

KOMPARASI MODEL SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN APLIKASI POLRI SUPER APP

Luban Abdi Susanto^{1*}

¹Alumni, Fakultas Teknik, Program Studi Teknik Informatika, Universitas Muhammadiyah Jember

Riwayat artikel:

Received: 24 Februari 2024

Accepted: 30 Maret 2024

Published: 2 April 2024

Keywords:

K-Nearest Neighbor, Naïve Bayes, Polri Super App, Sentiment Analysis

Correspondent Email:

lubanabdi9@gmail.com

Abstrak. Perkembangan teknologi informasi memacu inovasi dalam pengolahan data dan informasi. Inovasi *e-government* menjadi langkah progresif dalam menyediakan layanan publik yang efektif dan efisien. Polri menjadi contoh lembaga yang sukses menerapkan inovasi ini melalui "Polri Super App", sebuah aplikasi seluler yang meningkatkan layanan kepolisian dengan teknologi digital. Penelitian analisis sentimen terhadap aplikasi *e-government* di Google Play menarik perhatian. Dalam penelitian sebelumnya, SVM dan Naive Bayes memberikan hasil akurasi yang baik. Penelitian ini membandingkan kinerja KNN dan SVM dalam analisis sentimen terhadap Polri Super App. Hasil pengujian menunjukkan SVM dengan kernel Linear mencapai akurasi tertinggi, yaitu 89,67%, diikuti oleh SVM dengan kernel RBF dan KNN. Analisis ini memberikan pemahaman lebih baik tentang kinerja algoritma dalam mengelompokkan sentimen ulasan pengguna terhadap aplikasi tersebut.

Abstract. The advancement of information technology drives innovation in data processing. E-government innovation has become a progressive step in providing effective and efficient public services. The Indonesian National Police (Polri) serves as an example of an institution successfully implementing this innovation through the "Polri Super App," a mobile application that enhances police services with digital technology. Research on sentiment analysis of e-government applications on Google Play has attracted attention. In previous studies, SVM and Naive Bayes provided good accuracy results. This research compares the performance of KNN and SVM in sentiment analysis of the Polri Super App. Test results show that SVM with Linear kernel achieved the highest accuracy, 89.67%, followed by SVM with RBF kernel and KNN. This analysis provides a better understanding of algorithm performance in categorizing user sentiment towards the application.

1. PENDAHULUAN

Tuntutan akan efisiensi dan kecepatan dalam pengolahan data dan informasi mendorong perusahaan dan lembaga untuk bersaing dalam penciptaan inovasi terbaru [1]. Dengan perkembangan teknologi informasi yang pesat, inovasi dapat lebih mudah diwujudkan, mempengaruhi berbagai aspek kehidupan seperti transportasi, pendidikan, perdagangan, dan bahkan sektor pemerintahan [2].

Inovasi *e-government* merupakan langkah progresif yang diadopsi oleh pemerintah dan instansi untuk beradaptasi dengan era digital [3]. Melalui *e-government*, tercipta suatu sistem interaktif yang menghubungkan masyarakat, pemerintah, dan instansi menggunakan teknologi informasi, komunikasi, serta akses internet [4]. Implementasi *e-government* memungkinkan pemerintah dan instansi untuk menjalankan bisnis dengan lebih produktif dan

optimal, sehingga dapat memperhatikan kebutuhan masyarakat lebih baik. [5].

Salah satu contoh lembaga pemerintah yang telah sukses menerapkan *e-government* adalah Kepolisian Negara Republik Indonesia (POLRI). Polri aktif dalam memberikan layanan yang dibutuhkan masyarakat, seperti pembuatan SIM/STNK/SKCK, penanganan pengaduan masyarakat, serta pengurusan e-Tilang. Untuk mempermudah akses dan meningkatkan efisiensi proses tersebut, Polri mengembangkan sebuah inovasi berbasis teknologi digital yang dikenal sebagai "Polri Super App" [6].

Polri Super App adalah sebuah aplikasi seluler yang dikembangkan oleh Kepolisian Negara Republik Indonesia (POLRI) untuk meningkatkan layanan kepolisian kepada masyarakat. Aplikasi ini dirancang dengan menggunakan teknologi digital dalam rangka mendukung pelayanan publik secara efisien dan memberikan akses yang lebih mudah kepada masyarakat. Polri Super App menyediakan berbagai fitur dan layanan, termasuk perpanjangan SIM, pembayaran registrasi kendaraan, layanan keluhan publik, dan informasi layanan lainnya dari POLRI [6].

Data dari Aplikasi Polri Super App tidak hanya tentang layanan kepolisian, tetapi juga meliputi data lain yang diperoleh dari proses distribusi aplikasi tersebut [6]. Aplikasi ini didistribusikan melalui platform digital, terutama Google Play, yang memungkinkan pengguna memberikan ulasan [6]. Ulasan tersebut dapat diakses secara bebas dan berpotensi besar untuk memberikan kontribusi berharga terhadap pengembangan aplikasi [7]. Pengolahan data ulasan dapat memberikan wawasan penting bagi perbaikan aplikasi, karena ulasan pengguna menjadi sumber informasi yang berharga. Pendapat dari ulasan di Google Play dapat memengaruhi keputusan pengguna potensial. Namun, dengan jumlah ulasan yang besar, diperlukan metode otomatis untuk menganalisis tren pengguna terhadap aplikasi tersebut [8].

Untuk memperhatikan dampak ulasan pengguna terhadap kelangsungan suatu aplikasi, masuk akal dalam melakukan evaluasi perasaan atau opini terhadap ulasan Aplikasi Polri Super App di *platform* Google Play [9]. Analisis sentimen ini bertujuan untuk

mengelompokkan pendapat dari ulasan tersebut ke dalam kategori positif, negatif, atau netral, sebagai bagian dari proses penelitian dalam bidang *text mining* yang juga memiliki potensi nilai ekonomis [10].

