

---

## IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA KLASIFIKASI SENTIMEN REVIEW LAYANAN TELEMEDICINE HALODOC

Reynalda Nabila Cikania<sup>1</sup>

<sup>1</sup>Departemen Statistika Bisnis, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember

e-mail: [reynalda.cikania@gmail.com](mailto:reynalda.cikania@gmail.com)

---

### Abstrak

Halodoc merupakan aplikasi layanan kesehatan berbasis *telemedicine* yang menghubungkan pasien dengan praktisi kesehatan seperti dokter, apotek, dan laboratorium. Ada beberapa komentar masyarakat pengguna aplikasi Halodoc, baik komentar yang bersifat positif maupun bersifat negatif. Hal ini menunjukkan adanya perhatian masyarakat terhadap aplikasi Halodoc sehingga perlu dilakukan analisis sentimen atau komentar yang muncul pada layanan aplikasi Halodoc terutama di masa pandemi COVID-19 agar layanan aplikasi Halodoc menjadi lebih baik. Algoritma *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM) digunakan untuk menganalisis sentimen masyarakat pengguna aplikasi layanan *telemedicine* Halodoc. Hasil klasifikasi sentimen kategori negatif sebesar 12,33%, sedangkan sentimen kategori positif adalah 87,67% dari 5.687 review yang artinya sentimen *review* positif lebih banyak daripada sentimen *review* negatif. Performa ketepatan klasifikasi Algoritma *Naïve Bayes Classifier* menghasilkan tingkat akurasi sebesar 87,77% dengan nilai AUC sebesar 57,11% dan *G-Mean* sebesar 40,08%, sedangkan Algoritma SVM dengan Kernel RBF memiliki nilai akurasi sebesar 86,1% dengan nilai AUC sebesar 60,149% dan nilai *G-Mean* sebesar 49,311%. Berdasarkan nilai akurasi model dapat diketahui model SVM Kernel RBF lebih baik daripada NBC pada pengklasifikasian *review* sentimen pengguna layanan *telemedicine* Halodoc.

**Kata Kunci:** COVID – 19, Halodoc, Klasifikasi Naive Bayes, *Support Vector Machine*

---

### Abstract

*Halodoc is a telemedicine-based healthcare application that connects patients with health practitioners such as doctors, pharmacies, and laboratories. There are some comments from halodoc users, both positive and negative comments. This indicates the public's concern for the Halodoc application so it is necessary to analyze the sentiment or comments that appear on the Halodoc application service, especially during the COVID-19 pandemic in order for Halodoc application services to be better. The Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) algorithms are used to analyze the public sentiment of Halodoc's telemedicine service application users. The negative category sentiment classification result was 12.33%, while the positive category sentiment was 87.67% from 5,687 reviews which means that the positive review sentiment is more than the negative review sentiment. The accuracy performance of the Naive Bayes Classifier Algorithm resulted in an accuracy rate of 87.77% with an AUC value of 57.11% and a G-Mean of 40.08%, while svm algorithm with KERNEL RBF had an accuracy value of 86.1% with an AUC value of 60.149% and a G-Mean value of 49.311%. Based on the accuracy value of the model can be known SVM Kernel RBF model better than NBC on classifying the review of user sentiment of halodoc telemedicine service.*

**Keywords:** COVID – 19, Halodoc, *Naïve Bayes Classifier*, *Support Vector Machine*

---

## 1. PENDAHULUAN

*Telemedicine* Halodoc adalah layanan kesehatan yang memanfaatkan teknologi informasi medis melalui media audio visual dan data dengan melibatkan beberapa pihak seperti dokter, pasien dan praktisi kesehatan lain secara jarak jauh (Coelho 2011). Aplikasi ini menyediakan berbagai fitur-fitur konsultasi medis melalui *video call* (*teleconsultation*), pemeriksaan lab secara *on-demand*, dan pembelian obat melalui apotek berbasis layanan *online*. Secara keseluruhan fitur-fitur yang ditawarkan telah mengikuti beberapa layanan yang ada di rumah sakit. Aplikasi Halodoc ini dapat diunduh dan diinstal melalui *Google Play* atau *App Store*.

