Yogyakarta, Indonesia
 *e-mail *Corresponding Author.* ctian9133@gmail.com

Abstract

Current technological advances have an impact on various fields, including music applications. Guitar Tuna and other popular music applications can have an impact on public perception, customer decisions, and company reputation. This study analyzes user sentiment for Guitar Tuna using the Naive Bayes and Support Vector Machine (SVM) algorithms. Play Store is the place where data is obtained consisting of 626 neutral reviews, 234 positive reviews, and 2 negative reviews that have gone through several stages, namely preprocessing, classification, labeling, crawling, and evaluation models. The analysis resulted in a Naive Bayes accuracy of 0.95 and SVM of 0.97. The results of the Naive Bayes classification showed the highest value in the neutral sentiment precision parameter of 0.94. In the (SVM) method, positive sentiment gets the highest value at a precision of 1, while the highest recall and f1-score values are achieved by neutral sentiment with values 1 and 0.98 in the SVM method.

Keywords: *Naive Bayes; Support Vector Machine; Sentiment Classification; Opinion Analysis; Guitar Tuna.*

Abstrak

Kemajuan teknologi saat ini berdampak pada berbagai bidang, termasuk aplikasi musik. Gitar Tuna dan aplikasi musik populer lainnya dapat berdampak pada persepsi publik, keputusan pelanggan, dan reputasi perusahaan. Studi ini menganalisis sentimen pengguna untuk Gitar Tuna dengan menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM). Play Store ialah tempat perolehan data terdiri dari 626 ulasan netral, 234 ulasan positif, dan 2 ulasan negatif yang sudah melakukan sejumlah tahapan yakni *preprocessing*, *classification*, *labeling*, *crawling*, serta *evaluation* model. Penganalisisan tersebut menghasilkan akurasi *Naive Bayes* sebesar 0,95 dan SVM sebesar 0,97. Hasil klasifikasi *Naive Bayes* menunjukkan nilai tertinggi pada parameter *precision* sentimen netral sebesar 0,94. Pada metode (SVM), sentimen positif mendapatkan nilai tertinggi pada *precision* sebesar 1, sedangkan nilai tertinggi *recall* dan *f1-score* dicapai sentimen netral dengan nilai 1 dan 0,98 pada metode SVM.

Kata Kunci: *Naive Bayes; Support Vector Machine; Klasifikasi Sentimen; Analisis Opini; Gitar Tuna.*

1. Pendahuluan

Dengan kemajuan teknologi transformasi digital saat ini, analisis sentimen menjadi salah satu bidang yang telah berkembang pesat. Tujuan analisis sentimen ialah guna melihat dan mengkategorikan opini yang terdapat pada teks ulasan[1], apakah itu positif, negatif, atau netral. Penggunaan algoritma pembelajaran mesin misalnya *Naive Bayes* serta *Support Vector Machine* (SVM) sangat penting dalam hal ini. Aplikasi Gitar Tuna sangat populer untuk menyetem alat musik seperti gitar. Pengguna aplikasi ini sering memberikan ulasan atau opini tentang kinerja dan fiturnya. Pengembang dapat menggunakan ulasan ini untuk mendapatkan wawasan yang berharga tentang bagaimana meningkatkan kualitas aplikasi[2].

Untuk sejumlah individu yang telah lama mengenal frekuensi nada dasar serta suara nada, mereka dapat menjalankan menyetem (*tuning*) dengan manual berdasar penggunaan instingnya [3]. Namun, menyetem gitar sangat sulit untuk mendapatkan nada yang tepat karena pendengaran manusia terbatas, terutama bagi pemain gitar pemula. Aplikasi tersebut memiliki banyak keluhan dari pengguna, termasuk fitur yang melakukan *charge* secara tiba-tiba, tuner yang tidak akurat, dan fitur yang terkunci untuk pengguna non-premium. Kadang-Kadang seorang pengguna gitar, terutama pemula, terkadang mengalami kesulitan dalam menemukan

nada yang tepat. Terutama, melakukan setem gitar dengan benar adalah hal yang sulit untuk dilakukan karena harus mendengarkan setiap nada tanpa bantuan aplikasi[4].

Support Vector Machine (SVM) serta *Naïve Bayes* ialah 2 cara yang umum diterapkan dalam pengelompokan teks. *Naïve Bayes* didasarkan pada teori *bayes* dan sangat efektif dalam menangani kumpulan data besar dengan waktu komputasi yang relatif singkat[5]. Sebaliknya, SVM dikenal karena kemampuannya menghasilkan margin klasifikasi minimum, sehingga sangat efektif dalam menangani kumpulan data yang kompleks dan besar[6].

Studi peneliti memiliki tujuan guna mengimplementasikan metode SVM serta *naïve bayes* pada klasifikasi sentimen opini dari pemakai aplikasi Gitar Tuna. Dengan demikian, peneliti mengharapkan bisa mengetahui mana yang lebih efektif metode untuk mengklasifikasikan sentimen positif dan negatif dari ulasan pengguna[7]. Temuan dari studi ini peneliti harapkan bisa berkontribusi bagi pengembang aplikasi guna memahami feedback pengguna dan melakukan perbaikan yang diperlukan[8].

Studi ini memiliki tujuan guna menjalankan klasifikasi sentimen untuk aplikasi gitar tuner berdasarkan resensi pemakai di Google Play dengan menerapkan metode *naïve bayes* serta SVM, mengevaluasi persepsi pengguna terhadap aplikasi gitar tuner berdasar metode SVM serta *naïve bayes*, memberikan rekomendasi kepada pengembang aplikasi terkait hasil analisis sentimen untuk aplikasi gitar tuner untuk menambah atau mengurangi fitur yang ada

2. Tinjauan Pustaka

Ayu Sri Rahayu pada penelitiannya “Komparasi Algoritma *Naïve Bayes* Dan *Support Vector Machine* (SVM) Pada Analisis Sentimen Spotify”[9], komunitas jejaring sosial yang memiliki pendapat dan perasaan yang berbeda memperhatikan aplikasi Spotify. Analisis sentimen memungkinkan Anda secara otomatis mengelompokkan pendapat atau peringkat menjadi pendapat yang dianggap positif atau negatif. Studi tersebut menerapkan *Naïve Bayes* serta SVM. Keunggulannya sederhana, cepat, dan sangat akurat. Sebaliknya, SVM dapat menemukan hyperplane yang berbeda yang memungkinkan jarak antara dua kelas dimaksimalkan. Hasil studi menghasilkan dua label: negatif dan positif. Hasil melihat model uji yang paling cocok untuk klasifikasi sentimen. Untuk mengukur ketepatan, matriks konfusi digunakan. Hasilnya melihat bahwa algoritma SVM memiliki nilai akurasi 84,4%, sedangkan algoritma *Naïve Bayes* memiliki nilai akurasi 86,4%.

Arkiza Ariq pada penelitiannya “Analisis Sentimen Positif Aplikasi GitarTuna Dan Fender Guitar Tuner Dengan Menggunakan Metode Algoritma *Naïve Bayes* Dan *Support Vector Machine*”[10]. Pada pengklasifikasian data, sebelum memasuki tahap text processing, pelabelan dan praproses data diperlukan untuk mengklasifikasi data. Selanjutnya, metode TF-IDF (Term Frequency Inverse Document Frequency) dipilih guna menyajikan berat di tiap kata. Berdasar penerapan metode algoritma SVM serta *Naïve Bayes*, analisis sentimen dijalankan dari memecah data yakni 80% dan 20% dari data latih serta uji. Studi ini menemukan nilai akurasi 92% untuk kedua aplikasi dengan metode *Naïve Bayes* serta nilai akurasi 97% dan 95% untuk kedua aplikasi menggunakan metode SVM. Nilai akurasi dari kedua aplikasi menunjukkan bahwa algoritma SVM bernilai melebihi metode algoritma *Naïve Bayes*.

