

KOMPARASI METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR TERHADAP ANALISIS SENTIMEN PENGGUNA APLIKASI PEDULILINDUNGI

Ari Putra Wibowo, Wachid Darmawan, Nurul Amalia

STMIK Widya Pratama

ariputra.stmikwp@gmail.com, wachiddw.stmikwp@gmail.com, amalia.0121@gmail.com

Abstrak

Penggunaan aplikasi PeduliLindungi sebagai sarana tracing mendapat berbagai respon dari masyarakat yang disampaikan melalui komentar (ulasan) di playstore. Ada berbagai jenis komentar yang menyatakan aplikasi PeduliLindungi mudah digunakan dan ada juga komentar yang menyatakan aplikasi PeduliLindungi belum siap untuk digunakan. Dengan adanya komentar tersebut dapat dilakukan analisis sentimen untuk mengetahui sentimen atau opini dari pengguna aplikasi PeduliLindungi. Berdasarkan uraian diatas peneliti melakukan analisis sentimen terhadap penggunaan aplikasi PeduliLindungi dengan menggunakan data yang diperoleh dari komentar pengguna di playstore. Data yang digunakan sebanyak 321 data yang dikelompokkan kedalam kelas positif dan negatif. Penelitian ini membuat model klasifikasi dengan algoritma Naïve Bayes dan K-Nearest Neighbor, selain itu pada penelitian ini juga dilakukan komparasi kinerja pada algoritma klasifikasi Naïve Bayes dan K-Nearest Neighbor dengan menggunakan tools Rapid Miner versi 9.9. Berdasarkan eksperimen yang dilakukan diperoleh hasil akurasi sebesar 70,46%, untuk algoritma Naïve Bayes dan akurasi sebesar 73,33% untuk algoritma K-Nearest Neighbor. Dari kedua metode yang digunakan untuk mengetahui tingkat akurasi yang digunakan, maka bisa diambil kesimpulan bahwa metode K-Nearest Neighbor memiliki tingkat akurasi yang lebih baik dibandingkan dengan metode Naïve Bayes. Untuk penelitian selanjutnya diharapkan menambah jumlah dataset (data ulasan) yang ada, serta bisa menambahkan metode klasifikasi lainnya ataupun bisa menggunakan metod seleksi fitur yang ada di data mining.

Kata kunci: PeduliLindungi, Sentimen Analisis, Naïve Bayes, K-NN.

1. Pendahuluan

Aplikasi PeduliLindungi merupakan salah satu aplikasi mobile yang dibuat oleh pemerintah untuk membantu dalam mempermudah proses tracing dan tracking untuk mengurangi penyebaran *Coronavirus Disease* (COVID-19) (Kominfo, 2021). Pada dasarnya aplikasi PeduliLindungi ini sangatlah membantu masyarakat dalam melakukan aktivitas kegiatan karena aplikasi PeduliLindungi dapat memberikan informasi seperti tracing (penelusuran), tracking (pelacakan) dan fencing (pembatasan ruang gerak). Namun semakin banyaknya masyarakat yang menggunakan aplikasi PeduliLindungi, semakin banyak opini yang muncul dikalangan masyarakat yang menggunakan aplikasi PeduliLindungi.

Opini dari masyarakat terkait penggunaan aplikasi PeduliLindungi perlu juga diketahui. Salah satu platform yang biasa digunakan masyarakat untuk memperoleh aplikasi PeduliLindungi adalah melalui google Play Store. Play Store merupakan salah satu platform digital yang berisi aplikasi-aplikasi berbasis Android yang dikembangkan dan dikelola oleh Google (Jatimtech, 2021). Selain menyediakan berbagai jenis layanan digital, Play Store juga sering digunakan oleh para pengguna aplikasi sebagai tempat untuk menyampaikan opini atau ulasan terkait fitur-fitur layanan yang terdapat pada aplikasi.

Studi literatur yang ada, seperti: penelitian yang dilakukan oleh Rani Pustita dan Widodo tentang sentimen analisis pengguna layanan

BPJS, dengan menggunakan tool Rapidminer dengan membandingkan beberapa metode klasifikasi untuk mengetahui performa metode yang paling baik (Puspita and Widodo 2021). Penelitian lainnya seperti Samsir, dkk yang juga meneliti sentimen analisis pembelajaran daring, menggunakan metode Naïve Bayes untuk mengetahui akurasi dari metode tersebut, untuk mengetahui tingkat akurasinya menggunakan tool Rapidminer (Samsir et al. 2021).

