

ANALISIS SENTIMEN LARANGAN PENGGUNAKAN OBAT SIRUP MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika

Oleh

Fitri Wulandari

NIM. 11950125070



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU

2023

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU



Hak Cipta Diindungi Undang-Undang

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

ANALISIS SENTIMEN LARANGAN PENGGUNAAN OBAT SIRUP MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

TUGAS AKHIR

Oleh

FITRI WULANDARI

NIM. 11950125070

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 22 Juni 2023

Pembimbing I,

DR. Elin Haerani, S.T., M.Kom

NIP. 19810523 200710 2 003

UIN SUSKA RIAU



Hak Cipta Diindungi Undang-Undang

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

ANALISIS SENTIMEN LARANGAN PENGGUNAAN OBAT SIRUP MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Oleh

FITRI WULANDARI

NIM. 11950125070

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 22 Juni 2023

Mengesahkan,

Ketua Jurusan,



DR. HARTONO, M.PD.

NIP. 19640301 199203 1 003

IRWAN ISKANDAR, M.T.

NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua : Novriyanto, S.T., M.Sc.

Pembimbing I : DR. Elin Haerani, S.T., M.Kom.

Penguji I : Muhammad Fikry, S.T., M.Sc.

Penguji II : Elvia Budianita, S.T., M.Cs.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SURAT PERNYATAAN

Saya yang bertandatangan di bawah iniL

Nama : Fitri Wulandari

NIM : 11950125070

Tempat/Tgl.Lahir : Sintong, 15 Januari 2001

Fakultas : Sains dan Teknologi

Prodi : Teknik Informatika

Judul Skripsi : Analisis Sentimen Larangan Penggunaan Obat Sirup Menggunakan Metode *Naive Bayes Classifier*

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah surat pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga

Pekanbaru, 4 Juli 2023
Yang membuat pernyataan



FITRI WULANDARI
NIM. 11950125070

UIN SUSKA RIAU



Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma *naive bayes classifier*

Fitri Wulandari^{*1}, Elin Haerani², Muhammad Fikry³, Elvia Budianita⁴

Email: ¹11950125070@students.uin-suska.ac.id, ²elin.haerani@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id, ⁴elvia.budianita@uin-suska.ac.id

^{1,2,3,4} Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

Diterima: 30 Maret 2023 | Direvisi: - | Disetujui: 28 April 2023

©2023 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Pemerintah Indonesia membuat kebijakan larangan penggunaan obat sirup sebagai bentuk pencegahan terhadap penyakit gagal ginjal akut yang banyak menyerang masyarakat di Indonesia. Namun kebijakan tersebut banyak menimbulkan komentar dari masyarakat. Komentar masyarakat tersebut dapat ditemukan pada media sosial YouTube, karena YouTube mempunyai sumber informasi yang dapat digunakan sebagai bahan penelitian. Komentar-komentar tersebut dapat diolah secara langsung tanpa bantuan mesin, tetapi kurang efektif dan efisien. Oleh karena itu, komentar tersebut diolah menggunakan metode pembelajaran mesin. Berdasarkan penelitian terdahulu, algoritma naive bayes classifier cenderung sederhana dan mudah digunakan. Selain itu, algoritma ini juga mempunyai hasil akurasi yang tinggi. Jumlah data yang digunakan pada penelitian ini yaitu 1000 data komentar YouTube terkait video mengenai kebijakan larangan penggunaan obat sirup, komentar tersebut dibagi menjadi 2 kategori yaitu kelas positif dan kelas negatif. Hasil pelabelan dari 1000 komentar tersebut mendapatkan 704 komentar negatif dan 296 komentar positif. Berdasarkan percobaan yang dilakukan menggunakan bahasa pemrograman python, akurasi tertinggi didapatkan sebesar 74% dengan perbandingan data 70:30 pada dataset awal. Kemudian pada dataset seimbang (296 komentar positif dan 296 komentar negatif), akurasi tertinggi didapatkan sebesar 64,70% pada perbandingan data 80:20. Hasil tersebut menunjukkan bahwa algoritma *naive bayes classifier* cukup bagus saat melakukan analisa sentimen mengenai kebijakan larangan penggunaan obat sirup.

Kata kunci: analisis sentimen, naive bayes, obat sirup, YouTube

Sentiment analysis of medicine syrup prohibition using naive bayes classifier algorithm

Abstract

The Indonesian government made a policy to stop consuming syrup as a form of prevention against acute kidney failure, which affects many people in Indonesia. However, the policy has caused a lot of comments from the public. These public comments can be found on YouTube, because YouTube has a large data source opportunity to be used as a research material. These comments can be processed directly without using a machine, but it is less effective and efficient. Thus, the comments are processed using machine learning methods. Based on the earlier research, the naive bayes classifier algorithm tends to be simple and easy to use. In addition, this algorithm also has a high accuracy. The amount of data used in this study is 1000 YouTube comment data related to videos regarding the policy of prohibiting the use of syrup medicine, the comments are divided into 2 category, which are positive class and negative class. The results of labeling 1000 comments obtained 704 negative comments and 296 positive comments. Based on the experiments conducted using python programming language, the highest accuracy was obtained at 74% in 70:30 data split. Furthermore, in the balanced dataset (296 positive and 296 negative comments), the highest accuracy was obtained at 64.70% with in 80:20 data split. These results represent that the naive bayes classifier algorithm is good enough at sentiment analysis about the policy of prohibiting the use of syrup drugs..