Menurut Andri Wijaya dkk[11], Pada penelitiannya “Analisis Sentimen Ulasan Aplikasi Shazam Di *Google Play Store* Menggunakan *Support Vector Machine*”. Penganalisisan sentimen berdasar resensi di *Google Play Store* khususnya aplikasi pemakai Shazam dipilih pada studi ini. Ulasan pengguna menunjukkan perasaan pengguna tentang aplikasi, baik positif maupun negatif. Para peneliti membagi ulasan pengguna ke dalam dua kategori sentimen berdasar pada model *Support Vector Machine* (SVM), yaitu negatif dan positif. Hasilnya menunjukkan bahwa SVM dapat memprediksi sentimen ulasan Shazam dengan akurasi 84%.

Cindi Wulandari pada penelitiannya “Analisis Sentimen Aplikasi Spotify Pada Ulasan Pengguna di *Google Play Store* Menggunakan Metode *Support Vector Machine*”[12]. Analisis sentimen ialah studi kasus dari emosi, pendapat, serta perasaan, yang dituangkan pada teks. Karena terdapat banyak resensi yang berbeda, ulasan dikategorikan ke dalam kelas negatif serta positif dengan menerapkan metode SVM. Studi ini mempunyai tujuan guna mengetahui seberapa banyak ulasan positif dan negatif tersebut, sehingga dapat digunakan sebagai referensi untuk meningkatkan aplikasi spotify. Untuk memudahkan, klasifikasi objek didasarkan pada data latihan dan kesamaan objek ataupun menerapkan jarak, berdasar 5000 data resensi terkait di Desember 2023 hingga Januari 2024. Hasil komentar menunjukkan 3193 hasil positif dan 1347 hasil negatif setelah proses *labeling* ke dalam kategori negative serta positif. Temuan uji analisis sentimen

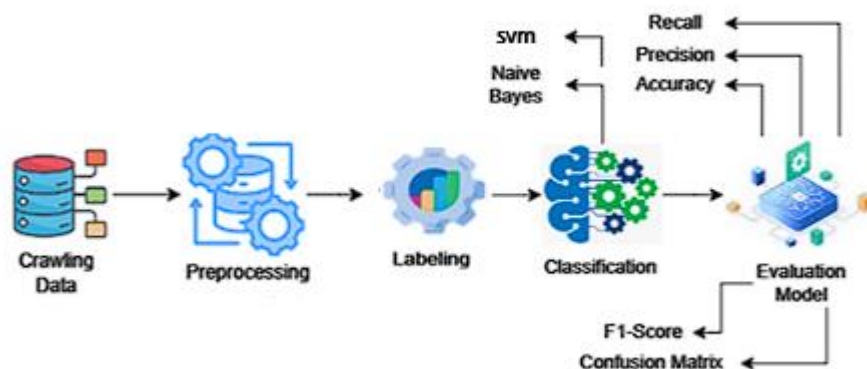
yang menggunakan metode SVM menunjukkan *precision* sebanyak 86%, *f1-score* sebanyak 89%, *recall* sebanyak 92%, serta akurasi sebanyak 85%.

Menurut Andriani Nurian dkk[13]. "Analisis sentimen ulasan Dana di Google play store dilakukan menggunakan metode klasifikasi algoritma *Naïve Bayes*". Hasilnya menunjukkan bahwa sebagian besar pemakai menyumbangkan resensi positif dalam aplikasi Dana berkisar 60%, resensi netral berkisar 28%, serta resensi negatif berkisar 12%. Jumlah terbanyak ialah klasifikasi dengan menerapkan *Naive Bayes Classifier* berbantuan pemilihan lambing TF-IDF, yang mencapai akurasi 85%, *precision* 79%, *recall* 85%, serta *f1-score* 80%. Pengujian dari data yang dipilih ialah 200 data, tepatnya 20% atas 3000 data yang diterapkan berbantuan metode acak sepanjang uji dijalankan.

Studi tersebut tidak sama dengan sejumlah studi terdahulu yang memfokuskan di analisis sentimen pada aplikasi musik. Studi ini menerapkan resensi di Google Play Store dan menerapkan perbandingan metode SVM serta *Naïve Bayes* untuk mengevaluasi pengguna terhadap aplikasi gitar tuner dan menyarakan pengembang aplikasi untuk menambah atau mengurangi fitur yang ada berdasarkan hasil analisis sentimen aplikasi gitar tuner.

3. Metodologi

Permasalahan ini akan mengevaluasi dan meneliti klasifikasi sentimen opini pada aplikasi gitar tuner. Dengan menggunakan pendekatan klasifikasi *naive bayes*, sentimen dapat diklasifikasikan menjadi negative, netral, serta positif. Pemrosesan analisis sentiment serta data dilakukan dengan perpustakaan *textblob*. Data yang dipilih bisa diperoleh atas hasil evaluasi google play store, yaitu bagaimana prosedur penelitian dapat ditentukan.



Gambar 1. Proses Penelitian

3.1 Crawling data

Google Play Store ialah tempat guna memperoleh data, dengan menerapkan data *crawling*/data scraping yang disebut juga data mining, yaitu pengumpulan data atau informasi secara spontan dari sumber data yang terbuka atau publik. Cara seperti ini dapat dipakai untuk memperoleh data yang tidak bisa dikumpulkan melalui cara yang lain, seperti data yang tersimpan di dalam format yang kompleks atau tidak teratur. Goals nya adalah pemotongan data adalah mengumpulkan data yang bisa diterapkan untuk analisis, pemodelan, dan pengambilan tindakan. Data yang digunakan untuk penelitian adalah review aplikasi gitar tuner yang tersaji pada *Google Play Store* di April 2024.

3.2 Preprocessing

Pada titik ini, proses *preprocessing* data dilakukan, yang mencakup membersihkan data, normalisasi, stopwords, tokenisasi, dan stemming, untuk mempersiapkan data guna diterapkan pada processing analisis.

3.2.1 Cleaning

Berdasar titik tersebut dilakukan tahapan pembersihan data pada ulasan content pada play store yang baris nya sama(duplikat) menghapus alamat website seperti ('https'www' url'), dan beserta angka angka, tanda baca dan seperti emoji emoji juga di cleaning

3.2.2 Stopword

Pada titik ini, teks dalam dataframe dinormalisasi untuk mengubah tulisan kata menjadi bahasa Indonesia karena menghilangkan kata sambung dalam bahasa Indonesia. menambahkan kata sambung lain yang ingin dihilangkan, seperti kata "dan", "yang", "lalu", "yg", "gk", "saya", "lagi", "di", dan sebagainya.

3.2.3 Tokenization

Tokenization ialah tahap membagi teks atas sejumlah unit kecil yang dikenal dengan nama "token". Token ini bisa berbentuk simbol, frasa, kata, atau materi penting lainnya. Tokenization merupakan langkah awal yang penting dalam pemrosesan teks alami (Natural Language Processing, NLP) dan sangat krusial untuk berbagai aplikasi NLP, seperti analisis sentimen, pemahaman bahasa, dan ekstraksi informasi[14].

