

## Penerapan Algoritma KNN Pada Analisis Sentimen Review Aplikasi Peduli Lindungi

Puji Astuti<sup>1</sup>, Nuzuliarini Nuris<sup>2</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri  
Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar, Jakarta Timur-13620, Indonesia

<sup>2</sup>Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika  
Jl. Kramat Raya No. 98 Senen, Jakarta Pusat-10420, Indonesia

e-mail: [Puji.pat@nusamandiri.ac.id](mailto:Puji.pat@nusamandiri.ac.id), [Nuzuliarini.nzn@bsi.ac.id](mailto:Nuzuliarini.nzn@bsi.ac.id)

Artikel Info : Diterima : 13-06-2022 | Direvisi : 21-07-2022 | Disetujui : 27-07-2022

**Abstrak** - Covid-19 membuat seluruh Indonesia bahkan seluruh Dunia menjadi negara yang mati akibat virus sars covid-19. Pemerintah Indonesia mengupayakan segala hal demi masyarakat agar terhindar dari wabah virus sars covid-19 ini. Membangun sebuah aplikasi pemerintahan Indonesia yang digunakan sebagai wadah dalam perjalanan, deteksi terjangkitnya virus covid-19 bahkan terdeteksi sertifikat vaksinasi yang diselenggarakan serentak oleh Pemerintah Indonesia. Aplikasi PeduliLindungi yang diunduh melalui aplikasi play store yang memberikan opini dari berbagai individu masyarakat. Opini tersebut penulis rangkum untuk melakukan penelitian dengan pengklasifikasian teks *review*, dimana jumlah data yang dirangkum sebanyak 200 data *review*, terdiri dari 100 data *review* positif dan 100 data *review* negatif, dimana sentiment tersebut berhubungan dengan kalimat: bagus, cepat, kecewa, bodoh, tidak layak. Melakukan klasifikasi dengan metode *K-Nearest Neighbor* terbukti bahwa dengan metode ini mendapatkan klasifikasi nilai akurasi baik dimana nilai akurasi yang didapat sebesar 81,72% dengan AUC 0.856, pengelompokan nilai AUC 0,856 termasuk kedalam kelompok *Good Classification*, sehingga dalam hal ini metode *K-Nearest Neighbor* mampu menganalisa sentiment review aplikasi PeduliLindungi. Tujuan dari penelitian yaitu untuk melihat apakah dengan dataset sebanyak 200 data review yang terdiri data positif dan data negatif memiliki nilai akurasi yang baik dalam pengklasifikasian data.

Kata Kunci : Sentimen Analisis, K-Nearest Neighbor, Peduli Lindungi

**Abstracts** - Covid-19 made all of Indonesia and even the whole world a country that died from the sars covid-19 virus. The Indonesian government is trying everything for the sake of the community to avoid the outbreak of the sars covid-19 virus. Building an Indonesian government application that is used as a container on the way, the detection of the covid-19 virus was even detected by a vaccination certificate held simultaneously by the Indonesian government. The PeduliLindungi application downloaded through the play store provides opinions from various individuals of the community. The author summarizes the Opinin to conduct research on classifying the text of the review, where the amount of data summarized as much as 200 reviews, consisting of 100 positive review data and 100 negative review data, where the sentiment is related to the sentence : good, fast, disappointed, stupid, not worth it. Doing classification with *K-Nearest Neighbor* method is proven that with this method to get a good accuracy value classification where the accuracy value obtained by 81.72% with AUC 0.856, grouping AUC value 0.856 included into the *Good Classification* group, so in this case the *K-NN* method is able to analyze sentiment review PeduliLindungi application.

Keywords : Sentiment analysis, K-nearest neighbor, Peduli Lindungi

### PENDAHULUAN

PeduliLindungi merupakan aplikasi yang sudah dibuat untuk membantu pemerintah dalam melacak penyebaran Covid-19. Aplikasi peduliLindungi berperan di masyarakat untuk membagikan data lokasi dimana kita singgah serta riwayat vaksin dan penelusuran kontak dengan penderita covid-19 dilakukan (<https://www.pedulilindungi.id/>). Dalam mendapatkan informasi pada internet kita dapat menyalurkan pendapat dari aplikasi PeduliLindungi. Pendapat yang diberikan bisa dalam bentuk pendapat positif maupun pendapat negative. Banyaknya informasi mengenai saran yang ada dimeida online kita dapat memanfaatkan data-data tersebut untuk diolah hingga mendapatkan pengetahuan yang baru.