Sedangkan penelitian sentimen analisis oleh Kurniawan dan Apriliani tentang analisis sentimen tentang kasus virus corona di Indonesia, disampaikan dalam penelitiannya menggunakan metode Naïve Bayes dan tidak disampaikan tool apa yang digunakan dalam penelitiannya (Kurniawan and Apriliani 2020). Sentimen analisis atau bisa juga disebut dengan opinion mining termasuk dalam text mining (Hotho, Nürnberger, and Paaß 2005). Salah satu ilmu yang dapat menghasilkan suatu informasi yang bersumber dari data teks dari situs web, artikel, ulasan dan lainnya atau biasa disebut Teks Mining. Untuk menghasilkan informasi pada text mining digunakan teknik tertentu (Witten 2014) (Zhang and Liu 2016). Ada banyak teknik atau metode yang bisa diterapkan dalam melakukan proses pencarian informasi.

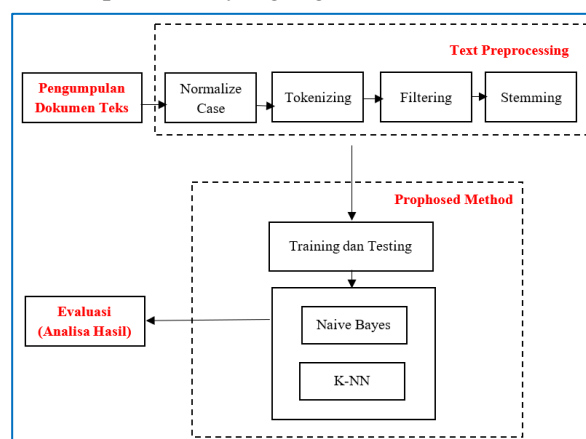
Penelitian yang dilakukan tentang analisis sentimen pada aplikasi PeduliLindungi akan melakukan komparasi dua metode yaitu metode naïve Bayes dan metode K-Nearest Neighbor (K-NN). Menurut Zaidi salah satu metode klasifikasi yang paling efektif dan efisien adalah Metode Naïve Bayes (Zaidi and Cerquides 2013). Selain itu metode tersebut masuk kedalam metode *machine learning* dengan penyelesaian berdasarkan probabilitas. Dan disebutkan jika metode Naïve Bayes memiliki kemampuan pemrosesan dengan kecepatan yang tinggi untuk melakukan klasifikasi teks dalam data besar (Fitriyyah, Safriadi, and Pratama 2019). Selain metode Naïve Bayes, metode yang digunakan adalah K-NN. Salah satu metode yang cukup populer diantaranya K-NN (Han, Kamber, and Pei 2012). K-NN termasuk ke dalam metode *supervised learning*, metode K-NN merupakan teknik *lazy-learning*. Dimana dalam melakukan proses klasifikasi berdasarkan pada data yang

jaraknya dekat (Cahyanti, Rahmayani, and Husniar 2020).

Berdasarkan permasalahan yang ada, maka penelitian analisis sentimen pada aplikasi PeduliLindungi akan melakukan komparasi antara metode Naïve Bayes dan metode K-Nearest Neighbor untuk membandingkan tingkat akurasi yang dihasilkan. Agar dapat mengetahui model terbaik dari kedua metode tersebut yang dapat digunakan analisis sentimen pada aplikasi PeduliLindungi.

2. Metode Penelitian

Dalam melakukan penelitian analisis sentimen terhadap aplikasi PeduliLindungi untuk mengetahui metode terbaik dari Naïve Bayes dan K-Nearest Neighbor. Pada gambar 1 menjelaskan metode penelitian yang digunakan.



Gambar 1. Metode Penelitian

2.1 Pengumpulan Dokumen Teks

Dokumen teks (*dataset*) yang digunakan diperoleh dari komentar (ulasan) pengguna Aplikasi PeduliLindungi dari play store (<https://play.google.com/store/apps/details?id=com.telkom.tracencare&hl=in&gl=US>). Dari 321 data ulasan yang ada akan di kelompokkan berdasarkan ulasan yang bernilai positif sebanyak 105 ulasan, sedangkan komentar atau ulasan yang bernilai negatif juga sebanyak 105 ulasan, sisanya merupakan data ganda atau data lainnya.