3.2.4 Stemming

Stemming ialah tahap yang merubah sejumlah kata bentuk atau terinfleksi turunannya menjadi bentuk akar atau dasar katanya. Stemming mempunyai tujuan guna memangkas variasi beberapa kata berbeda yang bermakna inti yang tidak berbeda, sehingga analisis teks bisa lebih sederhana dan efisien. Dalam memproses teks alami (Natural Language Processing, NLP), stemming dipakai untuk menyederhanakan data dan meningkatkan akurasi dalam berbagai aplikasi, seperti pencarian informasi, analisis sentimen, dan klasifikasi teks[15].

3.2.5 Case Folding

Sebuah tahap standar yang merubah semua karakter pada suatu teks sebagai bentuk yang konsisten, umumnya menjadi huruf kecil (*lowercase*). Langkah ini merupakan yang utama pada tahapan teks alami (Natural Language Processing, NLP) karena membantu mengurangi variasi teks akibat perbedaan kapitalisasi. Sebagai contoh, kata "Apple" dan "apple" akan dianggap sama setelah melalui proses case folding[16].

3.3 Labeling

Proses pelabelan di analisis sentimen dengan library *TextBlob*, guna melanjutkan analisis tersebut dalam aplikasi tuner gitar, diantaranya yakni:

- 1) Apabila nilai polaritas sentimen melebihi ($>$) 0 maka bertanda positif label sentimennya.
- 2) Apabila nilai polaritas emosi sama dengan ($=$) 0, maka bersifat netral label sentimennya.
- 3) Apabila nilai polaritas emosi kurang dari ($<$) 0 maka bersifat negatif label sentimennya.

Didalam metode tersebut, *TextBlob* dipilih guna mendapatkan nilai polaritas yang memperlihatkan apakah teks yang ada negatif, netral atau positif. Nilai polaritas ini selanjutnya diterapkan guna melihat label sentimen negatif, netral, atau positif, selaras menurut tahapan yang diuraikan sebelumnya.

3.4 Training Model

Metode klasifikasi Naïve Bayes ialah algoritma pembelajaran mesin yang didasarkan teorema Bayes. Teori tersebut dipilih guna probabilitas sebuah anggapan dasar menurut data atau bukti yang ada. Perbandingan probabilitas yang digunakan membantu dalam menentukan kemungkinan suatu hipotesis benar[17].

$$p(x/y) = \frac{p(y|x) \times p(x)}{p(y)} \quad \dots(1)$$

Dimana:
X= label
Y= fitur

3.5 Testing dan Evaluasi Model

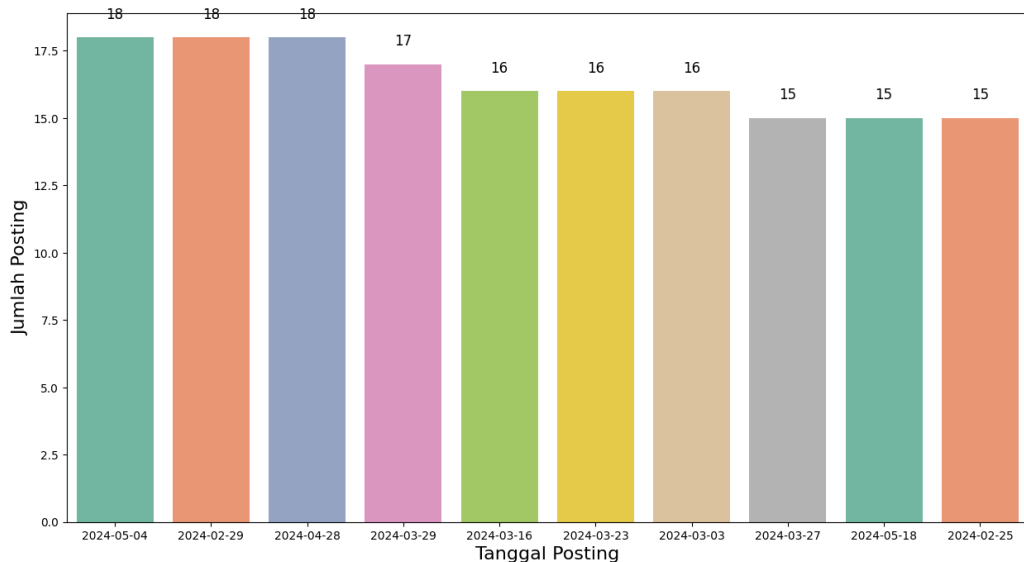
Model akan dinilai menggunakan cara memasukkan data baru. Kinerja model selama proses pengujian diukur menggunakan pendekatan *confusion matrix*. Pendekatan ini [18] harusnya akan menyajikan sejumlah parameter penting misalnya *recall*, *f1-score*, *precision*, serta akurasi.

4. Hasil dan Pembahasan

	date	time	userName	score	content
0	2024-05-20	16:04:13	Rachel Etougou	5	Recommended
1	2024-05-20	14:50:47	Julius Randik	4	Ok, tapi untuk lanjut audio training berbayar..
2	2024-05-20	06:59:58	Salim Pratama	5	Sangat membantu
3	2024-05-20	01:29:25	Wayan Darmawan	5	Good and easy to use 🙏
4	2024-05-19	16:30:04	Calista Sutarto 212	5	Top
5	2024-05-19	09:20:25	Ahmad Ardi	5	Good
6	2024-05-19	08:34:45	Rehan The big	3	Bagus
7	2024-05-19	08:30:44	Restutinengsih Restuti	5	Aplikasi ini sangat membantu bagi pemula Saya ...
8	2024-05-19	08:13:32	Bm 86	5	membantu
9	2024-05-19	05:01:34	TianbestdionZai	5	Top.. keren banget
10	2024-05-19	03:59:13	Taufik Rohidayat	5	Keren gak ada duanya
11	2024-05-18	19:04:13	Dandy Ginting (Sales Daihatsu)	5	Ini bagus
12	2024-05-18	15:36:17	Hans Naomi Wamaer	5	aplikasinya bagus, sangat membantu saya belaja...
13	2024-05-18	15:04:27	Iqka Official	5	Josss 🙏

Gambar 2. Hasil Crawling Data

Sebelum melakukan *Preprocessing* data berikut adalah data jumlah postingan per tanggal dari 2024-02-25 hingga sampai dengan 2024-05-04 yang bisa dilihat pada Gambar.

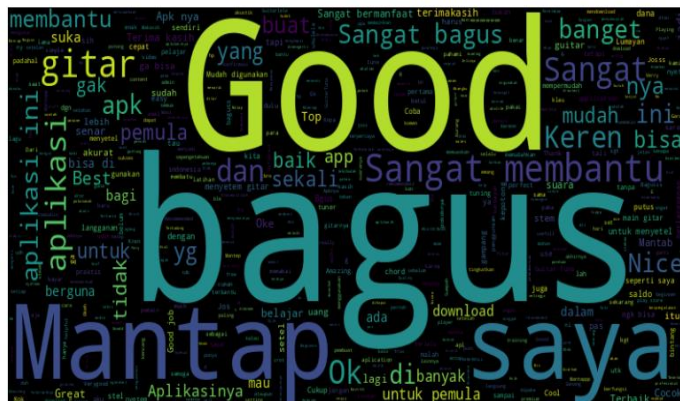


Gambar 3. Jumlah postingan per tanggal

Dataset jumlah postingan per tanggal diatas bisa dilihat bahwa terdapat jumlah postingan yang paling banyak dan paling kecil, seperti bisa kita lihat jumlah postingan terbanyak berjumlah 18 data pada tanggal 2024-05-04 sedangkan postingan yang sedikit yaitu 15 terdapat pada tanggal 2024-02-25.