Keywords: *naive bayes, sentiment analysis, syrup medicine, YouTube*

1. PENDAHULUAN

Berbagai tanggapan masyarakat muncul karena adanya kebijakan pemerintah terkait larangan penggunaan obat sirup. Kementerian Kesehatan (Kemenkes) menegaskan bahwa pengumuman dari Badan Pengawas Obat dan Makanan (BPOM) mengenai larangan penggunaan obat sirup yang memiliki kandungan *ethylene glycol* (EG) dan *diethylene glycol* (DEG). Mohammad Syahril yang bertugas sebagai juru bicara Kementerian Kesehatan Republik Indonesia menjelaskan bahwa pemerintah terus melakukan penyelidikan mengenai penyebab gangguan ginjal akut [1]. Tanggapan masyarakat tersebut dapat dijumpai pada komentar YouTube. YouTube termasuk media sosial yang paling populer di Indonesia, penggunaannya mencapai 170 juta atau sekitar 93,8% dari total populasi 181,9 juta penduduk Indonesia [2]. Berdasarkan penelitian yang sudah dilakukan oleh We Are Social dan Hootsuite, menjelaskan bahwa pengguna terbanyak yang menghabiskan waktunya untuk menonton video secara *online* di YouTube memiliki rentang usia 16-24 tahun [3]. Hal ini menunjukkan bahwa YouTube mempunyai sumber informasi yang dapat digunakan sebagai objek penelitian analisa sentimen mengenai suatu topik. Data berupa komentar YouTube tersebut membuat informasi yang ada menjadi menarik untuk dianalisa lagi [4].

Pada Oktober 2022, pemerintah melarang masyarakat menggunakan obat sirup dikarenakan bahan untuk membuat obat sirup yang dicurigai jadi pemicu gagal ginjal. Larangan tersebut berlaku sampai ada informasi terbaru yang diumumkan pemerintah sesuai dengan ketentuan yang berlaku. Kemudian sebagai alternatif, obat yang dapat digunakan oleh masyarakat Indonesia antara lain dalam bentuk herbal, kapsul, tablet, *suppositoria* (anal), suntik atau yang lainnya [5].

Analisis sentimen adalah proses memperoleh informasi berupa teks yang bertujuan untuk memperoleh informasi tentang kecenderungan penilaian seseorang terhadap suatu objek. Tanggapan atau opini masyarakat dapat dijadikan pedoman untuk membuat keputusan terhadap penilaian suatu produk. Tanggapan yang diberikan dapat berupa tanggapan positif, negatif ataupun netral [6]. Analisis sentimen adalah proses pemahaman data tekstual yang berbentuk opini masyarakat dan kemudian diklasifikasikan menjadi 3 yaitu kelas positif, kelas negatif dan netral [7]. Analisis sentimen dapat membantu menganalisa suatu fakta didalam organisasi besar seperti mengukur opini publik, menganalisa pasar, menampilkan merek dan popularitas produk, dan menilai pengalaman pelanggan [8]. Analisis sentimen merupakan teknik untuk mengolah data yang didapat dari luapan emosional seseorang, baik itu berupa teks atau percakapan. Fokus dari analisis sentimen adalah untuk menemukan arti yang sebenarnya dari data berupa teks atau dokumen, apakah sebuah data teks tersebut memuat informasi yang positif, negatif atau netral [9]. Masalah yang dihadapi saat menganalisis sentimen adalah atribut yang terlalu banyak didalam dataset. Pada umumnya, banyak atribut yang tidak relevan atau tidak berguna untuk penelitian, hal tersebut bisa menyebabkan proses analisis menjadi kurang optimal, bahkan membuat akurasi menjadi rendah [10].

Salah satu penelitian tentang analisis sentimen menggunakan media sosial, dilakukan oleh Kusriani & Mashuri (2019) yang menggunakan Twitter yang berdasarkan pada penggunaan kamus (*lexicon*). Penggunaan *lexicon* ini cenderung sederhana daripada saat menggunakan algoritma pada *machine learning* [11]. Metode *lexicon* memiliki kekurangan yaitu hanya berdasarkan kamus yang masih terbatas pada jumlah kata, akurasi yang dihasilkan lebih kecil dibandingkan dengan analisis sentimen menggunakan algoritma pada *machine learning*. Terdapat beberapa penelitian mengenai analisis sentimen komentar YouTube. Salah satunya yang dilakukan oleh Zhafira, Rahayudi, & Indriati (2021), mengenai kebijakan program kampus merdeka. Dataset yang digunakan adalah sebanyak 1000 komentar. Penelitian tersebut menggunakan pembobotan TF-IDF dengan beberapa pembagian data (*splitting data*). Akurasi tertinggi didapatkan sebesar 97% pada pembagian data latih sebanyak 900 dan 100 data uji [12].

Berdasarkan penelitian yang akan dilakukan menggunakan algoritma *naive bayes classifier*, penelitian terkait yang dilakukan oleh Surohman (2020) mengenai analisa sentimen penilaian Fintech menggunakan metode *naive bayes* dan KNN. Penelitian tersebut menggunakan dataset berupa 116 ulasan positif dan 116 ulasan negatif dari aplikasi Dana. Akurasi tertinggi didapatkan sebesar 85% pada algoritma *naive bayes*, sedangkan algoritma KNN sebesar 83% [13]. Pada penelitian Pintoko dan Muslim (2019) tentang analisis sentimen layanan ojek *online* pada media sosial Twitter yang menggunakan metode *naive bayes classifier*, untuk melihat opini masyarakat terhadap penggunaan layanan ojek *online* yang dalam hal ini adalah Gojek dan Grab. Data yang digunakan adalah *tweet* mengenai Gojek dan Grab, masing-masing berjumlah 1000 *tweet*. Penelitian tersebut menghasilkan akurasi sebesar 86,80%. Kelebihan dari *naive bayes classifier* antara lain yaitu prosesnya yang cepat dan sederhana, tetapi juga menghasilkan akurasi yang tinggi [14].