2.2 Text Preprocessing

Setelah dokumen teks didapatkan, dilanjutkan ketahap *text preprocessing*. *Text preprocessing* digunakan untuk menampilkan informasi yang

ada di dokument teks: *Normalize Case, Tokenizing, Filtering dan Stemming*.

2.3 Proposed Method

Setelah proses *text preprocessing* selesai dilanjutkan untuk melakukan klasifikasi menggunakan metode Naïve Bayes dan K-NN. Penggunaan metode tersebut digunakan untuk mengetahui tingkat akurasi dari metode tersebut. Dokument teks yang dihasilkan akan digunakan untuk data *training* dan data *testing*. Pembagian data ini menggunakan teknik *cross validation*.

2.4 Evaluasi (Analisa Hasil)

Berdasarkan hasil komparasi yang digunakan menggunakan Rapid Miner 9.9, akan digunakan untuk membuat analisa dari penelitian yang sudah dilakukan. Dari tingkat akurasi yang dihasilkan pada Rapid Miner 9.9 dapat digunakan untuk mengetahui metode terbaik.

3. Hasil dan Pembahasan

Pada penietian ini yang menggunakan data komentar (ulasan) pada aplikasi PeduliLindungi yang ada di Play Store dengan jumlah data ulasan sebanyak 321 data. Dari ulasan yang ada kemudian di filter secara manual dan selanjutnya di kelompokkan berdasarkan ulasan yang bernilai positif sebanyak 105 ulasan, sedangkan komentar atau ulasan yang bernilai negatif juga sebanyak 105 ulasan. Sedangkan sisanya sebanyak 111 ulasan tidak digunakan karena data ganda atau ulasanya lainnya. Setelah dokumen teks diperoleh, selanjutnya data tersebut masuk ketahap *Text Preprocessing* meliputi:

3.1 Normalize Case

Tahap *Normalize Case* digunakan untuk merubah teks (menjadi huruf kecil) pada Dokumen Teks yang akan digunakan.

Tabel 1. Tahap *Normalize Case*

Sebelum <i>Normalize Case</i>
App ini sangat bagus, ringan, mudah digunakan, mudah di akses, berisi informasi petunjuk perjalanan. Namun saat pengisian di menu EHAC beberapa pelabuhan domestik belum ada pada daftar pencarian pelabuhan kedatangan/ keberangkatan (port of departure/ arrival), untuk memudahkan pengisian EHAC kami mohon agar admin berkenan meng-input seluruh pelabuhan laut pada kolom pilihan

dimaksud, karena pelabuhan yang kami cari tidak ketemu, terima kasih

Sesudah *Normalize Case*

app ini sangat bagus, ringan, mudah digunakan, mudah di akses, berisi informasi petunjuk perjalanan. namun saat pengisian di menu ehac beberapa pelabuhan domestik belum ada pada daftar pencarian pelabuhan kedatangan/ keberangkatan (port of departure/ arrival), untuk memudahkan pengisian ehac kami mohon agar admin berkenan meng-input seluruh pelabuhan laut pada kolom pilihan dimaksud, karena pelabuhan yang kami cari tidak ketemu, terima kasih

3.2 Tokenizing

Tahap *Tokenizing* digunakan untuk memotong string input tiap kalimat berdasarkan kata-kata penyusunnya.

Tabel 2. Tahap *Tokenizing*

Sebelum <i>Tokenizing</i>
sudah lancar dan bisa masuk ke akun saya terima kasih kembangkan terus aplikasinya.
Sesudah <i>Tokenizing</i>
sudah lancar dan bisa masuk ke akun saya terima kasih kembangkan terus aplikasinya

3.3 Filtering

Tahap *Filtering* digunakan untuk mengambil kata-kata penting ataupun kata-kata kurang penting dari hasil dokumen teks, menggunakan algoritma *stop list* dan *word list*.

Tabel 3. Tahap *Filtering Stop Word List*

aju	benar	cukup	enggak
ada	berada	cuma	entah
adalah	berapa	dibuat	guna
adanya	beri	dijawab	hampir

adapun	berupa	dimaksud	harus
agak	boleh	disebut	inginkan
akhir	bukan	diucapkan	itulah
...