Selama pandemi COVID-19 unduhan Halodoc meningkat 10 kali lipat pada dua bulan awal masa pandemi yaitu bulan Maret dan April 2020 (Fitri 2020). Peningkatan unduhan terjadi setelah Kementerian Kesehatan RI dan Badan Nasional Penanggulangan Bencana (BNPB) menunjuk Halodoc sebagai platform layanan kesehatan digital untuk tes cepat (*rapid test*) dan tes *Polymerase Chain Reaction* (*PCR*) virus COVID-19. Hal ini sangat membantu masyarakat untuk memenuhi kebutuhan tes deteksi COVID-19. Layanan tersebut dapat diperoleh dengan melakukan pembuatan janji terlebih dahulu via Halodoc. Pengguna aplikasi Halodoc ada yang kurang puas terhadap layanan yang ada pada aplikasi Halodoc seperti pelaksanaan tes yang tidak sesuai jadwal sehingga menimbulkan pembatalan, tidak tepatnya jadwal pemberian hasil tes, kesulitan dalam menemukan rujukan apotek untuk menebus obat serta *customer service* yang dinilai kurang responsif dan keluhan yang lainnya. Keluhan masyarakat ini dapat diketahui dari komentar yang ada pada kolom *review* aplikasi Halodoc. Karena jumlah pengguna yang menggunakan aplikasi meningkat, opini dari pengguna selama pandemi ini diperlukan untuk meningkatkan layanan pada aplikasi Halodoc. Oleh karena itu, perlu dilakukan analisis sentimen terhadap komentar yang ada pada layanan aplikasi tersebut.

Analisis sentimen adalah analisis pengolahan bahasa alami, komputasi linguistic, dan teks mining yang bertujuan menganalisis sentimen atau penilaian pada suatu produk atau layanan tertentu (Liu 2010). *Text Mining* merupakan proses penggalian atau penambangan data pada suatu teks yang bertujuan untuk mengetahui kata-kata yang dapat mewakili isi dari sebuah dokumen sehingga dapat dilakukan proses analisa lebih lanjut (Aggarwal 2018). Analisis sentimen dapat digunakan untuk melihat bagaimana kecenderungan suatu sentimen layanan aplikasi (Halodoc) baik yang negatif maupun positif. Proses klasifikasi teks *review* terhadap layanan *telemedicine* Halodoc untuk sentimen positif maupun negatif dilakukan dengan beberapa metode yang paling populer diantaranya *Naïve Bayes Classifier*, *KNN*, *SVM*.

Penelitian mengenai *Naïve Bayes Classifier* untuk melakukan klasifikasi pada aplikasi Halodoc telah dilakukan oleh Neng Resti Wardani (Wardani and Erfina 2021) yang memperoleh nilai akurasi 82,86 %. merupakan metode klasifikasi dengan cara kerja menghitung probabilitas atau peluang berdasarkan rumus *Bayesian Rule* yang digunakan untuk memecahkan masalah klasifikasi pada *feature – feature* data bernilai nominal maupun numerik (Faisal and Nugrahadi 2017) dan dikenal memiliki tingkat akurasi tinggi pada pengaplikasian database dengan data yang besar (Prasetyo 2012). Berdasarkan penelitian Muhammad Ranga Aziz Nasution (Nasution and Hayaty 2019) diperoleh bahwa jika dibandingkan dengan *KNN* model *SVM* memiliki tingkat akurasi yang lebih baik

Oleh Karena itu, pada penelitian ini dilakukan perbandingan antara metode *Naïve Bayes Classifier* dan *Support Vector Machine* pada analisis sentimen *review* pengguna layanan *telemedicine* Halodoc pada masa pandemi COVID-19 berdasarkan performa kinerja masing-masing algoritma.