Penelitian Tyas et al (2021) tentang membandingkan metode *naive bayes* dan C.45 dalam melakukan klasifikasi untuk memprediksi kelulusan. Pengolahan data untuk algoritma *naive bayes* menggunakan *software* RapidMiner, menghasilkan akurasi sebesar 94%. Sedangkan pengolahan data untuk algoritma C.45 menggunakan *software* WEKA dan menghasilkan akurasi sebesar 92,60%. Ini membuktikan bahwa metode *naive bayes* lebih bagus daripada C.45. Metode *naive bayes* pada umumnya lebih stabil karena didasarkan pada probabilitas sebuah kata yang muncul dalam sebuah kalimat, hanya memerlukan jumlah data pelatihan yang kecil, mudah diimplementasikan dan pada banyak kasus memberikan hasil yang baik [15]. Kemudian penelitian oleh Kaparany et al (2021) mengenai penggunaan algoritma *naive bayes* dalam melakukan analisa sentimen terhadap kebijakan *new normal* pada saat pandemi *Covid-19*. Penelitian tersebut menggunakan data *tweet* yang berjumlah 10000 data. Kemudian data tersebut dikategorikan menjadi positif, negatif dan netral. Sentimen netral tersebut dihapus sehingga menyisakan 8000 data berupa sentimen positif dan negatif. Penelitian tersebut menghasilkan akurasi sebesar 80,37% [16].

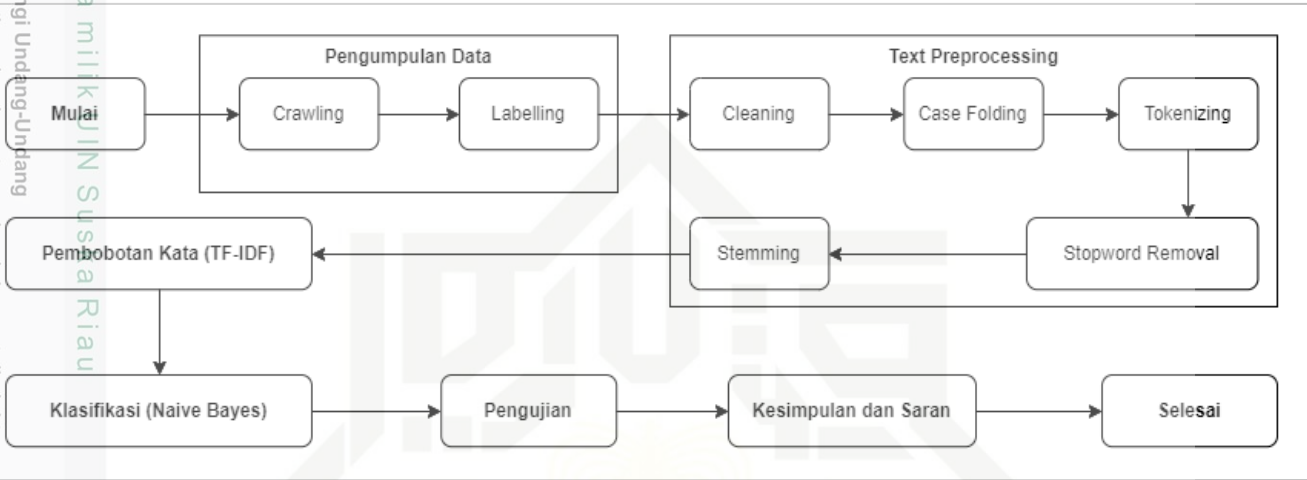


UIN SUSKA RIAU
Jurnal Computer Science and Information Technology (CoSciTech) Vol. 4, No. 1, April 2023, hal. 88-96

Oleh karena itu, analisis sentimen mengenai kebijakan larangan penggunaan obat sirup pada penelitian ini diharapkan dapat berpartisipasi dalam memberikan pemahaman atau edukasi kepada masyarakat Indonesia tentang bahaya penggunaan obat sirup pada saat ini, mencegah atau memperkecil kemungkinan masyarakat terkena penyakit gagal ginjal akut, memberikan *feedback* serta tanggapan masyarakat mengenai kebijakan yang dibuat oleh pemerintah tentang dilarangnya penggunaan obat sirup tersebut.

2. METODE PENELITIAN

Metodologi penelitian merupakan serangkaian langkah atau prosedur yang logis dan tersusun secara teratur dalam melaksanakan suatu riset atau penelitian yang bertujuan untuk tercapainya target penelitian. Langkah-langkah yang dilakukan dalam penelitian ini ditampilkan pada gambar 1 berikut.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pada tahap ini ada 2 tahap yang dilakukan yaitu *crawling* dan *labeling*. *Crawling* merupakan kegiatan dalam membuat duplikat dengan bagian yang sesuai yang diambil dari sebuah jaringan internet global [17]. Data yang diambil adalah komentar YouTube yang membahas tentang kebijakan larangan penggunaan obat sirup dengan jumlah 1000 data. Beberapa keyword pencariannya antara lain yaitu kebijakan larangan penggunaan obat sirup, komentar masyarakat terhadap dilarangnya sirup, pandangan masyarakat tentang kebijakan obat sirup, dan lain-lain. Selanjutnya tahap *labeling* yaitu aktifitas memberi kelas pada setiap komentar [4]. Label atau kelas yang akan diberikan adalah kelas positif dan negatif. Proses *labeling* dilakukan oleh tenaga pakar [18]. Label positif adalah komentar yang mendukung/setuju dengan kebijakan/aturan/himbauan tersebut. Sedangkan label negatif adalah komentar yang tidak mendukung/tidak setuju dengan kebijakan/aturan/himbauan tersebut, termasuk juga dengan komentar/opini yang mengarah ke perpecahan/mengandung SARA, ataupun memancing keributan antar warganet.