Penelitian tentang analisis sentimen oleh [7] pada aplikasi *e-government* berjudul "Sentimen Analisis *Review* Aplikasi Digital Korlantas di *Google Play* Menggunakan Metode SVM" telah menarik perhatian. Dalam penelitian ini, sebanyak 1.200 data ulasan dari masyarakat digunakan, dengan algoritma SVM sebagai metode klasifikasi. Temuan dari penelitian menunjukkan bahwa aplikasi digital Korlantas menerima 598 ulasan positif dan 511 ulasan negatif. Meskipun demikian, kinerja algoritma SVM menunjukkan variasi, dengan akurasi tertinggi 82% pada rasio data 90:10 dan terendah 74% pada rasio data 80:20. Hasil penelitian ini menegaskan bahwa pembagian data dalam pengujian memiliki dampak signifikan terhadap akurasi yang dihasilkan.

Studi berikutnya yang dilakukan oleh [5] berfokus pada "Analisis Sentimen Ulasan Pengguna Aplikasi Bibit dan Bareksa Menggunakan Algoritma KNN". Data sampel berasal dari *review* dua aplikasi tersebut, dengan total 998 ulasan untuk Bareksa dan 1063 ulasan untuk Bibit. Penelitian ini menggunakan rasio data 60:40, menghasilkan tingkat akurasi sebesar 85,14% untuk Bareksa dan 81,70% untuk Bibit. Selain itu, presisi mencapai 91,91% untuk Bareksa dan 87,15% untuk Bibit, dengan nilai *recall* sebesar 76,44% untuk Bareksa dan 75,73% untuk Bibit. Namun, kedua penelitian tersebut tidak menguraikan arah sentimen pengguna terhadap topik yang dibahas, akan tetapi hanya memusatkan perhatian pada pengelompokan ulasan dan mengevaluasi akurasi metode yang dipakai.

Dalam konteks pengembangan teknologi aplikasi, penting untuk membandingkan kinerja algoritma yang berbeda dalam analisis sentimen. Berdasarkan uraian diatas, penelitian ini dilakukan untuk membandingkan dua algoritma populer, yaitu *Support Vector Machine* dan *K-Nearest Neighbor* dalam analisis sentimen pada aplikasi Polri Super App. Hasil evaluasi kinerja keduanya akan memberikan wawasan yang berharga bagi pengembangan aplikasi di masa depan serta peningkatan kualitas layanan publik yang disediakan oleh Polri.

2. TINJAUAN PUSTAKA

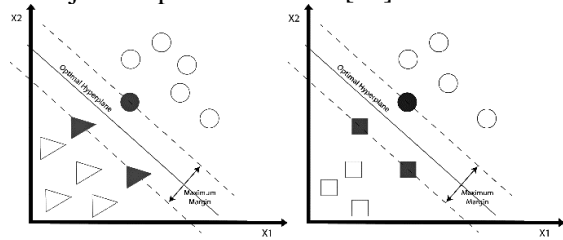
2.1. Analisis Sentimen

Studi analisis sentimen menjadi bagian penting dalam ranah riset yang berada di perbatasan berbagai disiplin ilmu, seperti Data Mining, *Natural Language Processing* (NLP), dan *Machine Learning* [11].

Pusat perhatian dalam penelitian sentiment adalah untuk mengklasifikasikan teks yang terdapat dalam suatu kalimat atau dokumen, kemudian mengidentifikasi apakah pendapat yang disampaikan di dalamnya bersifat positif, negatif, atau netral. Selain itu, analisis sentimen juga dapat mengungkapkan spektrum perasaan emosional, seperti kesedihan, kegembiraan, atau kemarahan [12].

2.2. Support Vector Machine

Teknik *Support Vector Machine* (SVM) digunakan untuk memprediksi dengan efektif, baik dalam konteks regresi maupun klasifikasi, dimana prinsipnya adalah menemukan hyperplane optimal dengan margin maksimal untuk menetapkan batas keputusan yang memisahkan kelas-kelas yang berbeda secara optimal [13]. Prinsip inti dari SVM adalah mencari *hyperplane* optimal untuk memisahkan dua kelas dalam ruang input, bisa berupa garis dalam dimensi dua atau bidang datar dalam dimensi lebih tinggi seperti ilustrasi yang ditunjukkan pada Gambar 1 [14].



Gambar 1. Hyperplane & Margin Maksimal

Support Vector Machine memiliki beragam jenis kernel yang dapat dipilih seperti kernel linear, RBF, Polynomial, dan Sigmoid [15]. Akan tetapi, hanya 2 kernel yang digunakan untuk penelitian ini yaitu kernel linear dan RBF. Kernel digunakan dengan tujuan untuk melakukan pemetaan terhadap dimensi awal dari sekumpulan data [13]. Berikut penjelasan mengenai kernel linear dan RBF.

2.2.1. Kernel Linear

Dalam analisis menggunakan kernel linear, fokus penyesuaian terletak pada parameter C

atau *Cost* [10]. Proses optimalisasi biasanya melibatkan percobaan berulang untuk menentukan nilai terbaik untuk parameter C . Berikut persamaan kernel linear [15].

$$K(x_i, x) = X_i^T X \quad (1)$$

2.2.2. Kernel RBF

Ketika mempertimbangkan kernel RBF dalam analisis, fokus pada pengoptimalan yang terletak pada dua parameter utama, yakni *Cost* (C) dan *Gamma* (γ) [15]. Seperti halnya dengan penggunaan kernel linear, upaya mengoptimalkan parameter-parameter ini biasanya melibatkan serangkaian percobaan untuk mencapai hasil terbaik [15]. Berikut persamaan kernel RBF.

$$K(x_i, x) = \exp(-\gamma |X_i^T X|^2) \quad (2)$$

Dimana pada persamaan (1) dan (2), x_i adalah data latih, x adalah data uji sedangkan K ialah konstanta [15].