## 2. METODE PENELITIAN

### 2.1 Sumber Data

Data yang digunakan adalah data sekunder yaitu sekumpulan *review* atau ulasan dari para pengguna layanan *telemedicine* Halodoc (<https://www.Halodoc.com>). Data penelitian adalah data *review* Halodoc mulai tanggal 1 November 2020 sampai dengan 8 Februari 2021 dengan menggunakan metode *scraping web*.

### 2.2 Variabel Penelitian

Variabel penelitian dapat ditunjukkan pada Tabel 1 sebagai berikut.

Variabel	Keterangan	Skala data
X	Jumlah kemunculan kata ( <i>Term Frequency</i> )	Rasio
Y	Kategori sentimen (Positif/Negatif) Syarat : Jika, skor $\geq 0$ , kategori positif Jika, skor $< 0$ , kategori negatif	Nominal

Adapun struktur data yang digunakan pada penelitian ini yang dapat ditunjukkan pada Tabel 2 sebagai berikut.

No.	Jumlah kemunculan <i>Term</i> (X)	Kategori Sentimen (Y)
1.	$X_1$	$Y_1$
2.	$X_2$	$Y_2$
$\vdots$	$\vdots$	$\vdots$
$n$	$X_n$	$Y_n$

Keterangan :

$X_i$  : Jumlah kemunculan *term*

$Y_i$  : Kategori sentimen (positif/ negatif)

dimana  $i = 1, 2$  sampai dengan  $n$

### 2.3 Metode Pengumpulan Data

Data keseluruhan *review* yang terdapat di kolom *review* Halodoc sejak tahun 2016 sampai saat ini (update terakhir) adalah sebanyak 290.343 *review*. *Scraping* pengambilan data dimulai tanggal 1 November 2020 sampai dengan 8 Februari 2021 sehingga dihasilkan data *review* sebanyak 5.687 data. Pemberian label sentimen berdasarkan pada perhitungan skor dengan rumus sebagai berikut.

$$\text{Skor} = (Z_{kp} - Z_{kn})$$

Keterangan :

$Z_{kp}$  : Jumlah kata positif

$Z_{kn}$  : Jumlah kata negatif

### 2.4 Langkah Penelitian

Adapun metode analisis data yang digunakan berdasarkan tujuan penelitian adalah analisis sentimen dan perbandingan hasil performa ketepatan klasifikasi sentimen pengguna layanan *telemedicine* Halodoc. Analisis dilakukan dengan menggunakan bantuan *software* program R. Langkah-langkah pengolahan dan analisis data adalah sebagai berikut.

1. Pemberian label sentimen *review* Halodoc
  - a. Perhitungan skor sentimen pada *text review* yang telah menjadi data bersih dengan menggunakan kamus *lexicon*.
  - b. Melakukan pemberian label sentimen *review* berdasarkan nilai skor. Jika nilai skor lebih besar daripada 0 maka dikategorikan positif sedangkan jika nilai skor lebih kecil dari 0 maka dikategorikan negatif.

## 2.5 Pre-Processing Data

- a. *Casefolding* yaitu tahapan merubah semua huruf kapital menjadi huruf kecil yang terkandung didalam *review text*.
- b. *Cleaning* yaitu tahapan menghapus tanda baca, angka, url dan karakter lain yang tidak diperlukan.
- c. *Spelling normalization* yaitu memperbaiki pengejaan kata yang salah atau disingkat seperti kata 'sama' ditulis 'sm', kata 'aku' ditulis 'ak' dan sebagainya.
- d. *Stemming* yaitu tahapan menghilangkan kata-kata imbuhan sehingga memperoleh kata dasar.
- e. *Tokenizing* yaitu tahapan menguraikan *review text* menjadi kata per kata.
- f. *Stopwords* yaitu tahapan menghilangkan kata-kata pada *review text* yang dianggap tidak penting dengan menggunakan *stopwords* yang telah dibuat.
- g. Melakukan pembobotan dengan TF-IDF.