2.2 Text Preprocessing

Text preprocessing merupakan serangkaian tahapan dalam melakukan pembersihan data teks (menghilangkan serta mengatasi *noise*) sebelum diolah agar hasil perhitungan menjadi optimal [17]. Data komentar tentang larangan penggunaan obat sirup pada tahap *text preprocessing* ini dibagi menjadi 5 proses yaitu:

1. *Cleaning*, yaitu proses menghapus atribut yang tidak digunakan, membuang *noise* (data yang mengganggu seperti emoji, tanda baca, multispati, angka, url, baris *enter* dan lain-lain). Adapun atribut yang digunakan pada penelitian ini adalah *comments* (komentar) dan label, sedangkan atribut yang tidak digunakan antara lain yaitu *username*, *likes*, *date*.
2. *Case Folding*, yaitu proses penyeragaman teks menjadi huruf kecil (*lowercase*).
3. *Tokenizing*, yaitu proses pemecahan atau memisah kalimat menjadi kata per kata.
4. *Stopword removal*, yaitu proses menghapus atau menghilangkan kata-kata yang sering muncul namun dianggap tidak penting dan tidak memiliki arti.
5. *Stemming*, yaitu proses mengubah kata-kata menjadi bentuk dasar dengan menghilangkan semua imbuhan yang melekat pada kata. *Library* Sastrawi digunakan pada proses *stopword* dan *stemming*.

2.3 Pembobotan TF-IDF

Term frequency (TF) adalah jumlah istilah (*term*) yang muncul dalam suatu dokumen, sedangkan *inverse document frequency* (IDF) adalah kebalikannya yaitu dokumen yang terdapat istilah (*term*) tersebut [19]. Fungsi IDF yaitu untuk memberi skor terendah dari *term* yang muncul pada dokumen-dokumen yang terdapat didalam dataset [20]. Setelah melalui *text preprocessing*, dataset harus dikonversi ke bentuk numerik agar dapat lanjut ke tahap klasifikasi. Salah satu metode yang dapat digunakan adalah

pembobotan TF-IDF. Bobot dari sebuah kata dihitung dengan menggabungkan nilai TF dan nilai IDF [21]. Persamaan TF-IDF dapat dilihat pada persamaan 1 dan 2 :

$$IDF = \log \log \frac{D}{DF} \tag{1}$$

$$TF - IDF = TF * IDF \tag{2}$$

Keterangan :

- D : jumlah dokumen pada data latih
- DF : jumlah dokumen yang memiliki *term*
- TF : jumlah kemunculan *term* pada suatu dokumen
- IDF : jumlah dokumen yang mengandung *term*

2.4 Naive Bayes Classifier

Naive bayes classifier yaitu suatu algoritma klasifikasi yang bekerja berdasarkan probabilitas yang kondisional, algoritma ini memprediksi probabilitas yang ada pada setiap kelas di dataset [22]. Metode ini pertama kali diperkenalkan pertama kali diperkenalkan oleh ilmuwan Inggris bernama Thomas Bayes. Metode *naive bayes* tersebut bekerja pada probabilitas dan statistik yang berguna untuk memprediksi kemungkinan di masa depan berdasarkan pembelajaran yang telah dilakukan di masa lalu. Oleh karena itu, metode ini disebut dengan Teorema Bayes. [21]. Rumus *algoritma naive bayes classifier* dapat dilihat pada persamaan 3.

$$P(X) = \frac{P(C) P(c)}{P(x)} \tag{3}$$

Keterangan :

- X : Data dengan kelas yang tidak diketahui
- C : Hipotesis data yang merupakan kategori spesifik
- $P(C|X)$: Probabilitas hipotesis C yang didasarkan pada probabilitas *posterior*
- $P(c)$: Probabilitas hipotesis C (probabilitas *prior*)
- $P(x|c)$: Probabilitas yang didasarkan pada kondisi pada hipotesis
- $P(x)$: Probabilitas c

2.5 Pengujian

Tahap pengujian dilakukan untuk melihat kinerja algoritma *naive bayes classifier* dalam melakukan klasifikasi. Adapun pengujian atau evaluasi menggunakan *confusion matrix*. *Confusion matrix* ialah matriks yang menyajikan informasi mengenai hasil prediksi dan nilai yang sebenarnya [8]. Pada tahap ini juga akan dilakukan penghitungan terhadap akurasi, *precision*, *recall* dan *F1-score*. Adapun persamaan dari akurasi, *precision*, *recall* dan *F1-score* adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100 \tag{4}$$