Text Mining dimanfaatkan untuk pengolahan dokumen yang bermanfaat dari berbagai data dimana sumber datanya yaitu berbentuk teks, yang memiliki format tidak terstruktur (Doloksaribu & Samuel, 2022). Tahapan pada text mining dalam preprocessing data bertujuan untuk mencari kata yang mewakili isi dari dokumen sehingga mampu melakukan analisis terhubung antar dokumen (Gata, 2017).

Pada metode KNN text mining yaitu tujuannya untuk mendapatkan nilai pada teks, dimana pada text mining kali ini adalah *classification* (Sahara & Permana, 2019).

Analisis sentiment yaitu teknik untuk mengesktrak data teks dalam mendapatkan informasi tentang sebuah sentiment bernilai positif, ataupun negatif (F. V. Sari & Wibowo, 2019). Analisis sentimen diperlukan dalam pengambilan keputusan yang terbaik. Klasifikasi sentimentil tujuannya untuk mengatasi masalah dengan otomatis mengklasifikasi ulasan-ulasan pengguna menjadi pendapat yang positif maupun pendapat yang negatif (Indrayuni, 2017). Analisis sentimen dapat memahami serta mengolah tekstual data secara otomatis untuk melihat sentiment yang terkandung dalam sebuah opini (Doloksaribu & Samuel, 2022). Penelitian terdahulu dengan judul analisis sentiment aplikasi peduli lindungi dengan algoritma Support Vector Machine dengan nilai akurasi sebesar 89,05% dan dengan Naïve Bayes dengan nilai akurasi sebesar 83,25% (Doloksaribu & Samuel, 2022).

Terdapat penelitian terdahulu dengan menggunakan algoritma K-Nearest Neighbor dimana review dengan Bahasa Indonesia pada review objek wisata dunia fantasi (R. Sari, 2020).

Pada penelitian sebelumnya tentang text mining sentiment review restoran menggunakan algoritma K-Nearest-Neighbor (KNN) mendapatkan akurasi sebesar 93% berdasarkan opini dari pribadi konsumen yang datang ke restoran (Indrayuni, 2017). K-Nearest Neighbor (K-NN) termasuk kedalam kelompok *instance-based learning*, dimana algoritma ini juga merupakan proses *lazy learning* (Cahyanti, Rahmayani, & Husniar, 2020). Dimana K-NN dilakukan untuk memecahkan kelompok k objek dalam data training yang paling dekat dengan objek pada data testing. Pada penelitian tentang review agent travel dengan metode K-NN mendapatkan nilai akurasi 87,00% dan AUC 0,916 sehingga kelompok AUC berada pada tahap klasifikasi *Excelent* dan menyatakan bahwa K-NN dalam pengklasifikasian memiliki hasil yang akurat dan ulasan perjalanan agen trabvel (Ernawati & Wati, 2018).

Penelitian yang berjudul analisis sentiment review terhadap aplikasi PeduliLindungi menggunakan algoritma *K-Nearest Neighbor* yang tujuannya untuk mengetahui nilai akurasi yang didapatkan. Dari beberapa penelitian *text mining* dengan metode *K-Nearest Neighbor* (K-NN) termasuk kedalam pengklasifikasian yang mempunyai nilai akurasi tinggi dengan algoritma tetangga atau teman terdekat untuk klasifikasi data review.

