

Penerapan *Text Mining* Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode *Naïve Bayes*

Fira Fathonah¹⁾, Asti Herliana²⁾

¹⁾²⁾ Fakultas Teknologi Informasi, Sistem Informasi, Universitas Adhirajasa Reswara Sanjaya Bandung Antapani, Jl. Terusan Sekolah No.1-2, Cicaheum, Kec. Kiaracandong, Kota Bandung

¹⁾ firafatonah18@gmail.com

²⁾ asti@ars.ac.id

Abstrak

Sejak wabah COVID-19 melanda dunia, media sosial merupakan media yang paling banyak digunakan oleh masyarakat untuk tetap beraktivitas dan berinteraksi dengan masyarakat lainnya. Selain masyarakat umum, pemerintah utamanya pemerintah Indonesia juga memanfaatkan media ini untuk memberikan berbagai informasi dan pelayanan yang dibutuhkan oleh masyarakat. Salah satu hal yang sedang hangat diperbincangkan di media sosial utamanya twitter adalah mengenai Vaksin COVID-19. Berbagai komentar dilontarkan oleh para pengguna baik positif dan negatif. Dalam rangka mengetahui tanggapan masyarakat terhadap adanya Vaksin COVID-19 ini, maka pada penelitian kali ini dilakukan analisis sentimen terhadap Vaksin COVID-19 dengan menggunakan metode *Naïve Bayes*. *Naïve Bayes* dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi Berdasarkan hasil pengujian terhadap 100 data latih yang kemudian dipilih kembali dengan menggunakan teknik data crawling menjadi 34 data, didapatkan bahwa analisis sentimen dari pengguna twitter untuk Vaksin COVID-19 ini didapatkan persentase *accuracy* sebesar 100%

Kata kunci: Vaksin, covid-19, analisis sentimen, *naïve bayes*, twitter

Abstract

Since the Covid-19 outbreak hit the world, social media has been the most widely used media by the public to stay active and interact with other people. In addition to the general public, the government, especially the Indonesian government, also uses this media to provide various information and services needed by the community. One of the hot topics being discussed on social media, especially Twitter, is the Covid-19 vaccine. Various comments made by users both positive and negative. In order to find out the public's response to the Covid-19 vaccine, in this study an analysis of sentiment towards the Covid-19 vaccine was carried out using the Naïve Bayes method. Naïve Bayes is considered to have good potential in classifying documents compared to other classification methods in terms of accuracy and efficiency. -19, the percentage of accuracy is 100%

Keywords: Vaccines, covid-19, sentiment analysis, naive bayes, twitter

1. PENDAHULUAN

Virus covid-19 sekarang sedang menjadi fokus utama masyarakat dunia, karena termasuk virus yang berbahaya. Munculnya vaksin covid-19 yang sekarang sedang diupayakan oleh pemerintah menjadi buah bibir di jejaring sosial media salah satunya adalah *twitter*. *Twitter* adalah jejaring sosial yang memungkinkan pengguna mengirim karakter hingga 140 karakter yang sering disebut *tweet* atau kicauan [1].

Dengan jejaring sosial kita dapat mengetahui opini positif ataupun negatif adapun teknik yang digunakan. *Text mining* sendiri merupakan teknik yang digunakan untuk memenuhi kebutuhan dalam proses klasifikasi dokumen dengan konten apapun. Cara kerja dari teknik ini

merupakan upaya dalam memunculkan variasi dari kumpulan data yang tersedia dalam jumlah besar, untuk kemudian berusaha menemukan pola yang sesuai dengan apa yang diharapkan dari kumpulan data teks yang ada. Analisa sentimen menjadi proses yang sangat penting dalam memahami isi data dengan tujuan [2] mengekstrak informasi atau menggali pengetahuan dari sumber data tekstual yang ada dalam jumlah besar secara otomatis. Hal ini diperuntukan pada dataset dengan opini positif dan negatif pada tweet [3]. Sedangkan model yang digunakan dalam menguji data tersebut adalah Metode Naive Bayes, algoritma tersebut dianggap cocok digunakan untuk menganalisa sentimen yang ada pada dataset. Naive Bayes pun memiliki performa yang baik dalam melakukan klasifikasi tanpa adanya penambahan metode lain, hal ini yang menjadikan landasan bagi penulis untuk memilih metode Naive Bayes pada penelitian ini.