4.2 Hasil Preprocessing Data

Sebelum memproses data, kita dapat memvisualisasikan kata dalam bentuk wordcloud. Gambar 4 di bawah ini menunjukkan perbedaan gambar visual terkait beberapa kata yang sering dilihat, sedangkan plot bisa menyajikan penjelasan terkait penyaluran frekuensi kata.



Gambar 4. Worldcloud

Gambar diatas merupakan visualisasi kata kedalam bentuk worldcloud, bisa dilihat bahwa kata yang umumnya memiliki tingkat kemunculan tinggi ialah huruf yang berukuran paling besar karena kata ini digunakan paling banyak pada ulasan, serta huruf yang paling kecil adalah kata yang sangat sedikit digunakan.

4.2.1 Hasil Cleaning

Pada tahap ini diterapkan aktivitas mengecek mutu data menggunakan strategi merubah atau data yang kurang akurat dihapus misalnya (data duplikat, hastag, mention, tanda baca, data kosong, emoticon, hyperlink, simbol, atau hasil data dengan nilai null/NaN). Data bagian ini bisa di ketahui di Gambar 5. dibawah ini.

content	cleaning
Recommended	Recommended
Ok, tapi untuk lanjut audio training berbayar..	Ok tapi untuk lanjut audio training berbayar
Sangat membantu	Sangat membantu
Good and easy to use 🙌	Good and easy to use
Top	Top
Good	Good
Bagus	Bagus
Aplikasi ini sangat membantu bagi pemula Saya ...	Aplikasi ini sangat membantu bagi pemula Saya ...
membatu	membatu
Top.. keren banget	Top keren banget

Gambar 5. Hasil Screenshot proses cleaning

4.2.2 Hasil Tokenization

Tokenization ialah tahap membagi teks atas sejumlah unit kecil yang dikenal dengan nama "token". Token ini bisa berbentuk simbol, frasa, kata, atau materi penting lainnya. Tokenization merupakan langkah awal yang penting dalam pemrosesan teks alami (Natural Language Processing, NLP) dan sangat krusial untuk berbagai aplikasi NLP, seperti analisis sentimen, pemahaman bahasa, dan ekstraksi informasi bisa dilihat Gambar 6.

content	tokenize
Recommended	[recommended]
Ok, tapi untuk lanjut audio training berbayar..	[ok, tapi, untuk, lanjut, audio, training, ber...]
Sangat membantu	[sangat, membantu]
Good and easy to use 🙌	[good, and, easy, to, use]
Top	[top]
Good	[good]
Bagus	[bagus]
Aplikasi ini sangat membantu bagi pemula Saya ...	[aplikasi, ini, sangat, membantu, bagi, pemula...]

Gambar 6. Hasil Screenshot proses tokenize

4.2.3 Hasil Case Folding

Pada bagian ini disebut proses yang mengubah huruf ke bentuk huruf kecil (lowercase)[19], hanya huruf dari "a" hingga "z" yang dapat digunakan. Karakter lainnya yakni huruf disisihkan serta diterapkan sebagai *delimiter* (pembatas)[20], hal ini sangat penting dalam pengolahan teks, terutama dalam pemrograman dan pengolahan bahasa alami contoh penggunaan case folding bisa diketahui di gambar 7.

content	cleaning	case_folding
Recommended	Recommended	recommended
Ok, tapi untuk lanjut audio training berbayar..	Ok tapi untuk lanjut audio training berbayar	ok tapi untuk lanjut audio training berbayar
Sangat membantu	Sangat membantu	sangat membantu
Good and easy to use 🙌	Good and easy to use	good and easy to use
Top	Top	top

Gambar 7. Hasil Screenshot proses casefolding

4.2.4 Hasil Stopwords

Stopwords suatu tahap penghapusan kata yang tidak begitu penting atau mungkin kurang berarti pada analisis sentimen data ini. Misalnya stopwords seperti kedalam bahasa Indonesia yakni sejumlah kata diantaranya "dan", "yg", "lalu", "yang", "gk", "saya", "lagi", "di", "ini". Hasil penghilangan stopwords bisa dilihat Gambar 8. dibawah ini

content	cleaning	case_folding	tokenize	Filtering/stopword removal
Recommended	Recommended	recommended	[recommended]	[recommended]
Ok, tapi untuk lanjut audio training berbayar..	Ok tapi untuk lanjut audio training berbayar	ok tapi untuk lanjut audio training berbayar	[ok, tapi, untuk, lanjut, audio, training, ber...]	[ok, audio, training, berbayar]
Sangat membantu	Sangat membantu	sangat membantu	[sangat, membantu]	[membantu]
Good and easy to use 🙌	Good and easy to use	good and easy to use	[good, and, easy, to, use]	[good, and, easy, to, use]
Top	Top	top	[top]	[top]
Good	Good	good	[good]	[good]
Bagus	Bagus	bagus	[bagus]	[bagus]
Aplikasi ini sangat membantu bagi pemula Saya ...	Aplikasi ini sangat membantu bagi pemula Saya ...	aplikasi ini sangat membantu bagi pemula saya ...	[aplikasi, ini, sangat, membantu, bagi, pemula...]	[aplikasi, membantu, pemula, bangga, aplikasi]

Gambar 8. Hasil Screenshot proses stopwords

4.2.5 Hasil Stemming

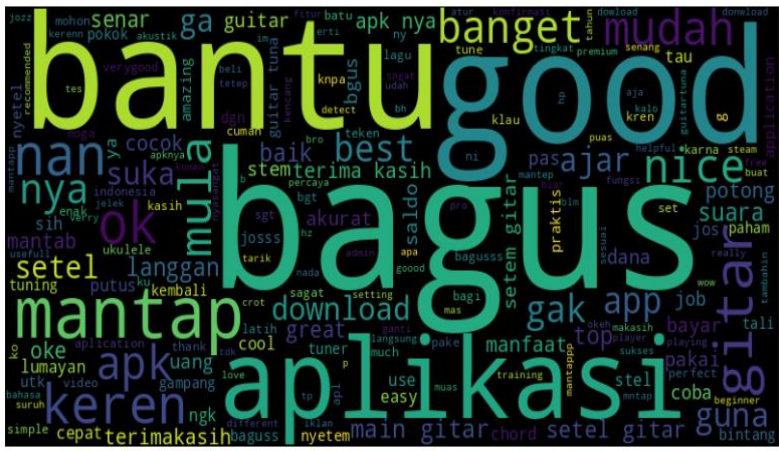
Hasil Stemming di ini yakni awal dan akhir kalimat dengan bentuk lain seperti meniggalkan kata dasar saja. Contohnya “Berguna” – “Guna” disederhanakan kata nya menjadi bentuk dasarnya bisa dilihat Gambar 9. Dibawah ini