## METODE PENELITIAN

Metode yang digunakan yaitu eksperimen dengan tahapan sebagai berikut :

1. Pengumpulan Data  
Peneliti mengambil data dari situs play sore review aplikasi PeduliLindungi. Data diambil terdiri dari 100 review positif dan 100 review negative.
2. Pengolahan Awal Data  
*Preprocessing* pada dataset yang dilakukan yaitu :
  - a. *Tokenize*  
*Tokenize* merupakan tahap *preprocessing* untuk menghapus tanda baca ataupun simbol yang bukan angka.
  - b. *Filter Stopword*  
Pada tahap *filter stopwords* menghapus kata yang tidak memiliki makna seperti “di, yang, ke, di”.
3. Metode yang diusulkan  
Pada penelitian ini, metode yang digunakan yaitu metode K-Nearest Neighbor (KNN) untuk mengetahui nilai akurasi yang dihasilkan. Metode K-NN memiliki beberapa keunggulan, diantaranya pelatihan sederhana, cepat, mudah dipahami, dan efektif apabila ukuran data terlalu besar (Cahyanti et al., 2020).
4. Eksperimen dan Pengujian Metode  
Pada penelitian ini penggunaan aplikasi dalam perhitungan nilai akurasi yaitu *software* RapidMiner dalam proses pengolahan data.
5. Evaluasi dan Validasi Hasil  
Penelitian ini dilakukan dengann menggunakan *fold cross validation*. Akurasi diukur dengan *Confusion Matrix*. Kurva ROC digunakan untuk mengukur nilai AUC. Dimana cross validation adalah action yang dilakukan dalam mencari akurasi dari masing-masing metode dengan membagi data training dan data testing (Puspita & Widodo, 2021). Cara kerja dari algoritma K-NN yaitu data yang diuji diklasifikasikan berdasarkan keanggotaan terdekat yang terbanyak dari data uji (Novita, Harsani, & Qur'ania, 2018)



yang muncul, gmna saya bisa melakukan perjalanan ke luar kota kalo apknya bermasalah kek bgni. Sedangkan semua di lihat dari apk ini. Trimakasih. Mohon segera di perbaiki, jika trus bgni saya merasa tidak puas atas pelayann dari apk ini, padahal data yang digunakan tdk sembarangn namun ada yang bermasalah dari data saya sendiri Mohon di PERBAIKI

bisa melakukan perjalanan ke luar kota kalo apknya bermasalah kek bgni Sedangkan semua di lihat dari apk ini Trimakasih Mohon segera di perbaiki jika trus bgni saya merasa tidak puas atas pelayann dari apk ini padahal data yang digunakan tdk sembarangn namun ada yang bermasalah dari data saya sendiri Mohon di PERBAIKI

Sumber : Peneliti (2022)

2. *Stemp (Snowball)*

Tabel 2. *Stemp (Snowball)*

Teks Sebelum di <i>Stemp (Snowball)</i>	Teks Sesudah di <i>Stemp (Snowball)</i>
Mohon di perbaiki, saya login di apk ini, sesuai identitas ktp dan sudah vaksin 2 kali, tapi di keterangan sertivikat vaksin saya nama orang lain yang muncul, gmna saya bisa melakukan perjalanan ke luar kota kalo apknya bermasalah kek bgni. Sedangkan semua di lihat dari apk ini. Trimakasih. Mohon segera di perbaiki, jika trus bgni saya merasa tidak puas atas pelayann dari apk ini, padahal data yang digunakan tdk sembarangn namun ada yang bermasalah dari data saya sendiri Mohon di PERBAIKI	Mohon di perbaiki saya login apk ini sudah sesuai identitas ktp dan sudah vaksin tapi di keterangan sertivikat vaksin saya nama orang lain yang muncul gmna saya bisa melakukan perjalanan ke luar kota kalo apknya bermasalah kek bgni Sedangkan semua dilihat dari apk ini Trimakasih Mohon segera diperbaiki jika trus bgni saya merasa tidak puas atas pelayann dari apk ini padahal data yang digunakan tdk sembarangn namun ada yang bermasalah dari data saya sendiri Mohon di PERBAIKI

Sumber : Peneliti (2022)