## 2.6 Pengolahan dan Analisis Data

1. Membagi data *review* yang telah diberi label atau kategori positif dan negatif menjadi data latih (*training*) dan data uji (*testing*) menggunakan *K-fold cross validation*. *K-fold cross validation* merupakan metode pembagian data antara data training dan data testing yang dilakukan untuk memvalidasi dataset agar menemukan nilai akurasi yang terbaik dengan cara membagi dataset sebanyak k subset (Mukharil,2017). Penggunaan *5-fold* dengan proporsi antara data *training* dan *testing* dengan proporsi 80% : 20%.
2. Klasifikasi data menggunakan pendekatan algoritma *Naïve Bayes Classifier*.
3. Klasifikasi data menggunakan *Support vector machine*.
  - a. Membuat model SVM dengan menerapkan fungsi kernel
  - b. Melakukan optimasi parameter pada tiap jenis kernel
  - c. Memilih SVM kernel yang optimum
4. Melakukan evaluasi performa hasil klasifikasi.  
Evaluasi yang dilakukan adalah dengan cara menghitung ketepatan klasifikasi dari metode *Naïve Bayes Classifier* dan *Support vector machine* yang berdasarkan pada tingkat *Accuracy*, *Recall (Sensitivity)*, dan *Specitivity* jika data *balanced* sedangkan jika data *imbalanced* menggunakan *AUC* dan *G-Mean*.
5. Membuat visualisasi *review* melalui *wordcloud*.
7. Menginterpretasikan hasil analisis

## 2.7 Support Vector Machine

*Support vector machine* adalah metode prediksi dalam permasalahan klasifikasi dan regresi (Santosa 2007). Prinsip dasar *Support vector machine* pada permasalahan klasifikasi yang secara linear maupun non linear dengan memasukkan konsep kernel pada ruang kerja yang berdimensi tinggi dengan cara kerja memisahkan kedua *class* atau kelompok suatu data dengan adanya garis *hyperlane* optimal yang tujuannya untuk mencari *hyperlane* terbaik (Faisal and Nugrahadi 2017). Penggunaan *Support vector machine* untuk klasifikasi karena SVM mempunyai keunggulan dalam menggeneralisasi data jika dibandingkan dengan teknik sebelumnya (Vapnik et al. 1997). Adapun jenis kernel yang digunakan yaitu Kernel RBF sebagai berikut.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i, \mathbf{x}_j\|^2), \gamma > 0$$

## 2.8 Ketepatan Klasifikasi

Salah satu pengukuran performa klasifikasi dengan menggunakan tabulasi silang atau *confusion matrix* (Prasetyo 2012) yang dapat menunjukkan klasifikasi berdasarkan data aktual terhadap data hasil prediksi dari model yang terbentuk yang ditunjukkan pada Tabel 3.

**Tabel 3** *Confusion Matrix* Klasifikasi Dua Kelas

		Hasil prediksi	
		Positif	Negatif
Aktual	Positif	<i>True positive (TP)</i>	<i>False Positive (FP)</i>
	Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

## 2.9 Wordcloud

*Wordcloud* adalah salah satu metode analisis dalam *Text Mining* yang dapat memberi visualisasi data teks yang menggambarkan frekuensi kata-kata yang ditampilkan dalam bentuk yang menarik namun informatif. Semakin besar ukuran kata pada tampilan *wordcloud* maka semakin sering kata tersebut digunakan (Fellows 2012).

## 3. HASIL DAN PEMBAHASAN

### 3.1 Karakteristik Data

Karakteristik data hasil *review* pengguna layanan *telemedicine* Halodoc berdasarkan hasil pemberian label sentimen kategori negatif dan positif menunjukkan bahwa persentase jumlah sentimen *review* pengguna layanan *telemedicine* Halodoc kategori negatif sebesar 12,33% sedangkan sentimen kategori positif adalah 87,67% dari 5.687 *review*. Hal tersebut menunjukkan bahwa data *review* dapat dikatakan *imbalance* karena jumlah kategori *review* kategori positif dan negatif tidak seimbang sehingga pengukuran ketepatan klasifikasi data *review* pengguna layanan *telemedicine* Halodoc menggunakan *G-Mean* dan *AUC*.