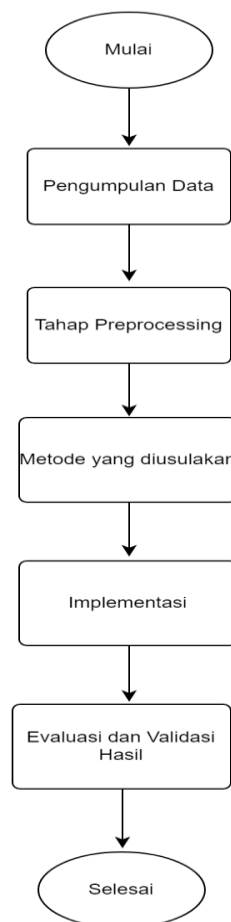
Alasan lain pemilihan metode *Naive Bayes* sebagai metode usulan pada penelitian kali ini didasarkan pada beberapa penelitian terdahulu yang telah membuktikan metode ini memberikan performansi yang baik. Adapun beberapa peneliti yang dimaksud adalah Suryono [4] melakukan penelitian dengan tema, Klasifikasi Sentimen pada *Twitter* dengan Naive Bayes. Akhmad Pandu Wijaya [2] melakukan penelitian dengan tema, Analisis sentimen dan Klasifikasi Komentar Positif Pada *Twitter* dengan *Naive Bayes Classification*. Muhammad Syarifuddin [1] melakukan penelitian dengan tema, Analisis Sentimen Opini Publik Mengenai COVID-19 pada *twitter* menggunakan Metode *Naive Bayes* dan KNN. Agustinus Excel Prasetya Aji [5] melakukan penelitian dengan tema, Analisis Pengaruh Pandemi COVID-19 pada Sentimen Pariwisata dan Analisis Jaringan Semantik pada media sosial. Adhi Viky Sudiantoro [6] melakukan penelitian dengan tema, Analisis Sentimen *Twitter* Menggunakan *Text Mining* dengan Algoritma *Naive Bayes Classifier*. Metode *Naive Bayes* juga digunakan dalam analisis sentimen opini masyarakat. Dimana metode *Naive Bayes* juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi.

2. TINJAUAN PUSTAKA

Sentiment analysis merupakan salah satu bidang dari *Natural Language Processing* (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi *review*. Analisis sentimen sendiri adalah proses memahami dan mengelompokkan emosi (positif, negatif, dan netral) yang terdapat dalam tulisan menggunakan teknik analisis teks [3]. Dan *Text mining* adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data [7]. Algoritma *Naive Bayes* merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian. Keuntungan penggunaan adalah bahwa metoda ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yg diperlukan dalam proses pengklasifikasian [8].

3. METODE PENELITIAN

Penelitian ini dilaksanakan melalui beberapa tahap yaitu pengumpulan data, pengolahan data, klasifikasi teks dengan metode *naive bayes classifier*, hasil klasifikasi dan pengujian data. Tahap-tahap yang akan dilakukan dalam penelitian ini.



Gambar 1. Metode penelitian

3.1 Pengumpulan Data

Pengumpulan data diambil dari *twitter* menggunakan API *twitter*. Menggunakan 100 data yang diambil dengan kata yang berkaitan dengan vaksin covid-19 dan hanya mengambil data dengan kata bahasa indonesia dengan *tools* *rapidminer*, setelah data di crawling dari *twitter* didapatkan 34 data latih.