2.3. K-Nearest Neighbor (KNN)

Salah satu pendekatan yang populer dalam analisis data, KNN merupakan teknik yang memanfaatkan konsep perbandingan jarak untuk mengklasifikasikan data [16]. Ide utama yang mendasari KNN adalah mengklasifikasikan data berdasarkan kedekatan spasialnya dengan data lain. [13]. Perbedaan jarak antara dua lokasi dapat diukur menggunakan salah satu dari berbagai metode jarak yang telah ditetapkan, seperti jarak *Euclidean*, jarak *Minkowski*, atau jarak *Mahalanobis* [5]. Adapun rumus jarak *Euclidean* sebagai berikut:

$$Distance(\vec{X}, \vec{Y}) = \sqrt{\sum_{n=1}^d (x_i - y_i)^2} \quad (3)$$

Dimana:

x_i : data uji X,

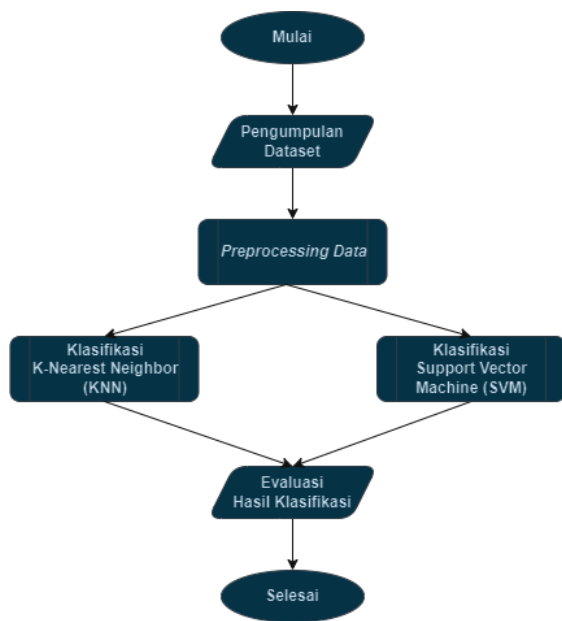
y_i : data uji Y,

$Distance(\vec{X}, \vec{Y})$: Jarak Euclidean data X & Y,

d : dimensi data dari variabel independen.

3. METODE PENELITIAN

Dalam pendekatan penelitian ini, akan diuraikan serangkaian langkah yang diambil untuk menganalisis sentimen terhadap aplikasi Polri Super App.



Gambar 2. Diagram Alir Penelitian

3.2. Pengumpulan Dataset

Proses pengumpulan dataset yang dilakukan menggunakan teknik mengambil data (*scrapping*) dari situs *Google Play Store* pada ulasan aplikasi *Polri Super App*. Proses *scrapping* ini menggunakan bahasa *python* dan *library google-play-scrapper*. Data yang diperoleh sejumlah 1.500 dalam bentuk ulasan (*text review*) dengan rentang waktu 07 Oktober 2022 s/d 05 Februari 2024.

3.3. Preprocessing Data

Tahapan ini dilakukan untuk mengoptimalkan dataset sebelum dilakukan pengujian pada sebuah model, seperti menghilangkan gangguan pada data serta mengorganisasi data yang awalnya tidak teratur menjadi data yang terstruktur. [4]. Berikut beberapa tahapan pada *preprocessing* data, yakni *case folding*, pelabelan dataset, *tokenizing*, normalisasi, *stopword removal*, *stemming*, dan pembobotan data [17].

3.4. Klasifikasi

Sebelum dilakukan klasifikasi, penelitian ini menerapkan strategi pembagian data 80:20 dari keseluruhan dataset yang tersedia. Dalam proses klasifikasi, digunakan perbandingan dua metode, yaitu KNN dan SVM, yang kemudian dievaluasi melalui metode validasi silang *k-fold* sebanyak *5-fold* & *10-fold*. Akan tetapi, khusus untuk metode SVM mengalami dua kali pengujian yakni pengujian menggunakan kernel

linear dan kernel RBF.

3.5. Evaluasi

Untuk mengetahui hasil dari kinerja perbandingan antara metode KNN dan SVM, dilakukan perhitungan menggunakan *confusion matrix* yakni menghitung nilai *accuracy*, *recall*, *precision*, dan *F1-Score*. Terdapat 4 parameter dalam pengujian menggunakan *confusion matrix* diantaranya *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN). Hasil akhir dari evaluasi yang dilakukan adalah membandingkan kinerja dari kedua metode KNN dan SVM.

4. HASIL DAN PEMBAHASAN

Di bagian ini, analisis dan pemaparan temuan dari penelitian sebelumnya akan diuraikan secara rinci, sebagai berikut.

4.1. Dataset

Teknik yang digunakan dalam mengumpulkan dataset adalah teknik web *scrapping* untuk mengambil dataset dari *website google-play* pada aplikasi *Polri Super App*. Dataset yang diperoleh berjumlah 1.500 data berbentuk ulasan dengan rentang waktu 07 Oktober 2022 s/d 05 Februari 2024 yang kemudian disimpan dalam bentuk format *.csv*. Berikut hasil *scrapping* data seperti yang terlihat dalam ilustrasi di Gambar 3.

	userName	score	at	content
178	Faisal Mudzawam	1	2024-02-08 07:55:39	Memverifikasi data diri sendiri aja sulit,pada...
430	Defi Rahmawati	1	2024-02-08 07:03:06	Pas isi data juga malah ga bisa di tengah teng...
0	Mochammad Irfan Fauzi	1	2024-02-08 05:47:37	Mohon diperbaiki kembali sistemnya sama sepert...
530	Silvi Nurcahyani	1	2024-02-08 05:41:53	Pengajuan SKCK selalu invalid request. Saran L...
1184	Ferdy Mahardian	1	2024-02-08 04:44:17	Invalid terus pas memasukkan data identitas
...
1075	asep firman	1	2023-04-04 03:56:50	ketika daftar baru ga ada otp SMS masuk mana L...
1349	Jonathan Franklin	2	2023-03-29 07:13:08	mau masukan identitas loading nya lama banget ...
1279	Raharjo Yees	1	2023-03-22 03:51:46	cara mengetahui data diri sudah terverifikasi ...
1060	Feberlaia 19	1	2023-01-17 01:49:55	Hari ini sy mencoba mengisi data tujuan utk me...
1174	Gall a	5	2022-10-07 09:46:15	Aplikasinya keren fitur nya juga mudah dipahami...