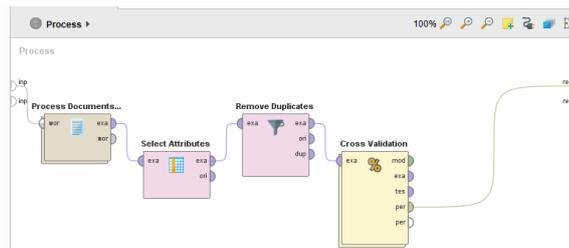
3. *Filter Stopwords (English)*

Tabel 3. *Filter Stopwords (English)*

Teks Sebelum di <i>Filter Stopwords (English)</i>	Teks Sesudah di <i>Filter Stopwords (English)</i>
Mohon di perbaiki, saya login di apk ini, sesuai identitas ktp dan sudah vaksin 2 kali, tapi di keterangan sertivikat vaksin saya nama orang lain yang muncul, gmna saya bisa melakukan perjalanan ke luar kota kalo apknya bermasalah kek bgni. Sedangkan semua di lihat dari apk ini. Trimakasih. Mohon segera di perbaiki, jika trus bgni saya merasa tidak puas atas pelayann dari apk ini, padahal data yang digunakan tdk sembarangn namun ada yang bermasalah dari data saya sendiri Mohon di PERBAIKI	mohon di perbaiki saya login di apk ini sesuai identitas ktp dan sudah vaksin kali tapi di keterangan sertivikat vaksin saya nama orang lain yang muncul gmna saya bisa melakukan perjalanan ke luar kota kalo apknya bermasalah kek bgni sedangkan semua di lihat dari apk ini trimakasih mohon segera di perbaiki jika trus bgni saya merasa tidak pua ata pelayann dari apk ini padah data yang digunakan tdk sembarangn namun ada yang bermasalah dari data saya sendiri mohon di perbaiki

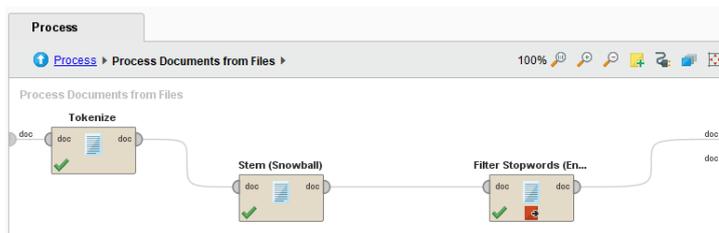
Sumber : Peneliti (2022)

Penelitian ini menggunakan metode *K-Nearest Neighbor* dimana menggunakan 200 data review. Disamping itu peneliti juga menggunakan 10-fold cross validation dalam pengujian model, yang mana akan terbentuk secara acak. Pada dasarnya 10-fold cross validation yaitu 1:9, terbagi antara 1 terbagi untuk testing dan 9 untuk training. Dibawah ini adalah tampilan dari pemrosesan data :



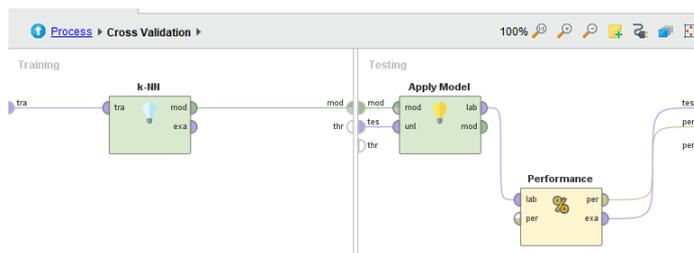
Sumber : Peneliti (2022)

Gambar 1. Tahap Proses Validasi Data



Sumber : Peneliti (2022)

Gambar 2. Proses Pemrosesan Klasifikasi Data



Sumber : Peneliti (2022)

Gambar 3. Proses Pemasukan Metode K-NN

Dibawah ini adalah tabel eksperimen dalam merubah nilai K :

Tabel 4. Eksperimen dengan Perubahan Nilai K

<b>K</b>	<b>Accuracy</b>	<b>AUC</b>
1	72,55%	0,500
3	75,61%	0,825
5	79,63%	0,858
7	78,11%	0,831
10	79,71%	0,860
13	80,74%	0,858
15	80,21%	0,853
17	79,68%	0,858
20	81,74%	0,861

Sumber : Peneliti (2022)

Berdasarkan dari tabel diatas dengan merubah nilai K, maka hasil accuracy tertinggi berada di posisi K=20 dengan nilai accuracy yaitu 81,74% dan nilai AUC 0,861.