3.2 Tahap Preprocessing

Merupakan tahap awal untuk mengolah data teks menjadi analisis sentimen dengan menggunakan *tools* *rapidminer*. Terdapat beberapa metode yang digunakan pada pengolahan data seperti:

1. Melakukan *filtering duplicate tweets* adalah tahapan dimana *tweet* yang mempunyai isi sama akan dihapus untuk menghindari duplikat isi dari *tweet*.
2. *Cleansing* adalah proses pembersihan data teks yaitu dengan menghilangkan data yang tidak konsisten atau tidak relevan.
3. *Parsing / Tokenizer* merupakan proses untuk memisahkan data teks menjadi beberapa token.
4. *Case folding* yaitu merubah bentuk kata menjadi bentuk dasarnya agar sebuah karakter dapat seragam (*lower case*).
5. *Filter StopWord(English)* merupakan proses untuk menghilangkan kata dengan bahasa inggris.

3.3 Metode yang Diusulkan

Metode yang diusulkan dalam penelitian ini adalah metode *naive bayes*. Metode penggolongan berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen). Metode ini merupakan salah satu algoritma dalam teknik klasifikasi yang mudah diimplementasikan dan cepat prosesnya.

3.4 Implementasi

Pada tahap ini metode yang diusulkan akan diterapkan pada *tweet* mengenai vaksin covid-19 menggunakan *rapidminer*.

3.5 Evaluasi dan Validasi Hasil

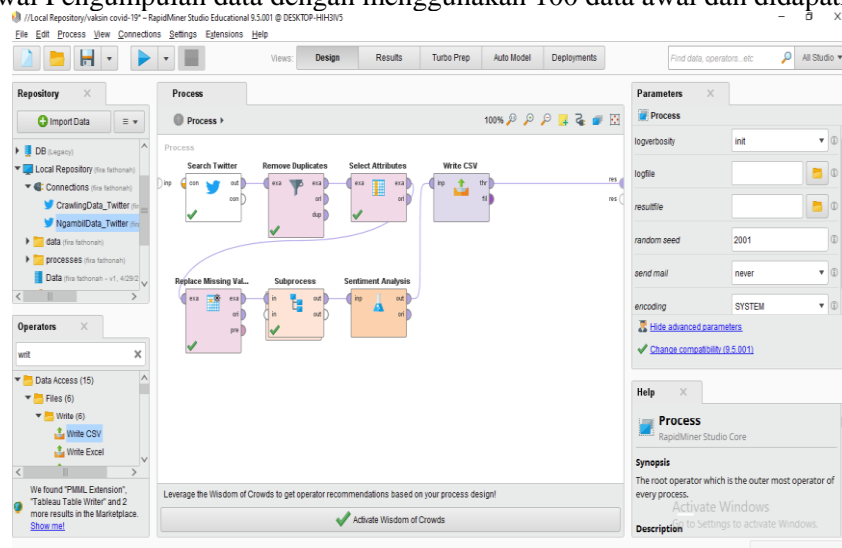
Evaluasi dilakukan dengan menganalisis hasil klasifikasi. Proses evaluasi menggunakan *confusion matrix*. Validasi yang digunakan yaitu membagi *data training* dan *data testing*. Performa pada model yang digunakan yaitu *naive bayes* diterapkan untuk mengetahui tingkat akurasi model.

4. HASIL DAN PEMBAHASAN

Penerapan algoritma *naive bayes* dimaksudkan untuk mengetahui tingkat keakuratan dan kecocokan klasifikasi dari analisis sentimen yang diperoleh dari *tweet* mengenai vaksin covid-19.

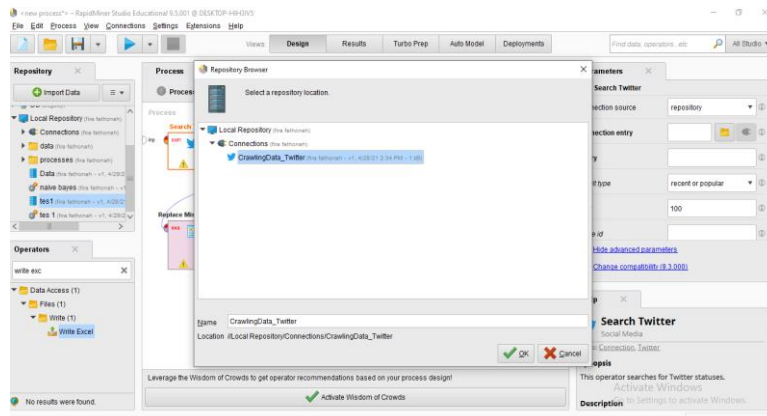
4.1 Pengumpulan Data

Tahap awal Pengumpulan data dengan menggunakan 100 data awal dan didapatkan 34 data.