content	cleaning	case_folding	tokenize	Filtering/stopword removal	steaming_data
Recommended	Recommended	recommended	[recommended]	[recommended]	recommended
Ok, tapi untuk lanjut audio training berbayar..	Ok tapi untuk lanjut audio training berbayar	ok tapi untuk lanjut audio training berbayar	[ok, tapi, untuk, lanjut, audio, training, ber...]	[ok, audio, training, berbayar]	ok audio training bayar
Sangat membantu	Sangat membantu	sangat membantu	[sangat, membantu]	[membantu]	bantu
Good and easy to use 🙌	Good and easy to use	good and easy to use	[good, and, easy, to, use]	[good, and, easy, to, use]	good and easy to use
Top	Top	top	[top]	[top]	top
Good	Good	good	[good]	[good]	good
Bagus	Bagus	bagus	[bagus]	[bagus]	bagus
Aplikasi ini sangat membantu bagi pemula Saya ...	Aplikasi ini sangat membantu bagi pemula Saya ...	aplikasi ini sangat membantu bagi pemula saya ...	[aplikasi, ini, sangat, membantu, bagi, pemula...]	[aplikasi, membantu, pemula, bangga, aplikasi]	aplikasi bantu mula bangga aplikasi

Gambar 9. Hasil Screenshot proses steaming

Setelah proses *preprocessing* misalnya tokenisasi, membenah data, stemming, serta menghilangkan stopwords, kita bisa mengetahui sejauh mana umumnya sejumlah kata muncul pada teks dengan menggunakan visualisasi seperti word cloud dan plot. Berdasar penerapan

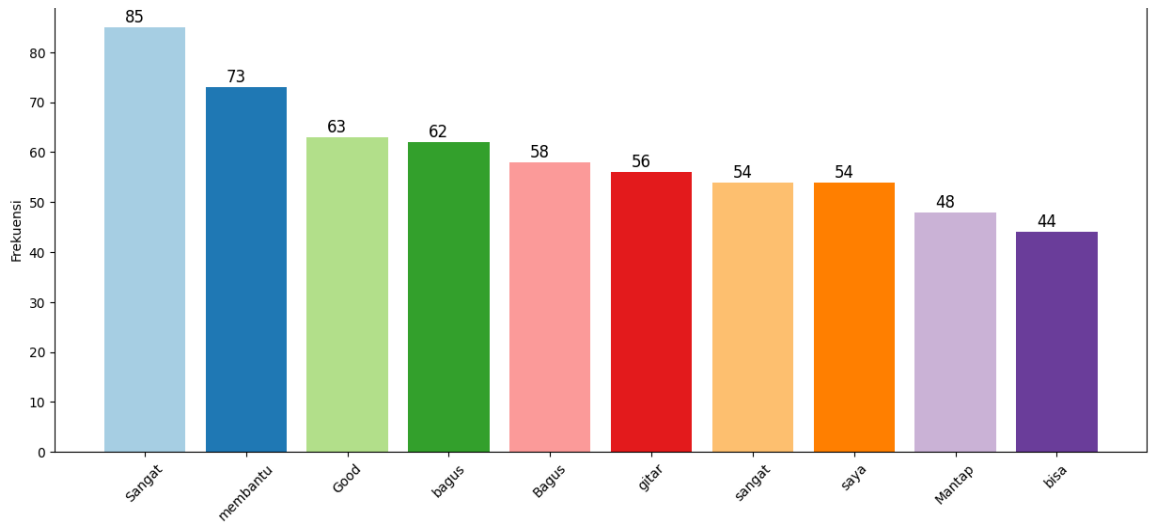
gambaran tersebut, Anda bisa secara gampang mengetahui tren serta pola sejumlah kata yang paling umum. Gambar 10. Dibawah ini



Gambar 10. Worldcloud setelah melewati data preprocessing

Mengikuti langkah-langkah pra-pemrosesan seperti pembersihan data seperti gambar diatas, *wordcloud* sebelumnya di proses dan dihilangkan kata kata sambung nya seperti 'yang', 'yg', 'ini', 'dan', didalam *wordcloud*, setelah dihilangkan kata sambungnya disitu juga terdapat hasil *crawling* dari *dataset* aplikasi gitar tuner, terdapat sebuah kata yang umum banyak digunakan akan mempunyai visualisasi berukuran lebih tinggi dan jika sejumlah kata umum yang sedikit digunakan hasil visualisasi akan terlihat kecil, sementara itu plot dapat memberikan penjelasan yang lebih detail terkait penyaluran frekuensi kata.

Setelah mengikuti sedemikian rupa pembersihan data ada juga terdapat kata-kata yang sering muncul atau sering digunakan dan akan ditampilkan kedalam bentuk diagram batang seperti Gambar 11. Dibawah ini.

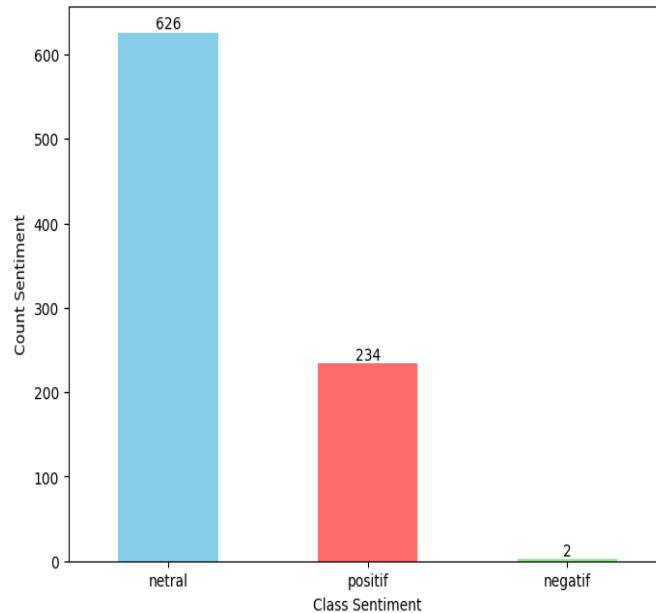


Gambar 11. Frekuensi kemunculan kata

Seperti gambar diatas terdapat kata yang sering muncul dan ada juga terdapat kata yang sedikit digunakan orang gunakan dalam melakukan komentar terhadap komentar yang ada di aplikasi play store yaitu terdapat kata yang nilai nya tinggi yaitu kata 'Sangat' dengan nilai 85 sedangkan kata paling sedikit adalah kata 'Bisa' dengan nilai 44.

4.2.5 Implementasi TextBlob

Setelah melalui tahap pembersihan data diatas hasil klasifikasi pada data sentimen dari data yang diuraikan kedalam bentuk diagram, implementasi *textblob* juga menampilkan hasil kedalam bentuk diagram yang bisa diketahui di gambar 12.