Tabel 5. Hasil accuracy menggunakan Algoritma K-Nearest Neighbor

Accuracy : 81,74% +/- 5,39% (micro average : 81,73%)			
	True data_negatif	True data_positif	Class precision
Pred. data_negatif	96	32	75.00%
Pred. data_positif	4	65	94.20%
Class recall	96.00%	67.01%	

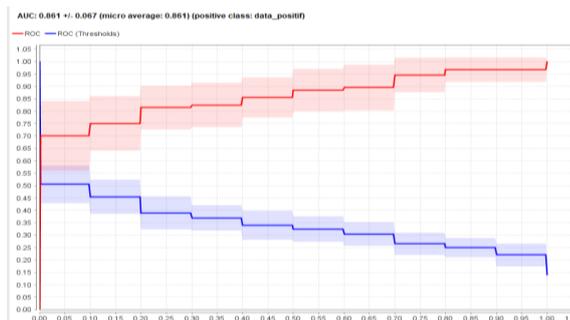
Sumber : Peneliti (2022)

Nilai Accuracy dari confusion matrix tersebut yaitu :

$$Accuracy = \frac{(TN + TP)}{(TN + FN + TP + FP)}$$

$$Accuracy = \frac{96 + 65}{96 + 4 + 65 + 32}$$

$$Accuracy = \frac{161}{197} = 0,81 = 81\%$$



Sumber : Peneliti (2022)

Gambar 4. Grafik AUC dengan Metode *K-Nearest Neighbor*

AUC atau *Area Under Curve* yang didapat yaitu 0,861 dimana klasifikasi keakuratan dalam penelitian ini masuk kedalam kelompok *Good Classification*.

Klasifikasi keakuratan diagnose menggunakan AUC (Gorunescu, 2011) yaitu :

1. 0.90-1.00 = Excellent Classification.
2. 0.80-0.90 = Good Classification.
3. 0.70-0.80 = Fair Classification.
4. 0.60-0.70 = Poor Classification.
5. 0.50-0.60 = Failure.

## KESIMPULAN

Pada penelitian ini dilakukan klasifikasi teks review aplikasi PeduliLindungi menggunakan metode *K-Nearest Neighbor*, dimana peneliti menggunakan 200 data *review*, yang terdiri dari 100 review data positif dan 100 review data negative, dimana sentiment tersebut berhubungan dengan kalimat : bagus, cepat, kecewa, bodoh, tidak layak. Nilai akurasi yang didapat dengan merubah nilai *K* menjadi *K*=20 nilai akurasi sebesar 81,74% dan nilai AUC sebesar 0,861 dimana termasuk kedalam *Good Classification*. Untuk meningkatkan nilai akurasinya, maka diperlukan data mungkin agak lebih banyak. Selain itu, dapat dilakukan untuk penelitian lanjutan untuk melakukan prediksi dengan algoritma yang lainnya seperti SVM, *decision tree* dan lain sebagainya.

## REFERENSI

- Cahyanti, D., Rahmayani, A., & Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2), 39–43. <https://doi.org/10.33096/ijodas.v1i2.13>
- Doloksaribu, H. P., & Samuel, Y. T. (2022). Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi. *Jurnal Teknologi Informasi: Jurnal Keilmuan Dan Aplikasi Bidang Teknik Informatika*, 16(1), 1–11.
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, 6(1), 64–69.
- Gata, W. (2017). Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS, 6, 1–13.
- Gorunescu, F. (2011). *Data Mining : Concepts, Models and Techniques*.
- Indrayanti, Sugianti, D., & Al Karomi, M. A. (2017). Optimasi Parameter K Pada Algoritman K-Nearest Neighbour Untuk Kklasifikasi Penyakit Diabetes Mellitus, 823–829.
- Indrayuni, E. (2017). Text Mining dalam Analisis Sentimen Review Restoran, III(2), 109–114.
- Novita, S., Harsani, P., & Qur'ania, A. (2018). Penerapan K-Nearest Neighbor (KNN) untuk Klasifikasi Anggrek Berdasarkan Karakter Morfologi Daun dan Bunga. *Komputasi*, 15(1), 118–125.
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN , Decision Tree , dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS, 5(4), 646–654.
- Sahara, S., & Permana, R. A. (2019). Methode K-Nn for Analys Sentiment Review Kids Apps. *JST (Jurnal Sains Dan Teknologi)*, 8(2), 127–137. <https://doi.org/10.23887/jstundiksha.v8i2.21240>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686.
- Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1), 10–17. <https://doi.org/10.31294/evolusi.v8i1.7371>