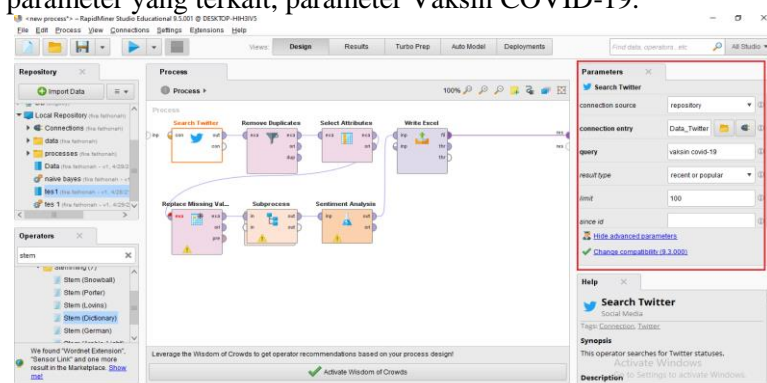
Gambar 2. Persiapan pengumpulan data

Membuat *Connection twitter* untuk menghubungkan ke twitter agar mendapatkan data latih.



Gambar 3. Membuat *connection* API twitter

Menginputkan parameter yang terkait, parameter Vaksin COVID-19.



Gambar 4. Menginputkan parameter vaksin Covid-19

Berikut ini beberapa data yang diambil dari data hasil *crawling tweet* dengan parameter Vaksin COVID-19.

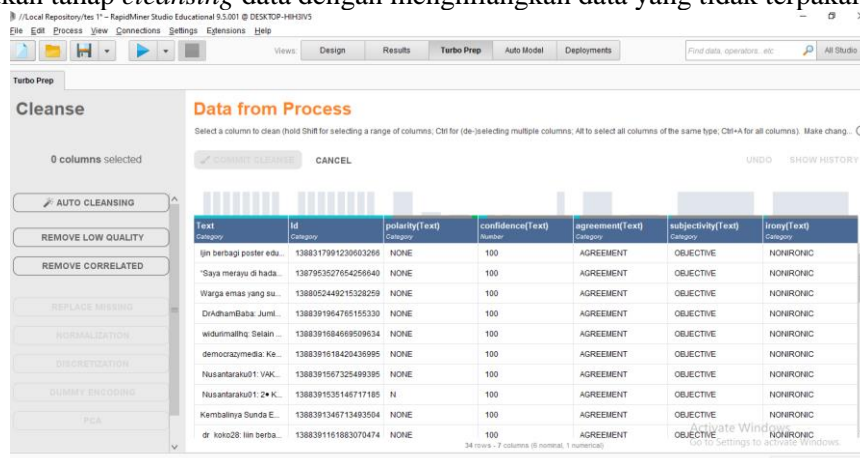
Tabel 1. Data *tweet*

<i>Komentar</i>	<i>Kategori</i>
Berapa Stok Vaksin Covid-19 Indonesia hingga Sekarang? https://t.co/p1719Xp5Sp https://t.co/VAq9KnLxJS	Negatif
notsoaidil: "Orang Asli Bateq di Kampung Aring 5, di sini yang belum mendaftar dalam Program Imunisasi Covid-19 Kebangsaan bukan kerana..."	Negatif
501Awani: Kementerian sedang meneliti kemungkinan membolehkan institusi penjagaan kesihatan swasta memperoleh stok bekalan semasa Singa...	Negatif
zarazettirazr: Aku HERAN 2 Juni masih ada pejabat yang HERAN padahal infonya sudah sejak 19 Mei Vaksin yang diterima hanya yang diset...	Negatif
datukhensem: MAKLUMAN: Kepada penerima vaksin COVID-19 Hospital Miri. Unit Vaksin hospital memerlukan kerjasama dari penerima vaksi...	Positif
Kita semua tetap perlu mematuhi SOP walaupun telah menerima suntikan vaksin COVID-19. KitaAkanMenang KomunikasiKita JabatanPenerangan https://t.co/ZOecKRmy2V	Positif
Ramai lagi yang masih ragu-ragu mengenai vaksin. "Sesuai ke vaksin jenis X ini dengan saya?" Macam-macam soalan dan ragam! Jika	Positif