$$Precision = \frac{TP}{TP + FP} * 100 \tag{5}$$

$$Recall = \frac{TP}{TP + FN} * 100 \tag{6}$$

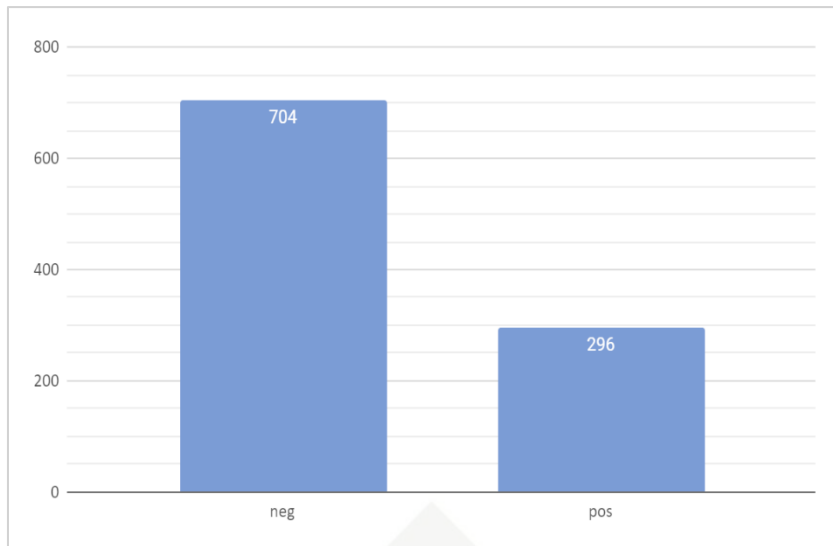
$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{7}$$

3. HASIL DAN PEMBAHASAN

Data yang telah dikumpulkan dengan cara *crawling* berjumlah 1000 data komentar YouTube mengenai larangan penggunaan obat sirup. Kemudian proses *labeling* dilakukan dengan menggunakan tenaga pakar sehingga memperoleh 704 komentar dengan kelas negatif dan 296 komentar dengan kelas positif yang dapat dilihat pada gambar 2.



Hak cipta © 2023 UIN Suska Riau. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber: a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah. b. Dilarang mengutip dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2. Jumlah komentar positif dan negatif

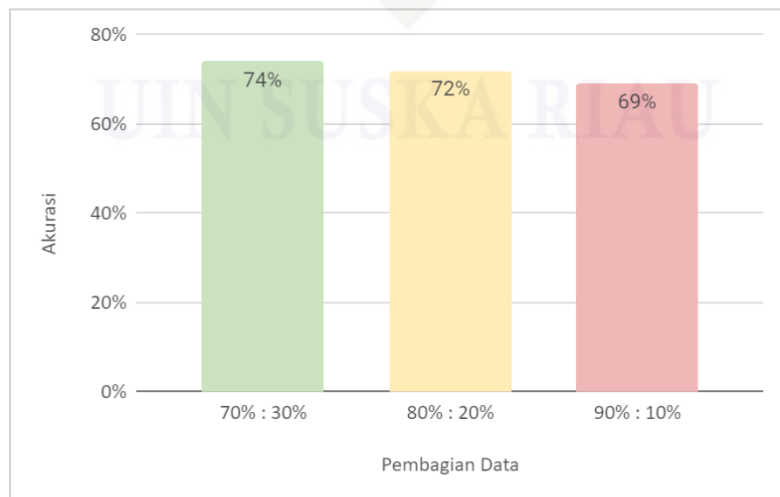
Bahasa pemrograman yang digunakan pada penelitian ini yaitu Python dan proses pengolahan dilakukan dengan Google Colab. Komentar-komentar atau data yang sudah diberi label selanjutnya masuk ke tahap *text preprocessing*, hasil dari setiap tahap pada *text preprocessing* dapat dilihat pada tabel 1 berikut.

Tabel 1. Tahap *text preprocessing*

Komentar Awal	Untuk sementara jgn pake sirup dulu, klo anak demam biar berkurang panasnya bisa kasih obat herbal aja... 😊
Cleaning	Untuk sementara jgn pake sirup dulu klo anak demam biar berkurang panasnya kasih obat herbal aja
Case Folding	untuk sementara jgn pake sirup dulu klo anak demam biar berkurang panasnya kasih obat herbal aja
Tokenizing	untuk, sementara, jgn, pake, sirup, dulu, klo, anak, demam, biar, berkurang, panasnya, kasih, obat, herbal, aja
Stopword Removal	sementara, sirup, anak, demam, berkurang, panasnya, obat, herbal
Stemming	sementara, sirup, anak, demam, kurang, panas, obat, herbal

Setelah melalui tahapan *text preprocessing*, selanjutnya dilakukan penghitungan seberapa sering setiap kata tersebut muncul dalam dokumen. Proses memberi bobot pada kata menggunakan TF-IDF adalah dengan menghitung nilai TF atau jumlah kata yang muncul pada seluruh dokumen, kemudian dilakukan penghitungan nilai IDF berdasarkan persamaan (1), lalu nilai TF-IDF pada dokumen dihitung berdasarkan persamaan (2).

Pada penelitian ini, pengujian dilakukan pada tiga percobaan. Percobaan pertama menggunakan pembagian data latih sebanyak 70%, percobaan kedua menggunakan pembagian data latih sebanyak 80% dan yang ketiga menggunakan pembagian data latih sebanyak 90%. Akurasi yang didapatkan dari masing-masing pembagian data ditampilkan pada gambar 3 berikut.