1500 rows × 4 columns

Gambar 3. Hasil Teknik Scrapping

4.2. Tahapan Preprocessing Data

Tahapan ini dilakukan mentransformasikan data yang sebelumnya tak terorganisir menjadi data yang terstruktur untuk siap diolah [17].

4.2.1. Pelabelan Data

Proses pemberian label pada kumpulan data dilakukan secara manual menggunakan *google colab* untuk mengklasifikasikan data ke dalam

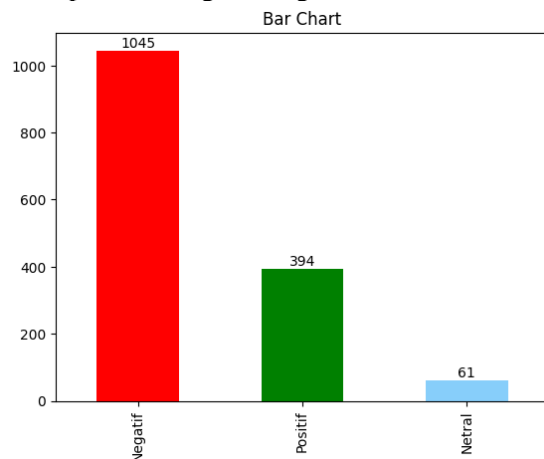
kategori positif, negatif, atau netral. Dibawah ini merupakan gambar hasil pelabelan data.

	score	at	content	sentimen
0	5	2024-01-22 04:19:42	Apk yang sangat membantu untuk membuat surat-s...	Positif
1	5	2024-02-04 11:59:00	Appsnya memudahkan dalam pembuatan SKCK secara...	Positif
2	4	2024-01-22 01:35:10	Apresiasi dulu buat developer pengembang aplik...	Positif
3	5	2024-01-30 11:17:00	Memudahkan kita kalau mau mengurus perpanjanga...	Positif
4	5	2024-01-21 03:47:44	Aplikasinya membantu banget, mudah di gunain j...	Positif
...
1495	2	2023-08-24 04:53:31	Terlalu lama dan sering keluar sendiri banyak ...	Negatif
1496	1	2023-08-28 13:00:19	Sangat buruk aplikasi ini.. Tidak bisa poto Ba...	Negatif
1497	1	2023-09-04 03:55:16	Mending buka web aja dulu pak kalo masih banya...	Negatif
1498	1	2023-08-30 20:30:10	Wah ini sih parah bgt., coba pak polisinya snd...	Negatif
1499	1	2023-08-16 12:05:03	Kalau meluncurkan sebuah aplikasi dipublik, la...	Negatif

1500 rows × 4 columns

Gambar 4. Output Proses Pelabelan Data

Dalam gambar 4, nilai 1-2 menunjukkan reaksi yang kurang positif, sedangkan nilai 3 mencerminkan sentimen netral, sedangkan skor 4-5 dinyatakan sebagai sentimen positif. Berikut gambar visualisasi dari pengelompokan kelas pada masing-masing sentimen.



Gambar 5. Visualisasi Pengelompokan Kelas

4.2.2. Case Folding

Setelah proses *cleansing* dilakukan, kemudian proses *case folding* dilakukan dengan tujuan mengubah huruf kapital ke huruf kecil [10]. Tabel 1 memperlihatkan output dari langkah proses case folding.

Tabel 1. Hasil Proses Case Folding	
Sebelum	Sesudah
Di saat urgent mau bikin SKCK sekarang diharus...	di saat urgent mau bikin skck sekarang diharus...
Saya sudah ajukan SKCK dan sudah transfer pake...	saya sudah ajukan skck dan sudah transfer pake...

ngga situs ngga aplikasi sama aja jelek semua....	ngga situs ngga aplikasi sama aja jelek semua....
Sangat buruk, padahal kode verifikasi nya suda...	sangat buruk, padahal kode verifikasi nya suda...
Masa aplikasi POLRI kek gini ?, gabisa...	masa aplikasi polri kek gini ?, gabisa...

4.2.3. Tokenizing

Proses *tokenizing* dilakukan untuk memisahkan ulasan menjadi kata-kata, menghapus karakter pembatas, dan menghilangkan angka dari teks [17]. Berikut gambar 6 hasil proses dari *tokenizing*.

	at	Tokenizing
0	2024-02-08 05:47:37	[mohon, diperbaiki, kembali, sistemnya, sama, ...
1	2024-01-22 04:19:42	[apk, yang, sangat, membantu, untuk, membuat, ...
2	2024-01-11 00:18:22	[mohon, untuk, aplikasinya, dikembangkan, lagi...
3	2024-01-05 13:17:34	[benar, ", lemot, ,, susah, masuk, ,, data, t...
4	2024-01-04 06:22:58	[tidak, bs, lanjut, input, data, alamat, ,, st...
...
1495	2023-08-24 04:53:31	[terlalu, lama, dan, sering, keluar, sendiri, ...
1496	2023-08-28 13:00:19	[sangat, buruk, aplikasi, ini, ,, tidak, bisa...
1497	2023-09-04 03:55:16	[mending, buka, web, aja, dulu, pak, kalo, mas...
1498	2023-08-30 20:30:10	[wah, ini, sih, parah, bgt, ,, ,, coba, pak, p...
1499	2023-08-16 12:05:03	[kalau, meluncurkan, sebuah, aplikasi, dipubli...