Komentar	Kategori
tidak tahu, baik bertanya. Jalankan tanggungjawab kita untuk menghapuskan Covid-19! Sertai webinar kami dengan mendaftar di https://t.co/Hm6pOYYaFh https://t.co/b6TPicHVHT H_Bakkaniy: Video Buya Yahya pada hari beliau ambil suntikan Vaksin Covid-19 beserta nasihat beliau kepada orang ramai. Beliau ini mem...	Negatif
Hari ini saya disuntik vaksin covid 19,pas mau disuntik saya ditanya bapaknya kader?saya jawab bukan bapaknya perangkat desa?saya jawab bukan juga	Negatif
Dalam hati saya jawab Aku penyelamat Bumi Sur,... LimaLapa1: Dunia sangat membutuhkan banyak vaksin COVID-19 untuk mengatasi kesenjangan akses yang sangat besar di seluruh dunia. Sinov...	Positif
KKMPutrajaya: Unit Bergerak untuk suntikan vaksin COVID-19 telah dilancarkan di daerah Padang Terap, Kuala Nerang, Kedah. Unit ini ak...	Positif
LimaLapa1: WHO hari ini memvalidasi vaksin Sinovac-CoronaVac COVID-19 untuk penggunaan darurat SinovacLolosUjiWHO https://t.co/VMkdq81...	Positif
Kita semua tetap perlu mematuhi SOP walaupun telah menerima suntikan vaksin COVID-19. KitaAkanMenang KomunikasiKita JabatanPenerangan https://t.co/piIHvXXvET	Positif
Ijin berbagi poster edukasi vaksin. Yang bikin posternya adalah dokter2 internship Rumah Sakit Darurat Covid-19 Wisma Atlet Kemayoran. Boleh didistribusikan dan dipakai bebas kata teman2 kita itu.	Positif
"Saya merayu di hadapan mereka untuk lepaskan saya, tapi mereka bersikap seperti binatang." AWANInews AWANIpagi HapusCOVID19	Negatif
Warga emas yang sudah menerima suntikan vaksin COVID-19 diingatkan untuk terus mematuhi prosedur operasi standard (SOP) memandangkan masih berisiko untuk dijangkiti,Ketua Pengarah Kesihatan, Tan Sri Dr Noor Hisham Abdullah.	Positif
DrAdhamBaba: Jumlah pemberian dos vaksin COVID-19 setakat 30 April 2021 - 1,448,066 dos. Jumlah terima dos pertama - 895,204 orang. Da...	Positif