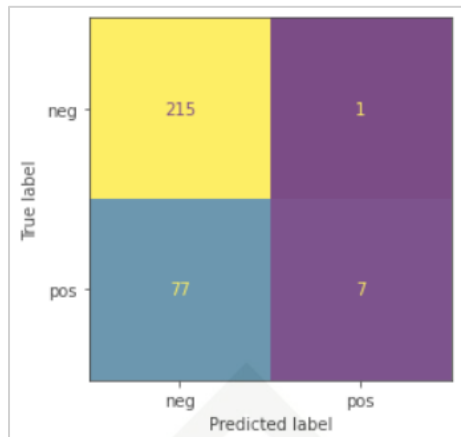
Gambar 3. Hasil Percobaan



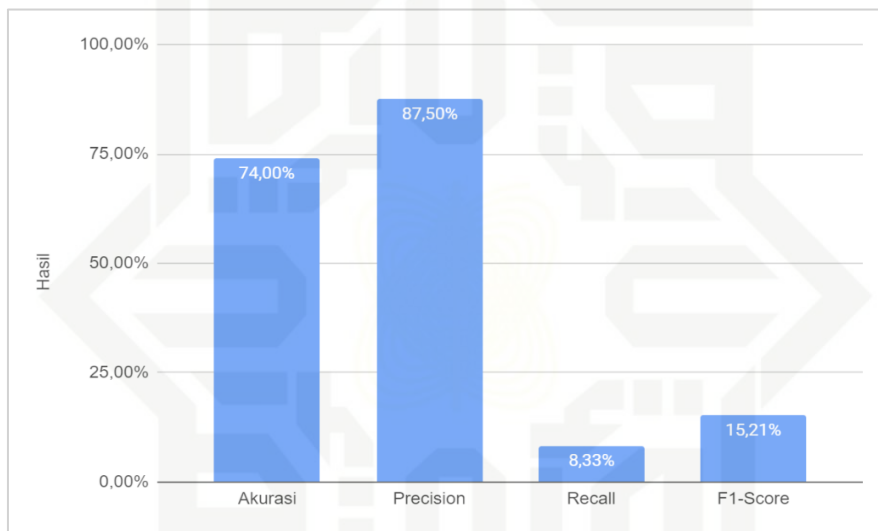
Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip sebagian atau seluruhnya tanpa mencantumkan sumber
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau
 University of Sultan Syarif Kasim Riau

Kemudian dilakukan tahapan pengujian kinerja, perhitungan nilai dan identifikasi model klasifikasi yang telah dibangun menggunakan algoritma *naive bayes classifier*, evaluasi dilakukan dengan melihat *confusion matrix* dan menghitung nilai *accuracy*, *precision*, *recall* dan *F1-Score* berdasarkan persamaan (4), (5), (6) dan (7). Hasil evaluasi pada akurasi tertinggi (74%) ditampilkan pada gambar 4 dan 5 berikut.

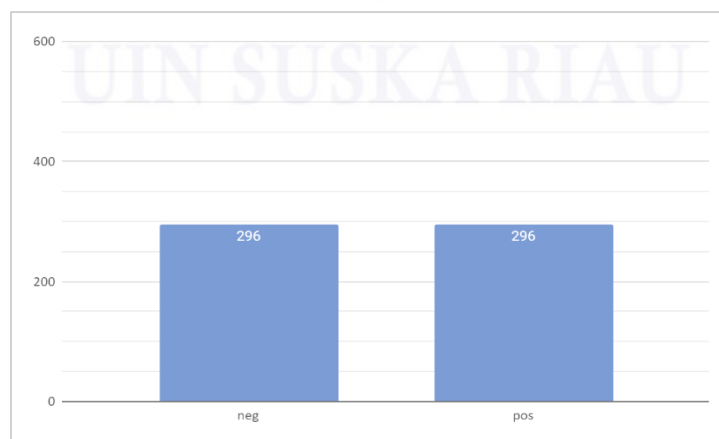


Gambar 4. Confusion Matrix



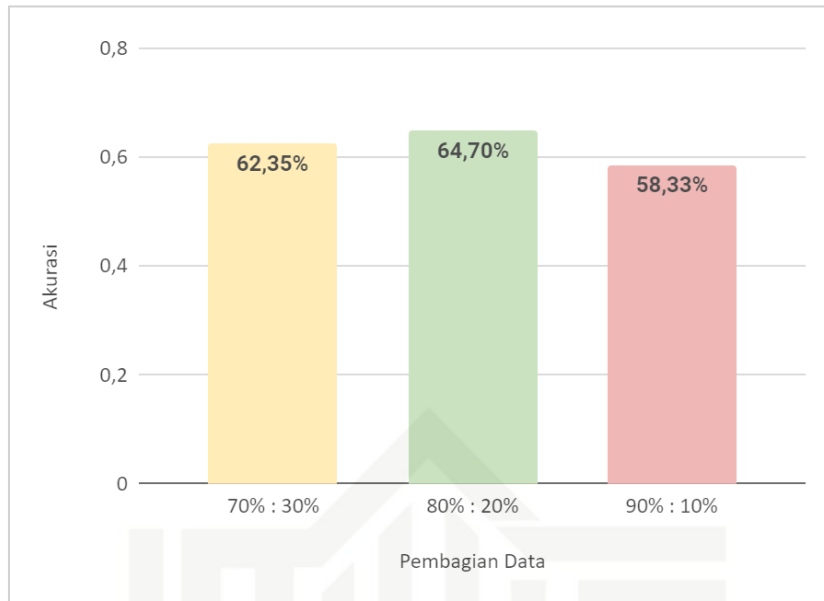
Gambar 5. Hasil Evaluasi Algoritma Naive Bayes Classifier

Karena dataset pada percobaan tersebut tidak seimbang (*imbalanced data*) yaitu lebih banyak komentar negatif, kemudian dilakukan juga percobaan dengan dataset yang seimbang yaitu 296 komentar positif dan 296 komentar negatif. Tujuan dari percobaan tersebut yaitu untuk mengamati performa algoritma saat melakukan klasifikasi pada kondisi data yang seimbang. Grafik dataset dapat dilihat pada gambar 6 dibawah ini.