1500 rows × 2 columns

Gambar 7. Output Proses Tokenizing

4.2.4. Normalisasi

Tahapan normalisasi bertujuan mengonversi frasa tidak resmi menjadi resmi sesuai standar dalam KBBI [10]. Melalui penggunaan dataset *slangwords* yang mencakup istilah-istilah slang yang akan disesuaikan sesuai KBBI, proses ini memanfaatkan pustaka RegEx untuk membantu dalam transformasi kata-kata tersebut menjadi baku [9]. Berikut gambar 8 hasil dari proses normalisasi.

	at	Normalisasi
0	2024-02-08 05:47:37	[mohon, diperbaiki, kembali, sistemnya, sama, ...
1	2024-01-22 04:19:42	[aplikasi, yang, sangat, membantu, untuk, memb...
2	2024-01-11 00:18:22	[mohon, untuk, aplikasinya, dikembangkan, lagi...
3	2024-01-05 13:17:34	[benar, ", lelet, ,, susah, masuk, ,, data, t...
4	2024-01-04 06:22:58	[tidak, bisa, lanjut, input, data, alamat, ,, ...
...
1495	2023-08-24 04:53:31	[terlalu, lama, dan, sering, keluar, sendiri, ...
1496	2023-08-28 13:00:19	[sangat, buruk, aplikasi, ini, ,, tidak, bisa...
1497	2023-09-04 03:55:16	[lebih baik, buka, situs, saja, dulu, pak, kal...
1498	2023-08-30 20:30:10	[wah, ini, sih, parah, banget, ,, ,, coba, pak...
1499	2023-08-16 12:05:03	[kalau, meluncurkan, sebuah, aplikasi, dipubli...

1500 rows × 2 columns

Gambar 8. Hasil Proses Normalisasi

Proses ini melibatkan penghapusan kata-kata umum yang sering muncul dalam data, dimana digunakan pustaka *Natural Language Toolkit* (NLTK) untuk menangani penghapusan tersebut [5]. Berikut gambar 9 hasil proses *stopword removal*.

1500 rows x 2 columns

Gambar 9. *Output stopwords removal*

Proses *stemming* adalah langkah dalam mencari akar kata dari token, di mana langkah ini melibatkan pembuatan aturan untuk menghapus imbuhan atau menemukan akar kata dalam sebuah file [12]. Berikut gambar 10 hasil dari proses *stemming*.

```
[1500 rows x 2 columns]
```

Gambar 10. Output Proses Stemming

Pembobotan data yang digunakan pada studi ini menerapkan teknik *Term Inverse Document Frequency* (TF-IDF) sebagai metode analisis [11]. Pembobotan data dilakukan setelah proses *stemming* dilakukan. Tujuan dari proses ini adalah untuk merekam kata-kata yang sering digunakan dengan mengukur seberapa sering kata tersebut muncul dan mengeliminasi kata-kata yang kurang relevan yang terdapat dalam kumpulan data [5]. Berikut gambar 11 hasil proses TF-IDF.



Gambar 11. Output dari Proses TF-IDF

Merepresentasikan informasi yang terdapat dalam teks atau data teks yang telah diolah melalui pemrosesan bahasa alami (*Natural Language Processing*/NLP) dengan menggunakan visualisasi grafis adalah suatu praktik yang penting [10]. Proses ini tidak hanya membantu dalam memahami secara lebih dalam, menganalisis dengan cermat, tetapi juga memudahkan dalam menyampaikan hasil dari algoritma pemrosesan bahasa alami kepada pemangku kepentingan secara efektif [9]. *Library* yang digunakan pada visualisasi data adalah *WordCloud*. Berikut gambar visualisasi data positif, netral dan negatif.



Gambar 12. Visualisasi Sentimen Negatif

Berdasarkan gambar 12, terdapat ukuran kata yang berbeda-beda. Jika diurutkan, ukuran yang paling besar adalah kata yang sering muncul pada kelas sentimen negatif, sedangkan ukuran kata yang paling kecil adalah kata yang sedikit muncul pada kelas sentimen negatif.

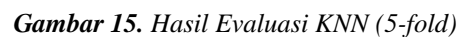


Gambar 14. Visualisasi Sentimen Positif

4.4. Klasifikasi dan Evaluasi

4.4.1. K-Nearest Neighbor (KNN)

Berikut adalah hasil klasifikasi dari pengujian pertama menggunakan *5-fold* yang ditunjukkan pada gambar 15.



Confusion Matrix

	0	1	2
0	817	1	14
1	44	1	3
2	52	0	268

True labels

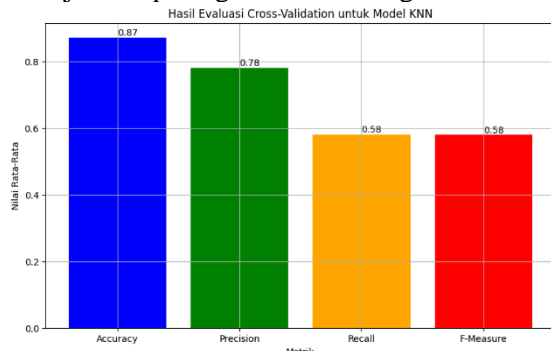
Predicted labels

Gambar 16. Confusion Matrix KNN 5-fold

1186

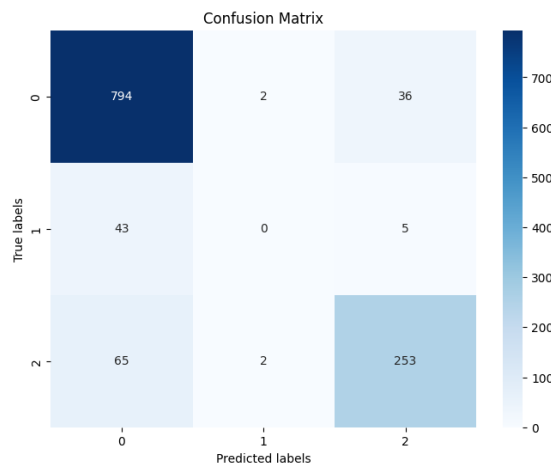
ketiga model mampu memprediksi dengan benar sebanyak 268 (dari 320) ulasan positif.