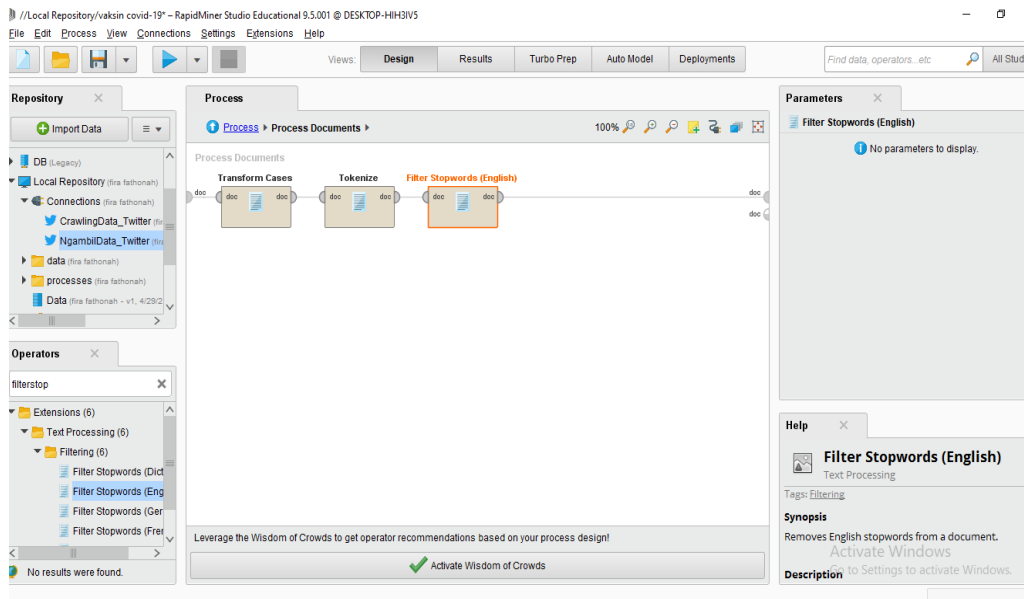
4.2 Tahap Preprocessing

Melakukan tahap *cleansing* data dengan menghilangkan data yang tidak terpakai.

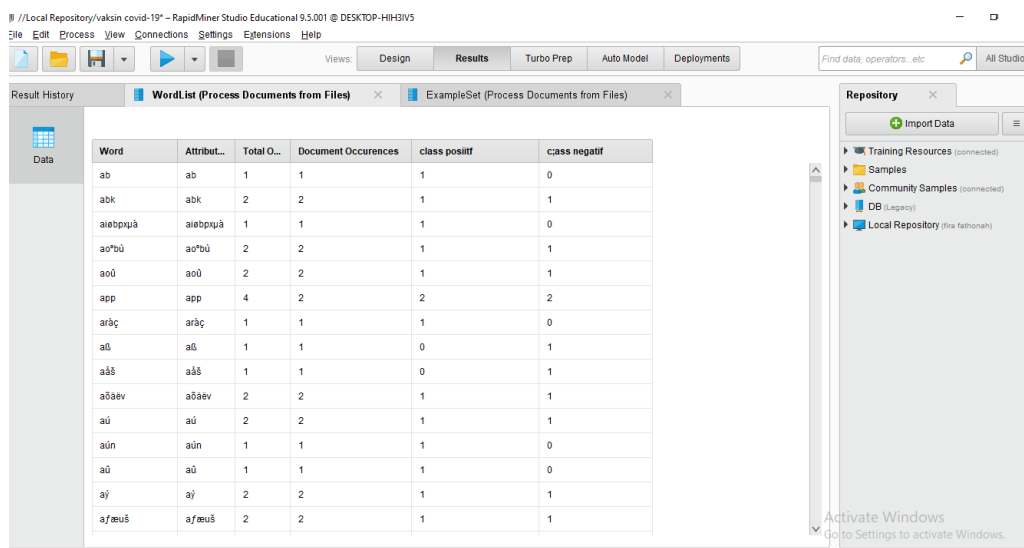


Gambar 5. Melakukan tahap *cleansing* data

Menambahkan operator *tokenizer*, *case folding*, *filter stopwords* untuk menghasilkan document list.