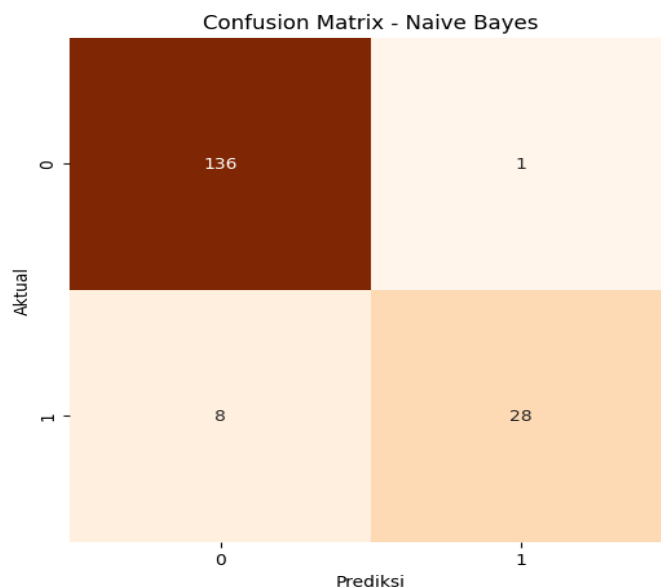


Gambar 12. Diagram data ulasan

Terlihat dari diagram diatas bahwa terdapat nilai paling tertinggi dan terendah dalam sentimen terhadap aplikasi gitar tuner bisa dilihat data terbanyak dari 2000 data, yaitu nilai yang presentase nilai netral lebih besar yaitu 626 dibandingkan dengan presentase positif dan negatif.

4.3 Implementasi Naïve Bayes Dan SVM

Hasil analisis latih *textblob* digunakan sebagai dataset pada model *naïve bayes* berhubung data dengan label negative hanya 2 maka data yang ada di dataset dibagi kedalam data pelatihan dan data penguji [21] dengan perbandingan 80 : 20. Temuan uji ini dapat digambarkan dengan confusion matrix di gambar 13.



Gambar 13. Confusion Matrix Naïve bayes

Berdasar gambar diatas label 0 merupakan sentimen netral dan label 1 merupakan sentimen positif, jumlah ulasan sentimen netral yang diinput pada pengujian sebanyak 137. Hasil pengujian menunjukkan bahwa sebanyak 136 tepat diprediksi sebagai netral dan sisanya sebagai positif. Sementara itu, jumlah sentimen positif sebanyak 36. Hasil menunjukkan jumlah sentimen netral memperoleh sebanyak 8 ulasan dan positif sebanyak 28 ulasan.

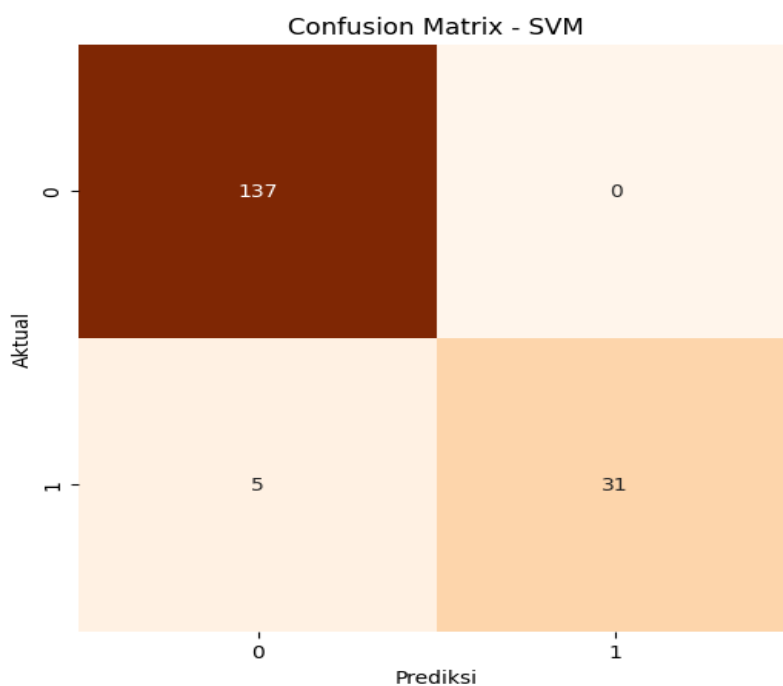
Hasil confusion matrix diatas menghasilkan nilai parameter yang dideskripsikan di tabel 1:

Tabel 1. *Naïve Bayes*

Label	Parameter			Support
	Precision	Recall	F1-score	
Netral	0.94	0.99	0.97	137
Positif	0.97	0.78	0.86	36

Berdasarkan tabel, sentimen positif memperoleh nilai *precision* tertinggi yaitu 0.97 dibandingkan dengan sentimen netral dengan nilai 0.94. Sementara itu nilai parameter *recall* dan *f1-score* memperoleh nilai tertinggi pada label netral dengan nilai masing-masing 0.99 dan 0.97. dan nilai terendahnya dengan masing-masing nilai 0.78 dan 0.86.

Berdasarkan table diatas pada pengujian metode SVM confusion matrix tertera di gambar 14.



Gambar 14. *Confusion matrix SVM*

Pada gambar diatas label 0 merupakan sentimen netral dan label 1 merupakan sentimen positif, Jumlah ulasan sentimen netral yang diinput pada pengujian sebanyak 137. Hasil pengujian menunjukkan bahwa sebanyak 137 diprediksi sebagai netral dan nilai positif nya 0. Sedangkan, jumlah sentimen positif sebanyak 36. Hasil menunjukkan jumlah sentimen memperoleh netral sebanyak 5 ulasan dan nilai positif 31 ulasan.

Hasil *confusion matrix* pada gambar 14 menghasilkan nilai parameter yang dideskripsikan di tabel 2:

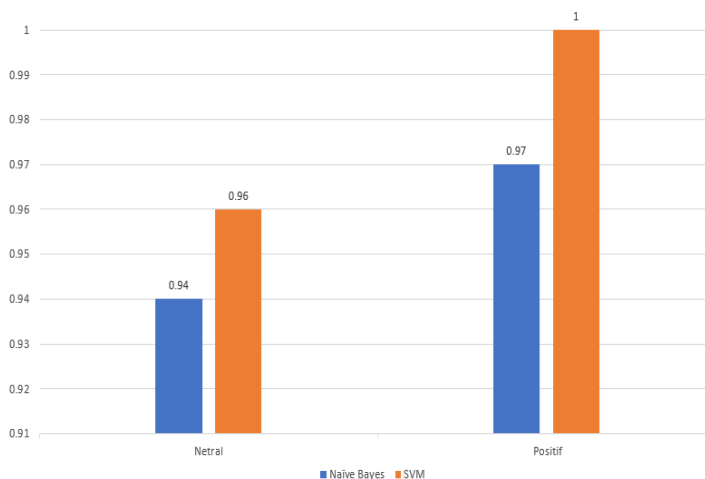
Tabel 2. *Confusion Matrix SVM*

Label	Parameter			Support
	Precision	Recall	F1-score	
Netral	0.96	1	0.98	137
Positif	1	0.86	0.93	36

Tabel diatas menunjukkan kinerja *support vector machine* (SVM) pada *precision*, sentimen positif bernilai terbanyak sebesar 1. kemudian, di parameter *recall* serta *f1-score* sentimen netral bernilai terbanyak masing-masing sebesar 1 dan 0.98. dan pada parameter *recall* dan *f1-score* memperoleh nilai terendah nya pada sentimen positif sebesar 0.86 dan 0.93.