Kemudian, pengujian yang kedua menggunakan *10-fold* diperoleh hasil yang ditunjukkan pada gambar 17 sebagai berikut:



Gambar 17. Hasil Evaluasi KNN (*10-fold*)

Pada gambar 17, klasifikasi model KNN menggunakan *10-fold* diperoleh nilai *accuracy* sebesar 87,25%, *precision* 78,14%, *recall* 58,13%, dan *f1-score* 57,97%. Kemudian, untuk hasil *confusion matrix* ditunjukkan pada Gambar 18.



Gambar 18. Confusion Matrix KNN *10-fold*

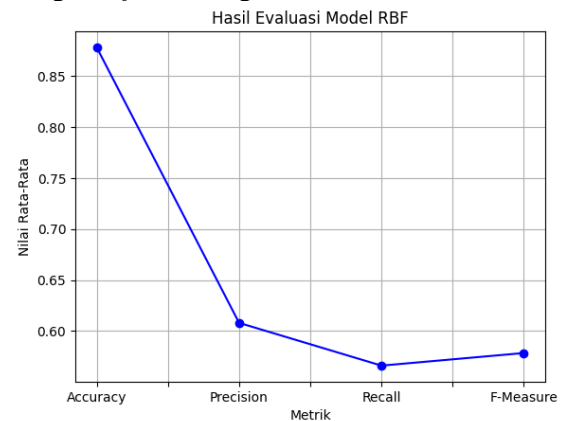
Dari gambar 18, dapat dijelaskan bahwasanya pada baris teratas model mampu memprediksi dengan benar sebanyak 794 (dari total 832) ulasan negatif, di baris kedua model mampu memprediksi dengan benar sebanyak 0 (dari total 48) ulasan netral, sedangkan di baris ketiga model mampu memprediksi dengan benar sebanyak 253 (dari 320) ulasan positif.

4.4.2. Support Vector Machine (SVM)

Pengujian pada model SVM dilakukan sebanyak 4 kali. Pengujian pertama, model SVM menggunakan kernel RBF dengan *5-fold*,

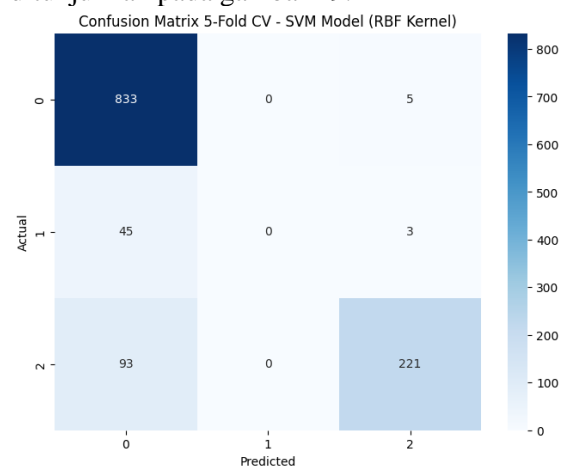
pengujian kedua model SVM menggunakan kernel RBF dengan *10-fold*, pengujian ketiga model SVM menggunakan kernel Linear dengan *5-fold*, dan pengujian keempat model SVM menggunakan kernel Linear dengan *10-fold*.

Berikut adalah hasil klasifikasi pengujian pertama model menggunakan kernel RBF dengan *5-fold*, sebagai berikut:



Gambar 19. Hasil Evaluasi SVM-RBF (*5-fold*)

Pada gambar 19, klasifikasi model SVM-RBF menggunakan *5-fold* diperoleh nilai *accuracy* sebesar 87,83%, *precision* 60,79%, *recall* 56,59%, dan *f1-score* 57,83%. Kemudian, untuk hasil *confusion matrix* ditunjukkan pada gambar 19.

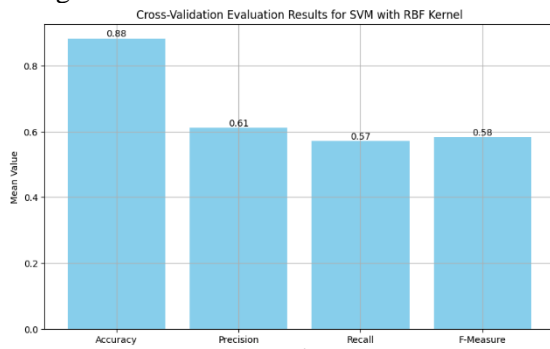


Gambar 19. Confusion Matrix SVM-RBF *5-fold*

Dari gambar 19, dapat dijelaskan bahwasanya pada baris teratas model mampu memprediksi dengan benar sebanyak 833 (dari total 838) ulasan negatif, di baris kedua model mampu memprediksi dengan benar sebanyak 0 (dari total 48) ulasan netral, sedangkan di baris

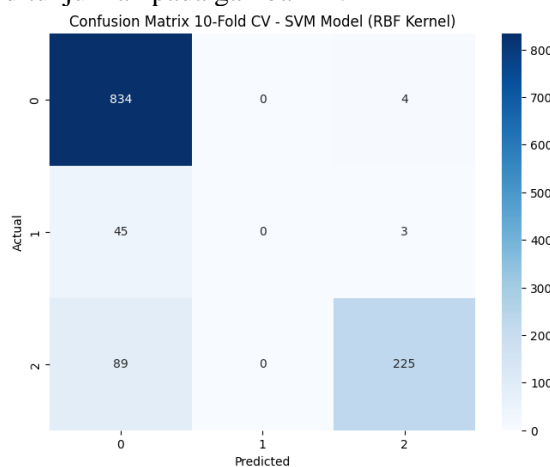
ketiga model mampu memprediksi dengan benar sebanyak 221 (dari 314) ulasan positif.