Gambar 6. Melakukan *tokenize*, *case folding*, dan *filter stopwords*

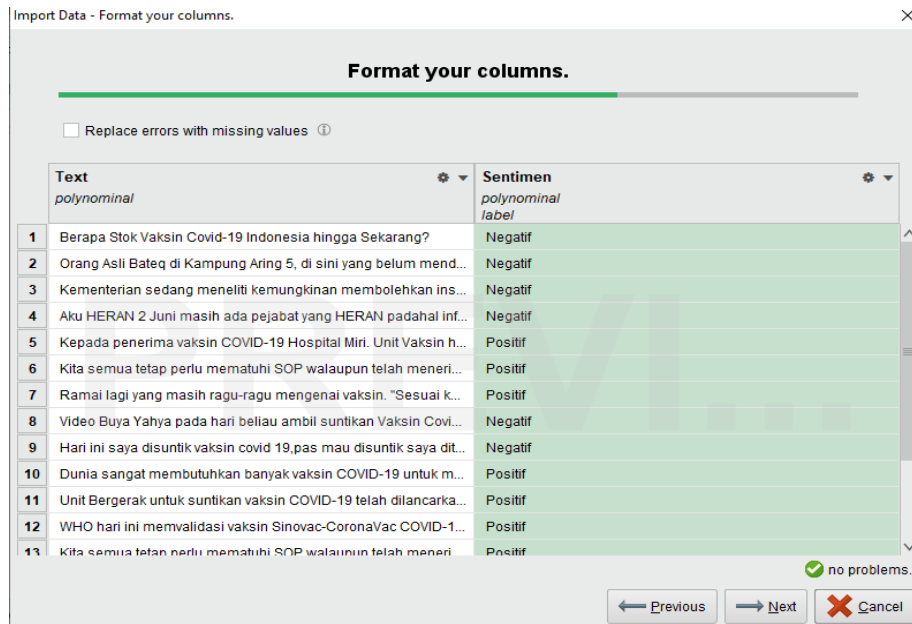


Gambar 7. Hasil *tokenize*, *case folding*, dan *filter stopwords*

4.3 Evaluasi dan Validasi Hasil

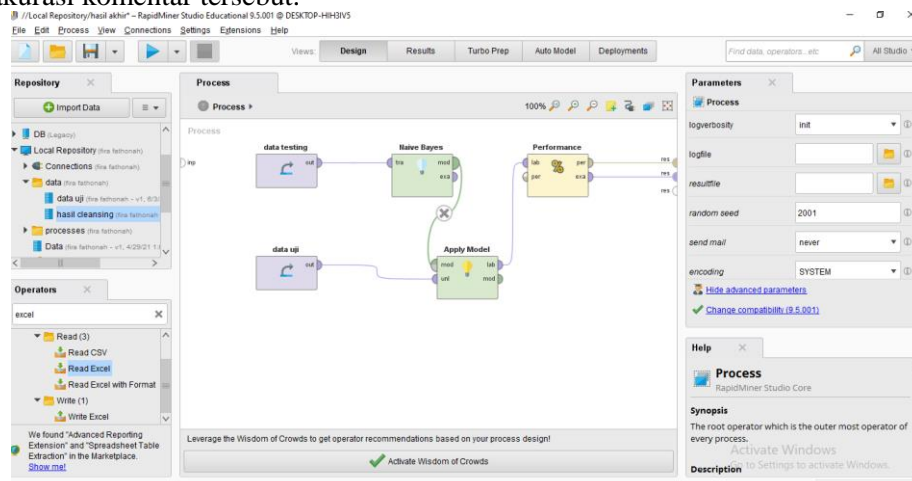
Setelah proses *preprocessing*, selanjutnya dilakukan proses klasifikasi tahap *validation*. Hasil yang diperoleh dari pengujian dengan menggunakan *naive bayes*. Hasil dari implementasi model *naive bayes* tersebut akan didapatkan *rule*. *Rule* yang dihasilkan akan digunakan sebagai dasar prediksi nilai yang akan dilakukan.

Model *confusion matrix* akan membentuk matriks yang terdiri dari *accuracy*, *true positive*, *true negative*, *recall*, dan *precision*. Berikut merupakan hasil dari *confusion matrix* menggunakan algoritma *naive bayes*.



Gambar 8. Import data tweet

Setelah melakukan import data tweet selanjutnya masuk ke proses naïve bayes untuk mencari tingkat akurasi komentar tersebut.