4.3 Perbandingan Kinerja Algoritma Naïve bayes dan SVM

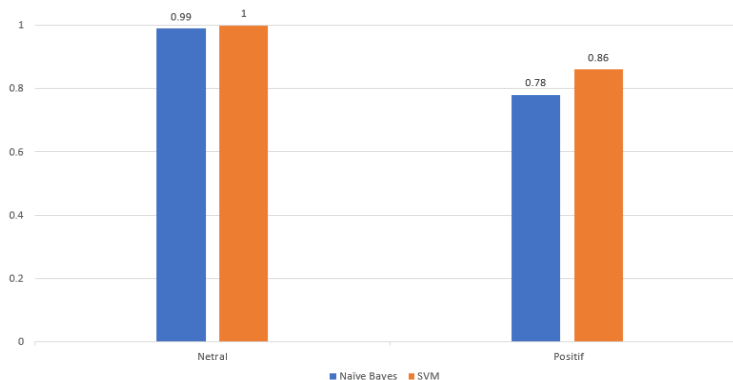
Guna melihat sejauh mana kinerja model pada *Naïve Bayes* dan SVM maka dilakukan perbandingan berdasarkan parameter *precision* yang ditunjukkan pada gambar.



Gambar 15. *Perbandingan model precision naïve bayes dan SVM*

Berdasarkan gambar diatas kinerja model *naïve bayes* dan SVM pada parameter *precision* menunjukkan model SVM memperoleh nilai tertinggi pada sentimen netral dan positif dengan nilai masing-masing sebesar 0.96 dan 1.

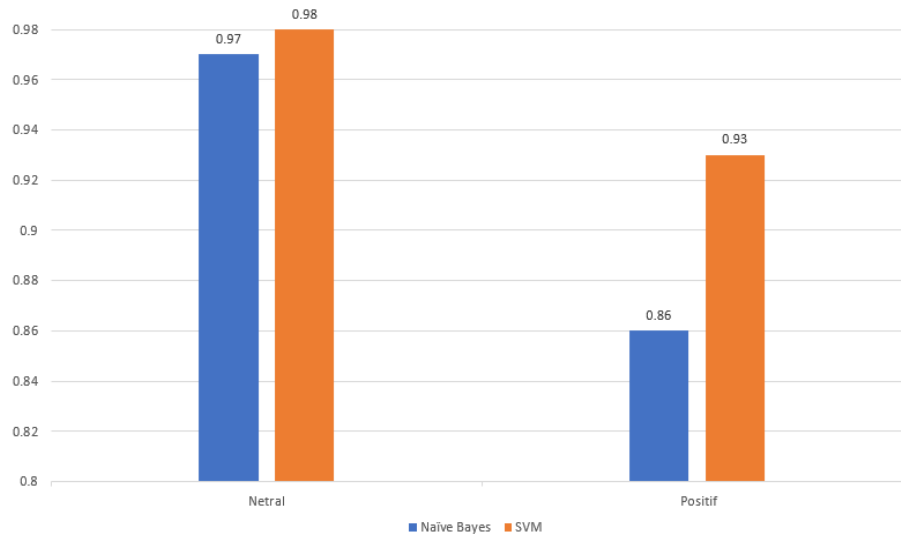
Pada parameter *recall* untuk mengukur kinerja model SVM serta *naïve bayes* ditunjukkan di gambar.



Gambar 16. *Perbandingan model recall naïve bayes dan SVM*

Berdasarkan gambar diatas kinerja model *naïve bayes* dan SVM pada parameter *recall* menunjukkan model SVM memperoleh nilai tertinggi pada sentimen netral dan positif dengan nilai masing-masing sebesar 1 dan 0.86.

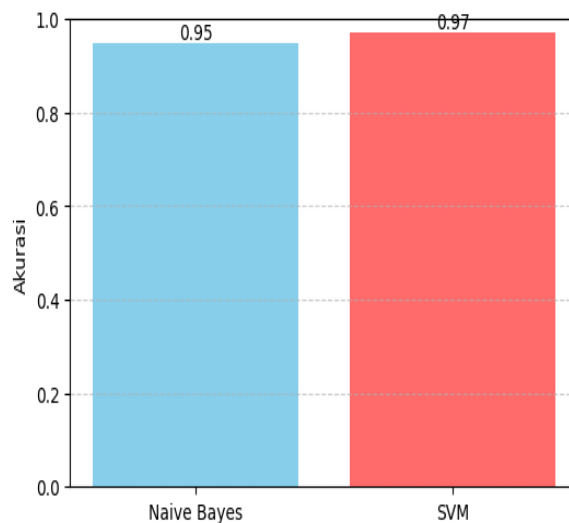
Pada parameter *f1-score* untuk mengukur kinerja model *naïve bayes* dan SVM ditunjukkan pada gambar.



Gambar 17. Perbandingan model *f1-score* *naïve bayes* dan SVM

Berdasarkan gambar diatas kinerja model *naïve bayes* dan SVM pada parameter *f1-score* menunjukkan model SVM memperoleh nilai tertinggi pada sentimen netral dan positif bernilai di setiap bagian sebanyak 0.98 serta 0.93.

Pada parameter akurasi perbandingan kinerja model *naïve bayes* dan SVM ditampilkan pada gambar 18. berikut



Gambar 18. Perbandingan *Naïve Bayes* dan SVM

Gambar diatas menunjukkan model SVM memperoleh nilai akurasi tertinggi sebesar 0.97. Secara umum, model SVM mempunyai kinerja lebih bagus berbeda pada *naïve bayes*.

Berdasarkan *confusion matrix Naïve Bayes* pada gambar 13. Jumlah ulasan sentimen netral yang diinput pada pengujian sebanyak 137. Menunjukkan bahwa sebanyak 136 diprediksi sebagai netral dan sebagai positif sebanyak 36. Berdasarkan table .1 sentimen positif memperoleh nilai *precision* tertinggi yaitu sebesar 0.97 dibandingkan dengan netral dengan nilai 0.94. Sementara itu nilai parameter *recall* dan *f1-score* memperoleh nilai tertinggi pada label netral dengan nilai 0.99 dan 0.97. dan nilai terendahnya dengan nilai masing masing 0.78 dan 0.86. Kemudian pada gambar *confusion matrix SVM* pada gambar 14. Jumlah ulasan sentimen netral yang diinput pada pengujian sebanyak 137. Hasil menunjukkan bahwa sebanyak 137 diprediksi sebagai nilai netral dan positif nya 0. Pada table 2. Sentimen di parameter *precision* sentimen positif bernilai terbanyak sebesar 1. Sementara itu, di parameter *f1-score* serta *recall* sentimen netral bernilai terbanyak masing masing sebesar 1 dan 0.98. dan pada parameter *recall* dan *f1-score* memperoleh nilai terendah pada sentimen positif sebesar 0.86 dan 0.93.

Sistem penilaian *confusion matrix* menggunakan metode SVM dengan rentang nilai diatas dengan metode *naïve bayes* dan yang berarti bahwa hasil penelitian ini dianggap memiliki kualitas yang sangat baik dibandingkan dengan metode *naïve bayes*. Penelitian ini memadukan metode klasik pembelajaran mesin dengan aplikasi modern dalam pemrosesan teks untuk menciptakan model klasifikasi sentimen yang handal. Penggunaan aplikasi Gitar Tuna sebagai studi kasus memberikan konteks yang relevan dan praktis untuk analisis ini.