Kemudian, pengujian kedua model SVM menggunakan kernel RBF dengan *10-fold* diperoleh hasil klasifikasi pada gambar 20 sebagai berikut:



Gambar 20. Hasil Evaluasi SVM-RBF (*10-fold*)

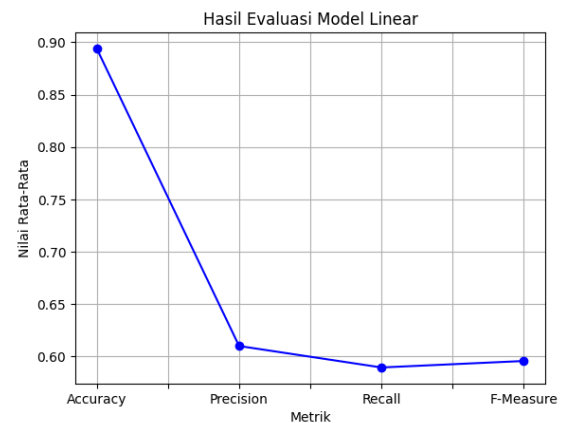
Pada gambar 20, klasifikasi model SVM-RBF menggunakan *10-fold* diperoleh nilai *accuracy* sebesar 88,25%, *precision* 61,11%, *recall* 57,06%, dan *f1-score* 58,23%. Kemudian, untuk hasil *confusion matrix* ditunjukkan pada gambar 21.



Gambar 21. Confusion Matrix SVM-RBF *10-fold*

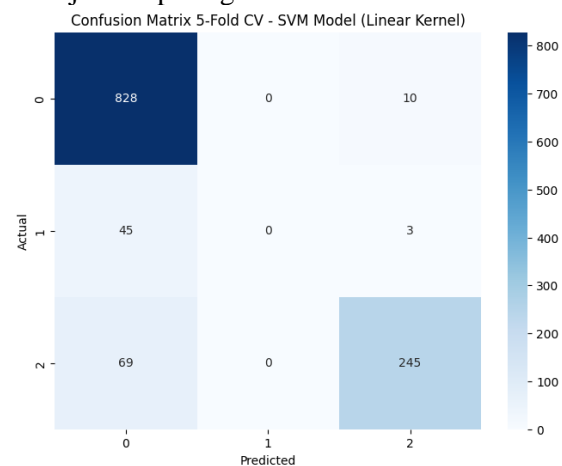
Dari gambar 21, dapat dijelaskan bahwasanya pada baris teratas model mampu memprediksi dengan benar sebanyak 834 (dari total 838) ulasan negatif, di baris kedua model mampu memprediksi dengan benar sebanyak 0 (dari total 48) ulasan netral, sedangkan di baris ketiga model mampu memprediksi dengan benar sebanyak 225 (dari 314) ulasan positif.

Selanjutnya, pengujian ketiga model SVM menggunakan kernel linear dengan *5-fold* diperoleh hasil klasifikasi yang ditunjukkan pada gambar 22.



Gambar 22. Hasil Evaluasi SVM-Linear (*5-fold*)

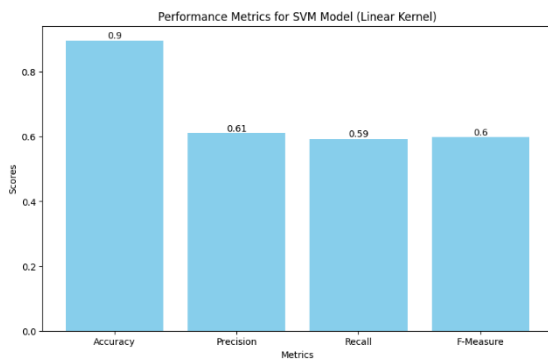
Pada gambar 22, klasifikasi model SVM-Linear menggunakan *5-fold* diperoleh nilai *accuracy* sebesar 89,41%, *precision* 60,99%, *recall* 58,94%, dan *f1-score* 59,56%. Kemudian, untuk hasil *confusion matrix* ditunjukkan pada gambar 23.



Gambar 23. Confusion Matrix SVM-Linear *5-fold*

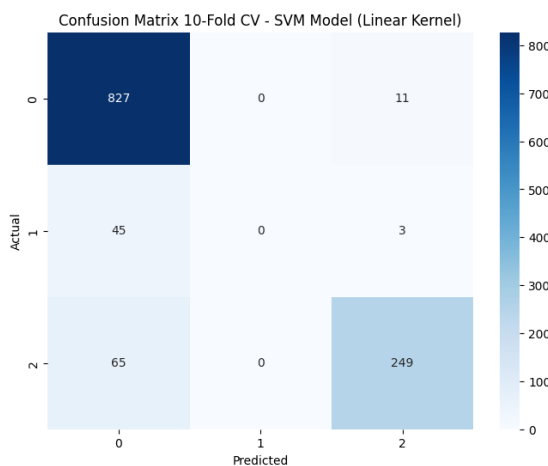
Dari gambar 23, dapat dijelaskan bahwasanya pada baris teratas model mampu memprediksi dengan benar sebanyak 828 (dari total 838) ulasan negatif, di baris kedua model mampu memprediksi dengan benar sebanyak 0 (dari total 48) ulasan netral, sedangkan di baris ketiga model mampu memprediksi dengan benar sebanyak 245 (dari 314) ulasan positif.