Gambar 9. Proses naïve bayes untuk mencari akurasi

Menghasilkan accuracy algoritmat dalam bentuk confusion matrix.

accuracy: 100.00%

	true Negatif	true Positif	class precision
pred. Negatif	7	0	100.00%
pred. Positif	0	10	100.00%
class recall	100.00%	100.00%	

Gambar 10. Hasil accuracy algoritma naïve bayes

1. *Precision* prediksi positif

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{10}{10 + 0} \times 100\% \\ &= 100\% \end{aligned}$$

2. *Precision* prediksi negatif

$$\begin{aligned} Precision &= \frac{TN}{TN + FN} \times 100\% \\ &= \frac{7}{7 + 0} \times 100\% \\ &= 100\% \end{aligned}$$

```
PerformanceVector

PerformanceVector:
accuracy: 100.00%
ConfusionMatrix:
True:   Negatif Positif
Negatif:    7    0
Positif:    0    10
precision: 100.00% (positive class: Negatif)
ConfusionMatrix:
True:   Positif Negatif
Positif:   10    0
Negatif:   0    7
recall: 100.00% (positive class: Negatif)
ConfusionMatrix:
True:   Positif Negatif
Positif:   10    0
Negatif:   0    7
AUC (optimistic): 1.000 (positive class: Negatif)
AUC: 1.000 (positive class: Negatif)
AUC (pessimistic): 1.000 (positive class: Negatif)
```

Gambar 11. *Performance vektor*

5. KESIMPULAN

Berdasarkan hasil pengujian dapat diambil beberapa kesimpulan sebagai berikut:

Metode *Naive Bayes* dapat mengelompokkan data untuk analisis sentimen data *twitter* dengan optimal sehingga menghasilkan analisis sentimen terbaik. Hasil analisis sentimen pada kata kunci vaksin covid-19 dengan menggunakan *confusion matrix* menggunakan algoritma *naive bayes* mendapatkan tingkat hasil akurasi 100%.

DAFTAR PUSTAKA

- [1] M. Syarifuddin, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes dan Knn," *Inti Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020.

- [2] A. P. Wijaya and D. Wardhani, “Analisa sentimen dan Klasifikasi Komentar Positif Pada Twitter dengan Naïve Bayes Classification,” *BRITech, J. Ilm. Ilmu Komputer, Sains Dan Teknol. Terap.*, vol. 1, no. 2, pp. 32–40, 2020.
- [3] G. N. Arviana, “Sentiment Analysis, Teknik untuk Pahami Maksud di Balik Opini Pelanggan,” *glints*, 2021. <https://glints.com/id/lowongan/sentiment-analysis/>.
- [4] S. Suryono, E. Utami, and E. T. Luthfi, “Klasifikasi Sentimen Pada Twitter Dengan Naive Bayes Classifier,” *Angkasa J. Ilm. Bid. Teknol.*, vol. 10, no. 1, pp. 89–96, 2018.
- [5] A. E. P. Aji, “Analisis Pengaruh Pandemi COVID-19 Pada Sentimen Pariwisata dan Analisis Jaringan Semantik Pada Media Sosial,” Universitas Atma Jaya Yogyakarta, 2020.
- [6] A. V. Sudiantoro and E. Zuliarso, “Analisis sentimen twitter menggunakan text mining dengan algoritma Naïve Bayes Classifier,” *J. Din. Inform.*, vol. 10, no. 2, pp. 69–73, 2018.
- [7] U. D. Arni, “Apa Itu Text Mining?,” *Garuda Cyber Indonesia*, 2018. <https://garudacyber.co.id/artikel/1254-apa-itu-text-mining>.
- [8] Anon, “Algoritma Naive Bayes,” *Informatikalogi*, 2017. <https://informatikalogi.com/algoritma-naive-bayes>.

Biodata Penulis

Fira Fathonah, S.Kom lahir di Bandung 1999. Meraih gelar Sarjana Komputer (S.Kom) di program studi Sistem Informasi Universitas Adhirajasa Reswara Sanjaya Bandung pada tahun 2021.

Asti Herliana, S.Kom, M.Kom Meraih gelar S2 Ilmu Komputer di STMIK Nusa Mandiri dan lulus pada tahun 2013. Saat ini berprofesi sebagai dosen di Program Studi Sistem Informasi Universitas Adhirajasa Reswara Sanjaya.