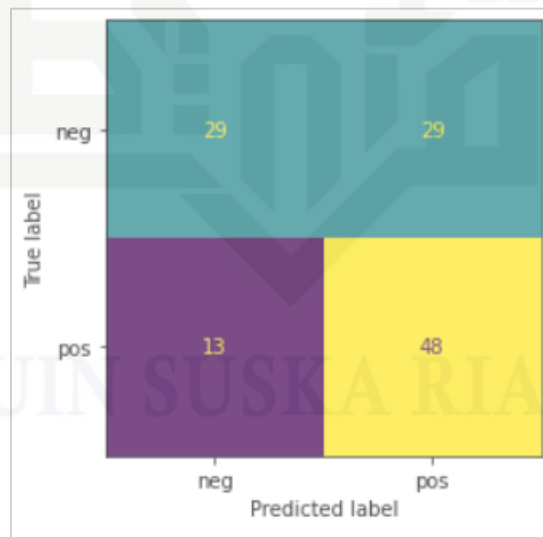
Gambar 6. Grafik Dataset Seimbang

Selanjutnya hasil akurasi pada setiap pembagian data (70:30, 80:20 dan 90:10) pada percobaan dengan data seimbang tersebut ditampilkan pada gambar 7 dibawah ini.



Gambar 7. Hasil Percobaan Pada Dataset Seimbang

Dengan total data sebanyak 592 pada dataset seimbang tersebut, akurasi tertinggi didapatkan sebesar 64,70% pada pembagian data 80:20. Perbedaan yang cukup signifikan dapat dilihat pada nilai *recall* dan *F1-score*. Pada percobaan menggunakan dataset awal, nilai *recall* diperoleh sebesar 8,33% dan *F1-score* sebesar 15,21%. Sedangkan pada percobaan menggunakan dataset seimbang (*balanced data*), nilai *recall* diperoleh sebesar 78,68% (tertinggi dibanding kriteria evaluasi lain) dan nilai *F1-score* sebesar 69,56%. Evaluasi pada dataset seimbang tersebut memiliki perbedaan nilai yang cukup berimbang dibandingkan dengan evaluasi pada dataset awal. Selanjutnya untuk evaluasi pada akurasi tertinggi menggunakan *confusion matrix* dan grafik perbandingan evaluasi pada akurasi tertinggi menggunakan dataset seimbang tersebut akan ditampilkan pada gambar 8 dan 9 dibawah ini.



Gambar 8. Confusion Matrix Pada Dataset Seimbang

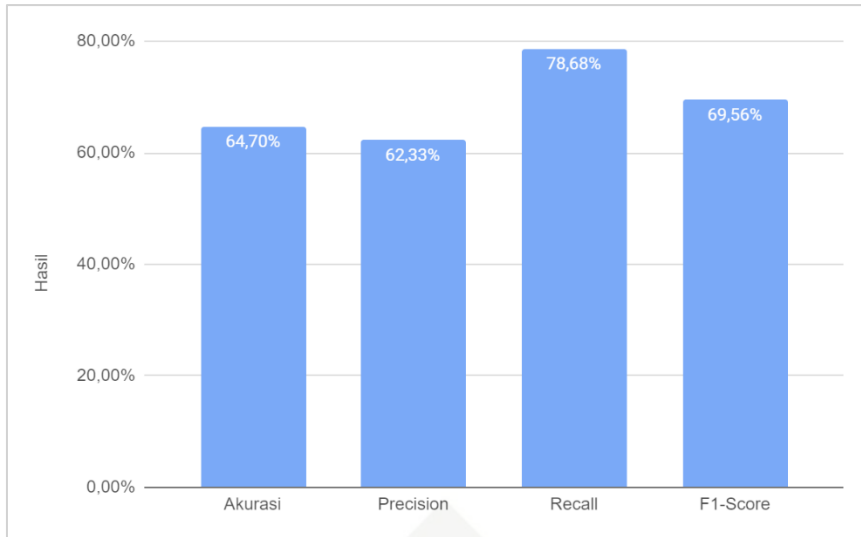


© Hak cipta milik UIN Suska Riau

Hak Cipta Dilindungi Undang-Undang

- Dilarang mengutip sebagian atau seluruh karya tulis atau hasil penelitian dan menyebutkan sumber:
 - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