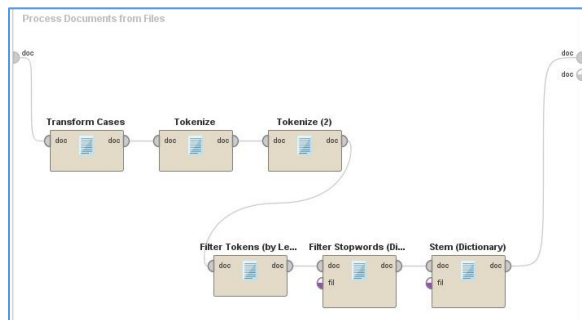
3.4 Stemming

Pada tahap *Stemming* digunakan untuk mencari kata dasar disetiap kata yang ada di dokumen teks atau mencari *root* di setiap kata yang ada.

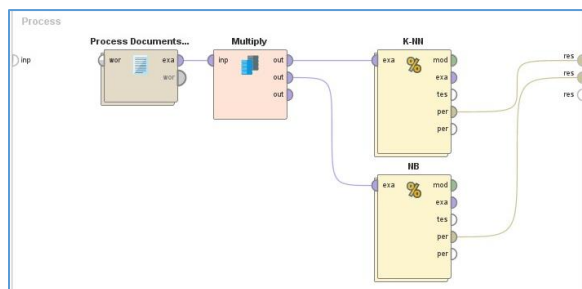
Tabel 4. Tahap *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
akhirnya	akhir
mengisi	isi
daftarin	daftar
persyaratanya	Syarat
selanjutnya	lanjut
...	...

Proses *text preprocessing* dan perhitungan akurasi terhadap metode klasifikasi yaitu Naïve Bayes dan K-Nearest Neighbor dilakukan secara bersamaan menggunakan software bantu Rapid Miner 9.9. Berikut proses *text preprocessing* dan perhitungan akurasi dari kedua metode, seperti terlihat digambar 4 dan 5.



Gambar 2. Tahapan *Text Preprocessing*



Gambar 3. Perhitungan Komparasi Metode NB dan K-NN

Dari proses *Text Preprocessing* yang sudah dilakukan, didapatkan kelompok kata dengan frekuensi kemunculan paling banyak. Berikut gambaran secara visual dari kata-kata tersebut, yang diperjelas ke dalam *Wordcloud*. Gambar visual tersebut bernilai *Positif* dan *Negatif*, seperti pada gambar 2 dan 3.



Gambar 4. *Wordcloud Positif*



Gambar 5. *Wordcloud Negatif*

Dari perhitungan yang sudah dilakukan di dapatkan hasil akurasi dari metode Naïve Bayes dan K-Nearest Neighbor, seperti terlihat digambar 6 dan 7.

accuracy: 70.48% +/- 11.40% (micro average: 70.48%)			
	true Positif	true Negatif	class precision
pred. Positif	72	29	71.29%
pred. Negatif	33	76	69.72%
class recall	68.57%	72.38%	

Gambar 6. Akurasi Metode Naïve Bayes

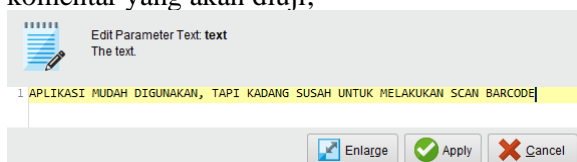
Dari gambar 6 dapat disimpulkan, dari jumlah dataset 105 ulasan yang diuji, prediksi positif yang masuk benar-benar kategori positif sebanyak 72, dan kategori negatifnya 29. Sedangkan prediksi negatif yang masuk benar-benar kategori positif 33 dan kategori negatifnya 76.

accuracy: 73.33% +/- 11.92% (micro average: 73.33%)			
	true Positif	true Negatif	class precision
pred. Positif	75	26	74.26%
pred. Negatif	30	79	72.48%
class recall	71.43%	75.24%	

Gambar 7. Akurasi Metode K-Nearest Neighbor

Dari gambar 7 dapat disimpulkan, dari jumlah dataset 105 ulasan yang diuji, prediksi positif yang benar-benar terdapat dikategori positif sebanyak 75, dan kategori negatifnya 26. Sedangkan prediksi negatif yang benar-benar berada dikategori positif 30 dan kategori negatifnya 79.

Berikutnya dilakukan pengujian untuk sebuah komentar yang akan diklasifikasikan ke dalam komentar positif atau negatif, berikut adalah komentar yang akan diuji;



Gambar 8. Data Uji

Setelah dilakukan pengujian diperoleh hasil bahwa algoritma Naïve Bayes mengklasifikasi data uji merupakan komentar negatif, sedangkan algoritma K-Nearest Neighbor merupakan komentar positif.