**Tabel 4** Contoh Hasil Pra-proses Data

Proses	Review
<i>Review awal</i>	Pelayanan yang baik dan cepat tanggapnya. Terima kasih holodokter
<i>Case folding</i>	<u>pelayanan</u> yang baik dan cepat tanggapnya <u>terimakasih</u> halodokter
<i>Cleaning</i>	pelayanan yang baik dan cepat tanggapnya terimakasih holodokter
<i>Spelling</i>	pelayanan yang baik dan cepat tanggapnya terimakasih halodoc
<i>Normalization</i>	
<i>Stemming</i>	“layan”, “yang”, “baik”, “dan”, “cepat”, “tanggap”, “terimakasih”, “halodoc”
<i>Tokenizing</i>	“layan”, “yang”, “baik”, “dan”, “cepat”, “tanggap”, “terimakasih”, “halodoc”
<i>Stopwords</i>	layan baik cepat tanggap terimakasih holodoc

### 3.2 Naïve Bayes Classifier

*Naïve Bayes Classifier* merupakan salah satu metode klasifikasi dengan menggunakan probabilitas untuk mengklasifikasikan *review* tersebut kedalam kategori sentimen positif maupun negatif. Adapun pembagian data menjadi sebanyak *k-fold cross validation* maka akan dilakukan beberapa percobaan sebanyak *5-fold cross validation* untuk memperoleh model yang memberikan hasil optimum. *Fold* yang terpilih yaitu k-1.

Pengukuran performa klasifikasi *Naïve Bayes Classifier* dalam klasifikasi sentimen *review* positif dan negatif yang ditunjukkan pada Tabel 5.

**Tabel 5** Hasil Ketepatan Klasifikasi *Naïve Bayes Classifier*

<i>k-fold</i>	Akurasi	AUC	<i>G-Mean</i>
k-1	0.8777	0.5711	0.4008

Tabel 5 menunjukkan bahwa performa model yang terbentuk pada *fold* k-1 diperoleh nilai *Area Under Curve* (AUC) sebesar 0.5711 dan nilai *G-Mean* sebesar 0.4008. Hasil ketepatan klasifikasi *review* kategori sentimen positif dan negatif sehingga hasil ketepatan klasifikasi sentimen dapat dibentuk melalui *confusion matrix* yang ditunjukkan pada Tabel 6.

**Tabel 6** *Confusion Matrix Naïve Bayes Classifier*

Prediksi	Aktual		Jumlah observasi
	Positif	Negatif	
Positif	975	117	1092
Negatif	22	23	45
Jumlah observasi	997	140	1137

Tabel 6 menunjukkan bahwa algoritma *naïve bayes classifier* memberikan prediksi *review* sentimen kategori positif sebanyak 1092 dan dapat mengklasifikasikan secara benar sebanyak 975 *review* sedangkan yang terklasifikasi salah sebanyak 117 *review*. Hasil prediksi *review* sentimen kategori negatif sebanyak 45 *review* dapat diklasifikasi secara benar sebanyak 23 sedangkan yang diklasifikasi salah sebanyak 22 *review*.

### 3.3 Support Vector Machine

*Support Vector Machine* memiliki konsep mencari *hyperplane* terbaik yang dijadikan sebagai pemisah dua kelas data dengan cara memaksimalkan margin pemisah. Adapun kernel trik yang mampu diterapkan pada data berdimensi tinggi. Kernel yang digunakan pada penelitian ini yaitu *Radial Basis Function (RBF)*. Adapun penggunaan pembagian data pada SVM menggunakan *5-fold cross validation* untuk memperoleh model yang memberikan hasil optimum.

SVM *radial basis function (RBF)* fungsi *Gaussian* dengan menggunakan parameter  $\gamma$ . Adapun parameter yang digunakan yaitu C menunjukkan jarak margin optimasi untuk menghindari kesalahan klasifikasi SVM dan  $\gamma$  menunjukkan percepatan fungsi pada kernel *RBF* agar mendapatkan akurasi dari model klasifikasi yang optimal. parameter C dan  $\gamma$  SVM Kernel *RBF* yang optimum adalah parameter C=1000 dan  $\gamma=0.1$ . Pengukuran performa klasifikasi *Support Vector Machine* dalam klasifikasi sentimen *review* positif dan negatif yang ditunjukkan pada Tabel 7.