5. Kesimpulan

Metode klasifikasi *naive bayes* digunakan dalam penelitian ini untuk mengkaji pendapat sejumlah orang terkait aplikasi gitar tuner. Perolehan data didapat pada Google Play Store dan diolah dengan menerapkan *textblob library* yang meliputi beberapa tahap *pra-pemrosesan* misalnya pembersihan data, penghilangan sejumlah kata yang *stemming*, tidak diperlukan (*stopwords*), serta tokenisasi. Hasil analisis sentimen dengan *textblob* menunjukkan jumlah sentimen netral sebanyak 626 ulasan, positif 234 ulasan dan negatif 2 ulasan. Dalam evaluasi model dengan metode *naïve bayes*. Hasil analisis dengan jumlah ulasan sentimen netral yang diinput pada pengujian sebanyak 137. Menunjukkan bahwa sebanyak 136 diprediksi sebagai netral dan sebagai positif sebanyak 36. Sementara itu dalam evaluasi model dengan metode SVM, dengan jumlah ulasan netral yang diinput pada pengujian sebanyak 137. Hasil menunjukkan bahwa sebanyak 137 diprediksi sebagai nilai netral dan positif nya 0. Kemudian hasil pengujian model SVM menunjukkan nilai tertinggi pada *precision* label positif sebesar 1, *recall* pada label netral sebesar 1 dan *f1-score* pada label netral sebanyak 0.98. Sementara itu pada model *naïve bayes* *precision* memperoleh nilai tertinggi pada label positif, sebanyak 0,97, *recall* pada label netral sebanyak 0.99 serta *f1-score* pada label netral sebanyak 0.97. Selanjutnya temuan perbandingan kedua model menunjukkan SVM mendapat nilai terbanyak di parameter *precision* akurasi dan *recall f1-score*. Diharapkan dapat membantu pengembang aplikasi dalam memahami feedback pengguna dan melakukan perbaikan yang diperlukan.

Untuk penelitian selanjutnya analisis lexicon vader (Sentimen Reasoner and Valence Aware Dictionary) ialah instrument analisis sentimen yang dirancang khusus guna menganalisis teks media sosial akan diterapkan pada data ulasan ini.

Daftar Referensi

- [1] L. Pang and S. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends® Inf. Retrieval*, vol. 2, no. 1, pp. 1-135, 2008.
- [2] E. K. and S.-M. Hovy, "Determining the sentiment of opinions," *Proc. 20th Int. Conf. Comput. Linguist. (COLING '04)*, Stroudsburg, PA, USA Assoc. Comput. Linguist. pp. 1367., 2004.
- [3] D.M. Priatama, "Perancangan Dan Implementasi Tuner Gitar Berbasis Fast Fourier Transform Pada Platform Ios," *e-Proceeding Eng. Vol.3, No.2 Agustus*, 2016.
- [4] D. T. L. Aah Sumiah, "Rancang Bangun Aplikasi Sinkronisasi Nada Pada Tuning Gitar Dengan Menggunakan Algoritma Fast Fourier Transform Berbasis Android," *JEJARING (Jurnal Teknol. dan Manaj. Inform. Vol. 6 Nomor 1,p-ISSN 2527-3930, e-ISSN 2614-5448 https://journal.uniku.ac.id/index.php/jejaring*, 2021.
- [5] F. E. Wibisono, R., & Gunawan, "Penerapan Algoritma Naïve Bayes dan SVM pada Klasifikasi Spam Email.," *J. Ilmu Komput. dan Informasi*, 13(3), 145-152., 2020.
- [6] D. Susanto, A. B., & Anggraini, "Implementasi Naïve Bayes dan SVM untuk Klasifikasi Review Produk E-Commerce.," *J. Teknol. Inf. dan Komunikasi*, 9(2), 120-128., 2018.
- [7] A. Pratama, I. D., & Widodo, "Penggunaan Metode Naïve Bayes dan SVM dalam Analisis Sentimen Pelanggan pada Aplikasi Transportasi Online.," *J. Rekayasa Teknol. Informasi*, 14(1), 65-72., 2021.
- [8] N. F. Hidayah, K. P. Kartika R., and S. N. Budiman, "Penerapan Metode Naive Bayes Dalam Analisis Sentimen Aplikasi Sentuh Tanahku Pada Google Play," *JATI (Jurnal Mhs. Tek. Inform., vol. 6, no. 2, pp. 679–683, 2022, doi 10.36040/jati.v6i2.5610.*, 2022.
- [9] A. S. Rahayu, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform. Hal 349–354 Vol. 4, Nomor 2, e-ISSN 2685-998X DOI 10.30865/json.v4i2.5398*, 2022.
- [10] A. Ariq, "Analisis Sentimen Positif Aplikasi GitarTuna Dan Fender Gitar Tuner Dengan Menggunakan Metode Algoritma Naïve Bayes Dan Support Vector Machine," *UPN "Veteran" Jakarta, Fak. Ilmu Komputer, S1 Sist. Inf. [www.upnvj.ac.id - www.library.upnvj.ac.id - www.repository.upnvj.ac.id]*, 2024.
- [11] M. M. A. Wijaya, Meilinda, "Analisis Sentimen Ulasan Aplikasi Shazam Di Google Play Store Menggunakan Support Vector Machine," *Zo. J. Sist. Informasi*, Vol. 6No. 1,page 197-207, 2024.
- [12] C. Wulandari, "Analisis Sentimen Aplikasi Spotify Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Support Vector Machine," *KLIK Kaji. Ilm. Inform. dan Komputer*ISSN 2723-3898 (Media Online)Vol 4, No 5, April. Hal 2588-2595DOI 10.30865/klik.v4i5.1762 <https://djournals.com/klik>, 2024.
- [13] B. N. S. A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naive Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter. Vol.11 No.3 S1, pISSN2303-0577 eISSN2830-7062 http://dx.doi.org/10.23960/jitet.v11i3%20s1.3348*, 2023.
- [14] G. Boudin, F., & Gallina, "A Comparison of Pre-trained Language Models for Multi-Documnt Summarization," *Ljubljana, Slov. © 2021 IW3C2 (International World Wide Web Conf. Committee), Publ. under Creat. Commons CC-BY 4.0 Licens. ACM ISBN 978-1-4503-8313-4/21/04. https://doi.org/10.1145/3442442.3451375*, 2021.
- [15] K. Kannan, R., Gurusamy, V., & Somasundaram, "Sentiment Analysis Using Deep Learning Techniques for Business Intelligence," *Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, pp. 1019-1024., 2021.
- [16] J. S. R.-M. et Al., "A systematic review of text mining techniques and applications in finance," *Expert Syst. with Appl. vol. 158, p. 113464*, 2020.
- [17] Y. Zhang, H., & Li, "An Improved Naive Bayes Algorithm for Text Classification," *IEEE Access*, vol. 9, pp. 81248-81256, 2021.
- [18] D. M. W. Powers, "Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation," *J. Mach. Learn. Technol. vol. 2, no. 1, pp. 37-63*, 2020.
- [19] D. Juang, "Analisis Spam Dengan Menggunakan Naive Bayes," *J. Teknol. Inf. dan Komunikasi*, vol. 2, no. 1, pp. 10-16, 2016.
- [20] C. Triawati, "Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia Statistical Concept Based Weighting Method

- for Indonesian Language Document Clustering and Categorisation,” *J. Inform. vol. 5, no. 2, pp. 90-100*, 2009.
- [21] S. S. Berutu, “Text Mining dan Klasifikasi Sentimen Berbasis Naive Bayes Pada Opini Masyarakat terhadap Makanan Tradisional,” *J. Teknol. Informasi, vol. 8, no. 3, pp. 45-51*, 2022.