Kemudian, pengujian keempat model SVM menggunakan kernel linear dengan *10-fold* diperoleh hasil klasifikasi yang ditunjukkan pada gambar 24.



Gambar 24. Hasil Evaluasi SVM-Linear (10-fold)

Pada gambar 24, klasifikasi model SVM-Linear menggunakan 10-fold diperoleh nilai *accuracy* sebesar 89,67%, *precision* 61,09%, *recall* 59,33%, dan *f1-score* 59,78%. Kemudian, untuk hasil *confusion matrix* ditunjukkan pada gambar 25.



Gambar 25. Confusion Matrix SVM-Linear 10-fold

Dari gambar 25, dapat dijelaskan bahwasanya pada baris teratas model berhasil menghasilkan prediksi yang akurat hingga 827 (dari total 838) ulasan negatif, di baris kedua model mampu memprediksi dengan benar sebanyak 0 (dari total 48) ulasan netral, sedangkan di baris ketiga model mampu memprediksi dengan benar sebanyak 249 (dari 314) ulasan positif.

4.4.3. Evaluasi

Berdasarkan hasil klasifikasi yang telah dipaparkan sebelumnya, maka pada tahap evaluasi ini akan telah dilakukan evaluasi terhadap hasil klasifikasi dari kedua model tersebut. Perhatikan data yang tercantum dalam tabel 2 di bawah ini:

Tabel 2. Perbandingan Hasil Klasifikasi

Model	Kernel	n-fold	Akurasi
KNN	-	5-fold	87,42%
		10-fold	87,25%
SVM	RBF	5-fold	87,83%
		10-fold	88,25%
	Linear	5-fold	89,41%
		10-fold	89,67%

Dari tabel tersebut, dapat diamati bahwa model SVM dengan kernel linear mencapai akurasi tertinggi, yaitu 89,67% dengan 10-fold, dibandingkan dengan model KNN dan model SVM dengan kernel RBF.

KESIMPULAN

Dari rangkuman temuan yang telah dijabarkan sebelumnya, dapat disimpulkan hal berikut:

1. Hasil pengujian menunjukkan SVM dengan kernel Linear mencapai akurasi tertinggi, yaitu 89,67%, diikuti oleh SVM dengan kernel RBF dan KNN.
2. Dari eksperimen menggunakan metode validasi silang *k-fold* yang telah dilakukan, diperoleh nilai akurasi tertinggi sebesar 89,67% pada model SVM-Linear dengan 10-fold, dan nilai akurasi terendah sebesar 87,25% diraih model KNN dengan 10-fold.

UCAPAN TERIMA KASIH

Penulis bersyukur kepada Sang Pencipta yang Mahakuasa dan kepada junjungan Nabi Muhammad SAW atas segala anugerahnya sehingga dapat menyelesaikan penelitian ini dengan baik.

DAFTAR PUSTAKA

- [1] L. A. Susanto, "Pemilihan Hyperparameter Pada Alexnet Cnn Untuk Klasifikasi Citra Penyakit Kedelai," *Indexia Inform. Comput. Intell. J.*, vol. 5, no. 02, p. 113, 2023, doi: 10.30587/indexia.v5i02.5508.
- [2] L. A. Susanto, A. Nilogiri, and L. Handayani, "Klasifikasi Citra Lesi Kulit Serupa Virus Monkeypox Menggunakan VGG-19 Convolutional Neural Network," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 1, pp. 1–9, 2023, doi: <https://doi.org/10.32528/justindo.v8i1.168>
- [3] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector

- Machine,” *Inf. Syst. Dev.*, vol. 5, no. 2, pp. 7–13, 2020.
- [4] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [5] A. D. Adhi Putra, “Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [6] J. Rizqullah and A. Reza, “‘Super App Precision’ Sebagai Bentuk Pelayanan Publik di Era Masyarakat 5.0,” *Justin (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 1, pp. 1–7, 2024, doi: 10.26418/justin.v12i1.65162.
- [7] E. R. Kaburuan and N. R. Setiawan, “Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, 2023, doi: 10.32736/sisfokom.v12i1.1614.
- [8] S. Ernawati and R. Wati, “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel,” *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.
- [9] P. R. Alvita Wagiswari D, I. Susilawati, and A. Witanti, “Analisis Sentimen pada Komentar Aplikasi MyPertamina dengan Metode Multinomial Naive Bayes,” *ForAI J. Informatics Artif. Intell. J.*, vol. 1, no. 1, pp. 10–19, 2023, [Online]. Available: <https://jurnal.forai.or.id/index.php/forai/article/view/4>
- [10] E. Fitri, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *J. Transform.*, vol. 18, no. 1, p. 71, 2020, doi: 10.26623/transformatika.v18i1.2317.
- [11] D. Rustiana and N. Rahayu, “Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 8, no. 1, pp. 113–120, 2017, doi: 10.24176/simet.v8i1.841.
- [12] N. R. Wardani and A. Erfina, “Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritma Naive Bayes,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform. Univ. Nusa Putra)*, pp. 11–18, 2021.
- [13] M. N. Muttaqin and I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, pp. 22–27, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [14] M. F. Naufal, “Analisis Perbandingan Algoritma SVM, KNN, dan CNN untuk Klasifikasi Citra Cuaca,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 311, Mar. 2021, doi: 10.25126/jtiik.2021824553.
- [15] S. Styawati, N. Hendrastuty, and A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [16] I. Habib Kusuma and N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023.
- [17] A. N. Hasanah, B. N. Sari, U. S. Karawang, T. Timur, and J. Barat, “Analisis Sentimen Ulasan Pengguna Aplikasi Jasa Ojek Online Maxim Pada Google Play Dengan Metode Naïve Bayes Classifier,” *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 12, no. 1, pp. 90–96, 2024.