**Tabel 7** Hasil Ketepatan Klasifikasi SVM Kernel *RBF*

<i>k-fold</i>	Akurasi	AUC	<i>G-Mean</i>
k-1	0.861	0.60149	0.49311

Tabel 7 menunjukkan bahwa klasifikasi menggunakan SVM kernel *RBF* yang terbaik dengan nilai parameter C=1000 dan  $\gamma=0,1$  pada fold k-2 menghasilkan nilai AUC sebesar 0.5531308 dan *G-Mean* sebesar 0.3367 sehingga hasil ketepatan klasifikasi sentimen dapat dibentuk melalui *confusion matrix* yang ditunjukkan pada Tabel 8.

**Tabel 8** Confusion Matrix SVM RBF

Prediksi	Aktual		Jumlah observasi
	Positif	Negatif	
Positif	943	104	1047
Negatif	54	36	90
Jumlah observasi	997	140	1137

Tabel 8 menunjukkan bahwa algoritma *Support Vector Machine* Kernel *RBF* memberikan prediksi *review* sentimen kategori positif sebanyak 1047 dan dapat mengklasifikasikan secara benar sebanyak 943 *review* sedangkan yang terklasifikasi salah sebanyak 104 *review*. Hasil prediksi *review* sentimen kategori negatif 90 *review* dapat terklasifikasi secara benar sebanyak 36 sedangkan yang terklasifikasi salah sebanyak 54 *review*.

### 3.4 Perbandingan Hasil Klasifikasi Algoritma *Naïve Bayes Classifier* dan *Support Vector Machine*

Perbandingan hasil ketepatan klasifikasi antara kedua Algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* dapat ditunjukkan pada Tabel 9.

**Tabel 9** Hasil Perbandingan Algoritma NBC dan SVM

Algoritma	Akurasi	AUC	G-Mean
NBC	0.8777	0.5711	0.4008
SVM Kernel RBF	<b>0.861</b>	<b>0.60149</b>	<b>0.49311</b>

Tabel 9 menunjukkan bahwa Algoritma NBC memiliki nilai akurasi sebesar 0.8777, AUC sebesar 0.5711 dan G-Mean sebesar 0.4008 sedangkan Algoritma SVM Kernel *RBF* memiliki nilai akurasi sebesar 0.861, AUC sebesar 0.60149 dan nilai G-Mean sebesar 0.49311 yang artinya berdasarkan nilai akurasi, AUC dan G-Mean SVM Kernel *RBF* lebih baik daripada NBC pada pengklasifikasian *review* sentimen pengguna layanan *telemedicine* Halodoc.

### 3.5 Visualisasi *Wordcloud*

Visualisasi *wordcloud* dapat memberikan informasi mengenai kata-kata apa yang sering muncul pada data *review* pengguna layanan *telemedicine* Halodoc sesuai dengan kategori sentimennya. Frekuensi kemunculan kata-kata yang sering muncul berdasarkan ukuran *font* yang terbentuk. Semakin besar ukurannya maka kata tersebut sering muncul yang dapat ditunjukkan pada Gambar 1 dan Gambar 2 sebagai berikut.



**Gambar 1** *Wordcloud* Sentimen Negatif



**Gambar 2** *Wordcloud* Sentimen Positif

Gambar 1 menunjukkan bahwa kata-kata sentimen kategori negatif yang sering muncul adalah Kata “dokter” karena menurut beberapa pengguna terdapat dokter yang

kurang informatif dalam menjawab pertanyaan pengguna. Kata “bayar” beberapa pengguna mengeluhkan sistem pembayaran Halodoc. Kemudian Kata “apotek” karena beberapa pengguna merasa layanan apotek Halodoc kurang lengkap dan kata “customer service” yang dirasa beberapa pengguna kurang responsive dalam menangani masalah pengguna. Sehingga diharapkan customer service lebih responsif dalam membantu menangani masalah para pengguna layanan *telemedicine* Halodoc.