State Islamic University of Sultan Syarif Kasim Riau



Gambar 9. Grafik Evaluasi Pada Dataset Seimbang

4. KESIMPULAN

Berdasarkan hasil penelitian tersebut, maka analisis sentimen mengenai larangan penggunaan obat sirup menggunakan komentar YouTube pada dataset awal mendapatkan nilai akurasi sebesar 69% pada pembagian data 90:10, 72% pada pembagian data 80:20 dan akurasi tertinggi yaitu 74% pada pembagian data 70:30, sementara nilai *F1-score* yang diperoleh pada akurasi tertinggi tersebut yaitu 15,21%. Nilai *F1-score* tersebut memiliki perbedaan yang signifikan dibandingkan dengan nilai evaluasi lainnya. Hal ini terjadi karena ketidakseimbangan data (*imbalanced data*). Kemudian dilakukan pengujian dengan data yang seimbang sehingga menghasilkan akurasi tertinggi sebesar 64,70% pada pembagian data 80:20 dan nilai *F1-score* yang cukup berimbang yaitu sebesar 69,56%. Hasil tersebut menunjukkan algoritma *naïve bayes classifier* cukup baik dalam melakukan analisis sentimen pada komentar YouTube mengenai larangan penggunaan obat sirup. Oleh karena itu, pada penelitian selanjutnya diharapkan melakukan pengembangan seperti menggunakan data yang seimbang dan jumlah yang lebih banyak, juga menggunakan algoritma lain untuk melihat performa algoritma yang lebih baik dalam melakukan klasifikasi sentimen tentang larangan penggunaan obat sirup.

DAFTAR PUSTAKA

- [1] I. Suhendra and D. L. Sahputri, "BPOM Tarik 73 Obat Sirup Terkait Etilen Glikol, Kemenkes: Dilarang Dipakai!," *November, 09th 2022, 2022*. [Online]. Available: <https://www.viva.co.id/gaya-hidup/kesehatan-intim/1542165-bpom-tarik-73-obat-sirup-terkait-etilen-glikol-kemenkes-dilarang-dipakai>. [Accessed: 09-Nov-2022].
- [2] Newswire, "Dipakai 93,8% Pengguna Internet Indonesia, YouTube Jadi Medsos Terpopuler," *February, 16th 2022, 2021*. [Online]. Available: <https://www.solopos.com/dipakai-938-pengguna-internet-indonesia-youtube-jadi-medsos-terpopuler-1107625>. [Accessed: 09-Nov-2022].
- [3] I. Zharfan and Ferlangga, "Youtube Jadi Raja Media Sosial Di Indonesia, Diakses 94 Persen Warga," *August, 25th 2022, 2022*. [Online]. Available: <https://student-activity.binus.ac.id/himti/2022/08/25/YouTube-jadi-raja-media-sosial-di-indonesia-diakses-94-persen-warga/>. [Accessed: 11-Nov-2022].
- [4] I. Adhal, R. Kurniawan, I. Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 1, pp. 122–130, 2022.
- [5] A. Ardwi, "5 Fakta Larangan Penggunaan Obat Sirup Menurut Edaran Kemenkes," 2022. [Online]. Available: <https://satukanal.com/baca/fakta-larangan-obat-sirup/92926/>. [Accessed: 05-Nov-2022].
- [6] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, "Comparison Of Classification Algorithm And Feature Selection In Bitcoin Sentiment Analysis," *J. Tek. Inform.*, vol. 3, no. 3, pp. 739–744, 2022.
- [7] D. A. Fitri and A. Putri, "Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine," *J. Comput. Sci. Inf. Technol. (CoSciTech)*, vol. 3, no. 3, pp. 472–478, 2022.
- [8] P. Karthika, R. Murugeswari, and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," *2019 IEEE Int. Conf. Intell. Tech. Control. Optim. Signal Process.*, pp. 1–5, 2019.
- [9] G. Khanvilkar and D. Vora, "Smart Recommendation System based on Product Reviews using Random Forest," *2019 Int. Conf. Nascent Technol. Eng.*, pp. 1–9, 2019.



Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mengutip sumber.
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

[10] D. Pramadhana, "Klasifikasi Penyakit Diabetes Menggunakan Metode CFS dan ROS dengan Algoritma J48 Berbasis Adaboost," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 89–98, 2021.

[11] Kusrini and M. Mashuri, "Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication," IEEE, Yogyakarta, Indonesia, 2019.

[12] D. F. Zhafira, B. Rahayudi, and Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes Dan Pembobotan Tf-Idf Berdasarkan Komentar Pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55–63, 2021.

[13] Surohman, S. Aji, Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," *Evolusi J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020.

[14] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.

[15] S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, and W. A. Arifin, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *Temat. - J. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, pp. 96–103, 2021.

[16] S. G. Kaparang, D. R. Kaparang, and V. P. Rantung, "Analisis Sentimen New Normal Pada Masa Covid-19 Menggunakan Algoritma Naive Bayes," *JOINTER – J. INFORMATICS Eng.*, vol. 02, no. 01, pp. 16–23, 2021.

[17] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naive Bayes," *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 267–281, 2022.

[18] E. P. Nirwandani, Indriati, and R. C. Wihandika, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naive Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, pp. 1039–1047, 2021.

[19] H. Prasetyo, G. A. Buntoro, D. Mustikasari, "Analisis Sentimen Pada Channel Autonetmagz Terhadap Review Mobil Almaz 2019 Dengan Metode Naive Bayes Classifier Dan Lexicon Based," *J. Tek. Univ. Muhammadiyah Ponorogo*, vol. 4, no. 1, pp. 58–73, 2020.

[20] A. Shiddicky and S. Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Logistic Regression," *J. Comput. Sci. Inf. Technol.*, vol. 3, no. 2, pp. 99–106, 2022.

[21] A. Deolika, Kusrini, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, pp. 179–184, 2019.

[22] C. Dabas, P. Kaur, N. Gulati, and M. Tilak, "Analysis of Comments on YouTube Videos using Hadoop," *Fifth Int. Conf. Image Inf. Process.*, pp. 353–358, 2019.

Hak Cipta Dilindungi Undang-Undang

State Islamic University of Sultan Syarif Kasim Riau

UIN SUSKA RIAU

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mengutip sumber:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.