Row No.	prediction(label)	confidence(Positif)	confidence(Negatif)
1	Negatif	0	1

Gambar 9. Hasil Klasifikasi Naïve Bayes

Row No.	prediction(la...	confidence(...	confidence(...
1	Positif	1.000	0

Gambar 10. Hasil Klasifikasi K-Nearest Neighbor

Berdasarkan hasil perhitungan yang dihasilkan dari software bantu Rapid Miner 9.9 didapatkan tabel komparasi dari hasil akurasi ,

recall dan precision yang dihasilkan, seperti terlihat di tabel 5.

Tabel 5. Hasil Komparasi Metode NB dan K-NN

Metode	Accuracy	Recall		Precision	
		Positif	Negatif	Positif	Negatif
NB	70,48 %	68,57%	72,38%	71,29%	69,72%
K-NN	73,33 %	71,43%	75,24%	74,25%	72,48%

Dari tabel 5 berdasarkan dokumen teks (dataset) komentar (ulasan) pada aplikasi PeduliLindungi yang ada di Play Store, dapat disimpulkan bahwa metode K-Nearest Neighbor (K-NN) memiliki nilai akurasi 73,33%. Sedangkan metode Naïve Bayes (NB) nilai akurasinya 70,48%. Jadi metode K-NN merupakan metode terbaik dibandingkan dengan NB berdasarkan dataset yang ada.

4. Kesimpulan dan Saran

Dari hasil eksperimen yang telah dilakukan bisa disimpulkan bahwa:

1. Tingkat akurasi yang lebih baik dihasilkan oleh model klasifikasi dengan algoritma *K-Neares Neighbor*
2. Algoritma K-Nearest Neighbor menghasilkan tingkat akurasi sebesar 73,33%, sedangkan algoritma Naive Bayes tingkat akurasinya sebesar 70,48%.

Untuk penelitian selanjutnya disarankan untuk perkembangan kedepan, seperti: menambahkan jumlah dokumen teks (dataset) yang ada, serta bisa menggunakan metode klasifikasi atau seleksi fitur pada dokumen teks (dataset) yang ada.

DAFTAR PUSTAKA

Cahyanti, Dewi, Alifah Rahmayani, and Syafira Ainy Husniar. 2020. "Analisis Performa Metode K-NN Pada Dataset Pasien Pengidap Kanker Payudara." *Indonesian Journal of Data and Science* 1 (2): 39–43. <https://doi.org/10.33096/ijodas.v1i2.13>.

Fitriyyah, Sitti Nurul Jannah, Novi Safriadi, and Enda Esyudha Pratama. 2019. "Analisis Sentimen Calon Presiden Indonesia 2019 Dari Media Sosial Twitter Menggunakan Metode Naive Bayes." *Jurnal Edukasi Dan*

- Penelitian Informatika (JEPIN) 5 (3): 279.
<https://doi.org/10.26418/jp.v5i3.34368>.
- Hotho, A, A Nürnberger, and G Paaß. 2005. "A Brief Survey of Text Mining." In *Ldv Forum* 20(1): 19–62.
- Kurniawan, Robi, and Aulia Apriliani. 2020. "Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper." *Jurnal INSTEK (Informatika Sains Dan Teknologi)*.
<https://doi.org/10.24252/instek.v5i1.13686>.
- Puspita, Rani, and Agus Widodo. 2021. "Perbandingan Metode K-NN, Decision Tree, Dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS." *Jurnal Informatika Universitas Pamulang* 5 (4): 646.
<https://doi.org/10.32493/informatika.v5i4.7622>.
- Samsir, Ambiyar, Unung Verawardina, Firman Edi, and Ronal Watrianthos. 2021. "Analisis Sentimen Pembelajaran Daring Pada Twitter Di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes." *Jurnal Media Informatika Budidarma* 5: 157–63.
<https://doi.org/10.30865/mib.v5i1.2604>.
- Witten, Ian H. 2004. "Text Mining." *The Practical Handbook of Internet Computing*, 14-1-14–22.
<https://doi.org/10.1201/9780203507223>.
- Zhang, Lei, and Bing Liu. 2016. "Sentiment Analysis and Opinion Mining." *Encyclopedia of Machine Learning and Data Mining*, no. May: 1–10.
https://doi.org/10.1007/978-1-4899-7502-7_907-1.