Gambar 2 menunjukkan bahwa kata-kata sentimen kategori positif yang sering muncul adalah kata “bantu” karena pengguna merasa sangat terbantu dengan layanan aplikasi Halodoc terutama di masa pandemi dan praktis melalui *online*. Kemudian kata “dokter”, “cepat”, “ramah” dan respon” karena pengguna merasa dokter Halodoc dapat melayani dengan respon cepat dan ramah.

#### 4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan dapat diperoleh kesimpulan bahwa persentase jumlah sentimen *review* pengguna layanan *telemedicine* Halodoc kategori negatif sebesar 12,33% sedangkan sentimen kategori positif adalah 87,67% dari 5.687 *review*. yang artinya jumlah *review* kategori positif lebih banyak daripada kategori negatif. Performa ketepatan klasifikasi Algoritma NBC memiliki nilai akurasi sebesar 0.8777, AUC sebesar 0.5711 dan G-Mean sebesar 0.4008 sedangkan Algoritma SVM Kernel *RBF* memiliki nilai akurasi sebesar 0.861, AUC sebesar 0.60149 dan nilai G-Mean sebesar 0.49311 yang artinya berdasarkan nilai akurasi, AUC dan G-Mean SVM Kernel *RBF* lebih baik daripada NBC pada pengklasifikasian *review* sentimen pengguna layanan *telemedicine* Halodoc. Kata-kata sentimen kategori negatif yang sering muncul adalah “bayar”, “apotek” dan “customer service” sedangkan Kata-kata sentimen kategori positif yang sering muncul adalah “dokter” “bantu” “ramah”, “cepat”, “aplikasi”, “obat” dan “konsultasi”.

#### DAFTAR PUSTAKA

- Aggarwal, C. C. (2018), *Machine learning for text, Machine Learning for Text*.  
<https://doi.org/10.1007/978-3-319-73531-3>.
- Coelho, K. R. (2011), “Identifying Telemedicine Services to Improve Access to Specialty Care for the Underserved in the San Francisco Safety Net,” *International Journal of Telemedicine and Applications*, 2011, 1–14. <https://doi.org/10.1155/2011/523161>.
- Faisal, M. R., and Nugrahadi, D. T. (2017), *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*, Kalimantan Selatan: Scripta Cendekia.
- Fellows, I. (2012), “Wordcloud: Word clouds,” *R package version*.
- Fitri, A. (2020), “Selama pandemi covid-19, instalasi Halodoc meningkat 10 kali lipat,” *Kontan*, Available at <https://kesehatan.kontan.co.id/news/selama-pandemi-covid-19-instalasi-halodoc-meningkat-10-kali-lipat>.
- Liu, B. (2010), “Sentiment Analysis and Subjectivity in: Handbook of Natural Language Processing, Second Edition,” *Handbook of Natural Language Processing, Second Edition*, 2.
- Nasution, M. R. A., and Hayaty, M. (2019), “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *Jurnal Informatika*, 6, 226–235. <https://doi.org/10.31311/ji.v6i2.5129>.
- Prasetyo, E. (2012), *Data Mining: konsep dan aplikasi menggunakan MATLAB*, Yogyakarta: Andi.
- Santosa, B. (2007), *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu.
- Vapnik, V., Golowich, S. E., and Smola, A. (1997), “Support vector method for function



approximation, regression estimation, and signal processing,” in *Advances in Neural Information Processing Systems*.

Wardani, N. R., and Erfina, A. (2021), “Analisis Sentimen Masyarakat Terhadap Layanan Konsultasi Dokter Menggunakan Algoritma Naive Bayes,” in *Seminar Nasional Sistem Informasi dan Manajemen Informatika*, Universitas Nusa Putra, pp. 11–18.